

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

ADRIANA BEZERRA BESSA

ENSAIOS SOBRE PREVISÃO DE VENDAS NO VAREJO DE MODA

SÃO PAULO

2021

ADRIANA BEZERRA BESSA

ENSAIOS SOBRE PREVISÃO DE VENDAS NO VAREJO DE MODA

Tese apresentada à Escola de Economia da Fundação Getúlio Vargas (FGV EESP) como requisito para obtenção do título de Doutor em Economia Aplicada

Campo de conhecimento:
Economia Aplicada

Orientador: Prof. Dr. Marcelo
Fernandes

SÃO PAULO

2021

Bessa, Adriana Bezerra.

Ensaio sobre previsão de vendas no varejo de moda / Adriana Bezerra Bessa. - 2021.

248 f.

Orientador: Marcelo Fernandes.

Tese (doutorado CDEE) – Fundação Getúlio Vargas, Escola de Economia de São Paulo.

1. Previsão de vendas. 2. Comércio varejista. 3. Análise de séries temporais. 4. Aprendizado do computador. 5. Processo decisório. I. Fernandes, Marcelo. II. Tese (doutorado) – Escola de Economia de São Paulo. III. Fundação Getúlio Vargas. IV. Título.

CDU 658.8.012.2

ADRIANA BEZERRA BESSA

ENSAIO SOBRE PREVISÃO DE VENDAS NO VAREJO DE MODA

Tese apresentada à Escola de Economia da Fundação Getúlio Vargas (FGV EESP) como requisito para obtenção do título de Doutor em Economia Aplicada.

Campo de conhecimento:
Economia Aplicada

Data de aprovação: ____/____/____

Banca Examinadora:

Prof. Dr. Marcelo Fernandes
(Orientador)
FGV EESP

Prof. Dr. Luiz Henrique Moraes
(Coorientador)
FGV EESP

Prof. Dr. Emerson Marçal
FGV EESP

Prof. Dra. Elia Matsumoto
FGV EESP

Prof. Dr. Mauricio Zevallos
IMECC/UNICAMP

AGRADECIMENTOS

Agradeço, primeiramente, a Deus pela saúde e pela confiança para produzir esta pesquisa.

Agradeço aos meus Pais e aos meus Irmãos por todo amor e compreensão pela minha maior ausência.

Agradeço aos meus Colegas do curso, em especial, ao Raphael Ferreira por toda a parceria e a troca efetiva de conhecimento, permitindo consolidar o meu aprendizado ao longo do Doutorado.

Agradeço a cada um dos meus Professores, em particular, a Elia Matsumoto, Edson Kitani, Gerson Faria e Luiz Moraes pelo conhecimento compartilhado na minha área de concentração de estudos.

Agradeço ao meu Orientador Marcelo Fernandes por me guiar nesta pesquisa e sempre estar disponível a me ajudar.

Agradeço ao meu Coorientador Luiz Moraes por me apoiar sempre que precisei.

Agradeço a colaboração singular de Mario Filho pelas dicas preciosas e pela contribuição essencial de Alan Lahoud para tornar meu código mais organizado e eficiente.

“I never think of the future – it comes soon enough.”

Albert Einstein

RESUMO

A previsão de vendas no varejo de moda é essencial no gerenciamento de negócios, afetando significativamente toda a cadeia de abastecimento, e pode ser vista como a informação básica para que outras atividades de planejamento possam ser desenvolvidas dentro das organizações que exercem esta atividade comercial.

Uma série de ineficiências são provocadas por uma previsão com baixo nível de precisão, tais como: aumento dos níveis de estoques, números maiores de remarcações, menores margens financeiras e um efeito chicote maior.

O objetivo deste trabalho foi verificar os aspectos relevantes ligados ao processo de previsão de vendas no curto prazo para produtos básicos, com foco maior em métodos de aprendizagem de máquina. Desta forma, esta pesquisa buscou responder os seguintes questionamentos: a aplicação de tratamentos sazonais nos dados contribui para gerar previsões mais precisas? O uso de tratamentos de seleção de variáveis, considerado o amplo universo de variáveis disponíveis, contribui para reduzir o erro das estimativas geradas? Quais variáveis foram mais relevantes para gerar as previsões calculadas? Qual o método de previsão é o mais preciso? (Tendo em vista uma ampla gama de métodos testados, contemplando métodos estatísticos, de inteligência computacional e *ensembles*.) Por fim, qual método de previsão de vendas deve ser adotado pelo setor em foco, considerando a análise AHP realizada?

Com base no estudo das quatro categorias de produtos mais relevantes para o Lojão do Brás, considerando as vendas semanais do período de 2007 a 2019 e tendo como foco a previsão de curto prazo da próxima semana, constatou-se que: não houve uma unanimidade sobre a efetividade da aplicação de tratamentos sazonais; o uso de tratamentos de seleção de variáveis permitiu obter previsões 13% mais precisas; as variáveis exógenas contempladas no estudo foram pouco relevantes; o método de previsão de vendas com melhor nível de acurácia para o setor de varejo de moda foi o XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA. Por fim, com base na análise AHP realizada, que considerou vários critérios e não somente o aspecto da precisão, a recomendação final como melhor modelo de previsão de venda para o setor em questão foi o método Naive ajustado, apesar da superioridade do XGBOOST no critério de acurácia.

Palavras-chave: Previsão de vendas. Varejo de moda. Séries temporais. Métodos de aprendizagem de máquina. Sazonalidade. Seleção de variáveis. Métodos de previsão por combinação. Análise AHP.

ABSTRACT

Forecasting sales in fashion retail plays an essential role in business management, significantly affecting the entire supply chain. Imprecise forecasts cause a series of problems, such as increased stock levels, higher numbers of markdowns, lower financial margins and a greater bullwhip effect. The objective of this work was to verify the relevant aspects related to the short-term sales forecast process for basic products, with a greater focus on machine learning methods. Thus, this research sought to answer the following questions: does the application of seasonal treatments in the data contribute to generate more accurate forecasts? Does the use of variable selection treatments, considering the wide universe of available variables, contribute to reducing the error of the generated estimates? Which variables were most relevant to generate the calculated forecasts? Which forecasting method is the most accurate? (In view of a wide range of tested methods, including statistical methods, computational intelligence and ensembles.) Finally, which sales forecasting method should be adopted by the sector in focus, considering the AHP analysis performed?

Based on the study of the four most relevant product categories for Lojão do Brás, considering weekly sales for the period 2007 to 2019 and focusing on the short-term forecast for next week, it was found that: there was no unanimity on the effectiveness of the application of seasonal treatments; the use of variable selection treatments allowed obtaining 13% more accurate predictions; the exogenous variables included in the study were not very relevant; the sales forecasting method with the best accuracy level for the fashion retail sector was the XGBOOST with seasonal treatment by curve fitting and with RFA variable selection treatment. Finally, based on the AHP analysis performed, which considered several criteria and not just the accuracy aspect, the final recommendation as the best sales forecast model for the sector in question was the adjusted Naive method, despite the superiority of XGBOOST in the criterion of accuracy.

Keywords: Sales forecasting. Fashion retail. Time series. Machine learning methods. Seasonality. Feature selection. Ensemble learning. AHP analysis.

LISTA DE FIGURAS

CAPÍTULO 1 – INTRODUÇÃO

Figura 1.1 – Estrutura da cadeia de abastecimento do setor têxtil.....	23
Figura 1.2 – Evolução do faturamento do <i>e-commerce</i> no Brasil.....	25
Figura 1.3 – <i>Framework</i> para análise de métodos de previsão.....	33
Figura 1.4 – Processo de previsão.....	34
Figura 1.5 – Categorias de métodos de previsão.....	36
Figura 1.6 – Etapas do modelo de planejamento e operações do varejo de moda...	36
Figura 1.7 – Categorias de métodos de previsão setor de moda.....	38
Figura 1.8 – Agrupamentos de métodos de previsão do setor de moda.....	40
Figura 1.9 – Categorias de métodos de previsão.....	41
Figura 1.10 – Tipos de penalização por método de regularização.....	50
Figura 1.11 – Árvore de decisão.....	55
Figura 1.12 – Crescimento da árvore de decisão.....	58
Figura 1.13 – Rede Neural MLP.....	60
Figura 1.14 – Arquitetura genérica de rede FTDNN.....	61
Figura 1.15 – <i>1-D Convolution for Time Series</i>	63
Figura 1.16 – Célula de memória LSTM.....	65
Figura 1.17 – Etapas do processo KDD.....	70
Figura 1.18 – Metodologia CRISP-DM.....	72
Figura 1.19 – Visão geral do processo de limpeza dos dados.....	77
Figura 1.20 – Quadro-resumo de técnicas de seleção de variáveis.....	86
Figura 1.21 – Visão geral das técnicas de redução de dimensionalidade.....	89
Figura 1.22 – Categorias dos algoritmos de previsão.....	120
Figura 1.23 – Hierarquia de cadastro dos produtos do Lojão do Brás.....	126
Figura 1.24 – Variáveis utilizadas para previsão de vendas.....	130
Figura 1.25 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série camiseta de homem.....	138
Figura 1.26 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série camiseta de homem.....	139
Figura 1.27 – Histograma e medidas descritivas da série camiseta de homem.....	139

Figura 1.28 – Histograma e medidas descritivas da série camiseta de homem em diferença	140
Figura 1.29 – FAC e FACP da série camiseta de homem.....	140
Figura 1.30 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série blusa de mulher	141
Figura 1.31 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série blusa de mulher.....	141
Figura 1.32 – Histograma e medidas descritivas da série blusa de mulher	142
Figura 1.33 – Histograma e medidas descritivas da série blusa de mulher em diferença	142
Figura 1.34 – FAC e FACP da série blusa de mulher	143
Figura 1.35 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série camiseta de menino.....	143
Figura 1.36 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série camiseta de menino	144
Figura 1.37 – Histograma e medidas descritivas da série camiseta de menino	144
Figura 1.38 – Histograma e medidas descritivas da série camiseta de menino em diferença	145
Figura 1.39 – FAC e FACP da série camiseta de menino	145
Figura 1.40 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série blusa de menina	146
Figura 1.41 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série blusa de Menina.....	147
Figura 1.42 – Histograma e medidas descritivas da série blusa de Menina.....	147
Figura 1.43 – Histograma e medidas descritivas da série blusa de menina em diferença	147
Figura 1.44 – FAC e FACP da série blusa de menina.....	148
Figura 1.45 – Gráficos da evolução das variáveis exógenas macroeconômicas	150
Figura 1.46 – Gráficos da evolução das variáveis exógenas climáticas.....	151

CAPÍTULO 2 – Análise do tratamento sazonal nos métodos de previsão de vendas baseados em aprendizagem de máquina

Figura 2.1 – Variáveis utilizadas para previsão de vendas	168
Figura 2.2 – Métodos de previsão e tratamentos sazonais utilizados	170

CAPÍTULO 3 – Seleção de variáveis nos métodos de previsão de vendas baseados em aprendizagem de máquina

Figura 3.1 – Métodos de previsão e tratamentos de seleção de variáveis utilizados	190
---	-----

CAPÍTULO 4 – Métodos de combinação de previsão: uma análise comparativa para o varejo de moda

Figura 4.1 – Métodos de previsão <i>ensembles</i> e não <i>ensembles</i> aplicados	215
--	-----

LISTA DE TABELAS

CAPÍTULO 1 – INTRODUÇÃO

Tabela 1.1 – Equações dos métodos Holt-Winters	46
Tabela 1.2 – Similaridade entre hiperparâmetros.....	68
Tabela 1.3 – Hiperparâmetros escolhidos por algoritmo de aprendizagem de máquina	69
Tabela 1.4 – Sumário de correspondências entre KDD, SEMMA e CRISP-DM.....	73
Tabela 1.5 – Quadro de modelos aplicados.....	122
Tabela 1.6 – Quadro das variáveis exógenas	125
Tabela 1.7 – Quadro-resumo variáveis estacionárias e não estacionárias	130
Tabela 1.8 – Alternativas de modelos para cada algoritmo de previsão	136
Tabela 1.9 – Medidas descritivas das variáveis exógenas macroeconômicas.....	150
Tabela 1.10 – Medidas descritivas das variáveis exógenas climáticas	152

CAPÍTULO 2 – Análise do tratamento sazonal nos métodos de previsão de vendas baseados em aprendizagem de máquina

Tabela 2.1 – Resultado das métricas de erro como uso do tratamento sazonal.....	172
Tabela 2.2 – Resultado das métricas de erro por tipo de tratamento sazonal	173
Tabela 2.3 – Resultado métricas de erro por tipo de algoritmo e por tipo de tratamento sazonal.....	174
Tabela 2.4 – Resultado métricas de erro por algoritmo e por tipo de tratamento sazonal.....	175
Tabela 2.5 – Resultado do teste MCS por algoritmo e por tipo de tratamento sazonal	176

CAPÍTULO 3 – Seleção de variáveis nos métodos de previsão de vendas baseados em aprendizagem de máquina

Tabela 3.1 – Resultado das métricas de erro como uso de tratamento de seleção de variáveis.....	192
Tabela 3.2 – Resultado das métricas de erro por tipo de tratamento de seleção de variáveis	193

Tabela 3.3 – Resultado das métricas de erro por tipo de algoritmo e por tipo de tratamento de seleção de variáveis.....	194
Tabela 3.4 – Resultado das métricas de erro por algoritmo e por tratamento de seleção de variáveis.....	195
Tabela 3.5 – Resultado de teste MCS por algoritmo e por tipo de tratamento de seleção de variáveis.....	196
Tabela 3.6 – Análise do custo computacional – tempo em segundos.....	198
Tabela 3.7 – Análise de frequência das variáveis por série de produto	199
Tabela 3.8 – <i>Ranking</i> das 30 variáveis mais frequentes para todas as séries de produtos	200

CAPÍTULO 4 – Métodos de combinação de previsão: uma análise comparativa para o varejo de moda

Tabela 4.1 – Escala fundamental	217
Tabela 4.2 – Pesos dos critérios para avaliação das previsões	218
Tabela 4.3 – Resultado das métricas de erro como uso de métodos <i>ensembles</i> ...	219
Tabela 4.4 – Resultado das métricas de erro por tipo de método <i>ensemble</i>	220
Tabela 4.5 – Resultado das métricas de erro por tipo de algoritmo	222
Tabela 4.6 – Análise de frequência dos algoritmos mais precisos	223
Tabela 4.7 – Resultado do teste MCS total.....	225
Tabela 4.8 – Vetor prioridade por critério e por tipo de produto	228
Tabela 4.9 – Resultado da análise AHP por categoria de produto.....	231

LISTA DE ABREVIATURAS

3F	Fast Fashion Forecasting
AD-ELM	Extreme Learning Machine com Métricas Adaptativas
ADF	Augmented Dickey-Fuller
ADL	Autoregressive Distributed Lag
AHFCCX	Automatic Hybrid Forecasting Model with Fuzzy Estimation of Corrective Coefficients of the Explanatory Variables Influence
AHP	Analytic Hierarchy Process
AIC	Akaike Information Criteria
ANFIS	Adjustment of Mean-term Forecast with Adaptive Neural Fuzzy Inference System
ANN	Artificial Neural Network
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving-average
AR MIDAS	Autoregressive Mixed-data sampling
BIC	Bayesian information criterion – Schwarz Criterion
BPTT	Backpropagation Through Time
CAGR	Compound Annual Growth Rate
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CRISP-DM	Cross Industry Standard Process for Data Mining
DFGLS	Dickey-Fuller Test Statistic Using Generalized Least Squares
DNN	Deep Neural Networking
EELM	Expanded Extreme Learning Machine
EFB	Exclusive Feature Bundling
ELM	Extreme Learning Machine Aprendizagem de Máquina Extrema
ENN	Evolutionary Neural Network Rede Neural Evolucionária
FAC	Função de Autocorrelação
FACP	Função de Autocorrelação Parcial

FECOMERCIO SP	Federação do Comércio de Bens, Serviços e Turismo do Estado de São Paulo
FIS	Fuzzy Inference Systems
FTDNN	Focused Time Delayed Neural Network
GBDT	Gradient Boosted Decision Trees
GBT	Gradient Boosted Tree
GHSOM	Growing Hierarchical Self Organizing Map
GM	Grey Model
GOSS	Gradient Based One-Side Sampling
HC	Hanna-Quinn
HFCCX	Hybrid Forecasting Model with Fuzzy Estimation of Corrective Coefficients of the Explanatory Variables Influence
HIS	Híbrido Intelligence System
HWA	Holt Winter Additive Model
HWM	Holt Winter Multiplicative Model
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IPCA	Índice de Preços ao Consumidor
KDD	Knowledge Discovery in Databases
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
LGBM	LightGBM
LSTM	Long Short-Term Memory
M-AR-MIDAS	Multiple Autoregressive Mixed-data Sampling
MA	Moving Average
MAE	Mean Absolute Error
MAEP	Erro Absoluto Médio Ponderado
MAPE	Mean Absolute Percentual Error
MARIMA	Multivariate ARIMA
MASE	Mean Absolute Scaled Error
MCS	Model Confident Set Conjunto de Modelos Confiáveis

MF-VAR	Mixed Frequency VAR
MDL-NN	Minimum Description Length Neural Network
ME	Mean Error
MIDAS	Mixed-data Sampling
MLP	Multilayer Perceptron
MPE	Mean Percentual Error
MQO	Mínimos Quadrados Ordinários
MS	Markov Regime Switching
MCS	Model Confident Set
PAC	Pesquisa Anual do Comércio
PCA	Principal Component Analysis
PED	Pesquisa de Emprego e Desemprego
pGBRT	Parallel Gradient Boosted Regression Trees
PIB	Produto Interno Bruto
POF	Pesquisa de Orçamentos Familiares
PPD	Pure Panel Data
PP	Phillips-Perron
p-value	Menor nível de significância para o qual se rejeita a hipótese nula
RF	Random Forest
RFA	Recursive Feature Addition
RMAPE	Revised Mean Absolute Percentual Error
RMSE	Root Mean Squared Error
RMSEP	RMSE Ponderado
RMSLE	Root Mean Squared Log Error
RN	Rede Neural
RNA	Rede Neural Artificial
RNN	Rede Neural Recorrente
RNP	Rede Neural Pulsada
RSMEP	Raiz do Erro Quadrático Médio Ponderado

RSS	Residual Sum of Squares
RU-MIDAS	Reverse Unrestricted MIDAS
SAMANFIS	Short-term Forecasting Model by Adjustment of Mean-term Forecast with Adaptive Neural Fuzzy Inference System
SARIMA	Seasonal Autoregressive Integrated Moving Average – ARIMA Sazonal
SARIMAX	Seasonal Autoregressive Integrated Moving Average with Exogenous Factors
SBVC	Sociedade Brasileira de Varejo e Consumo
SEADE	Fundação Sistema Estadual de Análise de Dados
SEMMA	Sample, Explore, Modify, Model, and Assess
SFS	Sequential Feature Selector
SHAP	Shapley Additive Explanations
SKU	Stock Keeping Unit
SMAPE	Symmetric Mean Absolute Percentage Error
SOM	Self Organizing Map
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TCN	Temporal Convolutional Networks
TPU	Tensor Processing Unit
U-MIDAS	Unrestricted Mixed-data Sampling
VAR	Vector autoregression

SUMÁRIO

I. INTRODUÇÃO.....	16
1. SUMÁRIO EXECUTIVO.....	16
1.1. Tema e objetivo	17
1.2. Motivações	18
1.3. Estrutura do trabalho	19
1.4. Resultados alcançados	20
2. INTRODUÇÃO À PREVISÃO DE VENDAS NO SETOR DE VAREJO DE MODA	22
2.1. O setor de varejo de moda no Brasil	23
2.2. O Lojão do Brás.....	26
2.3. Importância da previsão de vendas.....	27
2.4. Previsão de vendas no varejo de moda	29
3. INTRODUÇÃO AOS MÉTODOS DE PREVISÃO	42
3.1. Características das séries temporais.....	42
3.2. Métodos lineares	45
3.2.1. Métodos Naive	45
3.2.2. Métodos de Suavização Exponencial.....	46
3.2.3. Métodos Econométricos	47
3.2.4. Métodos de Regressão	48
3.2.5. Métodos RU-MIDAS.....	50
3.3. Métodos Não Lineares	53
3.3.1. Métodos de Árvores de Decisão	54
3.3.1.1. <i>Random Forest</i> (RF)	54
3.3.1.2. XGBoost	56
3.3.1.3. LightGBM.....	57
3.3.2. Métodos de Redes Neurais.....	59
3.3.2.1. MLP FTDNN	60
3.3.2.2. CNN	62
3.3.2.3. LSTM	64
3.3.2.4. Hiperparâmetros	66
3.4. Aspectos relevantes dos métodos não lineares	69
3.4.1. Processo KDD e CRISP-DM	69

3.4.1.1.	Preparação dos dados.....	74
3.4.1.2.	Limpeza dos dados.....	76
3.4.1.3.	Transformação dos dados	77
3.4.1.4.	Tratamento de sazonalidade	78
3.4.1.5.	Feature Engineering	81
3.4.1.6.	Seleção de variáveis.....	83
3.4.1.7.	Redução de dimensionalidade.....	88
3.5.	Métricas de avaliação de previsões.....	89
3.6.	Teste Wilcoxon e Model Confident Set (MCS)	93
3.7.	<i>Ensembles</i>	94
3.8.	Análise AHP	98
4.	REVISÃO DA LITERATURA.....	100
5.	METODOLOGIA.....	116
5.1.	Definição do problema de previsão	117
5.2.	Definição dos métodos de previsão.....	119
5.3.	Base de dados e variáveis	123
5.4.	Forma de avaliação das previsões	131
5.5.	Tratamentos sazonais aplicados	132
5.6.	Tratamento de seleção de variáveis.....	132
5.7.	Especificação dos hiperparâmetros dos métodos de previsão.....	133
5.8.	Pacotes computacionais.....	137
5.9.	Análise exploratória das séries temporais	137
5.9.1.	Análise exploratória das séries de quantidade vendida por categoria de produto	138
5.9.2.	Análise exploratória principais variáveis exógenas	149
REFERÊNCIAS.....		152
II. ANÁLISE DO TRATAMENTO SAZONAL NOS MÉTODOS DE PREVISÃO DE VENDAS BASEADOS EM APRENDIZAGEM DE MÁQUINA.....		164
1.	INTRODUÇÃO	166
2.	TIPOS DE TRATAMENTO SAZONAIS	168
3.	O TRATAMENTO SAZONAL FAZ DIFERENÇA?	171
4.	CONCLUSÕES	177
REFERÊNCIAS.....		179

III. SELEÇÃO DE VARIÁVEIS NOS MÉTODOS DE PREVISÃO DE VENDAS BASEADOS EM APRENDIZAGEM DE MÁQUINA.....	181
1. INTRODUÇÃO	183
2. MÉTODOS DE SELEÇÃO DE VARIÁVEIS	185
3. VALE A PENA SELECIONAR VARIÁVEIS?	191
4. CONCLUSÕES	201
REFERÊNCIAS.....	203
IV. MÉTODOS DE COMBINAÇÃO DE PREVISÃO: UMA ANÁLISE COMPARATIVA PARA O VAREJO DE MODA	205
1. INTRODUÇÃO	207
2. MÉTODOS DE COMBINAÇÃO DE PREVISÃO	210
3. ANÁLISE AHP	216
4. QUAL O MODELO DE PREVISÃO É O MAIS ADEQUADO PARA O SETOR DE MODA?	218
5. CONCLUSÕES	231
REFERÊNCIAS.....	235
V. CONSIDERAÇÕES FINAIS.....	237
REFERÊNCIAS.....	242

I. Introdução

“Fashion is not something that exists in dresses only. Fashion is in the sky, in the street. Fashion has to do with ideas, the way we live, what is happening.”

Coco Chanel

1. Sumário executivo

O estudo está situado no de conhecimento de aprendizagem de máquina para o setor de varejo de moda. Mais especificamente, na verificação dos ganhos em termos de acurácia da aplicação de métodos de inteligência computacional, contemplando os seguintes aspectos: aplicação de tratamentos de sazonalidade, inclusão de variáveis exógenas e criação de variáveis por meio de *feature engineering* juntamente com o uso de tratamento de seleção de variáveis e aplicação de modelos híbridos, o que foi feito através de modelos *ensembles*.

É fundamental fazer um posicionamento sobre o setor onde este trabalho está inserido, o varejo, especificamente, o varejo de moda, tendo como estudo de caso a rede Lojão do Brás. Varejo é toda atividade econômica da venda de um bem ou um serviço para o consumidor final, ou seja, uma transação entre um CNPJ e um CPF. No Brasil, o varejo é o maior empregador, sendo responsável por 26% dos empregos formais do país e representa 11,5% do Produto Interno Bruto (PIB) nacional, de acordo com a Sociedade Brasileira de Varejo e Consumo (SBVC, 2020a). O varejo de moda representa 6,8% do varejo total e é altamente competitivo, uma vez que as cinco maiores empresas detêm 16% do mercado. Além disso, apresenta alto grau de informalidade – cerca de 40% – e é afetado por uma série de variáveis exógenas, entre elas, renda, emprego, crédito e confiança, de acordo com a SBVC (2020a).

Em um cenário cada vez mais competitivo, no qual há cadeias globais de suprimentos estabelecidas para o atendimento das demandas das redes varejistas têxteis, a previsão de demanda tem ganhado cada vez mais relevância dentro das empresas e organizações desse segmento. A justificativa reside no fato de uma previsão mais assertiva possibilitar o desenvolvimento de uma cadeia de fornecimento

mais integrada e eficiente, levando a menores níveis de estoques, menos remarcações, menores rupturas¹, maiores margens e um efeito chicote menor, o qual é definido como um fenômeno que ocorre quando as ordens de compra para os fornecedores tendem a apresentar uma variância maior do que as vendas do elo mais próximo do consumidor, mais especificamente dos varejistas.

A importância da previsão de demanda está diretamente ligada à racionalização dos recursos envolvidos e a meta do nível de atendimento aos clientes. Qualquer atividade de planejamento pressupõe previsões sobre o futuro, as quais serão mais precisas dependendo das hipóteses utilizadas. O processo de previsão pode ser definido como o conjunto de todas as atividades envolvidas para que seja possível a geração dos números finais da previsão, o que inclui desde a etapa de coleta de dados até a última revisão da previsão (DIAS, 1999).

Cabe ressaltar que a complexidade e a especificidade do varejo têxtil tornam o processo de previsão mais difícil do que em outros setores varejistas. Nesse sentido, as características que aumentam o grau de complexidade de previsão no setor, de acordo com Thomassey (2010) são as seguintes: (i) a definição do horizonte de previsão em si, se é de curto ou longo prazo; (ii) o ciclo de vida de produtos, pois há produtos mais básicos e outros produtos com atributos mais *fashion*, que são renovados com maior frequência, diminuindo o volume de dados disponível para previsão; (iii) o impacto da sazonalidade e da temperatura; (iv) o efeito efêmero da moda e seu impacto nos hábitos de consumo e a influência de uma série de outras variáveis, tais como variáveis macroeconômicas, promoções e preços da concorrência, entre outros. Todos esses fatores tornam o processo de previsão no setor extremamente complexo e específico de ser realizado.

1.1. Tema e objetivo

O tema desta Tese é a previsão de vendas no varejo de moda com foco nos métodos de aprendizagem de máquina e o objetivo principal é tornar as previsões de vendas mais precisas em termos de quantidade para setor em estudo, o que pode ser

¹ A ruptura de estoque é a falta de um determinado produto no momento de compra do consumidor, afetando tanto o nível de serviço prestado como o nível de satisfação do consumidor junto a uma determinada marca.

feito por meio de diversas abordagens. Neste trabalho foram selecionadas as seguintes perspectivas: (1) aplicação de tratamentos sobre dados, especificamente o tratamento da sazonalidade; (2) seleção de variáveis e (3) consideração de novos métodos de previsão. Cada uma das abordagens é tema de um capítulo da presente tese, sendo que há um quarto e último capítulo se dedica a uma análise comparativa.

As previsões realizadas foram de curto prazo, isto é, para a próxima semana, a partir de dados semanais da rede de varejo de moda Lojão do Brás em cada uma das dezessete lojas da companhia para as seguintes categorias de produtos: blusa de mulher, camiseta de homem, blusa de menino e camiseta de menino, uma vez que essas categorias representam 35% do faturamento da rede. Neste estudo, foi contemplado o histórico de dados das vendas semanais de 2007 a 2019 das séries de produtos selecionadas.

1.2. Motivações

O varejo tem impacto significativo no PIB, tanto em termos de participação como em termos de trabalhadores formalmente empregados, como já mencionado na introdução. Especificamente sobre o varejo de moda, vale destacar seu crescimento global e a expressividade que os negócios de moda têm galgado nos últimos anos, o que foi notado no trabalho de Ipar e Almeida (2011). Tal fato por si só já se coloca como uma motivação bastante relevante como objeto de estudo.

Adicionalmente, o aumento da competitividade, tanto nacional como internacionalmente, vem produzindo cadeias globais de produto e obrigando as companhias do setor a melhorarem a sua gestão do estoque, o que passa necessariamente pelo desenvolvimento de previsões mais precisas, as quais podem ser obtidas pela implantação de melhorias ao longo de todo o processo de previsão.

Duas outras motivações para a realização desta Tese são: o baixo volume de trabalhos acadêmicos para previsão na indústria da moda, tendo em vista as suas especificidades, e o baixo volume de trabalhos com métodos de aprendizagem de máquina, também aplicados ao setor, com ênfase em previsão. Quando se restringe ao âmbito nacional, a produção acadêmica com foco no setor e para a temática de previsão pode ser considerada irrisória.

As motivações apresentadas até o momento podem ser consideradas secundárias, uma vez que a principal justificativa para a realização deste ensaio é encontrar modelos de previsão mais assertivos, com o intuito de garantir uma margem de erro menor do que os modelos econométricos e heurísticos que, conseqüentemente, assegurarão resultados superiores para as companhias desse setor. Nesse sentido, espera-se que os tratamentos propostos, juntamente com os métodos de previsão aplicados, alcancem esse objetivo.

O aumento de precisão da previsão de vendas tem impacto financeiro significativo para as companhias, já que permite reduzir os investimentos em estoques, elevar as margens de contribuição e a experiência de compra do consumidor e obter uma cadeia de suprimento mais eficiente. Esses aspectos contribuem para ampliar a lucratividade das organizações varejistas.

Por fim, a motivação pessoal e não menos importante, que é a atuação da autora na área de planejamento comercial no varejo de moda a cerca de 10 anos, que traz como desafio a busca constante por métodos de previsão mais precisos.

1.3. Estrutura do trabalho

A Tese está organizada em formato de artigos científicos, sendo constituída por cinco capítulos, cada um com suas respectivas referências bibliográficas.

O capítulo I, **Introdução**, apresenta o sumário executivo com os objetivos, as motivações e as delimitações deste ensaio; a introdução sobre o processo de previsão de vendas no setor de varejo de moda e a revisão teórica sobre os métodos de previsão utilizados, juntamente com as métricas e testes aplicados, permitido, assim, a avaliação dos resultados produzidos. Traz ainda uma revisão da literatura sobre o tema em estudo e aborda a metodologia aplicada ao longo do trabalho.

O capítulo II traz o primeiro artigo do ensaio, “**Análise do tratamento sazonal nos métodos de previsão de vendas baseados em aprendizagem de máquina**”, que apresenta o impacto da aplicação de tratamentos sazonais para tornar as séries de dados dessazonalizadas, buscando a geração de estimativas mais precisas.

O capítulo III contém o segundo artigo, “**Seleção de variáveis nos métodos de previsão de vendas baseados em aprendizagem de máquina**”, que examina o

impacto, em termos de precisão e custo computacional, do uso de métodos de seleção de variáveis sobre modelos de previsão de vendas baseados em aprendizagem de máquina, sobretudo face ao crescimento exponencial de dados e atributos disponíveis.

O capítulo IV contempla o terceiro e último artigo, “**Métodos de combinação de previsão: uma análise comparativa para o varejo de moda**”, com a apresentação dos resultados em termos de acurácia e do teste MCS de modelos de previsão lineares e não lineares, cujos representantes da última categoria são os métodos de inteligência computacional, e de métodos *ensembles*, sendo considerados também as possíveis combinações, tanto de tratamento sazonal como de seleção de variáveis. Na literatura, os modelos *ensembles* são agrupados tanto como métodos de inteligência computacional como métodos híbridos, sendo os últimos mais amplamente aplicados nos últimos anos, o que vai ao encontro do crescimento do volume de pesquisa sobre o tema em detrimento do uso de métodos de previsão puros. Adicionalmente, nesse mesmo capítulo, é apresentada uma análise AHP, cujo objetivo é considerar outros critérios, além da acurácia, para recomendar o melhor modelo para o varejo de moda.

Por fim, o capítulo V, **Considerações finais**, evidencia as principais conclusões, as limitações encontradas e as recomendações para trabalhos futuros.

1.4. Resultados alcançados

No que tange a resultados, o capítulo introdutório evidenciou o ambiente em que esta Tese está situada. Especificamente, apresentou o setor e os métodos de previsão utilizados, detalhou as técnicas de previsão empregadas, tanto lineares como não lineares, assim como os tratamentos de pré-processamento aplicados nas séries de dados contempladas. Ademais, indicou a metodologia usada no trabalho.

No segundo capítulo, verificou-se o impacto do uso do tratamento sazonal com base em dois tipos de tratamentos: por diferenciação semanal e por ajuste da curva através de um polinômio de terceiro grau. Tendo em vista os resultados obtidos, o tratamento sazonal não é recomendado para modelos de árvores de decisão, pois não trouxe ganhos de desempenho em termos de precisão, devendo ser aplicado somente nos modelos baseados em redes neurais, especificamente para o algoritmo CNN e

LSTM. Apesar da indicação do uso do tratamento sazonal para parte dos algoritmos construídos com algoritmos de redes neurais, é fundamental ressaltar, mediante os resultados do teste MCS, que apenas os modelos baseados em árvores de decisão apresentaram maior capacidade preditiva.

No terceiro capítulo foi analisado o impacto da aplicação de três técnicas de seleção de variáveis em termos de precisão, as quais são: filtragem do p-valor das variáveis RFA e SHAP sobre o conjunto das 128 variáveis disponíveis, entre variáveis exógenas e variáveis criadas. Adicionalmente, foi feita a avaliação do custo computacional, devido ao uso de tratamentos de seleção de variáveis, e foram observadas as variáveis mais frequentes nos modelos de previsão propostos. Dessa forma, considerando o desempenho médio em termos de acurácia para as quatro categorias de produtos, constatou-se que o uso das técnicas de seleção de variáveis foi capaz de produzir estimativas mais precisas, em torno de 13%. A técnica de seleção de variáveis com melhor performance foi a RFA.

Em relação ao custo computacional, mensurado pelo tempo de treinamento dos modelos de previsão em segundos, as técnicas de seleção de variáveis, sobretudo a RFA, em vez de reduzir o custo computacional, aumentaram significativamente o tempo de processamento, o que foi observado em todas as categorias de produtos.

Sobre a análise de frequência das variáveis mais presentes nos modelos de previsão de vendas propostos, considerando o emprego das diversas técnicas de seleção de variáveis, verificou-se, para todas as séries, uma alta frequência das variáveis criadas em detrimento das variáveis exógenas, tendo em vista o subconjunto de 30 variáveis selecionadas. Na categoria de variáveis exógenas, aquelas que apareceram um maior número de vezes foram: temperatura máxima e binária de Natal. Já no grupo de variáveis criadas vale salientar a presença mais frequente das seguintes variáveis: a primeira diferença defasada, a média móvel das quatro últimas semanas, a defasagem 3, a defasagem 4, a defasagem 22 e a defasagem 49.

Em suma, para a previsão de vendas no varejo de moda, lançar mão de um conjunto menor de variáveis, sendo que as exógenas não foram relevantes, foi capaz de gerar modelos de previsão baseados em aprendizagem de máquina com maiores níveis de previsão, porém com maior custo computacional. Assim, recomenda-se o uso de técnicas de seleção de variáveis para a previsão de vendas do setor em foco.

No quarto capítulo, buscou-se efetuar duas recomendações: qual o modelo de previsão de vendas é mais preciso, diante dos 79 modelos propostos, e qual o melhor modelo para o varejo de moda, a partir da análise multicritério AHP de Saaty (1987).

Sobre a recomendação de modelo mais preciso para a geração das previsões de vendas do setor, foram obtidos resultados distintos para cada métrica de precisão e para cada categoria de produto. Apesar disto, tendo em vista a necessidade de fazer uma sugestão final de um único modelo de previsão de vendas para o varejo de moda, a sugestão seria a escolha do modelo XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA. Tal escolha é justificado pelo fato dessa especificação de modelo de previsão aparecer em 37,5% dos casos como modelo de previsão com maior acurácia, considerando o espaço de busca analisado.

De forma geral, os modelos de previsão apresentaram bom nível de acurácia, já que o resultado médio alcançado, com base nos níveis de precisão dos modelos mais assertivos para cada categoria de produto, foi de 12,76% de RMSLE e de 10,07% de MAPE. As estimativas geradas variaram cerca de 11,70% do valor real.

De forma unânime, o melhor modelo para previsão de vendas no varejo de moda com base no resultado da metodologia AHP, para os quatro tipos de categoria de produtos analisados, é o *Naive* ajustado, um modelo *random walk*. Tal fato reforça a excelência dos modelos *Naives* e que não é simples superá-los. Ademais, salienta-se que esse resultado foi alcançado devido ao modelo de previsão *Naive* ter um desempenho de médio a superior na maior parte dos critérios. Caso os modelos de previsão *Naives* fossem eliminados como opção, a sugestão de modelo final para realização da previsão de vendas no setor de varejo de moda, seria o modelo *ensemble* de média simples, contemplando toda a diversidade de modelos propostos.

2. Introdução à previsão de vendas no setor de varejo de moda

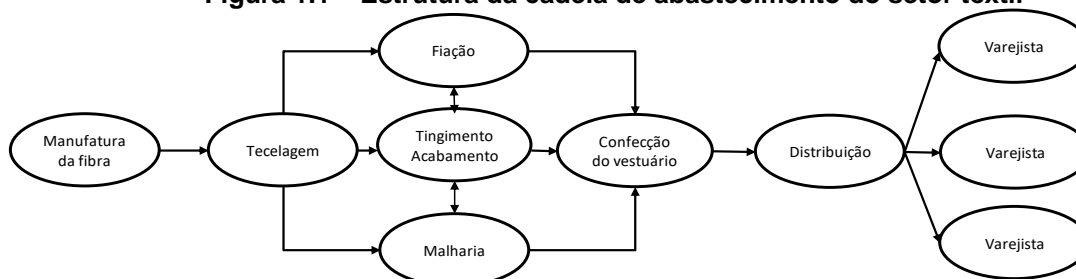
Esta parte do trabalho evidencia os seguintes aspectos: (1) apresenta e descreve o setor em que a Tese foi desenvolvida; (2) expõe resumidamente a empresa que concedeu os dados para execução do estudo; (3) descreve o processo de previsão de vendas para o varejo de moda, considerando suas peculiaridades.

2.1. O setor de varejo de moda no Brasil

O primeiro aspecto é a complexidade da indústria têxtil, que envolve uma série de etapas para o produto estar disponível para a compra pelo consumidor final. Na Figura 1.1, é possível verificar como se dá a interligação entre as partes para a produção de uma peça de moda até ela chegar à ponta final, que é o varejo.

Thomassey (2010) destaca que, devido ao ambiente competitivo, empresas e, principalmente, distribuidores de artigos de moda têm que reduzir seus custos de produção para se manterem eficientes. Dessa forma, a maioria dos processos de fabricação é feita em países distantes e de baixo custo na Ásia ou América do Sul.

Figura 1.1 – Estrutura da cadeia de abastecimento do setor têxtil



Fonte: Adaptado de CHOI *et al.*, 2014, p. 10.

A globalização tem como resultado o aumento do tempo de espera da entrega dos pedidos e o aumento do tamanho de cada lote de suprimentos. Os distribuidores precisam ter uma gestão rigorosa da cadeia de suprimentos para evitar atrasos e indisponibilidade de estoque. Na última década, novas ferramentas de gerenciamento da cadeia de suprimentos permitiram que os distribuidores melhorassem a programação e a sincronização dos fluxos de materiais e informações. A eficiência dessas ferramentas depende, sobretudo, da precisão de suas previsões de vendas.

No âmbito nacional, a cadeia de abastecimento têxtil e de moda opera de forma tradicional, caracterizando-se da seguinte forma: apresenta pouca coordenação entre os vários elos da cadeia; possui prazos de entrega extensos; demanda estoques volumosos com risco de obsolescência; possui elevado número de SKUs (*stock keeping unit*) em função de tamanhos, cores e detalhes; tem produtos impactados pela sazonalidade e pelo aspecto de moda de forma diferentes entre si, e geralmente, com ciclo de vida curto.

Diante desse contexto, pode-se afirmar que a cadeia de suprimentos do setor de varejo de vestuário não opera de modo eficiente, apresentando um estágio de desenvolvimento inferior ao da cadeia de abastecimento do varejo alimentício, composta por redes de hipermercados e supermercados. Entretanto, parte das empresas estão progredindo e intensificando esforços para construir uma cadeia têxtil mais ágil e enxuta, o que produziria uma operação mais eficiente e com maior integração entre as partes, gerando redução de custos, sobretudo através da redução dos níveis de estoques necessários, e atendendo melhor às necessidades do consumidor. Em síntese, essa visão busca a constituição de uma cadeia de abastecimento integrada e flexível, que seja capaz de colocar o produto certo no local certo, respondendo assim às mudanças e necessidades do seu consumidor final.

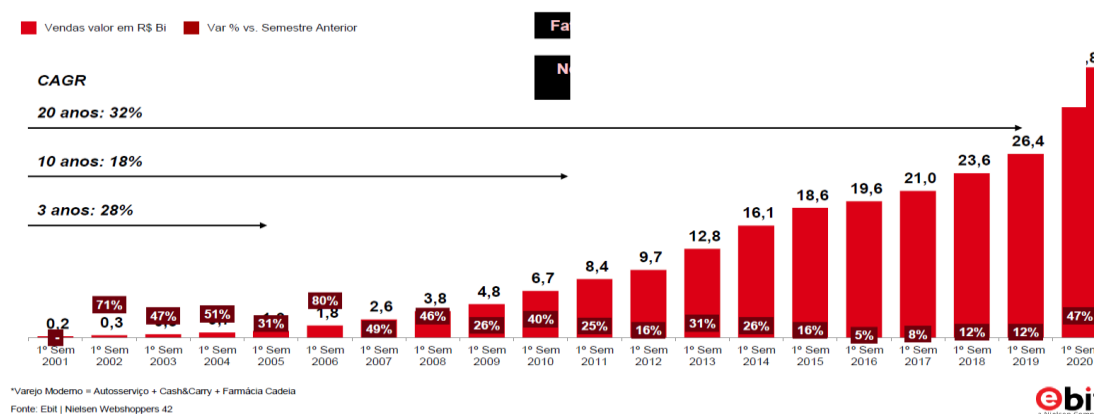
Após evidenciada a cadeia de produção do setor de moda, é importante compreender de forma mais detalhada o varejo em si, que é o foco deste trabalho.

Para Berman e Evans (1998), o varejo se distingue dos outros negócios, devido a três características: (1) as vendas são pequenas dado o foco no consumidor final; (2) clientes costumam comprar por impulso, sem planejar; e (3) geralmente, as pessoas necessitam frequentar pontos de vendas para adquirirem produtos/serviços.

Em relação à última característica, é fundamental destacar uma mudança que vem ocorrendo nos hábitos de consumo dos clientes de forma global: o aumento da participação das vendas do *e-commerce*, que elimina a necessidade da presença física nos pontos de venda, trazendo comodidade ao cliente. No Brasil, o *e-commerce* representou 5% do volume de vendas total e a categoria de vestuário teve participação de 6% do faturamento total das vendas on-line (RENNER, 2018). Na Figura 1.2 é possível constatar a evolução do faturamento do *e-commerce* no Brasil e a variação do crescimento em relação ao período anterior. Verifica-se a mudança do hábito do consumidor, que tem aumentado seu consumo on-line. Certamente, um dos fatores que intensificou essa mudança foi a pandemia da Covid-19, que teve início em 2020.

O CAGR presente na Figura 1.2 é a taxa de crescimento anual composta.

Figura 1.2 – Evolução do faturamento do e-commerce no Brasil



Fonte: Ebit (2020)

Para Eduardo Terra, presidente da SBVC, a Covid-19 teve impacto negativo no varejo por ter fechado a maioria das lojas por um longo período, com exceção do comércio essencial, que corresponde a 37% do total de lojas e inclui supermercados, farmácias e pet shops (SBVC, 2020b). Por outro lado, ainda segundo Terra, a pandemia trouxe como aspecto positivo a aceleração da digitalização. Por necessidade, as empresas mudaram mais rápido do que o usual e a transformação digital que levaria, pelo menos, cinco anos para ocorrer, foi realizada em cinco meses.

Apesar de ser visto como um intermediário, o varejo possui papel fundamental na cadeia de valor. Parente (2010) ressalta que, ao manter contato direto com o consumidor final, o varejo é responsável por identificar suas necessidades e desejos, definir o que será produzido e oferecer experiências que fortaleçam a sua marca.

O consumo das famílias é a principal referência do volume que o varejo movimenta no país. Em 2019, representava cerca de 65% do PIB brasileiro, como pode ser observado no relatório do papel do varejo na economia (SBVC, 2020a).

O varejo pode ser subdividido entre varejo restrito e varejo ampliado – esse último incorpora veículos e material de construção ao varejo restrito. O varejo restrito teve um impacto de 19,24% no PIB, em 2019, sendo medido pela simples divisão entre faturamento obtido pelo varejo em relação ao valor do PIB em real no mesmo período. Já o varejo ampliado representou 26,20% do PIB no mesmo ano. Em termos de valor agregado, em que é considerada uma margem de 40%, o valor agregado do varejo no PIB correspondeu a 11,5%, segundo dados da SBVC (2020a).

O varejo de moda está inserido na categoria de comércio varejista e na subcategoria de tecidos, artigos de armarinho, modas e calçados, segundo classificação do Instituto Brasileiro de Geografia e Estatística (IBGE).

O faturamento do comércio varejista de vestuário e complementos foi de R\$124,5 bilhões, em 2018, o que representa 6,8% do total apurado pelo comércio varejista, de acordo com a Pesquisa Anual do Comércio (PAC), realizada pelo IBGE. Segundo a pesquisa, o varejo de moda brasileiro tem baixa concentração, com um total de 163.000 empresas. As cadeias de lojas mais relevantes são Renner, C&A, Riachuelo, Pernambucanas e Marisa, com cerca de 16% do faturamento total do setor, o que reflete um setor fragmentado e com grande oportunidade de consolidação, sobretudo após a pandemia da Covid-19. Em 2019, o varejo de moda faturou R\$ 229,5 bilhões, consoante IEMI em Mercado e Consumo (2019).

A partir do estudo IEMI (2018), aproximadamente 70% do volume produzido é consumido pelas classes B e C, que representam 62% da população brasileira. Já a classe A (5% da população) responde por 18% do total. As 150 maiores cidades do país consomem 61% do total. No que diz respeito à distribuição de gastos, o brasileiro despende cerca de 5% da sua renda com itens de moda, segundo Pesquisa de Orçamentos Familiares (POF) 2017-2018 e, aproximadamente, 59,2% do consumo nacional de moda corresponde às classes C, D e E.

Além de extremamente pulverizado, o varejo de moda é muito segmentado, sendo que cada empresa busca atuar em um determinado nicho de mercado, focado em um público-alvo específico, de forma a identificar e atender melhor às necessidades de seus consumidores, garantindo maior aderência e competitividade.

2.2. O Lojão do Brás

Lojão do Brás é a empresa que concedeu os dados para a realização do estudo. A companhia iniciou suas atividades em 21 de outubro de 1995, depois uma bem-sucedida experiência de seus proprietários no setor têxtil com a marca Lojas Iracema, que chegou a ter seis unidades. A primeira e maior loja da rede está localizada no Largo da Concórdia, no Brás, o maior centro comercial de compras da América Latina.

A empresa sempre buscou oferecer qualidade, variedade de produtos e um conjunto de serviços para uma melhor experiência de compra para o consumidor. Ao longo do tempo, tem procurado se reinventar para se manter competitivo no mercado. Atualmente, o Lojão do Brás é uma rede com 17 pontos de venda nas principais cidades do estado de São Paulo. Em 2019, faturamento de cerca de R\$ 340 milhões.

O público-alvo são famílias das classes C e D. A empresa oferta grande variedade de produtos agrupados em departamentos, subdepartamentos e categorias. Os departamentos são: moda, cama, mesa e banho, moda íntima, moda praia, acessórios, beleza, utilidade doméstica e o departamento diversos. Cerca de 70% das suas vendas são originadas no departamento de moda, subdividido em moda adulto feminina, adulto masculina, infantil feminino e infantil masculino. Os principais concorrentes da rede são Torra-Torra, Caedu, Eskala e Marisa.

2.3. Importância da previsão de vendas

Em um ambiente de negócios cada vez mais competitivo e em constante mudança, a acurácia do processo de previsão de vendas é elemento decisivo nos resultados de uma organização, sobretudo para o varejo de moda.

Lee, Padmanabhan e Whang (1997) associam o aumento da competitividade na indústria de varejo de moda, incluindo a crescente implementação de cadeias globais de suprimentos para o atendimento das demandas das redes varejistas têxteis, à relevância da previsão de vendas para as empresas envolvidas. Uma previsão mais assertiva possibilita o desenvolvimento de uma cadeia de suprimento mais integrada e eficiente, gerando menores níveis de estoques, menos remarcações, menores rupturas, maiores margens e um efeito chicote² menor. Os mesmos apontamentos são feitos por Thomassey (2010) ao dizer que a implantação de técnicas avançadas de previsão de vendas gera os seguintes benefícios: redução do efeito chicote; incremento da eficiência de produção por parte dos fornecedores da cadeia de produção; redução da ruptura, que implica em minimização da perda de vendas, redução do nível das remarcações de preços e aumento da margem de lucro.

Makridakis, Wheelwright e Hyndman (1998) destacam três pontos que levaram ao crescimento da importância do processo de previsão, independente do setor em foco. Primeiro, o aumento da complexidade das corporações e do ecossistema em que elas estão inseridas, tornando extremamente difícil averiguar o impacto dos

² Efeito chicote é definido como um fenômeno que ocorre quando as ordens de compra para os fornecedores tendem a apresentar uma variância maior do que as vendas do elo mais próximo do consumidor, mais especificamente dos varejistas, a partir de Lee, Padmanabhan e Whang (1997).

diversos fatores que as afetam. Segundo, o crescimento da sistematização do processo de tomada de decisão. Terceiro, o desenvolvimento contínuo de diversas técnicas de previsão e de suas respectivas aplicações, juntamente com a sua difusão em diversos níveis dentro das companhias, não ficando sua utilização restrita somente aos cargos superiores.

Dias (1999) afirma que a importância da previsão de demanda está diretamente ligada à questão da racionalização dos recursos envolvidos e a meta previamente estabelecida do nível de atendimento aos clientes. Qualquer atividade de planejamento tem como requisito a realização de previsões sobre o futuro, as quais serão mais assertivas dependendo das hipóteses utilizadas. Assim, o processo de previsão pode ser definido como o conjunto de todas as atividades necessárias para a produção dos números finais da previsão, o que inclui desde a etapa de coleta de dados até o último ajuste efetuado na previsão em si.

A previsão de vendas é crucial para o varejo de moda, no qual a demanda do produto é muito volátil e o ciclo de vida do produto é extremamente curto. Nesse sentido, o planejamento de estoque é parte fundamental das operações de varejo de moda. A gestão dos estoques de forma adequada, capaz de equilibrar a oferta e a demanda, pressupõe que seja realizada uma previsão precisa da demanda futura. Um requisito fundamental de todo processo de previsão de vendas é que os fatores que afetaram a demanda no passado e estão afetando o presente ainda continuam a ter influência no futuro, permitindo assim gerar suas estimativas.

Giustiniano, Nenni e Pirolo (2013) explicam que a previsão de demanda desempenha um papel crítico no planejamento da gestão de operações eficientes, especialmente no caso da indústria da moda, em que coexistem os seguintes aspectos: demanda incerta, falta de histórico de dados e presença de tendências sazonais. Previsões ruins geram os seguintes efeitos: falta ou alto estoque, obsolescência, baixo nível de serviço, pedidos urgentes, uso ineficiente de recursos e grande efeito chicote se propagando a montante da cadeia de suprimentos.

Observando o setor varejista brasileiro de vestuário, Garcia *et al.* (2006) e Werner (2005) pontuam que ele se tornou muito competitivo, sobretudo, pela chegada de grandes *players* internacionais. Assim, técnicas de gestão da produção e de estoques têm sido aplicadas para assegurar a qualidade da operação e evitar perdas com estoques, logística e compras. É imprescindível manter os níveis de estoques

dimensionados adequadamente, já que estoques elevados geram altos custos de operacionalização e de oportunidade de capital parados. Sob outra perspectiva, manter baixos níveis de estoques pode levar à perda de economias de escala, custos altos de falta de produtos e baixo nível de serviço prestado junto ao consumidor.

Alicerçada em dados históricos de vendas, tendências de mercado e outros fatores relacionados, a previsão de vendas tem como objetivo estimar as vendas futuras (DU; LEUNG; KWONG, 2015). Em contrapartida, no estudo de Lindberg e Zackrisson (1991) é possível constatar os quatro maiores problemas associados ao uso de previsões no apoio à tomada de decisões: (1) a incerteza sobre o futuro; (2) a definição do método em si para gerar a previsão; (3) a qualidade e confiabilidade do conjunto de dados de entrada do método; e (4) a realização de uma interpretação adequada sobre a previsão gerada e forma efetiva de seu uso para o processo de tomada de decisão.

A partir do exposto, afirma-se que o principal objetivo da previsão de demanda é examinar incertezas e riscos, tendo em mente que a incerteza não pode ser totalmente suprimida e que o futuro pode sempre mudar e ser diferente do previsto.

2.4. Previsão de vendas no varejo de moda

As empresas utilizam a previsão de vendas, geralmente, para determinar dois níveis de demanda: o atual e o futuro. Além disso, o processo de previsão de vendas afeta as áreas de gestão de estoques e de marketing das empresas varejistas e, prioritariamente, o nível de satisfação do consumidor.

O processo de previsão impacta significativamente uma organização, pois afeta diversos planejamentos dentro da empresa, entre eles: o planejamento de vendas em si; o planejamento da demanda e da cadeia de suprimento; o planejamento financeiro; o planejamento de alocação e distribuição; e o planejamento agregado, que está em um patamar superior ao planejamento detalhado até o último nível da demanda.

Parente (2000) afirma que as empresas varejistas desenvolvem atividades para garantir o abastecimento para seus clientes, depois que o *mix* de produtos adequado para cada loja é estabelecido. O processo de previsão de vendas é fundamental para que as atividades de compras e abastecimento possam ser efetuadas. Consequentemente, erros nas previsões de vendas podem causar faltas ou excessos

de alguns produtos, gerando problemas e impactos financeiros aos varejistas. Como apontam Nunnari e Nunnari (2017), previsões de vendas imprecisas no varejo provocam demandas de clientes insatisfeitos e ruptura e/ou excesso de estoques.

Liu *et al.* (2013) evidenciam que, em comparação com os segmentos varejistas, a previsão de vendas no varejo de moda é uma tarefa difícil de ser feita por uma série de peculiaridades do setor, fato também percebido por Thomassey (2010). O autor ressalta a complexidade do processo de produção e distribuição da indústria de moda, que contempla várias companhias, é subdividido em múltiplas etapas e grande parte está inserido nas cadeias de produção globais. Tal complexidade é intensificada e difícil de ser administrada, pois o atributo de moda tem se tornado cada vez mais efêmero e os hábitos de consumo da população mudam mais rapidamente.

Algumas peculiaridades dos produtos de moda que devem ser incorporados aos sistemas de previsão de vendas, de acordo com Thomassey (2010), são: a sazonalidade existente; muitas variáveis exógenas afetam o volume de vendas, tais como promoções, poder de compra dos consumidores, entre outras; as vendas dependem das tendências de moda, o que leva a ausência de histórico de informação para a maior parte das previsões; os itens possuem muitas variações, em termos de cor, tamanho e outros atributos, gerando uma complexidade maior para a geração da previsão, dependendo do nível em que a mesma será feita.

Thomassey (2010) destaca que algumas definições devem ser feitas, previamente, para que o processo de previsão seja realizado no setor de moda. Tais definições passam pela resposta aos seguintes questionamentos e impactam na seleção do método de previsão mais adequado. Elas estão descritas a seguir.

- **Horizonte de previsão.** Existem dois horizontes mais comuns: o de longo prazo, que considera um planejamento de, pelo menos, um ano, e o de curto prazo, que considera poucas semanas, e é utilizado, normalmente, para a realização de reposições. A presença desses dois horizontes demanda modelos de previsão de vendas distintos para cada um deles.
- **Ciclo de vida de produto.** A maior parte dos produtos de moda possui um ciclo de vida curto, apesar disso os produtos desse setor podem ser agrupados em três grandes tipologias: os itens básicos, que são vendidos ao longo de todo o ano; os itens de moda, que são comprados uma única vez, não possuem reposição e estão disponíveis durante um curto espaço

de tempo; e os itens mais vendidos, que são vendidos ao longo de todo o ano com pequenas modificações para fazer frente a novas tendências.

- **Nível de agregação dos produtos.** Aqui há desde o nível maior de agregação, que é a marca/empresa, até o menor nível, que é o SKU, que contempla a variação de cor e tamanho de um determinado artigo.
- **Sazonalidade.** Essa característica é crítica para o varejo de moda e pode ser verificada em diversas magnitudes, dependendo do produto, mas basicamente existem os produtos sazonais, vendidos somente em determinado período do ano, e os não sazonais, vendidos ao longo do ano.
- **Variáveis exógenas.** Um conjunto extenso de variáveis exógenas afeta a decisão de compra do consumidor, como, por exemplo: feriados, promoções, aspectos macroeconômicos, estratégia de marketing etc.

O impacto dessas variáveis exógenas é difícil de ser mensurado e, muitas vezes, elas estão altamente correlacionadas entre si, gerando multicolineariedade.

O aspecto do ciclo de vida de produto é essencial para o processo de previsão de vendas no setor de moda, pois a sua definição interfere no tipo de método que deve ser selecionado. Rink e Swan (1979) estabelecem que o ciclo de vida de um produto representa a curva de vendas unitárias de um produto, estendendo-se desde o momento em que é colocado no mercado até que seja removido. O ciclo de vida de produtos pode ser aproximado a uma curva do tipo sino e é dividido nos seguintes estágios: lançamento, ascensão, maturação e declínio (THOMASSEY, 2010). Outro tipo de divisão é proposto por Sen (2008), que classifica os produtos de moda em: *fashion*, sazonal e básico. Detalhadamente, os produtos *fashion* são os que têm, com maior intensidade, o viés de tendência de moda; os produtos sazonais são afetados diretamente pelo clima e vendidos em épocas determinadas do ano; e os produtos básicos são aqueles ofertados durante todo o ano. Szozda (2010) observa que o ciclo de vida dos produtos *fashion* difere de um produto convencional, pois possuem tanto um alto pico de introdução como um rápido declínio das vendas.

Liu *et al.* (2013) reforçam que, no varejo de moda, a demanda dos produtos é altamente volátil devido ao gosto do consumidor estar em constante mudança e pelos produtos terem um ciclo de vida curto, como também observado por Szozda (2010).

Dada a alta obsolescência dos produtos de moda, seu valor é reduzido progressivamente ou até mesmo perdido em um curto prazo de tempo e muitas vezes

não é recuperado (MEHRJOO, 2014). Assim, em sua maioria, os produtos de moda devem ser tratados com prazo de validade. Além disso, as vendas de produtos de moda são afetadas "estocasticamente" por diversos fatores, tais como: sazonalidade, tendências de moda, clima, estratégia de *marketing*, política, aspectos macroeconômicos etc. Em termos de volume de itens, geralmente, as organizações possuem muitos SKUs, devido variações de cor e tamanho com baixo histórico de dados, tendo em vista a grande quantidade de lançamentos feitas em cada coleção.

Há outra agravante: o fato de a cadeia de suprimentos da indústria de moda ser longa, já que inclui fábricas de algodão, de fibras e de roupas, além de distribuidores, atacadistas e varejistas. Lee, Padmanabhan e Whang (1997) dizem que cadeias desse tipo podem ter um amplo efeito chicote, o qual ocorre quando as ordens de compra para os fornecedores tendem a apresentar uma variância maior do que as vendas do elo mais próximo do consumidor, mais especificamente, dos varejistas.

O efeito chicote é uma distorção na percepção da demanda pelos participantes de uma cadeia de distribuição, sendo a intensidade dessa distorção maior para os fornecedores mais distantes dos consumidores finais, nos elos iniciais da rede de fornecimento. Liu *et al.* (2013) concluem que, uma vez que a previsão é fator crítico que afeta a magnitude do efeito chicote no setor, melhorar a previsão ajuda a minimizar o impacto, aumentando a eficiência da cadeia de abastecimento da moda.

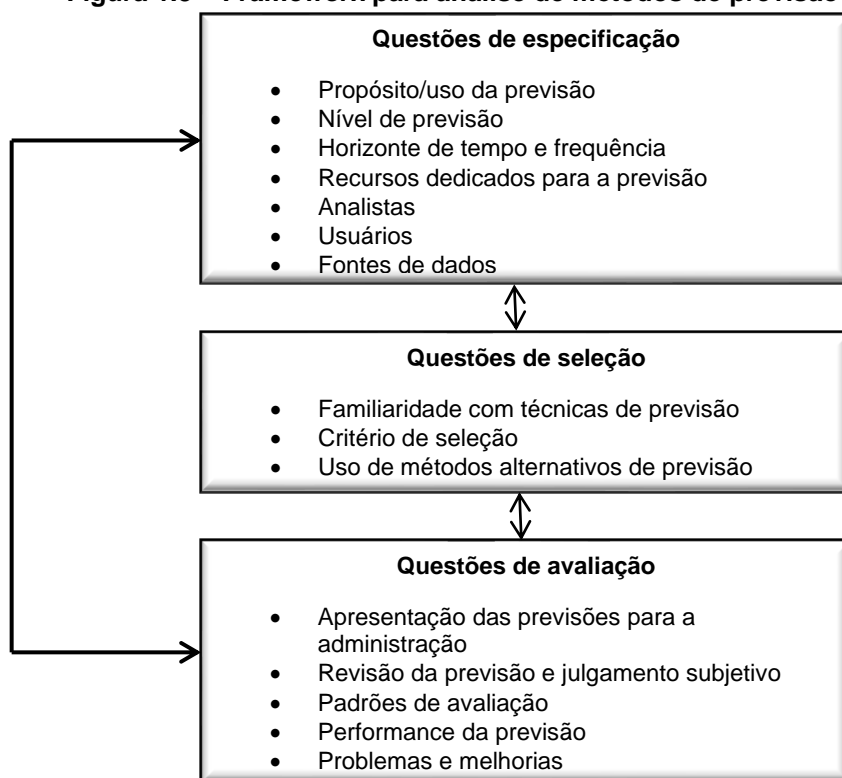
A complexidade do processo de previsão é inerente ao setor de moda. Ademais, é um mercado em constante transformação e com aumento crescente do seu nível de competição entre os participantes, o que contribui efetivamente para atrair um número cada vez maior de pesquisadores e interessados com o objetivo de compreender e ajudar as empresas do setor se tornarem mais eficientes.

Verificadas as especificidades intrínsecas ao setor da moda que impactam a previsão de vendas em si, deve-se apresentar outros aspectos que afetam qualquer processo de previsão. Nesse sentido, Winklhofer, Diamantopoulos e Witt (1996) desenvolveram uma estrutura com aspectos a serem levados em conta para que uma previsão seja feita. A estrutura também serve para classificar diversos estudos feitos na área. O *framework* apresentado na Figura 1.3 agrupa os fatores que afetam o processo de previsão em três blocos: questões de especificação, no qual são estabelecidos o objetivo, o tipo de previsão, os recursos envolvidos, as pessoas que

farão parte do projeto e as fontes de dados a serem utilizadas; questões de seleção, ponto em que são definidos os métodos a serem aplicados no projeto; e questões de avaliação, em que são determinadas as formas de apresentação, revisão e mensuração do desempenho das previsões desenvolvidas.

Além dos aspectos que devem ser contemplados e afetam o processo de previsão, Makridakis, Wheelwright e Hyndman (1998) indicam três requisitos mínimos para realização de uma previsão quantitativa: a informação sobre o passado deve estar disponível; a informação deve ser quantificável em termos de um padrão numérico; ter como pressuposto fundamental que ocorrerá a repetição de padrões observados em instantes anteriores no futuro. A última condição é denominada como suposição de continuidade e é uma premissa básica para utilização de todos os métodos quantitativos de previsão e de vários métodos qualitativos.

Figura 1.3 – Framework para análise de métodos de previsão



Fonte: Winklhofer, Diamantopoulos e Witt, 1996, p. 204.

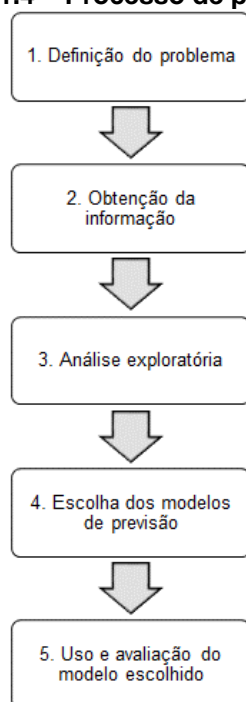
Em relação à metodologia para aplicar um sistema de previsão de vendas, Pellegrini e Fogliatto (2000) instruem que a implantação de um sistema de previsão de demanda depende de quatro etapas sequenciais: (1) identificação e definição do problema de previsão, (2) aplicação dos métodos de previsão, (3) definição dos

procedimentos para a escolha dos métodos adequados para o objeto a ser previsto, e (4) suporte organizacional para que os métodos selecionados sejam implantados.

Hyndman e Athanasopoulos (2013) definem de forma um pouco distinta os passos essenciais para executar tarefa de previsão: (1) definição do problema, em que é determinado o escopo do que será previsto e dos envolvidos, (2) obtenção da informação necessária, a qual pode ser classificada como estatística ou qualitativa, que se refere a experiência acumulada dos envolvidos na tarefa de previsão em executá-la e em conseguir os dados básicos para a mesma, (3) realização de uma análise exploratória, com o objetivo de fazer uma investigação inicial dos dados, (4) escolha dos métodos que serão aplicados para previsão em si e realização dos seus respectivos ajustes, sendo recomendado a comparação com, pelo menos, dois ou três métodos, e (5) uso e avaliação dos modelos de previsão por meio da acurácia.

A definição dos métodos de previsão a serem implementados depende da disponibilidade dos dados, da força da relação entre as variáveis exploratórias, da variável a ser prevista e da forma como a previsão será utilizada. A Figura 1.4 traz um quadro-resumo com os passos para a realização do processo de previsão.

Figura 1.4 – Processo de previsão



Fonte: adaptado de Hyndman e Athanasopoulos, 2013.

Levando em conta as etapas para elaboração de uma previsão evidenciadas por Pellegrini e Fogliatto (2000) e por Hyndman e Athanasopoulos (2013), é usada a

metodologia proposta pelos últimos autores, uma vez que não é foco de discussão neste trabalho a forma como se dá a implantação do método de previsão escolhido.

Na primeira etapa, o objetivo foi realizar a previsão das vendas semanais em unidades para as categorias de produtos escolhidas da empresa Lojão do Brás, sendo eleitas as mais representativas para o faturamento da companhia. O horizonte da previsão foi de curto prazo, sempre a próxima semana.

Na segunda etapa, obtenção da informação, houve facilidade de acesso à base de dados da empresa e para a captura da série de dados das variáveis exógenas contempladas neste trabalho, as quais são de domínio público. O detalhamento dos dados está evidenciado na parte da metodologia.

Na terceira etapa, aplicação de uma análise exploratória inicial, também está na parte de metodologia. Desde já, é fundamental destacar a importância do aspecto sazonal e da variabilidade existente nos dados, o que ficará mais claro posteriormente.

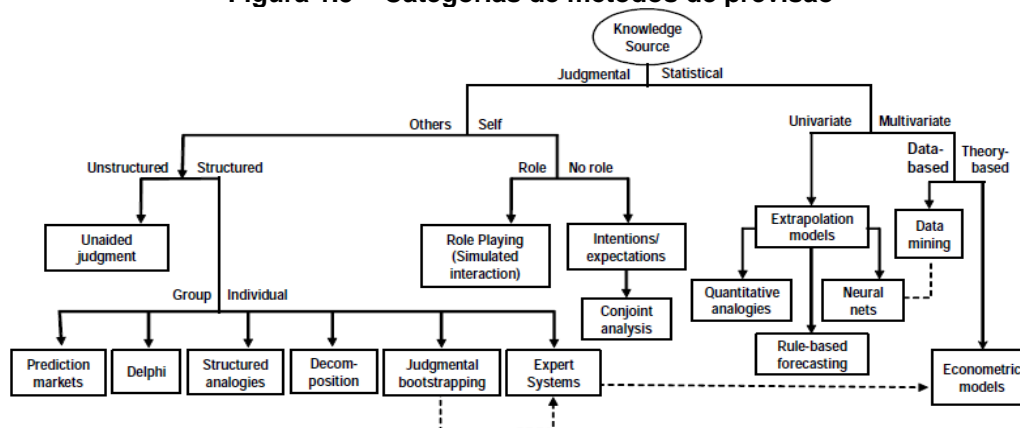
Na quarta etapa ocorre a definição dos modelos de previsão a serem adotados. Antes da apresentação dos métodos selecionados e do seu detalhamento teórico na próxima seção deste trabalho, é importante evidenciar as possibilidades existentes.

Existe uma diversidade de métodos disponíveis que podem ser agrupados por distintas abordagens. Makridakis, Wheelwright e Hyndman (1998) destacam que a escolha do método de previsão é afetada pelo horizonte de tempo da previsão, pelos fatores que afetam a variável a ser prevista, pelo tipo de padrão de dado existente e por outros aspectos. Como já colocado, existe uma série de outros pontos que afetam a previsão de vendas para a moda, os quais devem ser levados em consideração.

Uma primeira abordagem de agrupamento de métodos de previsão é a feita por Armstrong (2008). Para ele, as técnicas devem ser subdivididas, primeiramente, em técnicas de julgamento, que contempla técnicas com ênfase qualitativa ou subjetiva, e em técnicas estatísticas, com ênfase quantitativa. A principal diferença desses agrupamentos é que o primeiro é orientado pela opinião de um indivíduo ou de um grupo de indivíduos e, o segundo, pelos dados históricos. O autor defende que a combinação dos dois tipos de técnicas pode levar a previsões mais precisas.

Considerando a categorização de Armstrong (2008), mostrada na Figura 1.5, o foco deste trabalho é a utilização das técnicas estatísticas, ou seja, são utilizados dados históricos e a suposição da continuidade para que a previsão seja feita.

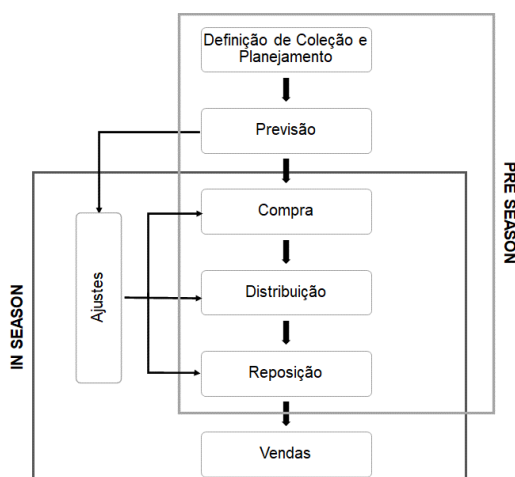
Figura 1.5 – Categorias de métodos de previsão



Fonte: Armstrong, 2008, p. 2.

O planejamento da cadeia de suprimentos para a indústria da moda é dividido em duas fases, geralmente: (1) o planejamento feito antes da temporada de vendas, denominado *pre season*, construído com base em dados históricos de venda, em que é estabelecida a previsão total da demanda e definido os totais a serem comprados para uma dada temporada de vendas, e (2) fase de acompanhamento, denominada como *in season*, que acontece durante a temporada de venda, onde as previsões feitas na etapa *pre season* são revisadas e atualizadas em conformidade com as vendas realizadas (THOMASSEY, 2014). A abordagem em duas fases também é compartilhada por Iannone *et al.* (2015) e pode ser visualizada na Figura 1.6.

Figura 1.6 – Etapas do modelo de planejamento e operações do varejo de moda



Fonte: Iannone *et al.*, 2015, p. 1291.

O planejamento *pre season* ocorre, normalmente, seis meses antes de uma determinada temporada de vendas e deve ser visto como uma previsão de vendas de médio e longo prazo, que ajuda a definir o planejamento comercial e operacional no que diz respeito à distribuição dos itens pelas lojas. Já a fase *in season* tem início

entre a segunda e a quarta semana pós-temporada de vendas e é quando ocorre uma recalibragem do plano feito na fase *pre season* para ajustar possíveis superestimativas ou subestimativas no planejamento inicial (IANNONE *et al.*, 2013).

Definição a ser evidenciada é o de temporada, período no qual um conjunto de produtos de moda é comercializado (HINES; BRUCE, 2007). Esse conceito se confunde, parcialmente, com sazonalidade, mas é importante ter em mente que um produto pode ser sazonal ou não e estar presente em várias temporadas de vendas.

Em termos de abordagem de agrupamento de técnicas de previsão para varejo de moda, Thomassey (2010) divide as técnicas em: estatísticas, usuais e avançadas.

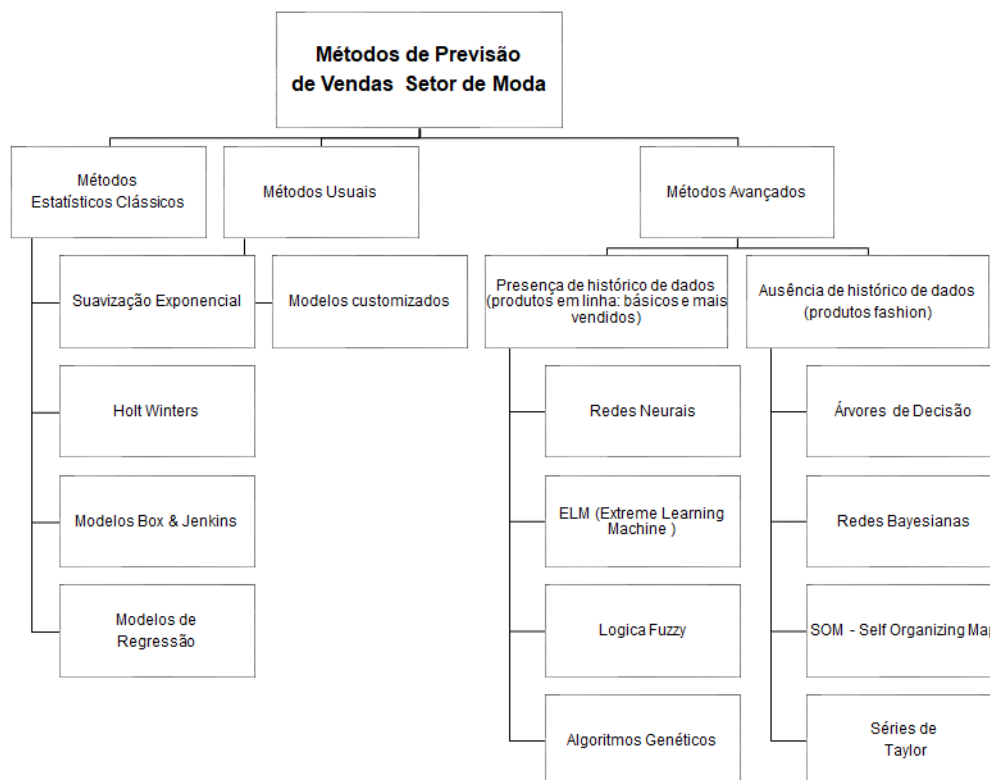
As técnicas estatísticas tradicionais, sendo a denominação mais apropriada técnicas econométricas, compreende os métodos de previsão baseados em suavização exponencial, em modelos Box e Jenkins e em regressões lineares.

As técnicas usuais são métodos customizados pelas empresas, devido, principalmente, à complexidade e ao custo computacional presentes nos métodos estatísticos e nos métodos avançados com diferentes intensidades entre eles. Assim, o que temos são métodos heurísticos embasados na realidade de cada empresa.

As técnicas avançadas são métodos de previsão baseados, sobretudo, em inteligência computacional. A maior parte dessas técnicas lida com a não linearidade e pode ser aplicada por meio de métodos híbridos, isto é, emprega a aplicação de mais de um método de previsão para elaborar a previsão em si.

A denominação de “métodos avançados” não é considerada ideal, pois o que é avançado depende do contexto e do momento visto. Certamente, a nomenclatura de métodos não lineares seria mais adequada para esse agrupamento de métodos. Contudo, será utilizada essa nomenclatura, tendo em vista a referência bibliográfica adotada de Thomassey (2010). Na Figura 1.7, pode-se visualizar um quadro ilustrativo com a abordagem recomendada pelo autor.

Figura 1.7 – Categorias de métodos de previsão setor de moda



Fonte: elaboração própria.

Dentro de cada um desses agrupamentos principais, há subcategorias com as variações de cada uma das técnicas de previsão, os quais são explicitados a seguir.

- **Métodos tradicionais:** incluem os métodos de séries temporais como ARIMA, Holt Winters e os modelos de regressão.

Entretanto, esses métodos não são implementados de forma eficiente no varejo de moda, por demandarem um grande volume de dados e a experiência dos operadores na definição da configuração do modelo mais adequado.

- **Métodos usuais:** considerando o custo e a complexidade de alguns métodos tradicionais, muitas empresas desenvolvem modelos customizados tendo como base as vendas do último ano, realizando alguns ajustes em termos de sazonalidade e da inclusão de variáveis exploratórias que impactam a decisão de compra e o fluxo de visitantes nas lojas. Little (1998) destaca as seguintes variáveis exploratórias: dados de temperatura, nível de concorrência, feriados, promoções e propagandas realizadas pelos varejistas, capilaridade de distribuição das redes de lojas, variedade em termos de cor e tamanho dos diversos itens comercializados e fatores macroeconômicos.

Na implementação desses métodos, a dificuldade reside na definição de variáveis exploratórias que deveriam ser empregadas e os resultados dependem, novamente, da experiência do operador, tornando o processo de previsão ineficiente.

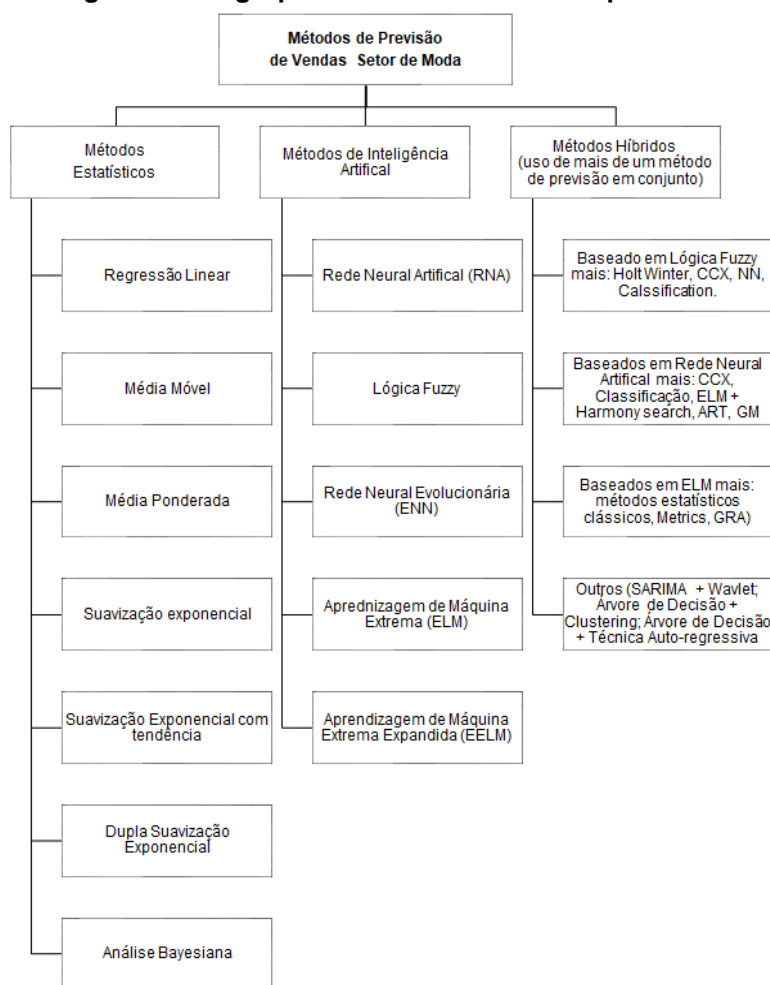
- **Métodos avançados de previsão de vendas:** esse agrupamento deve ser subdividido em dois grupos, tendo como base a disponibilidade de dados. Na presença de histórico, a rede neural é a técnica mais usada para previsões de curto prazo. Outras técnicas são: a *extreme learning machine* (ELM), que em alguns estudos tem resultados superiores aos das redes neurais, e *fuzzy inference systems* (FIS), modelos com lógica *fuzzy*, que lidam bem com a inclusão dos impactos das variáveis exploratórias.

Nos estudos apresentados, as técnicas mais avançadas de predição apresentaram previsões mais precisas do que os modelos estatísticos clássicos.

Na ausência de dados histórico – o que geralmente ocorre com novos produtos –, primeiramente, o novo produto é clusterizado em grupos de produtos previamente criados levando em conta suas características e, posteriormente, é feito um reajuste da previsão após as primeiras semanas de vendas. Nesse caso, são aplicados os seguintes modelos: redes neurais, redes Bayesianas e árvores de decisão, que lidam melhor com a não linearidade e com relacionamentos mais complexos.

Última abordagem é a de Liu *et al.* (2013), mais semelhante à de Thomassey (2010) e que pode ser vista na Figura 1.8. Os autores agrupam as diversas técnicas em: métodos estatísticos, métodos baseados em inteligência artificial e métodos híbridos. Apesar da simplicidade da implementação e velocidade dos métodos estatísticos, eles apresentam algumas deficiências que fazem com que os métodos embasados em inteligência artificial ou os métodos híbridos sejam os preferidos. As principais limitações dos métodos estatísticos para os autores são: necessidade do conhecimento de um especialista para definição do modelo, desempenho em termos de acurácia inferior em relação a outras técnicas mais avançadas e incapacidade de lidar com os diversos fatores que afetam a previsão de itens de moda.

Figura 1.8 – Agrupamentos de métodos de previsão do setor de moda



Fonte: elaboração própria.

Os métodos híbridos são preferidos, muitas vezes, em detrimento dos métodos de inteligência artificial puros por conseguirem aliar as principais forças de mais de um método de previsão, o que em alguns estudos gerou resultados mais assertivos.

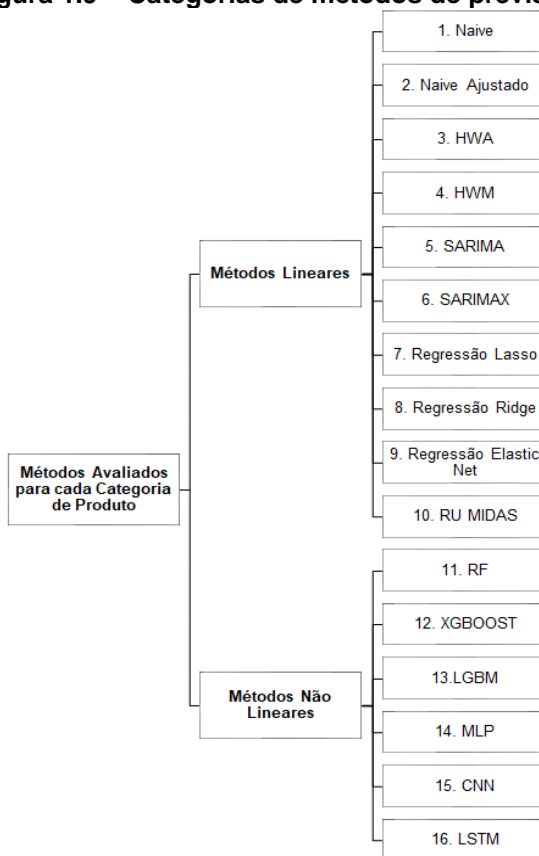
Neste trabalho são abordados tanto métodos estatísticos clássicos, por meio dos modelos de suavização exponencial de regressão e dos modelos Box-Jenkins, como métodos baseados em inteligência artificial especificamente, por meio dos modelos baseados em árvores de decisão e em redes neurais. Ademais, no último capítulo, traz os métodos *ensembles*, que podem ser categorizados tanto como métodos de inteligência artificial como métodos híbridos, dependendo do tipo de *ensemble*.

Estudo realizado por Batista (2017), que fez um levantamento do referencial teórico produzido sobre previsão de vendas no varejo de vestuário, constatou que cerca de 48% dos artigos analisados tiveram como objetivo evidenciar a utilização dos

métodos híbridos. Tal fato, mais uma vez, reforça a importância dos mesmos.

Em termos de categorização, neste trabalho as técnicas estão agrupadas em duas grandes categorias: métodos lineares, em que são implementadas técnicas estatísticas, e métodos não lineares, no qual são implantadas técnicas baseadas em inteligência computacional. Especificamente, os métodos lineares são: *Naive*, *Naive* ajustado, suavização exponencial com sazonalidade tratada de forma aditiva (HWA), suavização exponencial com sazonalidade tratada de forma multiplicativa (HWM), SARIMA, SARIMAX, que contempla variáveis exógenas, regressão Lasso, regressão Ridge, regressão Rede elástica e RU-MIDAS. Já os métodos não lineares são: RF, LGBM, XGBOOST, MLP, CNN e LSTM. A Figura 1.9 mostra as técnicas aplicadas.

Figura 1.9 – Categorias de métodos de previsão



Fonte: elaboração própria.

Os métodos RF, LGBM e XGBOOST são baseados em árvores de decisão e apresentam como diferenças entre eles a forma como a reamostragem dos dados é feita e a forma de crescimento das árvores de decisão. Por outro lado, os métodos MLP, CNN e LSTM são baseados em redes neurais, sendo que MLP é o mais antigo, o LSTM é uma evolução do MLP, um tipo de rede neural recorrente, e o CNN é uma

rede neural com arquitetura espacial completamente diversa do MLP e do LSTM.

Interessante destacar que nem todos os métodos conseguem contemplar as variáveis exógenas e, além disso, no caso dos métodos baseados em inteligência computacional, foi aplicada a variação de tratamento de sazonalidade. Esse aspecto está detalhado na seção de metodologia.

A última fase de um sistema de previsão é a avaliação e o uso em si dos métodos envolvidos. Para a realização da avaliação de desempenho das diversas técnicas de previsão, Passari (2003) coloca que podemos analisar os modelos através de vários critérios, tais como: acurácia, custo, habilidade do analista, características desejadas de previsão, características específicas do problema, facilidade do uso, requerimento de dados, disponibilidade de *software*, velocidade, facilidade de interpretação, eficiência, viés e a capacidade de incluir a experiência do gestor.

Conforme o autor, não há consenso na literatura sobre o critério ou o conjunto de critérios mais relevantes para selecionar uma técnica de previsão analisada em detrimento das demais. Mesmo assim, o mais utilizado é acurácia. Neste trabalho, o principal critério é a precisão, medida através das seguintes métricas: MAE, RMSE, RMSLE e MAPE. Ademais, foi aplicada uma análise AHP a fim de ponderar e considerar outros critérios, como: tempo de previsão, facilidade de uso, poder interpretativo e requisição de dados.

3. Introdução aos métodos de previsão

Esta parte do trabalho traz a revisão dos principais conceitos e modelos teóricos quantitativos utilizados para a previsão de séries temporais, incluindo as descrições das características presentes nas séries temporais e dos métodos utilizados.

3.1. Características das séries temporais

Uma série temporal é qualquer conjunto de observações ordenadas no tempo, sendo essa ordenação cronológica que faz surgir o efeito de autocorrelação entre os dados da série (MORETTIN; TOLOI (2006)). Os dados podem ser coletados em intervalos regulares – de forma diária, semanal, mensal, trimestral, semestral ou anual

– ou em intervalos extremamente curtos de forma quase contínua, a exemplo das cotações das ações. Ao analisarem uma série temporal, Morettin e Toloí (2006) assinalam que os principais objetivos são:

- Investigar o mecanismo gerador da série temporal;
- Criar previsões de valores futuros da série de curto ou longo prazo;
- Descrever graficamente o comportamento da série verificando a existência de tendências, ciclos e variações sazonais, construção de histogramas e diagramas de dispersão etc.;
- Verificar periodicidades relevantes através de análise espectral;
- Modelagem do fenômeno em análise;
- Obtenção de conclusões em termos estatísticos;
- Avaliação da adequação do modelo em termos de previsão.

Em termos de caracterização, os mesmos autores colocam que as séries podem ser de ordem univariada, baseada somente em uma série histórica, ou multivariada, baseada em mais de uma série histórica. Em relação ao tipo de dado, a série pode ser contínua, onde existe um intervalo contínuo de dados, ou discreta, quando há um conjunto de dados registrado em tempos específicos.

É fundamental determinar os componentes básicos para identificar um padrão de comportamento da série, permitindo uma melhor compreensão e, assim, tornando possível a geração de previsões. De acordo com Gujarati (2006), os principais tipos de componentes que podem estar presentes dentro de uma série temporal são:

- **Tendência:** determina o comportamento de longo prazo da série, indicando se ela sobe, desce ou permanece estável. Deve-se destacar que a tendência pode ser linear ou apresentar formas polinomiais.
- **Ciclos:** padrões que se repetem ao longo da série em períodos superiores a um ano.
- **Sazonalidade:** padrões que se repetem ao longo da série dentro de um período de um ano.
- **Termo errático:** ocorrências irregulares decorrentes de fatores aleatórios.

Fava (2000) constata que, até o início da década de 1970, a maior parte dos trabalhos sobre séries temporais tinha como objetivo decompor a série nos elementos

vistos anteriormente, considerando o mecanismo gerador da série de tempo como sendo de natureza determinista.

No estudo de séries temporais existem três características que precisam estar presentes para que o processo de previsão seja considerado válido: estocasticidade, estacionariedade e homocedasticidade, as quais estão especificadas a seguir.

Estocasticidade é um conceito ligado ao campo probabilístico. Neste caso, especificamente, salienta-se que o processo gerador de uma série temporal não é determinístico. Enders (2009) define que os valores de uma série estocástica são obtidos por meio de uma função de distribuição de probabilidade em cada instante do tempo e a série obtida reflete apenas um conjunto de realizações possíveis dentro do processo estocástico. Adicionalmente, isto se torna possível pela presença da autocorrelação entre os termos da série.

Uma série temporal é fracamente estacionária quando suas características estatísticas – média, variância e covariância – são constantes ao longo do tempo. Em outras palavras, uma série temporal é dita estacionária se sua média, variância e covariância permanecem constantes, independente do período de análise. Portanto, elas não variam no decorrer do tempo. Bueno (2008) reforça que a constatação de estacionariedade é que permitirá a realização de inferências estatísticas dos parâmetros estimados de um modelo, onde temos um processo estocástico.

Segundo Gujarati (2006), uma série temporal pode ser classificada mediante o comportamento da sua variância, a qual pode ser apresentada de formas diferentes em relação às dispersões, seja com homocedasticidade ou com heterocedasticidade, fator de suma importância para a escolha de modelos que a representem. Para o mesmo autor, a variância de u_i (erro) para cada X_i (variável explicativa), ou seja, a variância condicional de u_i é um número positivo constante, no caso de presença de homocedasticidade. O que significa que as populações Y correspondentes aos valores de X têm a mesma variância, ou ainda que a variação em torno da linha de regressão ou linha das relações médias entre X e Y é a mesma para todos os X .

É importante salientar que a presença de heterocedasticidade, embora não cause viés ou inconsistência dos parâmetros, invalida os erros-padrão, as estatísticas t e as estatísticas F . Sendo assim, os estimadores deixam de ser eficientes.

A seguir são apresentados os métodos de previsão usados no trabalho, que podem ser agrupados em duas categorias: métodos lineares, dominados pelos

métodos estatísticos clássicos, que contemplam os algoritmos mais antigos e conhecidos, e não lineares, baseados em inteligência computacional, que englobam os métodos mais novos, que surgiram e continuam evoluindo, tendo como base o desenvolvimento da capacidade computacional e a volumetria de dados disponíveis.

3.2. Métodos lineares

Os métodos lineares podem ser subdivididos em: métodos *Naive*, que consideram o último valor disponível como a previsão, os métodos de suavização exponencial, os métodos econométricos para séries temporais por meio da aplicação da metodologia Box Jenkins, os métodos de regressão e os métodos RU-MIDAS.

Os métodos de regressão também devem ser vistos como algoritmos de aprendizagem supervisionada, já que buscam encontrar a relação entre a variável de alvo e as variáveis independentes, sendo que os dados usados no treino do modelo devem ser necessariamente “rotulados”, isto é, nos dados de treino é necessário que tenhamos o resultado real da variável independente, tornando possível a sua comparação com a estimativa feita através do modelo. Assim, pode-se afirmar que a regressão, especificamente a linear, é um dos algoritmos mais conhecidos e bem compreendidos em estatística e aprendizado de máquina.

3.2.1. Métodos Naive

Em termos de métodos ingênuos, utilizamos o método tradicional e o método tradicional ajustado, considerando o aspecto sazonal.

A previsão *Naive* é um método extremamente simples, pois consiste em adotar os valores futuros da série como iguais ao valor atual: $y_t(h) = y_t$, para todo $h > 0$. Geralmente, é usada para fazer comparação com técnicas de predição mais elaboradas, a fim de verificar o ganho de performance e acurácia e é definida por:

$$\hat{y}_{t+1} = y_t$$

sendo \hat{y}_{t+1} é a previsão feita em t para o dia $t+1$ e y_t é o valor observado no dia t .

O método *Naive* ajustado aplicado nesta Tese considera o último valor como a previsão futura, exceto para as semanas 49 a 52, em que o valor apontado é o último verificado nessas semanas. Tal ajuste sazonal foi feito apenas para essas semanas,

dado o peso do aspecto sazonal para o conjunto de dados estudados, sobretudo nas semanas de dezembro. Simplificadamente, as semanas de dezembro consideram como previsão os valores vistos nas semanas do último mês de dezembro, que seria do ano anterior ao que está sendo previsto, e as demais semanas são previstas como igual respectivamente ao último valor existente até aquele instante.

3.2.2. Métodos de Suavização Exponencial

Método estatístico clássico bastante utilizado em diversos trabalhos de previsão é o de suavização exponencial. Essa técnica, proposta no final da década de 1950, conforme Hyndman e Athanasopoulos (2013), motivou alguns dos métodos de previsão mais bem-sucedidos e utilizados até os dias de hoje. Por meio dessa metodologia, as previsões produzidas são médias ponderadas de observações anteriores, com os pesos decaindo exponencialmente à medida que as observações envelhecem, sendo dada maior relevância para as observações mais novas.

Entre os métodos de suavização exponencial existentes, o mais recomendado é o de Holt Winters, pois leva em conta a presença de tendência linear e da componente sazonal nos dados a serem previstos (PELEGRINI; FOGLIATTO, 2000).

O método sazonal de Holt-Winters compreende a equação de previsão e três equações de suavização – uma para o nível bl_t , uma para tendência b_t e uma para componente sazonal s_t , com os seus respectivos parâmetros de suavização: α , β e γ .

Com base no tipo de sazonalidade presente na série, pode-se especificar dois tipos de modelos de Holt Winters, os quais são: HW aditivo, utilizado quando a amplitude da variação sazonal se mantém constante, e HW Multiplicativo, usado quando a amplitude da variação sazonal se altera ao longo do tempo

A Tabela 1.1 mostra as equações para cada modelo.

Tabela 1.1 – Equações dos métodos Holt-Winters

	HW Aditivo	HW Multiplicativo
Previsão	$\hat{y}_{t+h t} = l_t + hb_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (l_t + hb_t)s_{t+h-m(k+1)}$
Nível	$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$	$l_t = \alpha \frac{y_t}{s_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1})$
Tendência	$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1}$
Sazonalidade	$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \frac{y_t}{(l_{t-1} + b_{t-1})} + (1 - \gamma)s_{t-m}$

Fonte: elaboração própria.

3.2.3. Métodos Econométricos

Métodos econométricos têm como estratégia básica modelar comportamentos de séries de tempo estacionárias, usando dois conjuntos de informações: um primeiro conjunto do qual se procura extrair informações determinísticas presentes nos valores passados da série, estruturados na forma de equações em diferenças desses valores, cujo tratamento é conhecido como análise autorregressiva; e um segundo conjunto, por meio do qual se busca capturar efeitos de choques e incertezas, estruturados na forma de médias móveis dos erros produzidos pelo modelo (BUENO, 2008). Os métodos econométricos aplicados neste trabalho são o SARIMA e o SARIMAX.

A escolha do método SARIMA é justificada pelo fato de séries de tempo de vendas tipicamente apresentarem não estacionariedade (média de venda não constante ao longo do tempo) e sazonalidade (comportamento conhecido de picos de vendas em datas comemorativas), pois o método SARIMA, além de aplicar a metodologia para os dados da série estudada, pressupõe a verificação de condição de estacionariedade da série, e inclui tratamento de sazonalidades. A seguir é mostrada a fórmula desse método. A previsão do valor y_t para o período t , é dada por:

$$\phi_p(B) \Phi_P(B^S)(1 - B^S)^D(1 - B)^d Y_t = \theta_q(B) \Theta_Q(B^S) a_t$$

em que:

$\phi_p(B)$: é o polinômio autorregressivo não sazonal de ordem p ;
$\Phi_P(B^S)$: é o polinômio autorregressivo sazonal de ordem P com período sazonal S ;
$(1 - B^S)^D(1 - B)^d$: é o operador indicando a diferenciação sazonal de ordem D e uma diferenciação não sazonal de ordem d ;
$\theta_q(B)$: é o polinômio de médias móveis não sazonal de ordem q ;
$\Theta_Q(B^S)$: é o polinômio de médias móveis sazonal de ordem Q com período sazonal S ;
a_t	: é um processo ruído branco gaussiano.

O método SARIMAX é muito semelhante ao método SARIMA, com a diferença de que são adicionadas variáveis exógenas ao modelo. No método SARIMA apenas a própria série é considerada.

No caso dos métodos econométricos, diante da possibilidade de serem produzidas várias alternativas de modelos, é importante que seja estabelecida uma métrica para determinar o número adequado de parâmetros para o modelo e, dessa forma, tornar possível a seleção do mais adequado. Nesse trabalho foi adotado o critério de Akaike para seleção do modelo mais satisfatório. A ideia por trás do critério de informação é a de minimização de uma função baseada nos resíduos e penalizada pelo número de regressores (BUENO, 2008). O melhor modelo será o mais parcimonioso, satisfazendo a condição dos resíduos apresentarem os menores valores possíveis. Existem três principais critérios de informação: a estatística de Schwarz (BIC), a estatística de Akaike (AIC) e a estatística Hanna-Quinn (HC).

3.2.4. Métodos de Regressão

A regressão linear é um método estatístico, que utiliza a relação entre duas ou mais variáveis, de modo que uma variável pode ser estimada a partir das demais variáveis. Para a estimação dos parâmetros é geralmente utilizado o método dos mínimos quadrados. Adicionalmente, a aplicação desse método considera algumas suposições relevantes, tais como: a relação matemática entre a variável dependente e as independentes é linear, as variáveis independentes não são aleatórias, a média do erro é nula, o erro é homocedástico, o erro de uma observação é descorrelacionado com o erro de outra observação e os erros apresentam distribuição normal.

Na Econometria, o método mais tradicional e comum é a análise de regressão, cuja evolução está calcada no modelo clássico. Sua ampla aplicação pode ser justificada por dois aspectos: sua simplicidade e baixa complexidade computacional. Krajewski, Ritzman e Malhotra (2009) adicionam que o modelo de regressão linear é um dos modelos quantitativos mais conhecidos e usados para realização de previsão.

Para esta Tese, além da regressão linear com estimação pelo método dos mínimos quadrados, também foram consideradas outros métodos de estimação: Lasso, Ridge e a regressão Rede elástica.

Modelos de previsão adequados geralmente apresentam erros pequenos em termos de viés e variância. Tal configuração é difícil de ser verificada quando existe uma pequena quantidade de dados e uma grande quantidade de variáveis explicativas, praticamente equivalente à quantidade de amostras do conjunto de

dados. Esses pontos contribuem para geração de sobreajuste, o que pode ser solucionado com a aplicação de técnicas de regularização.

A regularização contribui para solução de problemas de ajuste excessivo, ou seja, quando o modelo apresenta um bom desempenho nos dados de treinamento, porém mostra desempenho inferior na amostra de teste. Por meio da inclusão de um termo de penalidade à função objetivo ou função custo, a regularização controla a complexidade do modelo criado. O procedimento de ajuste envolve especificamente a função de custo como soma residual dos quadrados (RSS). Assim, os coeficientes w são eleitos para minimizar essa função de custo com base nos dados de treinamento. De forma resumida, Devay (2019) estabelece que a regularização pode ser vista como a inserção de um viés na estimação, desencorajando a ocorrência do ajuste excessivo dos dados, com o objetivo de diminuir a sua variância.

A regularização dos coeficientes w implica necessariamente na restrição do seu respectivo tamanho. Para Tibishirani (1996), a regularização geralmente é útil nas seguintes situações: presença de grande número de variáveis, relação baixa de observações numéricas para número de variáveis e grande possibilidade de alta multicolinearidade. Nesse trabalho são utilizadas quatro técnicas de regularização: Lasso (L1), Ridge (L2), Rede elástica (L1 e L2) e *dropout*, sendo que a última será especificada na parte dos métodos de inteligência computacional.

Na regularização L1 ou Lasso, a função objetivo é minimizada adicionando um termo de penalidade à soma dos valores absolutos dos coeficientes. Quando há múltiplas variáveis altamente correlacionadas, a regularização Lasso seleciona apenas uma dessas variáveis e zera os coeficientes das outras, de forma a minimizar a penalização L1.

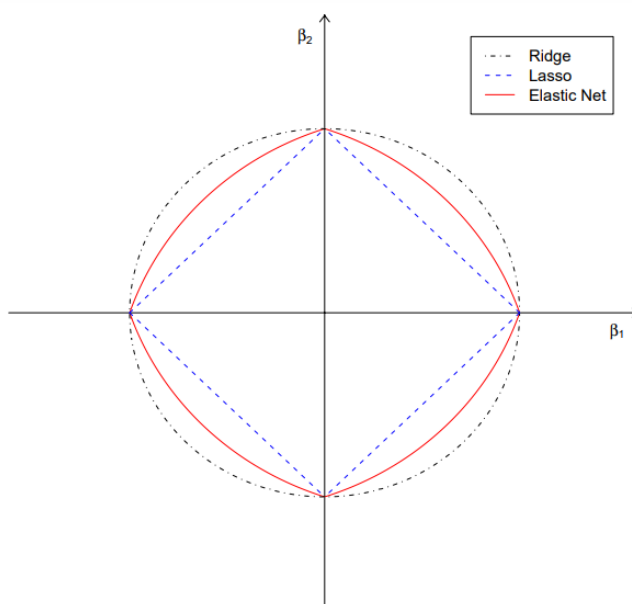
Na regularização L2, a função objetivo é minimizada adicionando um termo de penalidade à soma dos quadrados dos coeficientes. Regressão de Ridge ou regressão de encolhimento faz uso da regularização L2, não ocorrendo a eliminação total de determinada variável. Nesse caso, a penalização consiste nos quadrados dos coeficientes, ao invés dos seus módulos, fazendo com que na presença de variáveis altamente correlacionadas elas fiquem com coeficientes parecidos. No trabalho de Donald e Ronald (1975), foi possível observar nos exemplos estudados que, quando as variáveis preditoras são altamente correlacionadas, a regressão *Ridge* produz

coeficientes que predizem e extrapolam melhor que os mínimos quadrados puros e é um procedimento altamente recomendado para a seleção de variáveis.

A Rede elástica combina os termos de regularização L1 e L2. Assim, é possível extrair os benefícios das duas técnicas, porém surge o problema de determinação de dois hiperparâmetros para obter soluções ótimas. Foi possível constatar com base em dados reais e em um estudo de simulação que a Rede elástica geralmente supera o Lasso. Além disso, a rede elástica incentiva um efeito de agrupamento, onde previsores fortemente correlacionados tendem a permanecer ou sair do modelo de forma conjunta. Esse tipo de regularização é particularmente útil quando o número de preditores (p) é muito maior que o número de observações (n) (ZOU; HASTIE, 2005).

O gráfico da Figura 1.10 esboça tipos de penalidade por método de regularização.

Figura 1.10 – Tipos de penalização por método de regularização



Fonte: Zou e Hastie (2005), p.5

3.2.5. Métodos RU-MIDAS

Esta seção apresenta um resumo teórico dos métodos MIDAS (*mixed data sampling*), que possuem mais de uma variação e são aplicados neste trabalho. A inclusão da metodologia foi feita, pois os dados usados nesta Tese têm frequências distintas ao longo do tempo e os métodos MIDAS têm como foco lidar com este tipo de característica. Mais especificamente, as séries de dados de vendas e preços dos

produtos apresentam frequência semanal, enquanto a maior parte das variáveis exógenas, como, por exemplo, a taxa de desemprego é expressa de forma mensal.

Segundo Foroni e Massimiliano (2013), que desenvolveram pesquisa *survey* sobre modelos econométricos que aplicam dados de frequência mista para previsão, um desafio encontrado por pesquisadores reside no fato de realizar previsões utilizando variáveis que possuem frequências distintas de disponibilização. Para os autores, o tratamento mais comum diante dessa situação é a aplicação da agregação simples para garantir que as variáveis estejam na mesma periodicidade. Porém, essa solução tem como principal desvantagem a perda da informação contida nos dados. Nesse sentido, os modelos propostos para evitar a perda de informação são: os modelos de equações pontes, os modelos MIDAS e os modelos de espaço estado.

Os modelos de equações pontes, muito utilizados pelos bancos centrais para previsões de curto prazo, são regressões lineares, as quais associam variáveis de alta frequência com variáveis de baixa frequência, sendo necessário que todo o conjunto de regressores seja conhecido para o período de projeção.

Os modelos de espaço estado, tendo como principais representantes os modelos VAR de frequência mista (MF-VAR) e os modelos de fatores, são abordagens em que é verificada, conjuntamente, a dinâmica tanto da variável independente como da variável dependente, em que o uso do filtro de Kalman fornece não apenas previsões de observações futuras, mas também estimativas do estado latente atual.

Os modelos MIDAS, criados por Ghysels, Santa-Clara e Valkanov (2004) com base nos de defasagens distribuídas (ADL), usam as variáveis explicativas de alta frequência ponderadas por funções para obter maior nível de flexibilidade, dada a necessidade de menos parâmetros quando comparado com o modelo de defasagem distribuída. O MIDAS trabalha com variáveis de diferentes frequências, enquanto o de defasagens distribuídas usa variáveis exógenas defasadas, mas na mesma frequência da variável endógena e não faz uso da própria variável exógena defasada.

A principal vantagem da especificação MIDAS é permitir a aplicação de uma grande quantidade de atrasos, gerando uma quantidade limitada de parâmetros a serem previstos, tendo em vista o uso da função polinomial, que se torna ainda mais relevante nas aplicações multivariadas (GHYSELS; SANTA-CLARA; VALKANOV, 2004). Além disso, as regressões MIDAS possuem ampla aplicabilidade em finanças e macroeconomia, levando a estimativas mais eficientes do que a abordagem típica

de agregação de todas as séries para a menor frequência disponível e, em muitos casos, as regressões MIDAS são tão eficientes quanto modelos ADLs.

De forma prática, Ghysels e Wright (2006) propõem a aplicação do uso do método MIDAS para a previsão diária de uma série de indicadores macroeconômicos relevantes, o que torna possível mensurar os impactos dos eventos e das notícias diárias, resultando em uma informação mais relevante para o processo de tomada de decisão dos agentes econômicos. As previsões realizadas conseguiram antecipar uma parte significativa das pesquisas de expectativas e, assim, foi possível capturar o impacto efetivo de eventos e novas notícias sobre essas expectativas.

Além do modelo básico MIDAS, Foroni e Massimiliano (2013) apontam a existência de outros MIDAS, sendo os principais: AR MIDAS, com defasagens da variável independente ao modelo original MIDAS; M-AR-MIDAS, que considera uma abordagem com múltiplos regressores e as defasagens da variável independente; U-MIDAS (*unrestricted* MIDAS), em que é estimado um parâmetro para cada observação da variável de alta frequência, tendo como principal diferença em relação ao modelo MIDAS a não utilização de uma função polinomial para agrupar todos os regressores contemplados e a suposição de que a função polinomial não contém a flexibilidade adequada para representar os verdadeiros pesos geradores dos dados aleatórios; M-AR-UMIDAS, modelo com múltiplos regressores. Adicionalmente, existem outros modelos, como MIDAS não linear, MIDAS assimétrico, Smooth Transition MIDAS, Markov-Switching MIDAS, MIDAS com etapas etc.

Existe consenso dos ganhos ao explorar dados com diferentes frequências, no entanto não está claro qual método tem melhor desempenho, segundo Foroni e Massimiliano. Uma ressalva em relação ao uso da abordagem VAR: por ser feita por meio da aplicação do filtro do Kalman, ela apenas apresenta um bom funcionamento quando é considerado um número pequeno de variáveis. Tal fato pelo aumento expressivo de parâmetros, que devem ser calculados, gerando, assim, aumento expressivo de complexidade. Por fim, os modelos MIDAS parecem ser mais robustos a erros de má especificação, em comparação aos modelos de equações de ponte e abordagens de espaço de estado, e são computacionalmente mais simples.

Posteriormente, Foroni, Guérin e Massimiliano (2018) desenvolveram outro modelo MIDAS: o RU-MIDAS, o único cujo objetivo é incorporar informação em baixa frequência para realizar previsões de variáveis em alta frequência. Em todos os

demais é feita a previsão em baixa frequência por meio de variáveis em uma periodicidade maior. O modelo MIDAS aplicado nesta Tese é o RU-MIDAS, já que são utilizadas as variáveis macroeconômicas em baixa frequência mensal e as vendas são previstas em uma periodicidade superior, que é semanal. Abaixo, conforme Foroni, Guérin e Massimiliano (2018), é apresentada a estrutura periódica do modelo:

$$\begin{aligned}\tilde{c}_0(L)x_t &= g_0(Z)y_t + \tilde{\gamma}_0(L)e_{xt}, \\ t &= 0, 1, 2, \dots\end{aligned}$$

A variável x (HF) depende das suas próprias defasagens sobre as defasagens das variáveis y (LF) observadas e de um termo de erro que, geralmente, apresenta uma estrutura MA. Para estimação desses modelos é usado o método dos Mínimos Quadrados Ordinários (MQO) e a ordem da defasagem é determinada por algum critério de informação. No trabalho de Foroni, Guérin e Massimiliano (2018) foi feita uma comparação da metodologia RU-MIDAS com o MF-VAR para realizar previsões de indicadores mensais como inflação, crescimento da produção industrial e gastos pessoais com consumo, através de dados trimestrais de pesquisas sobre esses indicadores. Os resultados indicam que há informações adicionais que podem ser extraídas de pesquisas trimestrais logo após sua publicação, quando alguém estiver interessado em prever o indicador mensal correspondente. Então informações de baixa frequência podem realmente ser úteis ao prever variáveis de alta frequência. Entretanto, não foi possível afirmar que o modelo RU-MIDAS superou o MF-VAR.

3.3. Métodos Não Lineares

Esta seção apresenta os diversos métodos não lineares aplicados nesta Tese para a tarefa de previsão de vendas, que é o foco deste trabalho. Dentre não lineares, que são baseados em inteligência computacional, podemos dividi-los em dois grandes grupos: os que são baseados em árvores de decisão e os que são baseados em redes neurais. Como representantes dos métodos de árvores de decisão, temos: *random forest* (RF), XGBoost (XGB) e LightGBM. E como representantes dos modelos construídos com redes neurais, temos: rede neural multicamadas (MLP), rede neural convolucional (CNN) e rede neural *long short term memory* (LSTM).

Existe uma infinidade de algoritmos de aprendizagem de máquina e o foco aqui são os algoritmos mais avançados, em termos de árvores de decisão e os de *deep learning*, os quais geralmente utilizam redes neurais.

3.3.1. Métodos de Árvores de Decisão

Métodos baseados nas árvores de decisão são algoritmos simples e fáceis de compreender. Implicitamente já fazem o processo de seleção de variáveis, podem lidar com dados tanto numéricos como categóricos, requerem pouco esforço em termos de preparação do conjunto de dados em si e a presença de relações não lineares entre os parâmetros não afeta a performance do algoritmo.

3.3.1.1. *Random Forest* (RF)

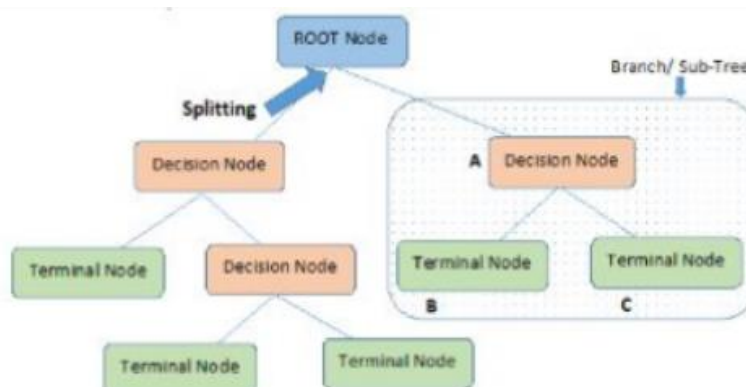
Antes de explicar o RF, é primordial definir o conceito de árvore de decisão, tipo de algoritmo de aprendizagem de máquina supervisionado, cujo princípio é a divisão em grupos homogêneos, sendo aplicado tanto em tarefas de classificação como de regressão. Ademais, esse algoritmo tem as seguintes características: é facilmente interpretável, aceita dados tanto categóricos como numéricos, apresenta alta probabilidade de sobreajuste e é bastante instável, o que significa que pequenas alterações nos dados de treino produzem árvores completamente distintas.

Para montagem da árvore, é feita uma divisão no conjunto de dados, em que cada divisão representa um nó da árvore. Existe o nó raiz, em que começa a árvore; os nós de decisão, que determinam o caminho da árvore; e os nós terminais/folha, representados por uma classe ou valor contínuo. O objetivo de uma árvore de decisão é obter a melhor divisão dos dados, isto é, um subconjunto de dados com a maior pureza possível, que pode ser medida por indicadores como índice de Gin, ganho de informação etc. Na Figura 1.11 é possível ver a estrutura de uma árvore de decisão.

O conceito por trás de RF, algoritmo de aprendizagem supervisionada, é a criação de uma floresta aleatória, sendo a combinação de árvores de decisão treinadas geralmente com o método de *bagging*. Esse método consiste em pegar repetidamente amostras aleatórias com reposição dos dados originais e realizar o treinamento do modelo, de modo a garantir que a combinação gere um modelo

superior (BREIMAN, 1996). Assim, o algoritmo RF cria várias árvores de decisão e as combina para uma previsão mais precisa, robusta e estável. Quando comparada com a árvore de decisão, a RF gera menor sobreajuste, o que está diretamente ligado à aleatoriedade para a criação das árvores. No entanto, é mais lenta dependendo do número de árvores contempladas no modelo.

Figura 1.11 – Árvore de decisão



Fonte: Santana (2020).

A RF reduz o erro da variância das árvores de decisão ao usar amostras distintas para treinamento, trabalha sobre subconjuntos de características aleatórios e considera a combinação e a construção de pequenas árvores. Diferença em relação às árvores de decisão simples é que as últimas são mais fáceis de interpretar.

Algoritmos de árvores de decisão baseados em *gradient boosting*, quando comparados com a RF, apresentam duas diferenças importantes: a forma como as árvores são construídas e como os resultados são combinados. Na RF, cada árvore é construída de forma independente e, no *gradient boosting*, é aplicado um modelo aditivo cujo objetivo é a introdução de novos preditores fracos para melhorar os erros verificados pelos preditores fracos já utilizados. Com relação à combinação dos resultados, as RF combinam os resultados somente no fim do processo, enquanto o *gradient boosting* realiza essa combinação ao longo do processo. Existe uma dificuldade maior para definição dos parâmetros do *gradient boosting*, mas se ele for feito de forma adequada, pode-se obter estimativas mais precisas em relação às obtidas por meio das RF. Ressalta-se, no entanto, que o *gradient boosting* não é recomendado na presença de dados muito ruidosos, sendo preferível as RFs.

3.3.1.2. XGBoost

Os criadores Chen e Guestrin (2016) descrevem o método como um sistema de aprendizagem de máquina baseado no algoritmo *gradient boosted tree* (GBT), cujo objetivo é criar vários modelos de árvore de decisão considerados mais simples (ou fracos) a fim de criar um modelo mais poderoso e robusto, por combinar resultados desses modelos fracos, o que é realizado por meio de um treinamento iterativo de um conjunto de árvores de decisão minimizando uma função de perda estabelecida.

O XGBoost minimiza uma função objetiva regularizada (L1 e L2), que combina uma função de perda convexa com base na diferença entre as saídas previstas e desejadas, e um termo de penalidade para a complexidade do modelo, ou seja, as funções de árvore de regressão. O treinamento é iterativo, sendo adicionadas novas árvores que preveem os erros das anteriores, que são, então, combinadas com árvores anteriores para a previsão final. É chamado *gradient boosting*, porque usa um algoritmo de descida de gradiente para minimizar a perda ao adicionar novos modelos.

O sistema é reconhecido em várias competições de aprendizagem de máquina, sendo que grande parte das melhores soluções implementaram o XGBoost como algoritmo. Tal fato demonstra que o método chega ao estado da arte para uma gama ampla de problemas, tais como: previsão de vendas, classificação de eventos de picos de energia, classificação de textos da *web*, classificação de *malware*, categorização de produtos etc. A principal justificativa para o êxito do algoritmo é a escalabilidade possível nos diversos problemas em que foi aplicado. O sistema produz resultados dez vezes mais rápido do que soluções populares que rodam em uma única máquina e consegue escalar para bilhões de exemplos de forma distribuída ou com configurações de memória limitada.

A escalabilidade do XGBoost é explicada pela implementação de várias otimizações algorítmicas, como: a criação de um novo algoritmo de árvore de decisão para lidar com dados esparsos; um procedimento quartílico com pesos teoricamente ajustados, que permite tratar os pesos de cada instância em uma aproximação da aprendizagem por meio de árvores; o uso da computação paralela e distribuída, que torna o aprendizado mais rápido. Em suma, o XGBoost permite que os cientistas de dados processem centenas de milhões de exemplos em um simples *desktop*.

A popularidade do XGBoost entre os cientistas de dados é justificada pelos seguintes motivos: velocidade, desempenho, possibilidade de paralelização, apresentação de solução tanto linear quanto algoritmos de árvores de decisão, o que disponibiliza a paralelização em uma máquina simples e, por fim, é referência de performance em diversas tarefas de aprendizagem de máquina.

Brownlee (2016) afirma que o XGBoost tem uma série de propriedades avançadas, em termos das características dos seus modelos, dos seus sistemas e do algoritmo. No que diz respeito às características do modelo, três formas de *gradient boosting* são suportadas: *gradient boosting* tradicional, *gradient boosting* estocástico e *gradient boosting* regularizada. As características do seu sistema permitem a paralelização utilizando todos os *cores* de CPUs disponíveis durante o treinamento, a computação distribuída, a computação *out-of-core*, que possibilita lidar com grandes conjuntos de dados e permite a otimização do *cache* das estruturas dos dados e do algoritmo para fazer melhor uso do *hardware*. Em relação às características do algoritmo, a sua implementação foi desenvolvida para gerar eficiência em termos de tempo computacional e recursos de memória, tornando possível fazer o melhor uso dos recursos disponíveis para a fase de treinamento. Isso se dá via tratamento dos dados esparsos de uma estrutura de bloco para efetivar a paralelização da construção das árvores de decisão e do treinamento continuado, tornando possível melhorar um modelo já ajustado em novos dados.

Dois trabalhos de previsão com aplicação do XGBoost relacionados a esta Tese são: Saradhi e Naresh (2018), que aplicaram várias técnicas de *data mining* para previsão de vendas no nível de lojas diariamente, sendo que o XGBoost teve o melhor desempenho, e o trabalho de Ji *et al.* (2019) com um novo modelo XGBoost, o C-A-XGBoost, que considera tanto as características como a tendência apresentada no conjunto de dados em um processo de duas etapas. O modelo superou o XGBoost tradicional, o ARIMA e cada um dos usados em cada etapa de forma separada.

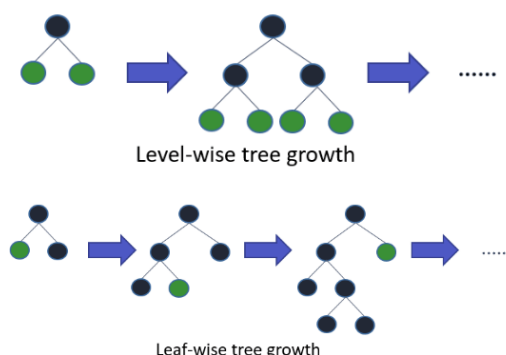
3.3.1.3. LightGBM

O LightGBM é um *gradient boosting* rápido, distribuído e de desempenho baseado no algoritmo da árvore de decisão, aplicado em uma série de atividades de aprendizagem de máquina. Ao comparar os algoritmos de *gradient boosting decision*

tree (GBDT) entre si, pode-se diferenciá-los pelo modo como as árvores de decisão crescem. Assim, o XGBoost utiliza a abordagem chamada de *level-wise tree growth*, ao passo que o LightGBM usa a abordagem denominada *leaf-wise tree growth*.

Segundo Ke *et al.* (2017), no algoritmo LightGBM, o crescimento da árvore ocorre verticalmente, enquanto os demais algoritmos de *boosting* geram o crescimento das árvores horizontalmente, o que significa que LightGBM cresce árvore em forma de folha, enquanto outro algoritmo cresce em nível. Ou seja, na abordagem *leaf-wise*, em cada nível, apenas um dos lados da árvore fica mais profunda. Dessa forma, a cada nível, uma quantidade cada vez menor de resíduos permanece para ser considerada na determinação do ponto de corte que maximize o ganho, permitindo, então, um treinamento mais rápido. A Figura 1.12 evidencia a diferença entre as abordagens de crescimento das árvores de decisão.

Figura 1.12 – Crescimento da árvore de decisão



Fonte: Mandot (2017).

Quando o crescimento ocorre em uma mesma folha no LightGBM, o algoritmo consegue reduzir mais perdas do que o algoritmo de profundidade e, portanto, resulta em uma precisão muito melhor que dificilmente pode ser alcançada por qualquer um dos algoritmos de *boosting* existentes.

As principais vantagens do LightGBM são: maior velocidade de treinamento e maior eficiência, possível devido ao uso de um algoritmo baseado em histograma, isto é, ele agrupa valores de recursos contínuos em caixas discretas, que aceleram o processo de treinamento; menor uso de memória; maior acurácia que outros algoritmos de *boosting*; e é muito apropriado para grandes conjuntos de dados. Uma recomendação relevante é que o LightGBM não deve ser aplicado em pequenos conjuntos de dados, devido à grande possibilidade de sobreajustamento dos dados.

Ke *et al.* (2017) criaram o LightGBM considerando que as implementações XGBoost e pGBRT não eram satisfatórias, em termos de eficiência e escalabilidade, na presença de conjunto de dados grandes e com muitas dimensões. Para lidar com essas limitações, os autores utilizaram duas novas técnicas: *gradient-based one-side sampling* (GOSS) e *exclusive feature bundling* (EFB). De forma resumida, o GOSS é uma modificação no método de gradiente *boosting*, que concentra a atenção nos exemplos de treinamento que resultam em um gradiente maior com o intuito de acelerar o aprendizado e reduzir a complexidade computacional. Já o EFB é uma abordagem para empacotar características esparsas mutuamente exclusivas, como entradas de variáveis categóricas que recebem o tratamento de *one-hot encoding*. Pode-se afirmar, então, que o EFB é um tipo de seleção automática de características, o que colabora para tornar o treinamento mais rápido e o modelo mais parcimonioso.

Nos experimentos feitos por Ke *et al.* (2017) em vários conjuntos de dados públicos, verificou-se que o LightGBM acelera o processo do *gradient boosting decision tree* (GBDT) convencional em até mais de 20 vezes, mantendo praticamente o patamar de precisão.

Por fim, apenas para comprovar o excelente desempenho desse algoritmo e considerando a temática deste trabalho, na competição M5 – que faz parte das competições Makridakis, que são abertas e organizadas por equipes lideradas pelo pesquisador Spyros Makridakis, e que têm como objetivo avaliar e comparar a precisão de diferentes métodos de previsão –, o LightGBM foi o método com o melhor desempenho em termos de RMSE. O objetivo da competição em questão era realizar a previsão para os próximos 28 dias de vendas das lojas da rede Walmart.

3.3.2. Métodos de Redes Neurais

As redes neurais devem ser aplicadas considerando as seguintes vantagens intrínsecas a elas: bom desempenho, robustez e capacidade de generalização (HAYKIN, 2001). Característica positiva dos MLP é sua capacidade de captar relações não-lineares entre as variáveis de entrada a ponto de serem considerados como potenciais aproximadores universais de funções.

Por outro lado, o mesmo autor pontua algumas desvantagens das RNs: dificuldade de justificar o comportamento das redes, podendo ser consideradas “caixas-pretas”; em RNs que utilizam o algoritmo de retropropagação, o tempo de

treinamento pode ser muito longo; dificuldade de determinar a topologia mais adequada para a RN, considerando os objetivos de convergência e generalização e, dependendo do tipo RN, pode haver necessidade de grande volume de dados para realização do treinamento.

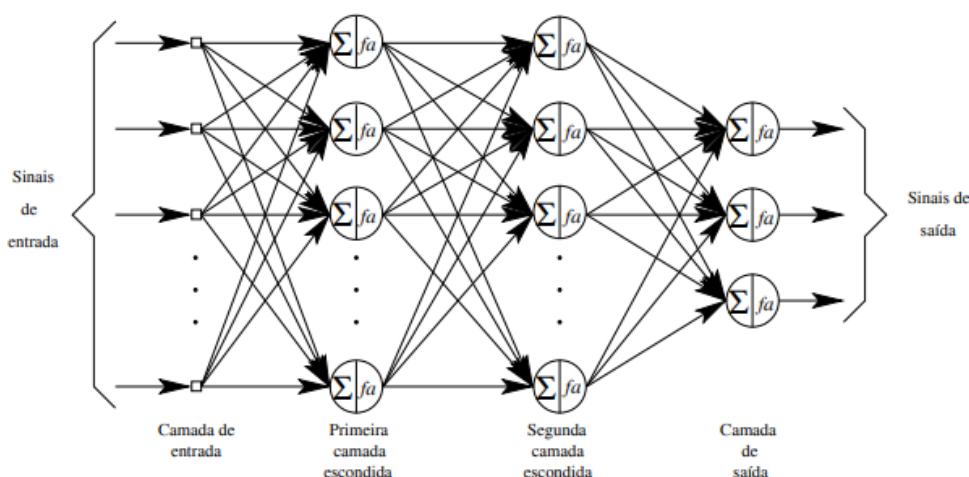
3.3.2.1. MLP FTDNN

O modelo de rede neural mais utilizado para previsão de séries temporais é o MLP, uma extensão do Perceptron, desenvolvido por Rosenblatt, sendo composto de várias camadas de neurônios, segundo Felisoni de Angelo *et al.* (2011). Vellasco (2007) destaca três características básicas presentes nas redes MLP:

- cada neurônio da rede inclui, necessariamente, uma função de ativação não-linear para lidar com as não linearidades presentes nos dados;
- a rede contém uma ou mais camadas de neurônios ocultos, que permitem que a rede possa aprender tarefas mais complexas;
- a rede possui alta conectividade por meio dos seus pesos.

Além disso, a autora menciona como principal deficiência a sua alta complexidade e conectividade, o que torna difícil a sua compreensão teórica e sua interpretabilidade. A Figura 1.13 evidencia uma representação de uma rede MLP.

Figura 1.13 – Rede Neural MLP



Fonte: Adaptado de Haykin, 2001, p. 186.

Existem vários modelos de rede neural, sendo que a base conceitual inicial tem como alicerce a estrutura da MLP. Os modelos vistos nesta Tese são: *focused time*

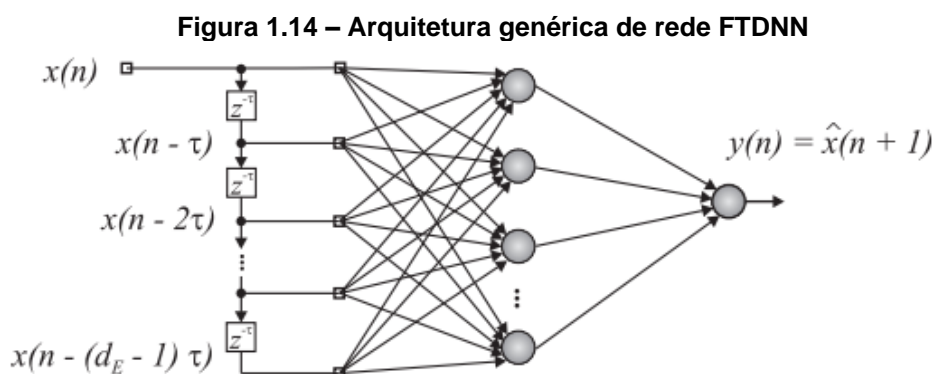
delay neural network (FTDNN), *convolutional neural network* (CNN) e *long time short memory* (LSTM). O primeiro modelo de rede neural – e o mais simples quando comparado com a CNN e a LSTM – é o MLP FTDNN.

Especificamente sobre a previsão de séries de tempo, as redes FTDNN e LSTM são redes MLP nas quais o aspecto temporal é incluído por meio da introdução de defasagens no fornecimento das variáveis de entrada, durante o processo de treinamento da rede para calibração dos parâmetros da rede. Mais recentemente, passou-se a utilizar a CNN para realização de previsões de séries temporais, tendo em vista que seu uso massivo estava ligado ao reconhecimento de imagens.

De acordo com Haykin (2001), as FTDNN são redes MLP, onde o aspecto temporal é incluído através da camada de entrada, por meio da introdução de várias defasagens da variável que se visa prever. Deve-se considerar que as entradas precisam ser submetidas sequencialmente à rede. Sendo assim, estamos incluindo uma memória de curto prazo na camada de entrada da rede.

$$y_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p})$$

A consideração das defasagens torna possível à rede identificar em que ponto ela está de um determinado intervalo de tempo, possibilitando a resolução de problemas, onde o tempo é uma variável relevante. A Figura 1.14 mostra a FTDNN.



Fonte: Menezes Junior, 2006, p. 64.

Em que $x(n)$ é um vetor que contém de elementos da série contados a partir do elemento atual $x(n)$, separados um do outro por t unidades de tempo, y é a saída do neurônio e x é a variável escalar predita baseada em atrasos no tempo.

Redes FTDNN podem ser interpretadas como modelos autorregressivos com formulação não-linear (KIM; SHIN, 2007), com a vantagem da identificação das

defasagens pela própria rede, por meio da consideração dos pesos, os quais são ajustados pela Regra de Delta. O método de treinamento desse tipo de rede é o retropropagação.

3.3.2.2. CNN

Rede neural convolucional (CNN) é um algoritmo de aprendizagem profunda, que consegue captar uma imagem de entrada, atribuir importância a vários aspectos/objetos da imagem e, assim, ser capaz de diferenciar as imagens entre si. Em termos de pré-processamento, exige menos que outros algoritmos. Já no que diz respeito à arquitetura, apresenta um padrão muito próximo ao de conectividade de neurônios do cérebro humano e foi inspirada na organização visual do córtex.

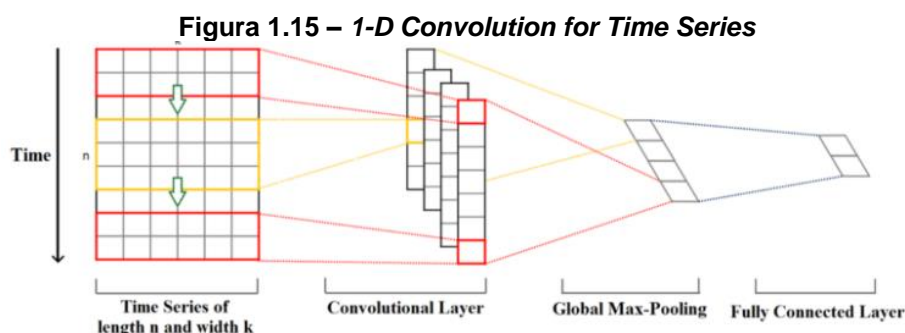
Segundo Deshpande (2019), os componentes de uma rede convolucional são: as convoluções, que funcionam como filtros e formam, no final da leitura da imagem, um mapa de características, sendo que a profundidade da saída de uma convolução é igual à quantidade de filtros aplicados; os filtros, formados por pesos inicializados de forma aleatória, sendo atualizados pelo processo de retropropagação; o *padding*, um parâmetro para evitar que as camadas não diminuam antes da realização do aprendizado; a camada de *pooling*, cujo objetivo é simplificar a informação da camada anterior e evitar sobreajuste, que pode ser feito de várias formas, sendo o *maxpooling* o mais aplicado; a camada *fully connected*, em que a entrada é a saída da camada anterior e a saída são N neurônios, que representa a quantidade de classes do modelo.

A CNN é bastante utilizada no reconhecimento de imagem, mas também pode ser aplicada na previsão de séries temporais. Sayyad (2020) explica o funcionamento da CNN para regressão. Considere uma série temporal de comprimento n e largura k , sendo n o número de defasagens temporais e a largura o número de variáveis, tendo em vista uma série temporal multivariada. Nesse caso, as camadas convolução sempre têm a mesma largura das séries temporais, embora seu comprimento possa ser variado. Assim, o *kernel* se move em uma direção desde o início de uma série temporal em direção ao seu final, realizando a convolução. Ele não se move para a esquerda ou para a direita como acontece quando a convolução 2-D usual é aplicada às imagens. A CNN aprenderá uma função que mapeia uma sequência de

observações passadas como entrada para uma de saída. Assim, a sequência de observações é transformada em exemplos com os quais o modelo pode aprender.

Uma CNN unidimensional é um modelo com uma camada oculta convolucional que opera em uma sequência 1D. Em alguns casos, pode ser adicionada uma segunda camada convolucional, como nas situações de sequências de entrada muito longas e, posteriormente, é incluída uma camada de *pooling*, cujo trabalho é gerar a saída da camada convolucional para os elementos mais relevantes. Após as camadas convolucional e de *pooling*, é acrescida uma camada *fully connected*, que interpreta os recursos extraídos pela parte convolucional do modelo. Uma camada *flatten* é usada entre as camadas convolucionais e densa para reduzir os mapas de recursos a um único vetor unidimensional.

Em termos de funcionamento, Sayyad (2020) fala que os elementos do *kernel* de convolução são multiplicados pelos elementos correspondentes da série temporal que cobrem em um determinado ponto. Em seguida, os resultados da multiplicação são somados e uma função de ativação não linear é aplicada ao valor. O resultado torna-se um elemento de uma nova série temporal univariada “filtrada” e, em seguida, o *kernel* avança ao longo da série temporal para produzir o próximo valor. Dependendo do comprimento do *kernel*, diferentes aspectos, propriedades e recursos da série temporal inicial são capturados em cada uma das novas séries filtradas. A próxima etapa é aplicar o agrupamento máximo a cada um dos vetores de série temporal filtrados: o maior valor é obtido de cada vetor. Um novo vetor é formado a partir desses valores e esse vetor de máximos é o vetor de recurso final, que pode ser usado como uma entrada para a camada final *fully connected*. Tal processo de funcionamento é ilustrado na Figura 1.15.



Fonte: Sayyad (2020).

Filho (2019b) afirma, assim como Brownlee (2018), que as redes neurais convolucionais foram originalmente constituídas para lidar com imagens, contudo seu uso vem se tornando cada vez mais popular para modelagem de séries temporais. De forma resumida, elas aprendem filtros que multiplicam os atributos originais dentro das janelas convolucionais. Quando comparadas com as redes neurais recorrentes, como é o caso da LSTM, que será vista a seguir, elas são treinadas mais rápido e possuem menos parâmetros a serem ajustados. Ademais, quanto maior o volume de dados presente, melhor será o funcionamento desse tipo da rede neural convolucional.

3.3.2.3. LSTM

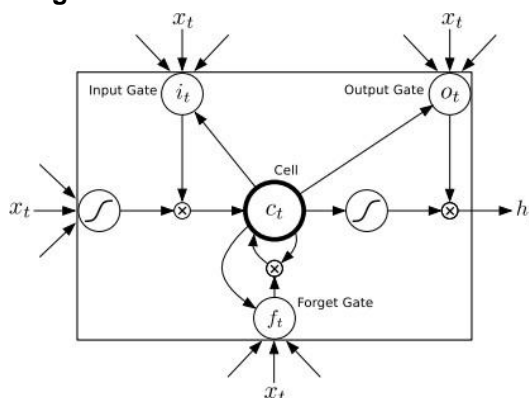
LSTM é uma rede neural do tipo recorrente que, por tal característica, apresenta retroalimentação. Desenvolvida, em 1997, por Sepp Hochreiter e Jürgen Schmidhuber, foi aprimorada, em 2000, pela equipe de Felix Gers, através da inclusão do portão de esquecimento. Tipo especial de rede neural recorrente, a LSTM é capaz de aprender dependências de longo prazo, o que não é possível nas redes neurais MLP tradicionais. Em termos de desvantagem, a LSTM exige volume expressivo de dados e alto poder de computação para alcançar bom desempenho. Greff *et al.* (2015) afirmam que as redes LSTM se tornaram os modelos de estado da arte para vários problemas de aprendizagem de máquina.

Ao comparar a LSTM com os demais tipos de redes recorrentes, a diferença mais relevante é ela não usar uma função de ativação dentro dos seus componentes recorrentes, consequentemente o valor armazenado não é modificado de forma iterativa. Para Gomes *et al.* (2017), o ganho relevante da rede LSTM é evitar problemas de propagação do erro ao longo de toda a estrutura. Já para Chung *et al.* (2014), outros tipos de redes neurais recorrentes têm dificuldade de capturar as dependências de longo prazo, pois os gradientes tendem a desaparecer ou a explodir.

Hochreiter e Schmidhuber (1997) e Gers, Schmidhuber e Cummins (2000) evidenciam que a topologia de um neurônio de uma LSTM é fundamentada em portões, cujo papel é controlar a quantidade de informação que entra em cada unidade e que, após memorizada, passa para as próximas unidades. A rede é treinada usando BPTT e supera o problema de dissipação do gradiente.

Diferentemente das redes MLPs convencionais, as LSTM possuem blocos de memória conectados através de camadas. Gers, Schmidhuber e Cummins (2000) constataram uma fraqueza das redes LSTM na versão original: maior dificuldade de aprendizado para séries longas que não são segmentadas em subsequências de treinamento. Diante do problema, os autores propõem a inclusão dos portões de esquecimento, que redefinem o conteúdo da célula de memória, excluindo o conteúdo não relevante. A Figura 1.16 mostra a ilustração da unidade básica da rede LSTM.

Figura 1.16 – Célula de memória LSTM



Fonte: Graves, 2014, p. 2.

Um bloco de memória tem um componente para armazenar sequências recentes, que não existe no neurônio padrão. Cada bloco contém portões que gerenciam o estado e a saída do bloco. Há três tipos possíveis de portão: portão de entrada, determina os valores da entrada para atualizar o estado da memória; portão de esquecimento, especifica quais informações jogar fora do bloco; e portão de saída, define o que produzir com base na entrada e na memória do bloco.

Cada unidade de bloco funciona como uma máquina de mini estado. Os portões das unidades têm pesos aprendidos durante o procedimento de treinamento, sendo esse aprendizado feito através do algoritmo de *backpropagation through time* (BPTT).

Com o objetivo de previsão para séries temporais, a possibilidade de descartar valores faz com que os dados ao longo da série de tempo possam ser tratados de forma distinta, dependendo da importância do volume de dados necessário para que o contexto temporal seja compreendido (GERS; SCHMIDHUBER; CUMMINS, 2000). Isso significa que as redes LSTM com portões de esquecimento concedem tratamentos distintos para as memórias de curto e de longo prazo.

3.3.2.4. Hiperparâmetros

A definição dos hiperparâmetros e sua otimização é um processo que ocorre, sobretudo, junto aos modelos de aprendizagem de máquina. Nos modelos estatísticos clássicos existem parâmetros a serem definidos, mas que não têm a mesma funcionalidade, impacto e volume, como ocorrem nos de inteligência computacional.

Distinção a ser feita é em relação aos parâmetros, que são as propriedades dos dados de treinamento apreendidos durante o próprio processo pelo modelo escolhido, como pesos, vieses, ponto de divisão da árvore de decisão etc. Já os hiperparâmetros são as propriedades que sustentam o processo de treinamento em si, sendo configurados antes de iniciado o treinamento, sendo exemplos a taxa de aprendizado, o número de neurônios, a função de ativação etc. Os parâmetros são aprendidos automaticamente, já os hiperparâmetros são definidos manualmente para guiar o processo de aprendizado.

Os hiperparâmetros são importantes por afetarem diretamente o processo de treinamento do algoritmo e o desempenho em si do modelo que está sendo treinado. Cabe destacar que existem hiperparâmetros mais relevantes que outros. Logo, entender os hiperparâmetros cruciais para cada algoritmo torna-se fundamental.

Um hiperparâmetro, em geral, tem seu efeito conhecido de forma ampla, não sendo possível saber previamente o seu impacto (BROWNLEE, 2020b). Nesse sentido, não existe uma solução exata de como definir um hiperparâmetro da forma mais adequada para um determinado conjunto de dados. No mais, alguns modelos de ML têm uma série de hiperparâmetros que podem interagir entre si de várias formas.

Muitos modelos têm parâmetros importantes que não podem ser estimados diretamente a partir dos dados, como, por exemplo, o valor de k do algoritmo KNN (KUHN; JOHNSON, 2013). Esse tipo de parâmetro do modelo deve ser visto como ajustável exatamente porque não há fórmula analítica disponível para calcular um valor apropriado. A maioria dos algoritmos de aprendizado por máquina apresenta hiperparâmetros, sendo que sua quantidade pode variar para cada um deles. Certamente, parte desses hiperparâmetros controla a complexidade do modelo, o que implica necessariamente que valores inadequados levem ao problema de sobreajustamento e, conseqüentemente, falta de capacidade de generalização.

O conjunto de possíveis hiperparâmetros é denominado espaço de busca, que representa o volume a ser pesquisado, em que cada dimensão representa um hiperparâmetro e, cada ponto, uma configuração para o modelo (BROWNLEE, 2020). Assim, um ponto no espaço de busca é um vetor que contempla um valor específico para cada hiperparâmetro existente no modelo. O objetivo do procedimento de otimização é encontrar o vetor que leve ao melhor desempenho do modelo.

Entre as várias estratégias de otimização de hiperparâmetros, as mais comuns são: definição manual dos hiperparâmetros; *Grid Search*, faz a varredura completa do espaço de busca para encontrar a arquitetura que gere o melhor desempenho, sendo extremamente custosa computacionalmente na presença de muitos parâmetros; *Random Search*, em vez de uma varredura de todas as combinações possíveis, uma busca aleatória considera somente parte das combinações possíveis, sendo a meta encontrar os melhores hiperparâmetros nesse subconjunto do espaço de busca; otimização Bayesiana, usa os resultados de cada iteração considerando um determinado conjunto de hiperparâmetros para melhorar ou guiar a próxima iteração para encontrar os melhores hiperparâmetros em termos de desempenho do modelo.

Neste trabalho é aplicada a otimização Bayesiana, considerando sua capacidade de encontrar de forma mais ágil os hiperparâmetros mais adequados para o treinamento do conjunto de dados em questão. A abordagem busca encontrar o valor que minimiza a função objetivo, construindo um modelo de probabilidade com base nos resultados anteriores da métrica de avaliação da função objetivo. A otimização procura o equilíbrio entre a exploração da melhor região que contém os melhores hiperparâmetros e a exploração dessa região para maximizar ou minimizar a métrica de desempenho. Ou seja, a otimização Bayesiana segue a direção de combinações possíveis de hiperparâmetros que melhoram o desempenho da métrica de avaliação do algoritmo selecionado, eliminando várias áreas do espaço de busca que reduzem esse desempenho. O ganho de desempenho e velocidade que a abordagem Bayesiana proporciona é enorme em relação à do *grid search* e do *random search*, sendo uma das melhores escolhas para grandes conjuntos de dados e modelos com muitos hiperparâmetros.

Apesar dos modelos clássicos não trabalharem com hiperparâmetros, algumas configurações podem ser feitas previamente, as quais apresentam um papel muito próximo do conceito-efeito visto até o momento de hiperparâmetros. Já nos casos dos

métodos baseados na metodologia Box Jenkins da década de 1970, a definição dos parâmetros p , d , q , em que p é a ordem (número de defasagens) do modelo autorregressivo, d é o grau de diferenciação e q é a ordem do modelo de média móvel, também poderiam ser equiparados ao conceito de hiperparâmetros. No caso dos modelos de regressão com regularização, os parâmetros alfa e beta desses métodos são hiperparâmetros que devem ser alvo de otimização.

Tabela 1.2 – Similaridade entre hiperparâmetros

Função	XGBoost	Light GBM
Hiperparâmetros importantes para controlar <i>overfitting</i>	<ol style="list-style-type: none"> 1. learning rate - taxa de aprendizado, valores ótimos entre 0.01-0.2; 2. max_depth - profundidade máxima das árvores; 3. min_child_weight - número mínimo de amostras que um nó deve ter para poder ser dividido, valor default é 1. 	<ol style="list-style-type: none"> 1. learning rate - taxa de aprendizado; 2. max_depth - máxima profundidade da árvore, sendo 20 o valor default e é fundamental destacar que a abordagem de crescimento é a <i>leaf-wise</i>, o que reforça a necessidade de ajustamento; 3. num_leaves - número de folhas, é o principal parâmetro para controlar a complexidade do modelo e como boa prática deve ser menor que $2^{\text{max_depth}}$; 4. min_data_in_leaf - é um parâmetro que ajuda a evitar o problema de <i>overfitting</i>, quanto maior o seu valor menor é a probabilidade de ocorrer este problema, seu valor padrão é igual a 20.
Parâmetros para variáveis categóricas	Não disponível.	categorical_feature - campo em que as variáveis categóricas são informadas para treinamento do modelo.
Hiperparâmetros para controlar velocidade de treinamento	<ol style="list-style-type: none"> 1. colsample_bytree - índice de subamostragem de colunas, quando cada árvore é construída; 2. subsample - índice de subamostragem para o treinamento; 3. n_estimators - número máximo de árvores de decisão. 	<ol style="list-style-type: none"> 1. feature_fraction - fração de variáveis que deve ser considerada para cada iteração; 2. bagging_fraction - volume de dados que será utilizado em cada iteração; 3. num_iterations - número de iterações boosting, sendo 100 o valor default.

Fonte: Adaptado de Swalin (2018).

O quadro comparativo dos principais hiperparâmetros para os dois importantes algoritmos utilizados neste trabalho, o LightGBM e o XGBoost, foi feito a partir de Swalin (2018) e é mostrado na Tabela 1.2. Existem muitos hiperparâmetros similares, sendo que, na parte de controle de velocidade de treinamento, os que precisam ser especificados são distintos. Assim como há semelhanças nos modelos de árvores de decisão em relação aos hiperparâmetros que devem ser especificados, o mesmo é observado nos modelos alicerçados em redes neurais contemplados na Tese.

A Tabela 1.3 traz todos hiperparâmetros ajustados por algoritmo neste trabalho. Ressalta-se a existência de outros hiperparâmetros que também poderiam ser ajustados, no entanto, por meio de heurística e consulta de outros trabalhos, constatou-se que esses seriam os mais relevantes. No capítulo de metodologia é feita a especificação de cada um deles e do universo de possibilidades testado.

Tabela 1.3 – Hiperparâmetros escolhidos por algoritmo de aprendizagem de máquina

Randon Forest	LGB	XGB	MLP	CNN	LSTM
Max_depht	Learning Rate	Learning Rate	Neurons amount first layer	Neurons amount first layer	Neurons amount first layer
Max_features	Max depth	Max depth	Neurons amount second layer	Learning Rate	Neurons amount second layer
Min_samples_leaf			Dropout	Activate function	Dropout
Min_samples_split			Learning Rate	Optmizers	Learning Rate
Criterion			Activate function		Activate function
			Optmizers		Optmizers

Fonte: elaboração própria.

O processo de calibração dos hiperparâmetros deve ser feito com o modelo contendo todas as variáveis. Realizar o processo dessa forma impede que a escolha de determinada variável afete a definição dos melhores hiperparâmetros, que são usados no modelo junto com a aplicação da seleção das variáveis mais importantes.

3.4. Aspectos relevantes dos métodos não lineares

Apesar de amplamente utilizados, os métodos lineares apresentam algumas limitações e os não lineares passaram a ganhar relevância como métodos de previsão a serem usados em situações em que exista alguma não linearidade no conjunto de dados estudados. A seguir, são apresentados aspectos importantes relacionados aos métodos não lineares.

3.4.1. Processo KDD e CRISP-DM

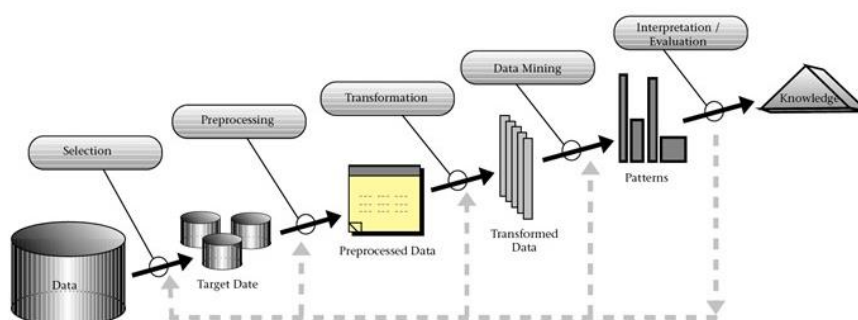
Esta seção evidencia as metodologias KDD, CRISP-DM, SEMMA, que possuem pontos em comum e buscam garantir por meio de várias etapas que seja possível realizar a descoberta de conhecimento útil com base em um conjunto de dados. A KDD é uma metodologia científica, enquanto as demais são amplamente usadas em projetos da área de ciências de dados, mas não têm cunho acadêmico.

O termo “*knowledge discovery in databases*” foi apresentado, em 1989, no primeiro *workshop* de KDD, tendo como principal objetivo evidenciar que o conhecimento é o resultado de um processo baseado em dados. A informação é de Fayyad, Piatetsky-Shapiro e Smyth (1996), que diferenciam os processos KDD e *data mining*. KDD refere-se a um processo mais amplo, composto por um conjunto de passos a serem seguidos, sendo possível a descoberta de conhecimento válido por meio dos dados. O *Data Mining*, por sua vez, seria apenas a etapa de modelagem dos

dados do processo KDD e seus principais propósitos são: previsão e descrição. Para atingir esses objetivos há uma ampla variedade de métodos, sendo que existe uma maior adequação de determinado método dependendo do tipo de problema e de dado.

A Figura 1.17, extraída de Fayyad, Piatetsky-Shapiro e Smyth (1996), mostra somente uma versão básica do KDD, já que todos os passos não são evidenciados.

Figura 1.17 – Etapas do processo KDD



Fonte: Fayyad, Piatetsky-Shapiro e Smyth, 1996, p.41.

O processo do KDD é iterativo e interativo. Em outras palavras, presume uma série composta por nove etapas, as quais podem ser repetidas inúmeras vezes até ser alcançado o objetivo estipulado. As etapas estão descritas a seguir.

- **Entendimento do problema:** identificação do objetivo da aplicação do KDD na perspectiva do cliente;
- **Seleção dos dados:** seleção dos dados brutos em si;
- **Limpeza e pré-processamento:** verificação da qualidade dos dados, sendo necessário, muitas vezes, a aplicação de alguma técnica de limpeza para tratar problemas de dados repetidos, dados anômalos e faltantes;
- **Transformação:** aplicação de um conjunto de técnicas com o intuito de permitir que os dados fiquem disponíveis para a aplicação dos modelos de análise, sendo assim são aplicados métodos que permitam alterar o tipo ou a distribuição do dado, métodos de agregação, de criação de novos atributos, de seleção de variáveis e de redução de dimensionalidade;
- **Correspondência:** correspondência entre os objetivos propostos na primeira fase e um método específico de mineração de dados;
- **Análise exploratória e seleção de modelos e de hipóteses:** análise exploratória mais profunda com o intuito de definir os modelos mais

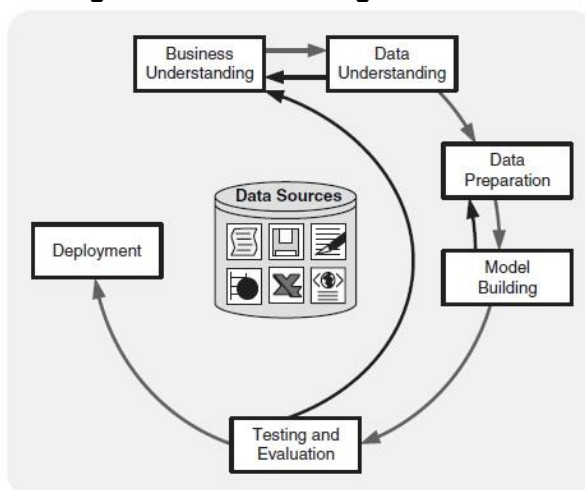
apropriados, juntamente com os seus respectivos parâmetros para o problema especificado;

- **Mineração dos dados:** modelagem em si e aplicação de técnicas de mineração de dados para realizar a descoberta de algum conhecimento, o que pode ser tanto descritivo como preditivo;
- **Interpretação e avaliação:** avaliação do desempenho do modelo a partir de um conjunto de dados fora da amostra de treinamento;
- **Ação sobre o conhecimento descoberto:** o conhecimento adquirido é colocado em prática, podendo ser incorporado em outros sistemas ou ser simplesmente documentado e reportado às partes envolvidas no projeto.

Metodologia bastante usada em projetos *Data Mining* é a CRISP-DM (*Cross industry standard process for data mining*) concebida, em 1996, por Daimler-Chrysler, SPSS e NCR para ser uma metodologia estruturada e robusta para planejamento e realização de projetos de mineração de dados (CHAPMAN *et al.*, 1996). São seis etapas: (1) entendimento do negócio, em que são especificados os objetivos do projeto com base no problema; (2) compreensão dos dados, que contempla a captura e a limpeza dos dados; (3) preparação dos dados, equivalente à etapa de transformação da metodologia KDD; (4) modelagem, visa aplicar as técnicas de modelagens adequadas para o problema proposto; (5) avaliação, busca realizar teste para validar se o modelo gerado é adequado para as necessidades de negócio; e (6) aplicação, que é a entrega final do projeto por meio da apresentação dos resultados obtidos, juntamente com a possibilidade de implementação em produção do modelo proposto, contribuindo para o processo de tomada de decisão dentro da companhia.

Assim como verificado na KDD, a CRISP-DM também é uma metodologia interativa, onde ocorre uma retroalimentação entre as fases, gerando novas descobertas e permitindo um aperfeiçoamento contínuo do processo. Na Figura 1.18 é possível acompanhar as fases do modelo CRISP-DM.

Figura 1.18 – Metodologia CRISP-DM



Fonte: Chapman *et al.* (1996), p.10.

A última metodologia é a SEMMA, acrônimo para as palavras *Sample, Explore, Modify, Model e Assess* e se refere a um tipo de processo para condução de projeto de mineração de dados. Foi criada pela empresa SAS como uma organização lógica para aplicação da ferramenta de DM desenvolvida por eles. Além disso, ela foca, sobretudo, nas tarefas de criação do modelo em si, dando menos importância para as questões de entendimento do negócio e do próprio problema a ser solucionado.

As etapas que compõem o processo SEMMA são: (1) amostra, seleção de uma amostra significativa para se obter a informação necessária por meio de uma análise de dados; (2) explorar, busca por tendência e anomalias nos dados através de recursos gráficos e estatísticos; (3) modificar, modificação e preparação dos dados, presentes também nas outras metodologias, permitindo deixar o dado disponível para a extração do conhecimento; (4) modelar, aplicação de técnicas de modelagem apropriadas para o propósito especificado; e (5) avaliar, validação dos resultados obtidos por meio da aplicação do modelo proposto na etapa anterior, o que deve ser feito, preferencialmente, em um conjunto de dados fora da amostra. A metodologia deve ser vista como uma estrutura simples e de fácil compreensão para o desenvolvimento adequado de projetos de *Data Mining*, permitido que sejam encontradas tanto soluções para problemas de negócios, assim como estabelecidas as estratégias de *Data Mining* para a empresa.

Tabela 1.4 – Sumário de correspondências entre KDD, SEMMA e CRISP-DM

KDD	SEMMA	CRISP-DM
Entendimento do problema		Entendimento do negócio
Seleção	Amostra	
Pré processamento	Explorar	Entendimento dos dados
Transformação	Modificar	Preparação dos dados
Correspondência, análise exploratória, seleção do modelo e hipóteses e Data Mining	Modelar	Modelagem
Interpretação e Avaliação	Avaliar	Avaliação
Ação sobre o conhecimento descoberto		Aplicação

Fonte: Adaptado de Santo e Azevedo (2008), p.185.

Uma análise comparativa, baseada em Azevedo e Santos (2008), permite constatar que as metodologias são similares. Na Tabela 1.4 é mostrado um quadro-sumário com as correspondências entre as três metodologias: KDD, SEMMA e CRISP-DM. Verifica-se que a metodologia SEMMA é equivalente à implementação prática do processo KDD básico de cinco etapas, enquanto que a metodologia CRISP-DM, é também a aplicação prática do processo KDD completo, considerando as seguintes fases não contempladas na metodologia simplificada: entendimento do negócio e dos objetivos do projeto para que seja desenvolvida a solução mais apropriada, análise exploratória, seleção do modelo mais condizente para o problema a ser solucionado e a fase de ação sobre o conhecimento obtido, que é a etapa final de aplicação.

Por fim, cabe expor uma metodologia mais recente e objetiva, descrita por Brownlee (2020), cuja base é o KDD. O processo de aprendizagem de máquina aplicado – ou de ciência de dados – consiste nas seguintes etapas: (1) definição do problema, fase do entendimento profundo do projeto, sendo necessária, muitas vezes, a aplicação de análise exploratória dos dados; (2) preparação dos dados, que compreende todas transformações necessárias para que eles possam ser incluídos na modelagem proposta; (3) avaliação dos modelos, em que é feita a validação dos mesmos, o que pressupõe a definição da métrica de comparação dos modelos e a especificação da divisão da base de dados em treino, validação e teste para sua completa avaliação; e (4) finalização do modelo, que consiste na escolha do modelo final que representará a solução para o problema do projeto.

Apresentadas as possíveis metodologias para desenvolvimento de um projeto de aprendizagem de máquina, é fundamental destacar que uma das etapas mais importantes é a de preparação dos dados, a qual visa avaliar as variáveis mais

relevantes, criar variáveis, fazer a seleção das mesmas e, por fim, apresentar as transformações mais adequadas para que os dados possam ser utilizados nos diversos algoritmos de previsão de vendas selecionados. Sendo assim, a seguir está detalhado todo o processo de preparação dos dados.

Há certo consenso entre os profissionais da área de ciência de dados que a maior parte do tempo em projetos de *machine learning* é gasto com *feature engineering* e com a limpeza dos dados, que são processos de preparação dos dados. Entretanto, apesar da sua relevância, esse tópico não é discutido na amplitude necessária.

3.4.1.1. Preparação dos dados

Em muitos projetos de aprendizagem de máquina, os dados não podem ser usados em sua forma bruta na modelagem em si por uma série de razões, tais como: alguns algoritmos não aceitam trabalhar com dados categóricos, somente numéricos; presença de ruídos e erros nos dados que precisam ser corrigidos previamente; alguns algoritmos de aprendizado por máquina impõem que os dados sejam apresentados de acordo com alguns requisitos, entre outros (BROWNLEE, 2020).

Assim, dados brutos necessariamente devem ser pré-processados para que possam ser aplicados em um determinado modelo. Em um projeto de aprendizado por máquina, essa é a etapa de preparação dos dados, também chamada de: organização dos dados, limpeza dos dados, pré-processamento dos dados e *feature engineering*.

A relevância do processo de preparação dos dados já se justificaria pelo fato em si dos dados não poderem ser aplicados em sua forma bruta, devido a problemas de presença de ruídos e/ou de requisitos dos algoritmos escolhidos para modelagem. Mais do que isso, por meio do processo de preparação dos dados torna-se possível escolher uma representação melhor dos mesmos, que consiga expor de forma efetiva a sua estrutura subjacente, garantindo que o algoritmo selecionado possa apreendê-la e alcance o melhor desempenho possível frente aos recursos disponíveis.

Bibliotecas de código aberto estão cada vez mais disponíveis para os mais diversos algoritmos de aprendizagem de máquina, podendo ser aplicados diretamente, sem grandes adaptações. Sendo assim, o que torna único cada projeto são os dados utilizados e o processo da sua respectiva preparação. Interessante notar

que em muitas das competições do Kaggle – a maior plataforma de hospedagem para projetos e competições de ciências de dados – existem duas grandes estratégias vencedoras para a maioria das suas competições de aprendizado por máquina: aplicação de *feature engineering* e/ou implementação de redes neurais profundas.

Estudo de 2020, da Cognilytica, registrou e analisou respostas de organizações, agências e empresas de usuários finais a fim de identificar a distribuição do tempo em projetos de inteligência artificial e aprendizagem de máquina. Verificou-se que tarefas de preparação dos dados consomem 80% do tempo dos cientistas, o que inclui rotulagem, anotação, limpeza, aumento e enriquecimento de dados. Por um lado, a constatação pode ser vista como positiva, já que dados com boa qualidade contribuem para a construção de bons modelos. Por outro, idealmente espera-se que cientistas de dados gastem a maior parte do tempo realizando análises avançadas, treinando e avaliando os modelos para, assim, colocá-los em produção. Apesar da grande concentração do tempo gasto em tarefas de preparação de dados, existe esforço crescente na indústria de ciência de dados para automatizar cada vez mais essas tarefas, permitindo que cientistas aloquem mais tempo a outras atividades.

Várias técnicas podem ser aplicadas em projetos de aprendizagem de máquina, o que depende dos dados em si e dos algoritmos selecionados para o projeto (BROWNLIE, 2020). Ademais, as técnicas podem ser agrupadas conforme o objetivo de tratamento a ser empregado nos dados. A seguir estão algumas técnicas:

- **Limpeza de dados:** visa identificar e corrigir erros nos dados;
- **Seleção de variáveis:** busca identificar as variáveis que são mais relevantes para o modelo;
- **Transformação dos dados:** muda a escala ou a distribuição das variáveis;
- **Engenharia de variáveis:** busca derivar novas variáveis a partir das variáveis existentes;
- **Redução de dimensionalidade:** compactam os dados através da sua projeção.

Ressalta-se que as técnicas de seleção de variáveis e de redução de dimensionalidade têm o objetivo comum de reduzir o número de variáveis de entrada do modelo, mas de formas distintas, o que está detalhado mais adiante. Além disso, é importante frisar que, neste trabalho, o foco está na aplicação de tarefas de

transformação dos dados, de seleção de variáveis e de engenharia dos dados, que possam contribuir para gerar um modelo de previsão de vendas para o varejo de moda mais parcimonioso e preciso.

Em termos de sequência das atividades pertencentes ao processo de preparação dos dados, pode-se estabelecer o seguinte encadeamento: (1) limpeza dos dados, (2) engenharia de variáveis, (3) transformação dos dados, (4) seleção de variáveis, e (5) redução de dimensionalidade. Importante destacar que nem todas as etapas aparecem de forma conjunta em um projeto de aprendizagem de máquina. Cada uma das atividades está detalhada a seguir.

3.4.1.2. Limpeza dos dados

O processo de limpeza dos dados lida com problemas de erros sistemáticos ou erros de dados sujos, segundo Brownlee (2020). Alguns exemplos de erros são: dados anômalos, dados faltantes, dados redundantes, dados vazios etc. Para cada um desses problemas, existe um tratamento adequado, que implica, necessariamente, algum ajuste nos dados. Na maioria dos casos, o tratamento aplicado ou eliminará o dado faltante ou completará esses dados com alguma estatística ou modelo de aprendizado. Isto se faz necessário, pois muitos algoritmos não conseguem lidar com dados faltantes. A Figura 1.19 mostra um quadro-resumo com os principais erros nos dados que devem ser ajustados.

Os dados usados neste trabalho apresentaram poucos problemas. Contudo, ressalta-se que parte dos dados têm frequências distintas, o que significa ausência de dados. Sendo assim, optou-se pela técnica de interpolação para preencher esses dados faltantes. Há várias técnicas de interpolação, como linear, polinomial e *spline* cúbica, sendo a última a implantada neste trabalho. Especificamente, a *spline* cúbica é uma técnica de aproximação, que busca dividir o intervalo de interesse em vários subintervalos e, posteriormente, interpolar, da forma mais suave possível, nesses subintervalos com polinômios de grau três.

Figura 1.19 – Visão geral do processo de limpeza dos dados



Fonte: Brownlee (2020), p.18.

Cabe salientar que a limpeza de dados deve ser a primeira etapa de todo o processo de preparação dos dados.

3.4.1.3. Transformação dos dados

Etapa bastante relevante no processo de preparação dos dados é a sua transformação, que consiste na mudança do tipo ou da distribuição dos dados. Essa transformação muitas vezes é pré-requisito de alguns algoritmos, com o intuito de garantir a convergência dos métodos de previsão empregados. Adicionalmente, a transformação do tipo do dado pode poupar memória computacional, quando, por exemplo, dados numéricos são transformados em inteiros.

Considerando que, neste trabalho, são tratados somente dados numéricos, temos dois grandes grupos de técnicas, as quais são:

- **Técnicas de mudança de escala (dimensionamento):** normalização, padronização e padronização robusta;
- **Técnicas de mudança de distribuição:** desratização, quartílica e gaussiana.

Em conjuntos de dados com distribuições distorcidas, antes de aplicar a técnica de mudança de escala, por vezes é sugerido aplicar técnicas para tornar a distribuição mais simétrica, tais como: transformação logarítmica, Box-Cox etc. Certamente, essa transformação inicial já produzirá um conjunto de dados mais simétrico.

Outra transformação de dados possível é o dimensionamento, método de transformar dados em um intervalo específico. Ao fazer isso, as alterações em diferentes variáveis se tornam mais comparáveis. A escala pode ser feita com normalização ou padronização. Essa transformação é essencial, sobretudo, na

aplicação de algoritmos de regressão ou de cálculo de distâncias, como, por exemplo *K-means* e *KNN*, pois são muito sensíveis à variação de magnitude e intervalo entre recursos, e nas próprias redes neurais, para garantir a convergência da rede, considerando o grande número de parâmetros que devem ser calculados.

Brownlee (2020c) diz que, para variáveis numéricas, a aplicação das técnicas de dimensionamento garante uma maior representatividade do conjunto de dados, quando comparado ao conjunto de dados na escala original. O dimensionamento para o intervalo de 0 a 1 é conhecido como normalização ou dimensionamento *min-max*.

Caso os dados apresentem uma distribuição gaussiana, pode ser mais útil mudá-los para um padrão gaussiano com média zero e desvio padrão de um, conhecida como técnica de padronização. Outra abordagem de padronização é a robusta, recomendada na presença de dados anômalos. Essa técnica de dimensionamento desconsidera do cálculo da média e do desvio padrão os valores dados anômalos e, em seguida, usa os valores calculados para dimensionar a variável. A variável resultante apresentará média e mediana zero e desvio padrão de 1 não distorcida por dados aberrantes e os mesmos continuam presentes com as mesmas relações relativas a outros valores.

Por último, deve ser considerado que o processo de transformação dos dados é aplicado de forma individual para cada variável, sendo possível aplicar mais de uma transformação em uma mesma variável.

3.4.1.4. Tratamento de sazonalidade

Os componentes de uma série de temporal são: tendência, sazonalidade, ciclo, erro e constante. Os três primeiros aspectos podem ser endereçados com tratamentos específicos facilitando o processo de modelagem e previsão em si. Adicionalmente, esses tratamentos compõem o processo de preparação dos dados, já que devem ser realizados antes que os mesmos possam ser modelados.

A sazonalidade presente nos dados pode ser definida como qualquer padrão de comportamento em que a frequência desse comportamento é estável em um momento do tempo (NIELSEN, 2019). A identificação e o tratamento da sazonalidade podem ser vistos de duas maneiras: (1) como uma das etapas do processo de

modelagem, e (2) como uma forma de limpeza dos dados, como ocorre na forma de divulgação de muitos dados governamentais, em que os dados são dessazonalizados.

Em termos do tratamento sazonal, são aplicadas duas metodologias neste trabalho: ajuste da sazonalidade por modelagem e diferenciação sazonal.

O ajuste da sazonalidade, conforme Brownlee (2020), é feito através da modelagem. Geralmente, o componente sazonal de uma série é uma onda senoidal com período e amplitude fixos, o que permite ser facilmente aproximado pelo método de ajuste de curva. Sendo assim, é feita a previsão da sazonalidade da série em estudo por meio de uma função polinomial. Na maioria dos casos, o uso de um polinomial de até quinto grau é suficiente para modelar os diversos padrões de sazonalidades existentes. Neste trabalho foi aplicado um polinômio de terceiro grau.

Com a curva de sazonalidade calculada, é possível utilizá-la como referência para realizar novas estimativas. Importante destacar, que para obter os valores dessazonalizados de um determinado conjunto de dados, basta realizar a subtração entre o valor real observado da série e o valor da curva prevista.

A presença da tendência em uma série de dados torna essa série não estacionária. De acordo com Bueno (2008), uma série temporal é estacionária se sua média, variância e covariância permanecem constantes, independente do período em que foram analisadas, ou seja, elas não variam no decorrer do tempo.

$$\text{Média: } E[y_t] = \mu$$

$$\text{Variância: } Var(y_t) = E[(y_t - \mu)^2] = \gamma = \sigma^2$$

$$\text{Covariância: } \gamma_k = E[(y_t - \mu)(y_{t-k} - \mu)]$$

Gujarati (2006) salienta que, na maioria dos procedimentos de análise estatística de séries temporais, supõe-se que essas séries sejam estacionárias. Caso a série não seja estacionária, deve-se aplicar as seguintes transformações com o intuito de eliminá-la: logaritmo, diferenciação e diferenciação sazonal.

Ao se falar do aspecto da estacionariedade para os métodos de aprendizagem de máquina, sobretudo os profundos, a principal promessa por parte das redes recorrentes é que a dependência temporal dos dados de entradas pode ser aprendida, não sendo necessário especificar uma janela fixa. No caso das redes recorrentes, uma

corrente de estudiosos defende que essas redes são capazes de aprender tais estruturas sistemáticas, não sendo preciso sua eliminação.

Por outro lado, vários autores sustentam que uma boa prática para modelagem com redes neurais para séries temporais é o ajuste para eliminação da tendência e da sazonalidade, sendo possível realizar a modelagem após esse tratamento inicial (NELSON *et al.*, 1999). Embora a estacionariedade não seja pré-requisito para esses métodos, assim como é nos métodos estatísticos clássicos na prática, mesmo as técnicas mais avançadas de aprendizagem de máquina não conseguem lidar bem com dados em que um componente de tendência esteja presente (NIELSEN, 2019). Portanto, há a necessidade da aplicação de um pré-processamento para ajustar os dados. Na parte prática deste trabalho são avaliadas tanto a realização do pré-processamento em termos de ajuste sazonal e de tendência como a não realização desses processos e os resultados obtidos.

De forma geral, um modelo de série temporal, cuja série não é estacionária, varia em sua precisão à medida que as métricas da série temporal variam. Assim, se o objetivo maior é determinar um modelo para estimar a média de uma série temporal com uma média e variância não estacionárias, a tendência e o erro do modelo irão variar ao longo do tempo, tornando o modelo criado discutível (NIELSEN, 2019).

Utiliza-se a transformação logarítmica quando a estacionariedade está ligada, de certa maneira, à variância da série. No final da modelagem é imprescindível realizar a transformação inversa.

A técnica mais usada para tornar uma série estacionária é a diferenciação, que consiste em tirar a primeira diferença da variável em questão y_t : Δy_t . O processo deve ser repetido até se atingir a estacionariedade. A seguir estão demonstradas, respectivamente, as equações da primeira e da n -ésima diferença:

$$\Delta y_t = y_t - y_{t-1} \quad \Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}$$

Quando autocorrelações muito grandes forem observadas em defasagens espaçadas n períodos adiante e com periodicidade, pode-se lançar mão da diferenciação sazonal para ajustar a série. Abaixo, as equações de diferenciação sazonal são apresentadas:

$$\begin{aligned}
\Delta_S^1 y_t &= \Delta_S y_t = y_t - y_{t-s} \\
\Delta_S^2 y_t &= \Delta_S y_t - \Delta_S y_{t-s} \\
&\dots \\
\Delta_S^D y_t &= \Delta_S^{D-1} y_t - \Delta_S^{D-1} y_{t-s}
\end{aligned}$$

Importante destacar que, muitas vezes, é possível constatar a estacionariedade de uma série apenas pela análise visual da mesma. No entanto, existe uma série de testes estatísticos para checar esse aspecto. Em relação a esse aspecto, neste trabalho são aplicados os seguintes testes: Dickey-Fuller aumentado (ADF), Phillips-Perron (PP) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS). É fundamental a aplicação de mais de um teste, pois o fato de apresentarem diferenças metodológicas pode levar a resultados distintos.

3.4.1.5. Feature Engineering

A denominação *feature engineering* está relacionada ao contexto específico de criação de novas variáveis, apoiado nas variáveis de entradas presentes no modelo. Esse processo também pode ser chamado de geração de variáveis.

Em muitos trabalhos, o termo *feature engineering* é equiparado à preparação de dados, termo usado nesta Tese. Sendo assim, englobaria todos os processos de tratamentos das variáveis vistos até o momento. Porém, segundo Brownlee (2020), o processo de engenharia de novas variáveis é algo específico do conjunto de dados estudado, o que demanda, muitas vezes, a colaboração de um especialista no assunto. Algumas técnicas de criação de variáveis são: (1) adicionar uma marcação fundamentada no valor de alguma variável; (2) fazer o agrupamento ou aplicar alguma estatística descritiva dos dados, como a média; (3) adicionar novas variáveis por meio de componentes de uma variável, como ocorre no caso de variáveis de data em que se pode, por exemplo, extrair semana, mês e ano; e (4) criar cópias de variáveis a partir da aplicação de uma operação matemática.

É fundamental destacar que a criação de novas variáveis é um dos melhores caminhos para aprimorar a performance dos modelos. Para esse processo é necessário um pouco de criatividade e experimentação para criar variáveis que sejam

relevantes para o modelo. Alguns exemplos: combinar variáveis, calcular frequências em um período definido e calcular prazo de ocorrência de determinado fato.

Filho (2019a) afirma que, para modelos de aprendizado por máquina de séries de tempo, existem quatro tipos de variáveis que devem ser criadas e testadas para qualquer modelo, as quais podem ser verificadas a seguir:

- **Lag:** uso de um valor anterior da variável alvo com uma variável preditora;
- **Diferenças:** utilização do cálculo de diferença de valores entre dois períodos anteriores. A operação mais comum é a subtração, mas também podem ser aplicadas somas, razões e multiplicações;
- **Rolagem estatística:** é possível criar variáveis com alguma estatística descritiva dos dados, tais como média, desvio padrão, máximo e mínimo;
- **Decomposição da data:** é possível criar variáveis como semana, mês, dia, hora através da variável data presente no modelo, o que colabora para especificação de padrões de sazonalidade. Outra possibilidade dentro dessa categoria, é a capacidade de calcular períodos entre duas datas.

Em casos de modelos que têm mais de uma categoria de produto ou cliente, por exemplo, os tipos de variáveis mencionadas podem ser criados para cada classe, sendo gerada uma subsérie, ou podem ser criadas de forma agregada.

O processo de geração de variáveis é limitado pela natureza do dado, criatividade, capacidade de programação e domínio de conhecimento específico de quem desenvolve o modelo, segundo Niesel (2019). Apesar dessas particularidades, a autora aponta algumas variáveis simples aplicadas em projeto de séries temporais: média, variância, diferença entre o primeiro e o último valor, número de pontos de máximo e de mínimo, suavização da série e periodicidade e autocorrelação das séries.

A mesma autora destaca três pontos relevantes para a realização do processo conjunto de criação e seleção das variáveis: (1) disponibilizar os dados de forma adequada para uso em algoritmos de aprendizado de máquina, já que a maioria é projetada para aceitar conjuntos de recursos por ponto de dados em vez de uma série temporal, (2) sumarizar os dados em indicadores relevantes para análise exploratória e armazenamento, e (3) resumir os dados de forma mais ampla e em métricas relevantes, podendo torná-los comparáveis com outras séries.

Considerando as técnicas expostas, praticamente todas são aplicadas neste trabalho, mais especificamente: extração de ano, mês e data; cálculo de média e desvio padrão móveis; uso de defasagens e uso de variáveis com diferença.

Aspecto interessante posto por Nielsen (2019) é a série de esforços para automação da criação das variáveis para séries temporais, cujos principais pacotes são: tsfresh, do Python, Cesium time series analysis platform e tsfeatures, do R. Muitas das variáveis geradas por essas bibliotecas são computacionalmente custosas e, considerando a extensão de variáveis possíveis, é necessário realizar um filtro das quais sejam relevantes para solucionar o problema definido previamente.

Com base no total de variáveis existentes, que contempla tanto as variáveis originais quanto as criadas, o próximo passo no processo de preparação dos dados consiste na seleção de variáveis e/ou redução de dimensionalidade.

3.4.1.6. Seleção de variáveis

A seleção de variáveis é o processo de redução do número de variáveis de entrada, considerando o projeto de desenvolvimento de um modelo preditivo. Busca-se reduzir o número de variáveis de entrada para diminuir o custo computacional da modelagem e, em muitos casos, melhorar o desempenho do modelo em si. Para Guyon e Elisseeff (2003), há três objetivos no processo: melhorar o desempenho da previsão do modelo em questão, determinar preditores mais rápidos e econômicos, e dar melhor compreensão do processo subjacente que gerou os dados.

Considerar características desnecessárias aumenta o tempo de treinamento do modelo, diminui a capacidade de interpretabilidade e, fundamentalmente, reduz a capacidade de generalização do modelo proposto no conjunto de dados de teste. Nesse sentido, Kuhn e Johnson (2013) dizem que o foco do processo de seleção de variáveis deve ser remover preditores não informativos ou redundantes do modelo.

As técnicas de seleção de variáveis eliminam as variáveis preditoras que não são úteis para reduzir a complexidade do modelo resultante (ZHENG; CASARI, 2018). O objetivo final é produzir um modelo parcimonioso com um baixo custo de computação e com pouca ou nenhuma degradação na precisão preditiva.

Do ponto de vista de benefícios, um modelo com menos variáveis preditoras tende a apresentar maior poder interpretativo e é computacionalmente menos

custoso. Em termos estatísticos, é mais atrativo prever um conjunto menor de parâmetros e considera-se que alguns algoritmos de aprendizado por máquina são mais prejudicados pela inclusão de variáveis preditoras não informativas (KUHN; JOHNSON, 2013).

No contexto da ciência de dados, o processo de seleção de variáveis é essencial face a todas variáveis disponíveis, devido a três pressupostos presentes nessa área de conhecimento: (1) a maldição da dimensionalidade, que implica ter um número muito superior de variáveis em relação ao volume de exemplos de dados a serem treinados, levando ao problema de sobreajuste, que, conseqüentemente, produz um modelo com baixo poder de generalização; (2) a navalha de Occam, também conhecida como a lei da parcimônia, a qual pressupõe que entre dois modelos com mesmo nível de desempenho, o modelo a ser selecionado é o mais simples; e (3) *garbage in, garbage out*, significa que a inclusão de variáveis com baixa capacidade informativa impacta negativamente o modelo de previsão.

Dentro dos agrupamentos de técnicas supervisionadas, é possível subgrupá-las nas seguintes categorias, conforme Zheng e Casari (2018):

- **Técnicas de filtragem:** as variáveis não informativas são removidas do modelo por meio de um processo de filtragem. Exemplo é o cálculo da correlação entre as variáveis de entrada e a variável resposta, sendo filtradas as variáveis que ultrapassem um determinado limiar;
- **Técnicas *wrapper*:** são técnicas mais custosas computacionalmente, que avaliam diversos subconjuntos de variáveis de entrada e selecionam as que geram o melhor desempenho, considerando o algoritmo e a métrica de acurácia definidos previamente;
- **Técnicas *embedded*:** os próprios algoritmos de modelagem já contemplam técnicas de seleção de dados de forma intrínseca no seu processo de treinamento. Esse tipo de técnica pode ser verificado nos seguintes modelos: árvore de decisão, regressão lasso, *random forest* etc.

Considerando o método de filtragem e que o modelo exposto neste trabalho compreende somente variáveis de entrada e de saída numéricas, as técnicas mais comuns para este caso são as de correlação. Nesse sentido, recomenda-se a aplicação do coeficiente de correlação de Pearson, no caso de correlação linear, e do

coeficiente de Spearman, no caso de correlação não linear. Salienta-se que, na presença de variáveis categóricas, outros métodos de filtragem são recomendados.

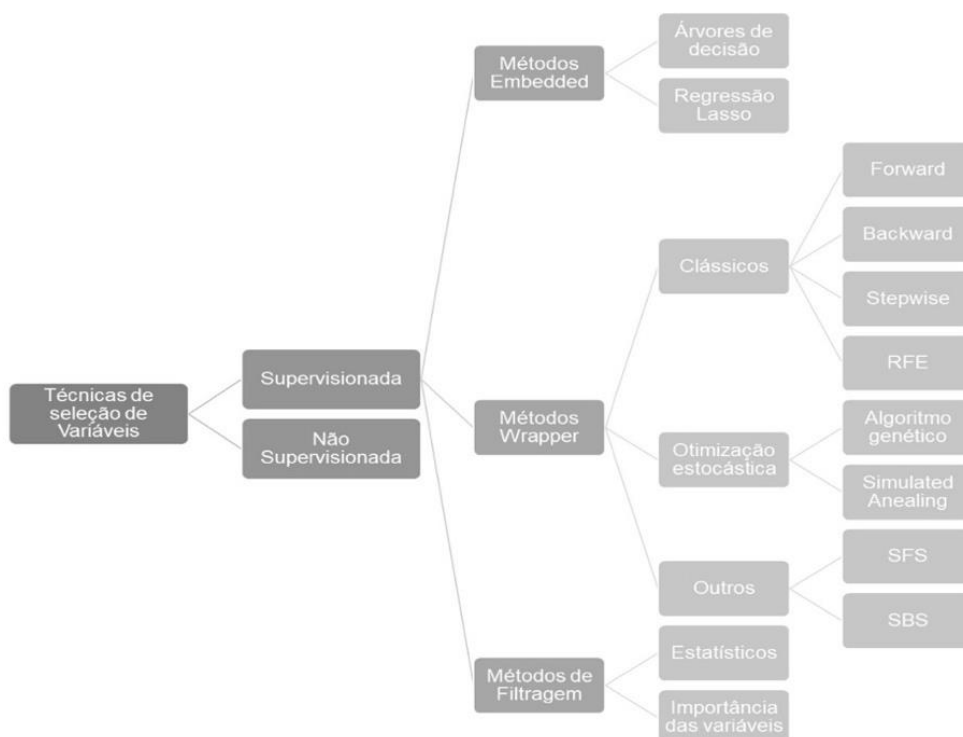
O pacote *tsfresh* abordado anteriormente também apresenta o algoritmo FRESH, cujo objetivo é a seleção de variáveis por meio de testes de hipóteses. O algoritmo avalia a importância de cada variável de entrada em relação a uma variável resposta por meio do cálculo do p-valor (uma estatística de significância) para cada variável preditora. Uma vez calculados, os p-valores por variável são avaliados em conjunto através do procedimento do Benjamini-Yekutieli, que determina quais recursos manter no modelo com base nos parâmetros de entrada a partir de taxas de erro aceitáveis e similares.

No caso dos métodos *wrapper*, há várias técnicas para executar o processo de busca entre as diversas variáveis e definir o subconjunto das mesmas que contribuem para gerar modelos melhores. Ou seja, é definido um modelo para fazer a seleção das variáveis. Os métodos mais simples e clássicos, que sistematicamente incluem e excluem variáveis até que nenhuma melhoria adicional possa ser obtida, são: seleção *forward*, seleção *backward*, *Stepwise* e seleção RFE (*recursive feature elimination*), cujo objetivo é remover recursivamente as variáveis e, assim, construir o modelo com base nos preditores remanescentes, sendo necessário levar em conta o coeficiente de importância de cada variável para determinar a sua eliminação do modelo.

O problema principal da abordagem *wrapper* é que o espaço de busca entre todas as variáveis possíveis é extremamente vasto e procurar a melhor entre todas as combinações é um processo lento e custoso computacionalmente. Nesse sentido, consoante a Brownlee (2020) e a Kuhn e Johnson (2013), é sugerido o uso de algum algoritmo de busca global estocástico, em que o espaço de busca não seja explorado de forma exaustiva, tais como algoritmo genético e *simulated annealing*. Em relação aos métodos mais simples, os métodos mencionados são mais eficientes, mas apresentam um alto custo computacional na presença de grandes volumes de dados para treinamento e na implementação de modelos mais sofisticados.

A Figura 1.20 apresenta um quadro-resumo com as principais técnicas de seleção de variáveis juntamente com a sua respectiva categorização.

Figura 1.20 – Quadro-resumo de técnicas de seleção de variáveis



Fonte: Adaptado de Brownlee (2020), p.19.

Além dos métodos citados, Brownlee (2020) adiciona o método em que um modelo é utilizado para determinar a pontuação das variáveis de entrada e um método de filtragem de seleção de variável que é aplicado posteriormente. Desarda (2018) acrescenta os métodos de seleção de variáveis de forma sequencial, em que um algoritmo de busca é aplicado para reduzir o espaço total de variáveis disponíveis. Existe tanto o método para frente e para trás e, assim, as variáveis são incluídas ou removidas com base na melhora do desempenho da métrica de avaliação do modelo até que se chegue em um subconjunto máximo de k variáveis previamente definidas. É fundamental destacar que qualquer variável apenas entrará ou permanecerá no modelo se contribuir positivamente para aumentar a qualidade de predição.

Dada a sua aplicação efetiva neste trabalho, a seleção para frente sequencial (SFS) é amplamente utilizada por sua simplicidade e por sua velocidade (MARCANO-CEDENTO *et al.*, 2010).

O algoritmo SFS é um procedimento de pesquisa de baixo para cima que começa com um conjunto vazio de variáveis S e, gradualmente, adiciona variáveis por meio de alguma métrica de desempenho, como, por exemplo o erro médio quadrático. Em cada iteração, a variável a ser incluída no conjunto S é selecionada entre as

preditoras disponíveis que ainda não pertencem a esse conjunto S . Dessa forma e necessariamente, o conjunto de variáveis selecionadas S completo deve produzir um erro menor quando comparado a outro conjunto de variáveis com mais variáveis.

Kuhn e Johnson (2013) apontam algumas desvantagens dos métodos de filtragem quando comparados aos *wrapper*: (1) a validação de cada preditor é feita de forma univariada, podendo levar a um problema de multicolineariedade, (2) as variáveis de entrada são avaliadas antes do treinamento do modelo, o que faz com que a avaliação das mesmas esteja apartada da avaliação do modelo em si, e (3) caso sejam considerados testes de hipótese, como o nível de significância para determinar quais variáveis apresentam relações estatisticamente significativas com a variável resposta, pode-se incorrer no problema de multiplicidade, que é o crescimento do erro tipo I, isto é, a probabilidade do falso positivo aumenta exponencialmente. Por outro lado, as técnicas de filtragem apresentam como vantagem baixo custo computacional quando comparadas com as técnicas *wrapper*, pois o critério de seleção não está diretamente relacionado com a eficácia do modelo selecionado.

As técnicas *embedded* não são tão eficazes quanto os métodos de *wrapper*, mas são mais acessíveis e menos custosas. Comparadas às técnicas de filtragem, os métodos *embedded* selecionam variáveis específicas ao modelo. Nesse sentido, atingem um equilíbrio entre custo computacional e qualidade dos resultados.

Independentemente do método de seleção de variável, Kuhn e Johnson (2013), alertam para o problema de viés de seleção, isto é, o processo de seleção de variáveis deve ser considerado como uma das etapas de determinação do modelo, o que implica que essa seleção deve ser verificada em uma amostra apartada dos dados com o intuito de constatar sua efetividade.

Essencial frisar que, assim como não existe o melhor algoritmo de aprendizagem de máquina, também não há a melhor técnica de seleção de variáveis. Deve-se procurar a técnica que seja mais adequada para o problema estabelecido, o que é descoberto através de um processo de exploração e experimentação.

Conclui-se, então, dessa seção, que o processo de seleção de variáveis tem sido amplamente utilizado para reduzir a dimensionalidade dos dados. Essa redução leva a uma melhora do desempenho de classificação, da função de aproximação e dos sistemas de reconhecimento de padrão em termos de velocidade e simplicidade. Adicionalmente, na grande maioria dos casos, gera um aumento de precisão.

3.4.1.7. Redução de dimensionalidade

A redução de dimensionalidade é uma ferramenta que pode ser implantada ao longo do processo de preparação dos dados, com o intuito de, efetivamente, permitir que o algoritmo selecionado consiga aprender a estrutura existente nos dados e, assim, alcançar um desempenho superior considerando o objetivo definido.

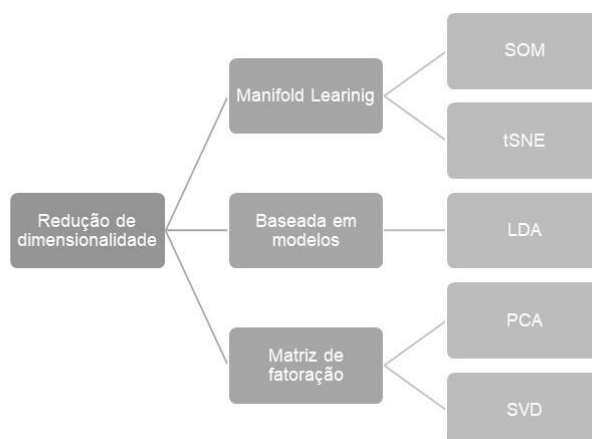
Segundo Brownlee (2020), quanto mais variáveis possui um espaço a ser modelado, mais provável é que o conjunto de dados apresente uma amostragem muito esparsa e, conseqüentemente, não representativa. Esse é o conceito subjacente à maldição da dimensionalidade. Tal fato reforça a redução do número de variáveis, o que pode ser feito por meio da criação de uma projeção dos dados em um espaço de dimensão inferior que, apesar de reduzido, ainda preserva as propriedades mais importantes dos dados originais. Isso geralmente é denominado como redução de dimensionalidade e oferece uma alternativa para as técnicas de seleção de variáveis vistas anteriormente. Ao contrário do processo de seleção de variáveis, no processo de redução de dimensionalidade, como ocorre o processo de transformação dos dados, a interpretação das variáveis transformadas é mais complexa.

A abordagem mais comum para a redução da dimensionalidade é o uso de uma matriz técnica de fatoração, tendo como principais métodos: análise de componente principal (PCA) e decomposição em valores singulares (SVD).

Consoante Kuhn e Johnson (2013), o método PCA muitas vezes é aplicado quando no conjunto de dados analisados há mais dimensões do que amostras de dados em si ou quando os preditores são extremamente correlacionados. Nesse sentido, a aplicação do PCA permite reduzir a dimensionalidade e gerar novas combinações de variáveis preditoras descorrelacionadas.

A Figura 1.21 apresenta um quadro-resumo considerando também outros mecanismos de redução de dimensionalidade.

Figura 1.21 – Visão geral das técnicas de redução de dimensionalidade



Fonte: Brownlee (2020), p.23.

Aspecto a ser exposto é que o processo de seleção de variáveis está diretamente relacionado ao propósito de redução de dimensionalidade, já que ambos os métodos têm como objetivo diminuir o número de variáveis de entrada para um modelo preditivo. A diferença é que os métodos de seleção de atributos, efetivamente, determinam a manutenção ou a retirada de uma característica do conjunto de dados, enquanto a redução da dimensionalidade cria uma projeção dos dados, resultando em variáveis totalmente alteradas, diferentes dos dados originais.

Diante de todos os processos expostos para a preparação dos dados, indo desde os processos de limpeza até os de redução de dimensionalidade, é claramente perceptível porque a etapa de preparação dos dados consome 80% do tempo de um projeto de aprendizado de máquina. E, mais do que isso, evidencia a importância dessa etapa para garantir a construção de um modelo com desempenho superior.

3.5. Métricas de avaliação de previsões

Há diferentes métricas para realizar a avaliação dos métodos de previsão usados (PASSARI, 2003). Alguns dos critérios são: acurácia, habilidade do analista, características desejadas de previsão, características específicas do problema, custo, facilidade do uso, requerimento dos dados, estabilidade, disponibilidade de *software*, velocidade, facilidade de interpretação, eficiência, viés e capacidade de incorporar a experiência do gestor. A escolha deve ser feita com base nas características da série

a ser prevista e no funcionamento de cada métrica existente, considerando o critério selecionado.

A acurácia é o critério mais usado para fazer a avaliação dos métodos de previsão, com base na revisão sistemática da literatura segundo Aczel (1993 *apud* PASSARI, 2003, p. 24).

Analisar se um modelo é bom o suficiente depende dos objetivos definidos previamente e da qualidade e natureza dos seus dados. Mesmo assim, Nielsen (2019) estabelece alguns pontos a serem analisados como oportunidades de melhoria em um modelo. Nesse sentido, as análises a serem feitas são: realize o gráfico comparativo do dado real em relação ao previsto no conjunto de teste para verificar grandes inconsistências; analise os resíduos do modelo ao longo tempo, o que permite constatar que a falta de homogeneidade leva a um modelo subespecificado; analise se o modelo proposto consegue ser superior a um modelo *Naïve*, sendo negativa a resposta, certamente o modelo possui algum problema crítico; verifique como o modelo lida com dados anômalos, se o modelo lidar bem com eles, pode ser um sinal de sobreajuste ou que a seleção da função de perda não foi adequada; por fim, verifique se o modelo proposto trata padrões temporais semelhantes da mesma forma.

Assim como Passari (2003), Zheng (2015) indica diversas métricas de avaliação para tarefas de regressão, todas com foco em acurácia, sendo as principais RMSE e os percentis de erro, representados pelo MAPE, tendo em vista que o RMSE é uma média extremamente afetada por dados anômalos, não sendo tão robusto. Ademais, o MAPE apresenta boa interpretação. Dica de Zheng (2015) é que a métrica mais adequada considere aspectos do negócio para o qual o modelo está sendo concebido.

Para Zhang, Patuwo e Hu (1998), RMSE e MAPE são métricas de avaliação de modelos, geralmente, utilizadas em trabalhos relacionados com redes neurais.

A acurácia foi estabelecida como principal métrica de avaliação para este trabalho. Adicionalmente, os pontos colocados anteriormente por Nielsen (2019) são levados em consideração para cada um dos algoritmos aplicados.

Medir o nível de precisão de uma previsão nem sempre é algo simples e não existe uma métrica que seja adequada para todos os tipos de situações. Primeiro ponto a ser colocado é que toda previsão tem uma magnitude de viés, que indica a

direção geral do erro médio, e tem um nível de precisão, que indica a intensidade do erro. Nielsen (2019) destaca como principais métricas de acurácia da previsão: o erro médio percentual absoluto (MAPE), o erro absoluto médio (MAE), a raiz do erro quadrático médio (RMSE) e o erro médio ao quadrado (MSE), que é o quadrado do RMSE, porém computacionalmente mais rápido.

Das métricas citadas no último parágrafo, somente o MAPE pode ser visto como medida padronizada, justificativa para seu amplo uso, porém não é tido com boa métrica em termos de precisão. O RMSE dá mais importância para erros maiores, enquanto MAE dá a mesma importância para cada erro. Dessa forma, uma otimização do RMSE busca fazer uma correção da média. Por outro lado, a otimização através do MAE tem como objetivo alcançar a mediana do total a ser previsto. Vandepu (2019) diz que para previsões de demanda com grande variância ou intermitentes, o RMSE é um melhor indicador, pois não irá gerar o viés que o MAE gera.

Os pontos negativos do MAPE que levam a não recomendação do uso da métrica em sua versão original, a partir de Hyndman e Athanasopoulos (2013), são: valores de erro iguais em relação a um valor real geram erros percentuais distintos, dependendo da base; quando o valor real é muito pequeno pode gerar erros percentuais muito grandes, ultrapassando 100% e essa métrica não pode ser aplicada diretamente no caso dos algoritmos *Naive*, em que a previsão é exatamente o valor anterior. No caso do *Naive* ajustado sazonalmente, essa métrica não pode ser aplicada na presença de valores zerados.

Hyndman e Koehler (2006) afirmam que o MAPE tem a desvantagem de colocar uma penalidade superior nos erros positivos em relação a erros negativos. Outro aspecto sobre o MAPE é que seu cálculo computacional não é eficiente quando comparado com as demais métricas elencadas nessa seção, pois não permite aplicar o cálculo por meio de derivadas para se obter a otimização.

Mais recentemente, uma nova métrica tem sido utilizada em vários projetos de dados, que é a raiz do erro logarítmico quadrado (RMSLE), em que a diferença matemática perante o RMSE é a inclusão do *log* tanto do lado do valor da previsão como do valor real. Pode-se considerar que o RMSLE é uma medida mais robusta que o RMSE, já que o seu valor não cresce significativamente à medida que aparece um número maior de amostras dados anômalos. Ademais, o RMSLE possui três outras vantagens em relação ao RMSE: considera apenas o erro relativo entre o valor

previsto e o real, não sendo a escala do erro significativa, o que a torna uma métrica padronizada e de mais fácil entendimento, assim como o MAPE; apresenta uma penalidade superior para subestimação do valor atual em relação à superestimação; é computacionalmente menos custosa que o RMSE, por considerar o logaritmo em sua fórmula, o que facilita o cálculo das derivadas no processo de otimização.

Diante dos aspectos vistos até o momento e considerando o problema de previsão de vendas no varejo de moda, cujos dados são explorados nos próximos capítulos, afirma-se que o RMSLE é a métrica mais apropriada, pois os dados apresentam grande variância, o que pode ser explicado, principalmente, pelas datas comemorativas presentes no varejo. Além disso, ela é uma métrica robusta e padronizada, permitindo uma interpretação relativa, assim como ocorre com o MAPE, o que é um facilitador para a gestão. Aspecto relevante é o maior peso dado à subestimação, que vai ao encontro da necessidade da empresa estudo de caso desta Tese, pois problemas de ruptura de estoque têm custo superior à sobra dos mesmos.

A métrica principal deste trabalho é o RMSLE, porém também são calculadas outras métricas como: MAE, RMSE e MAPE. Além disso, considerando o contexto de negócio da empresa estudada, o Lojão do Brás, em que custo de ruptura é maior que o de estocagem, foi feita uma adaptação nas métricas RMSE e MAPE, sendo criadas RMSEP e MAEP, que são métricas ponderadas para refletir a necessidade da área de gestão da companhia. Assim, a ruptura tem um peso duas vezes maior que a sobra.

Em negócios de moda, existe grande preocupação com as sobras de estoque devido ao aspecto fugaz da moda em si, o que acaba elevando o nível de remarcações. No entanto, no caso do Lojão do Brás, seus consumidores não podem ser considerados *early adopters* da moda, tendo esse atributo um destaque menor. Dessa forma, passa a se configurar uma intolerância maior no que diz respeito a rupturas, com o intuito de evitar perda de vendas.

A seguir, são mostradas as fórmulas das medidas de desempenho aplicadas:

- **MAE: Mean absolute error (erro absoluto médio)**

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_t - \hat{Y}_t|$$

- **MAPE: Mean absolute percentual error (média dos erros percentuais absolutos)**

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

- **RMSE: Root mean squared error (raiz do erro quadrático médio)**

$$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (Y_t - \hat{Y}_t)^2 \right]^{\frac{1}{2}}$$

- **RMSLE: Root mean squared logarithmic error (raiz do erro quadrático logarítmico médio)**

$$RMSLE = \left[\frac{1}{n} \sum_{i=1}^n (\log(Y_t + 1) - \log(\hat{Y}_t + 1))^2 \right]^{\frac{1}{2}}$$

Considera-se o erro como sendo: $Y_t - \hat{Y}_t$.

As fórmulas do RMSEP e do MAEP não são apresentadas, pois são iguais, respectivamente, a do RMSE e do MAE, com exceção que o erro positivo é multiplicado pelo peso de 0.75 e o erro negativo por 0.25, sendo que o erro é a diferença entre o valor real e o valor previsto calculado por determinado modelo.

Um último aspecto que merece ser evidenciado é que as métricas de precisão foram utilizadas para avaliar as previsões e para implementar a técnica de seleção de variáveis RFA.

3.6. Teste Wilcoxon e Model Confident Set (MCS)

No caso dos métodos de inteligência computacional, a análise do modelo com melhor desempenho estritamente de forma numérica não é recomendada por existir a realização do processo estocástico. Um modelo pode ter desempenho de previsão aparentemente superior a outro simplesmente pelo acaso e não por ter um poder preditivo superior. Tal fato está associado ao problema de *data snooping*, que ocorre quando um conjunto de dados é usado mais de uma vez para fins de inferência e de seleção de modelos (WHITE, 2000). Para resolver o problema de *data snooping*, pode-se realizar uma validação adicional, a qual pode ser feita pela abordagem do teste de Wilcoxon ou pela abordagem de avaliação Model Confidence Set (MCS).

Na abordagem do teste de Wilcoxon, é preciso fazer a replicação de um número grande de arquiteturas dos diversos métodos de previsão e capturar as métricas de

avaliação no conjunto de teste, com o intuito de efetivamente comprovar os resultados encontrados. Especificamente, a replicação foi feita por meio da alteração do valor da semente, o restante se manteve constante. Posteriormente, deve ser aplicado o teste de postos com sinais de Wilcoxon, um teste não-paramétrico, que usa os postos de dados amostrais compostos de pares combinados (GIBBONS; CHAKRABORTI, 2003). O teste é usado para testar diferenças nas distribuições populacionais de modo que a hipótese nula é que as duas amostras provêm de populações com a mesma média e a hipótese alternativa é que as duas amostras provêm de populações com médias diferentes. Tal teste é operacionalizado para comparar se as médias obtidas na métrica de performance selecionada, através processo de validação com as diversas replicações efetivadas, para dois tipos de métodos de previsão são realmente distintas. Necessariamente, o teste de Wilcoxon só permite realizar a comparação de dois métodos simultaneamente, pois é um teste pareado.

Na abordagem MCS, o objetivo é determinar o conjunto de modelos que contenha o maior poder de predição, sendo que o conceito de “melhor” é estabelecido por meio de um critério associado a performance de previsão. Estimar um conjunto dos melhores modelos – e não apenas o melhor – é a principal diferença do MCS em relação aos modelos convencionais propostos para esse tipo de validação, segundo Hansen, Lunde e Nason (2011), inclusive quando comparado ao teste de Wilcoxon.

3.7. **Ensembles**

Para Brownlee (2020c), *ensemble* é um modelo de aprendizado de máquina que combina as previsões de dois ou mais modelos, denominados membros do *ensemble*. Tais modelos podem ou não ser oriundos do mesmo tipo de algoritmo e ser treinados no mesmo conjunto de dados. Ademais, existem várias formas de combinação desses modelos, as quais são detalhadas posteriormente.

Métodos *ensemble* devem levar em conta a forma como agrupam os modelos a fim de minimizar as desvantagens individuais de cada modelo em um modelo final. Assim, o princípio que norteia os modelos combinados é que um grupo de modelos fracos é reunido para gerar um modelo forte.

A popularidade do uso de *ensembles* aumentou com o sucesso obtido na implementação dessas técnicas em diversas competições de aprendizagem de

máquina. Essa popularidade pode ser justificada por dois aspectos: (1) possibilidade de aumento de desempenho, já que um modelo *ensemble* faz previsões mais precisas e obtém desempenho superior do que qualquer modelo isoladamente, e (2) incremento de robustez, isto é, um modelo *ensemble* reduz o viés e a variância das previsões, tornando-se um modelo mais confiável. Deve-se ter em mente que esse aumento de robustez por si só garante uma melhora do desempenho preditivo.

O viés é a incapacidade de um modelo capturar a relação entre a variável-alvo e as variáveis preditoras, gerando *underfitting*. Já a variância diz respeito à sensibilidade de um modelo ao ser usado com conjunto de dados diferente daquele em que foi treinado, gerando, possivelmente, um problema de sobreajuste. Diante disso, fica claro o *trade off* entre viés e variância, pois um viés maior passa a ser aceito em modelos de aprendizado de máquina na medida em que reduz a variância e, assim, aumenta o poder de generalização do modelo. Assim, é positivo o uso de modelos *ensembles* por minimizar tanto problemas de subajuste quanto de sobreajuste. Esse aspecto vai ao encontro do que foi colocado por Dietterich (2000), que destaca três razões para o uso de modelos *ensembles*.

A primeira é a questão estatística, relacionada ao fato do espaço de hipóteses a ser explorado por um conjunto de dados ser muito grande, o que consequentemente possibilita a existência de várias hipóteses gerarem a mesma precisão no conjunto de treinamento, o que pode não necessariamente ocorrer no conjunto de teste, fazendo com que o uso de vários modelos, o que significa levar em conta várias hipóteses, minimize esse risco. A partir do descrito, há um problema de variância.

A segunda razão é a computacional, isto é, muitos algoritmos de aprendizagem de máquina aplicam algum tipo de pesquisa local que pode levar o modelo a ficar preso em ótimos locais, o que não necessariamente é a melhor solução. Ao lançar mão de modelos *ensembles*, esse risco é minimizado, pois o ponto de partida de cada modelo será diferente. Nesse caso, o problema é de variância computacional.

A terceira e última razão é a questão de representatividade. Em outras palavras, em várias tarefas de aprendizado de máquina, a verdadeira hipótese não poderia ser representada por nenhuma hipótese existente no espaço total disponível. Ao usar métodos *ensembles*, combina-se as hipóteses, o que permite ao algoritmo aproximar uma função que seja mais próxima da verdadeira hipótese desconhecida. Nesse caso, o problema que surge é o de viés.

Segundo Zhou (2012), através da análise de vários estudos, tanto a variância como o viés de algoritmos de aprendizagem de máquina podem ser reduzidos com o uso de modelos *ensembles*. Outro benefício relevante dos modelos *ensembles* é a diversidade que proporciona ao agregar vários modelos.

Um ponto negativo dos modelos *ensembles* é que eles elevam significativamente o custo e a complexidade computacional, tendo em vista o aumento de tempo para treinar vários modelos e não somente um.

As principais formas de produzir modelos *ensembles* é por meio de três técnicas: *bagging*, *boosting* e *stacking* (FILHO, 2019b). Antes de apresentar cada uma delas, é importante salientar que a forma mais simples de constituir um modelo *ensemble* é pela aplicação da média, simples ou ponderada, das previsões dos modelos envolvidos. Esse tipo de *ensemble* é denominado *blending ensemble*.

O método *bagging* é um treinamento de vários modelos de forma repetida de amostras aleatórias com reposição de um mesmo conjunto de dados, tendo como exemplo o algoritmo *Random Forest* (RF), que é um conjunto de árvores aleatórias.

Já o método *boosting* treina seus modelos de forma sequencial – e não de forma independente como ocorre no método *bagging* – visando aumentar o desempenho do modelo final, com enfoque nas observações que os modelos anteriores tiveram mais dificuldade, tornando a predição mais resistente a viés. Ou seja, o objetivo é incluir modelos que consigam gerar erros menores que os anteriores.

Em modelos baseados em árvores de decisão, as árvores são ajustadas consecutivamente e, a cada etapa, o objetivo é reduzir o erro observado na anterior. Quando uma entrada é classificada incorretamente ou a regressão gera um valor muito distante do real por uma hipótese, seu peso é aumentado para que a próxima hipótese seja mais provável de gerar a estimativa mais correta. Assim, a combinação de todos os modelos transforma modelos fracos em um de melhor desempenho.

Na categoria do método de *boosting*, o *gradient boosting* é uma técnica que tem sido aplicada atualmente em detrimento do *adaboost*. Nesse caso, o *boosting* passa a ser visto como um problema de otimização numérica, em que a meta é minimizar a função de perda do modelo através da adição de modelos mais simples e da descida do gradiente, conforme visto em Freund e Schapire (1997). Como o gradiente *boosting* tem como foco a minimização de uma função de perda, pode-se utilizar diversos tipos de função de perda, o que torna a técnica bem flexível, podendo

ser aplicada a vários tipos de problemas. A contribuição de cada modelo fraco para a previsão final é baseada em um processo de otimização gradiente para minimizar o erro geral do modelo forte.

Como exemplos de algoritmos que usam a técnica de gradiente *boosting*, temos: XGBOOST, LightGBM e o Cat Boost, os quais apresentam pequenas distinções entre si. Os dois primeiros representantes são aplicados neste trabalho.

Por fim, o método mais usado ultimamente é o *stacking*, que separa parte dos dados para treinar os modelos nos dados originais, prever em outro conjunto de dados e usar as próprias previsões como atributos para treinar novos modelos.

O *stacking* difere do *bagging* e do *boosting* em dois pontos específicos. O primeiro é que o *stacking* considera preditores heterogêneos, ou seja, as previsões são originadas de algoritmos distintos, enquanto *bagging* e *boosting* consideram preditores homogêneos, isto é, as estimativas são originadas de um mesmo tipo de algoritmo. O segundo aspecto diz respeito à forma de combinação dos modelos, já que o *stacking* combina os modelos básicos usando um metamodelo, enquanto o *bagging* e o *boosting* combinam os modelos por meio de algoritmos determinísticos.

A arquitetura de modelo *stacking* envolve dois ou mais modelos bases, geralmente denominados como modelos nível 0 e um metamodelo, que combina as previsões geradas pelos modelos base, sendo designado como modelo nível 1. O metamodelo pode ser um algoritmo distinto dos aplicados nos modelos níveis zero.

A principal característica que um modelo *ensemble* deve evidenciar é a diversidade, o que pode ser obtida através dos seguintes mecanismos para a construção dos modelos níveis 0: aplicação de algoritmos de treinamento, aplicação de diferentes conjuntos de dados de treinamento, uso de diferentes conjuntos de atributos e uso de diferentes configurações de hiperparâmetros (FILHO, 2019).

Por fim, vale destacar a conclusão obtida no trabalho Pavlyshenko (2019) de aplicação de modelos de aprendizagem de máquina para previsão de séries de vendas, que é o tema desta Tese, mas focado para moda. O autor coloca que o uso de modelos *ensembles* com a técnica *stacking* leva em consideração as diferenças nos resultados advindas de vários modelos de previsão com diferentes conjuntos de parâmetros e, com isso, melhora a precisão tanto nos dados de validação como nos dados de teste. Além disso, o autor constata que o problema de previsão de vendas é muito mais um problema de regressão do que de série temporal, o que faz com que,

em muitos casos, o uso da abordagem de regressão gere resultados melhores quando comparado com métodos de séries temporais. Outras limitações do uso da última abordagem é que demanda um volume maior de observações para capturar o padrão sazonal, não lida bem com a presença de dados anômalos e dados faltantes, e não consegue contemplar os diversos fatores exógenos que podem afetar as vendas.

3.8. Análise AHP

Saaty (1987) introduz a metodologia Analytic Hierarchy Process como uma teoria geral de mensuração. A análise AHP já foi bastante aplicada, sobretudo para a realização da tomada de decisão com a presença de uma ampla gama de critérios, para o planejamento e alocação de recursos e para a resolução de conflitos. A metodologia é embasada na comparação sequencial aos pares das opções existentes, tendo o peso atribuído a cada critério como um elemento fundamental.

O principal benefício do método é lidar com aspectos qualitativos e quantitativos de um problema de decisão. Isso é possível, pois os valores dos julgamentos das comparações paritárias são baseados em experiência e dados físicos (SAATY, 1994).

O diferencial mais relevante do método AHP é que ele divide o problema geral em avaliações de menor importância, mantendo a participação desses problemas menores na decisão global. Assim, diante de um problema complexo, é mais fácil subdividi-lo em outros menores, porque, quando solucionados individualmente e, depois, somados, representam a decisão adequada para o problema maior inicial.

Para Costa e Moll (1999), o método AHP está fundamentado em três princípios do pensamento analítico, os quais são:

- **Construção de hierarquias:** problema a ser solucionado é decomposto em níveis hierárquicos, como forma de buscar uma melhor compreensão e avaliação do mesmo.
- **Priorização:** o ajuste das prioridades no AHP é definido por meio da capacidade do ser humano na sua avaliação das relações existentes e nas situações observadas, comparando de forma pareada mediante um determinado conjunto de critério.
- **Consistência:** esse aspecto é contemplado na análise AHP.

Saaty (1991) afirma que a hierarquia é uma abstração da estrutura de um sistema para estudar as interações funcionais de seus componentes e seus impactos no sistema total. Essa abstração pode tomar várias formas interrelacionadas, todas descendentes de um objetivo geral, abrindo-se em subobjetivos, desmembrando-se nas forças influentes e até nas pessoas que influenciam essas forças.

Após a definição da hierarquia, a próxima fase é a de avaliação e de julgamento por parte dos participantes do processo, a qual consiste na comparação pareada, em relação ao seu impacto sobre um elemento acima deles na hierarquia. Na realização das comparações, os tomadores de decisão podem usar dados tanto quantitativos como qualitativos para justificar sua avaliação entre as opções existentes.

No AHP, essas avaliações são convertidas em valores numéricos normalizados, possibilitando realizar a comparação efetiva entre os critérios definidos e determinar a decisão final. Logo, um valor numérico é estabelecido para cada elemento da hierarquia, o que permite que os diversos critérios sejam comparados entre si de forma consistente, mesmo que sejam qualitativos. Finalmente, as prioridades numéricas são calculadas para cada uma das alternativas de decisão. Os resultados devem refletir como cada alternativa soluciona o problema/objetivo proposto considerando os diversos critérios escolhidos com os seus respectivos pesos e julgamentos determinados.

Assim, o passo a passo para aplicação da metodologia AHP, conforme Saaty (1994), é: (1) definição do problema/objetivo de estudo; (2) especificação do foco principal; (3) identificação das alternativas viáveis para a escolha; (4) definição do conjunto de critérios essenciais para a montagem da hierarquia; (5) seleção dos participantes do processo de avaliação das alternativas, considerando os critérios selecionados e sua respectiva relevância para atender o objetivo principal; (6) determinação dos métodos para apuração dos julgamentos; (7) emissão dos julgamentos de valor a partir de comparações pareadas com base em uma escala previamente estabelecida; (8) processamentos dos julgamentos obtido com objetivo de validar a sua consistência, o que é feito por meio do cálculo da prioridade de cada alternativa em relação ao foco principal; e (9) análise final dos resultados.

4. Revisão da Literatura

Esta parte traz a revisão bibliográfica do tema abordado nesta Tese. Assim, busca contextualizar uma síntese dos trabalhos produzidos na academia com o objetivo de verificar pontos em comum e divergentes e posicionar esta Tese em termos de sua contribuição efetiva para a comunidade acadêmica.

No contexto nacional, foram encontrados estudos sobre previsão de vendas e até mesmo de demanda para o setor de varejo de moda somente com a aplicação de técnicas estatísticas tradicionais de previsão. Porém, não foram localizados trabalhos para o setor com a aplicação de técnicas não lineares com ênfase em inteligência computacional, as quais são apresentadas no referencial teórico. Esse aspecto é relevante, pois evidencia o caráter inovador deste trabalho. Ademais, esta Tese tem como propósito avaliar o impacto do tratamento da sazonalidade, do uso das variáveis exógenas e da aplicação de alguns métodos de seleção de variáveis sobre a previsão de vendas em si para o setor estudado, o que também não foi observado no âmbito nacional e nem de forma conjunta no âmbito internacional, como propõe este trabalho.

A produção realizada no Brasil não tem trabalhos específicos com os objetivos citados. No que diz respeito à aplicação de redes neurais para previsão de vendas no varejo, destacam-se três trabalhos: Almeida e Passari (2006), Pasquotto (2010) e Felisoni de Angelo *et al.* (2011).

Almeida e Passari (2006) realizaram a previsão de vendas no nível individual de produto, diferentemente dos outros trabalhos, contemplando também a demanda simultânea de possíveis produtos substitutos, além de considerar os impactos de promoção e de preço, componentes da estratégia varejista. Em relação à acurácia, apoiado nos resultados das métricas MSE e MAPE, as redes neurais apresentaram um desempenho superior frente ao método *Naive* e ao método de regressão linear.

Pasquotto (2010) analisou a aplicação de redes neurais recorrentes de Elman para a previsão de três séries temporais agregadas do varejo de bens e serviços do Brasil: produtos farmacêuticos, adubo e tráfego aéreo. Foi verificado que a aplicação desse método também não gerou uma redução do erro da previsão gerada.

Felisoni de Angelo *et al.* (2011) compararam a previsão de vendas de forma agregada no mercado varejista através de duas técnicas: séries temporais ARIMA e modelos de Redes Neurais (RNs). A conclusão obtida foi que a diferença de precisão

entre as duas previsões geradas foi baixa e que as RNs apresentaram uma precisão maior. No entanto, elas possuem um grau maior de complexidade para definição da arquitetura adequada e requisitam uma série de testes e validações, que precisam ser feitas. Assim, observa-se, nitidamente, um *trade off* entre um pequeno ganho de precisão contra um alto investimento de tempo para se estabelecer o modelo apropriado no caso da rede neural em detrimento do método ARIMA.

A produção internacional é extensa, mesmo sob o filtro aspecto de moda, pois aborda vários assuntos interligados com a previsão de vendas no varejo de moda, como a gestão dos estoques e da própria cadeia de suprimentos *fast fashion*, dada as suas particularidades; métodos existentes, passando pelos clássicos e incluindo os baseados em inteligência artificial; aspectos que afetam o processo de previsão, como *big data*, entre outros temas. Nesse sentido, os esforços foram para a análise de artigos com uma abordagem voltada para o estudo de métodos de previsão de vendas para o varejo de moda, não deixando de analisar os aspectos relevantes que devem ser contemplados para a realização do processo de previsão nesse setor.

Previsões com altos níveis de acurácia para a indústria de varejo de moda tem como objetivo garantir maior rentabilidade para as empresas do setor, inseridas em um ambiente cada vez mais competitivo com redes globais se expandindo pelo mundo. Como já citado, uma previsão mais precisa tem os seguintes impactos: níveis adequados de estoque, evitando tanto rupturas como excessos de estoque, que levam a níveis menores de remarcações e um efeito chicote menor, contribuindo para uma maior eficiência da cadeia de suprimentos.

Zhang, Patuwo e Hu (1998) assinam estudo rico sobre aplicação das redes neurais para a realização da previsão de vendas, um dos métodos de aprendizado por máquina mais utilizados. Para os autores, uma das principais aplicações da RNN é o uso para previsão, sendo a primeira aplicação com esse objetivo feita em 1964.

Os principais benefícios de uma RNN são: ela aprende por meio de exemplos, podendo generalizar a partir de dados amostrais; ela é um aproximador universal, sendo mais genérica e flexível que os métodos tradicionais estatísticos, e não é linear. Do ponto de vista de limitações, os seguintes pontos são destacados: para processos lineares com poucos ruídos, a RNN não é um método adequado, tendo em vista sua baixa capacidade de interpretação; ela também possui uma maior probabilidade de apresentar problemas de sobreajuste, dado o grande número de parâmetros que são

necessários; não há um método determinístico para definir a melhor estrutura de rede, sendo necessário um procedimento heurístico para a realização dessa configuração.

Diante desses aspectos, recomenda-se que as redes sejam utilizadas para tratar grandes bases de dados, com problemas não lineares e/ou problemas de séries temporais com, pelo menos, mais de uma variável.

A configuração da arquitetura da RNN é o aspecto mais crítico e pode ser subdividida em: número de *inputs*, número de camadas ocultas e número de *outputs*. A maior parte das redes neurais tem apenas uma camada oculta, no entanto, em alguns casos, duas camadas podem ser mais eficientes, o que ocorre na presença de problemas mais complexos. Para séries temporais, o número de *inputs* corresponde ao número de defasagens necessário para entender o padrão da série e, assim, fazer a previsão. No caso do número de *outputs*, ele é determinado pelo horizonte da previsão, se a previsão for de um passo ou múltiplos passos. A função de ativação, na maioria dos casos, é a função logística, podendo ser usada a linear no caso de previsões de séries de tempo. Quanto ao treinamento, o mais usado é o retropropagação, funcionando como gradiente iterativo.

Zhang, Patuwo e Hu (1998) citam a importância da normalização dos dados para um desempenho superior e a própria convergência das redes neurais. O processo tem como objetivo mitigar os problemas oriundos do uso de unidade de dispersões distintas entre as variáveis e reduzir o tempo de treinamento da rede.

O trabalho de Alon, Qi e Sadowski (2001) foi um dos primeiros a abordar o uso das RNs para previsão de vendas no varejo de forma agregada comparando com os métodos estatísticos tradicionais. Com o objetivo da previsão mensal do varejo americano, as RNs apresentaram uma precisão maior, segundo o indicador MAPE, em relação aos métodos tradicionais ARIMA, suavização exponencial e regressão múltipla. Os autores fizeram testes em diversos cenários macroeconômicos e constataram que as RNs tiveram um desempenho muito superior frente às técnicas tradicionais de previsão quando a economia passa por períodos de maior volatilidade.

Outra conclusão relevante é que as RNs conseguem capturar as tendências não lineares e os padrões de sazonalidades. Nelson *et al.* (1999) discordam efetivamente da constatação em relação à sazonalidade, colocando a necessidade de dessazonalização dos dados para garantir a obtenção de redes neurais mais precisas.

Alon, Qi e Sadowski (2001) destacam a importância das empresas avaliarem o *trade off* entre a acurácia da previsão e os custos e limitações de implantação de cada método, como foi visto em Felisoni de Angelo *et al.* (2011). Não obstante, os estudos evidenciam que as RNs são mais precisas, na média. Por outro lado, as mesmas são menos parcimoniosas. Adicionalmente, a complexidade para definir a sua configuração adequada é muito maior, o que acaba por exigir mais esforço.

Tendo em vista o foco no setor de moda desta Tese, foram encontrados trabalhos dos autores Thomassey, Happiette e Castelain (2002 e 2005) que apontam peculiaridades do setor e apresentam modelos específicos de previsão de demanda.

O primeiro artigo dos autores, de 2002, indica que um sistema de previsão de vendas para o setor têxtil no médio prazo, um ano, deve conseguir lidar com os seguintes pontos: reagir rapidamente às variações de tendência e sazonalidade, identificar e suavizar eventos randômicos, gerar previsões com pouco histórico de dados e incluir os efeitos das variáveis exploratórias.

Dado o elevado número de parâmetros que afetam a demanda e alta incerteza da determinação da mesma, os autores acreditam que os modelos tradicionais não são adequados e propõem dois modelos: HFCCX (*hybrid forecasting model with fuzzy estimation of corrective coefficients of the explanatory variables influence*) e AHFCCX (*automatic hybrid forecasting model with fuzzy estimation of corrective coefficients of the explanatory variables influence*). Os modelos propostos aplicam lógica *fuzzy* para mapear e prever os efeitos das variáveis exploratórias sobre as vendas, e a distinção entre os modelos é que o AHFCCX não requer a intervenção de um especialista.

O modelo base de previsão HFCCX possui três estágios: os dados de vendas são dessazonalizados da influência das variáveis exploratórias, que nesse caso são preço, estações do ano e período de férias; os dados dessazonalizados são usados para prever as vendas considerando o ajuste sazonal; as vendas previstas são sazonalizadas novamente com a influência das variáveis exploratórias correspondentes da próxima estação. Por funcionar por meio do uso de regras linguísticas, o modelo de inferência *fuzzy* torna-se bastante intuitivo e lida bem com relações não lineares.

A análise da métrica de erro MSE detectou que AHCFCCX teve um desempenho melhor que as seguintes técnicas clássicas: Holt Winter com sazonalidade, ARIMA e regressão múltipla com variáveis exploratórias. Todavia,

destaca-se a complexidade e o aumento do tempo de aprendizado exigido pelo algoritmo genético do sistema *fuzzy* em comparação aos métodos clássicos.

No artigo “A short and mean-term automatic forecasting system – application to textile logistics”, Thomassey, Happiette e Castelain (2005) citam mais um modelo: o SAMANFIS para previsão de curto prazo, complementando o primeiro *paper*, que tinha foco no médio prazo. O novo modelo proposto consiste também em um sistema de inferência *fuzzy* com uma rede neural correspondente e tem como objetivo se adaptar aos acontecimentos reais com base em um pequeno volume de informação disponível. A técnica SAMANFIS também apresentou uma acurácia superior em relação às técnicas estatísticas tradicionais, só que considerando a métrica MAPE.

Celia *et al.* (2003) também realizaram um trabalho aplicado para previsão de vendas de moda e concluem que a maior parte das previsões se apoiam nos dados históricos, o que é mais difícil para o setor da moda. Eles dividem os modelos de previsão em três categorias: (1) univariados, que usam o tempo como única variável e são exemplos os modelos de média móvels Holt Winters, séries de Fourier e ARIMA; (2) multivariável, que usam um conjunto de variáveis e tentam estabelecer as relações de causa e efeito do sistema, sendo exemplos regressão múltipla e ARIMA multivariável (MARIMA); e (3) métodos qualitativos, incluindo pesquisa de mercado Delphi, painel, analogia histórica e *soft computing*, que conseguem reconhecer padrões não lineares e seus exemplares são RNs, lógica *fuzzy* e algoritmos genéticos. Os autores justificam a complexidade da previsão de vendas no setor de moda devido à presença de muitas variáveis endógenas e exógenas. Embora uma regressão linear consiga levar em conta as variáveis exógenas colocadas anteriormente, sua aproximação é estritamente linear. Dessa forma, os autores propõem empregar uma técnica não linear como a RN para verificar sua performance contra as técnicas clássicas de suavização exponencial e Holt Winters. A análise do coeficiente de determinação R^2 mostrou um desempenho proeminente das RNs para as diversas categorias de produtos analisadas perante as técnicas tradicionais. Contudo, os resultados apresentaram grande sobreajuste.

Sobre a questão sazonal para o setor da moda, verificou-se um artigo de Chu e Zhang (2003) comparando modelos lineares e não lineares para previsão de vendas no varejo de forma agregada. Ao compararem a acurácia de um modelo SARIMA com os modelos de redes neurais com dados sazonalizados e dessazonalizados, mediante

os resultados das métricas MAE, MSE e MAPE, a RN com dados dessazonalizados gerou previsões com erro menor, indo ao encontro do que foi observado no trabalho de Nelson *et al.* (1999). Chu e Zhang (2003) ressaltam que a dessazonalização dos dados é um ponto crítico para performance das redes neurais e que, historicamente, modelar e prever dados com sazonalidade é um dos grandes esforços das pesquisas que os métodos heurísticos têm tentado superar. Os autores relatam que vários estudos comparam modelos tradicionais e RNs e existe um misto de conclusões sobre a flexibilidade dos métodos não lineares, que é realmente um ponto compensatório em relação à simplicidade da previsão linear.

Não existe um consenso sobre a questão da dessazonalização dos dados para a utilização da técnica de RN para realização de previsões. Muitos pesquisadores defendem que uma das competências das RNs é ser capaz de identificar claramente os ciclos e padrões de sazonalidade presentes em uma série temporal, não sendo necessário eliminá-los da série para que a modelagem seja feita.

Continuando a análise dos trabalhos propostos para a realização da previsão para o setor de moda, Thomassey (2010) diz que, para se manterem competitivas, as companhias de moda têm investindo em sistemas de informação e logística sofisticados e, sobretudo, em sistemas de previsão mais precisos e confiáveis. Os sistemas de previsão para o setor devem levar em conta os seguintes aspectos: a presença de sazonalidade nas vendas, as diversas variáveis exógenas que impactam as vendas, a influência das tendências de moda, o que reduz o histórico de dados, a existência de várias opções de cores, tamanhos e de atributos *fashion*.

A maior parte das empresas não faz uso dos métodos clássicos de previsão, mas sim de métodos desenvolvidos internamente, baseados nas vendas passadas e em algumas poucas variáveis, não sendo aplicada nenhuma técnica mais atual. Thomassey (2010) acredita que o uso de lógica *fuzzy*, algoritmos genéticos e RNs é apropriado para previsões na indústria de moda. No entanto, para aplicação dessas técnicas é preciso um histórico mínimo de dados. Tal fato impacta diretamente no nível de agregação da previsão, tornando possível fazê-la somente no nível maior de agregação de categoria produto e não em níveis menores, como item, cor e tamanho.

No nível de item, Thomassey (2010) sugere que o sistema trabalhe com as seguintes características: ausência de histórico de vendas (substituição total dos itens), número grande de itens com histórico pertencentes a coleções equivalentes e

presença dos critérios descritivos. Dada a falta de informação suficiente para efetivação da previsão, o objetivo passa a ser encontrar descrições comuns entre produtos novos e antigos, o que, por efeito, possibilita realizar a previsão no nível de granularidade menor. Para encontrar produtos equivalentes, que poderiam pertencer ao mesmo grupo, é sugerido o uso de redes Bayesianas ou árvores de decisão.

O uso das redes Bayesianas, segundo Choi, Hui e Yu (2014), é feita de forma similar ao modelo aplicado por Thomassey (2010), em que a previsão de vendas de item novo é com base em item antigo similar e suas respectivas vendas passadas.

Wong e Guo (2010) abordam a evolução, as últimas aplicações, analisam os trabalhos de Thomassey e sugerem um modelo para previsão de vendas no varejo de moda ao proporem a implementação de um sistema inteligente híbrido (HIS) para superar as limitações dos modelos com RNs. Especificamente, é feito um aprimoramento da forma de definição dos pesos da RN. Considerando a métrica MSE, em testes mensais, trimestrais e anuais, o modelo HIS apresentou previsões mais precisas que os modelos ARIMA, RN e ELM, com exceção da periodicidade anual.

Ni e Fan (2011) propõem um modelo de previsão dinâmico de duas etapas, com previsões para longo e curto prazo, que combina a aplicação de árvores de decisão com redes neurais. O primeiro estágio utiliza dados históricos com fatores exógenos, tais como temporada, clima e região, para previsão de longo prazo. Já o segundo estágio, com foco na previsão de curto prazo, leva em conta vendas em tempo real e influências externas, como chuva, variação de fatores econômicos e feriados. Os experimentos constataram que o modelo apresenta boa acurácia e indica um modelo de ajuste que contemple outros fatores externos como trabalho futuro.

Tendo em vista a complexidade do setor de moda para a previsão de vendas, justificada pela grande quantidade de artigos e falta de dados históricos, Aksoy, Ozturk e Sucky (2012) recomendam o uso de um modelo ANFIS. A técnica alia a capacidade de raciocínio de um sistema de lógica *fuzzy* com a capacidade de aprendizado de RN, reforçando o estudo de Thomassey (2002). Os autores também apontam aspectos que afetam a indústria da moda: desenvolvimento das cadeias de globais de produtos, aumento da competição no setor e mudança frequente nos hábitos de consumo, o que aumenta o nível de volatilidade e incerteza para a previsão de demanda.

Face a esse cenário, Aksoy, Ozturk e Sucky (2012) atestam que o uso de sistemas híbridos de previsão apresenta desempenho superior em termos de precisão

do que a utilização de métodos individuais. Além disso, o fato de os produtos serem substituídos ao longo das coleções faz com que o histórico dos dados seja curto e perdido e, por isso, os métodos quantitativos devem ser descartados como metodologia de previsão. Assim, é proposto um modelo neuro fuzzy: o ANFIS, uma evolução do modelo de Thomassey, Happiette e Castelain. De forma conclusiva, a análise dos indicadores de acurácia MSE, MAPE e MAE verificou que o método híbrido teve erros menores que a RN pura.

A aplicação de métodos híbridos também foi observada por Xia *et al.* (2012). Especificamente foi aplicado o método *Extreme Learning Machine* com Métricas Adaptativas (AD-ELM), que visam reduzir o problema de determinação de tendência e diminuir o efeito de sobreajuste, quando há uma aderência muito grande no conjunto de dados de treino, a qual passa a não ocorrer no conjunto de dados de teste.

Os resultados dos testes em vários conjuntos de dados e a comparação com os modelos de RNs puras, AR, AD-ELM usando como forma de avaliação de precisão o MSE, evidenciaram que o modelo sugerido AD-ELM supera de forma consistente os demais modelos comparativos, sendo recomendado como o método mais apropriado para previsão de vendas no varejo de moda.

Considerando o objetivo de uso de métodos não lineares para a previsão de vendas no varejo de moda, três trabalhos abordam o uso de redes neurais para previsão de séries temporais e podem ser aplicados para qualquer setor: Dorffner (1996), Gers, Eck e Schmidhuber (2001) e Långkvist, Karlsson e Loutfi (2014).

Um resumo de todos os tipos de redes neurais que podem ser aplicados na modelagem para previsão de séries temporais é feito por Dorffner (1996), que afirma que o grande benefício da utilização das redes neurais MLP é a possibilidade de aproximar funções não lineares, o que permite criar aplicações mais poderosas.

Estudo feito por Gers, Eck e Schmidhuber (2001) apresenta testes da aplicação dos modelos de rede neural recorrente LSTM *versus* o uso de redes neurais MLP. Dada a promessa das LSTMs conseguirem modelar a dependência temporal para as observações de uma série, os autores quiseram constatar a superioridade desse tipo de rede. A conclusão foi que uma MLP de janela fixa tem um desempenho melhor que um modelo LSTM simples tipo AR e recomendam usar esse método somente no caso em que a técnica de MLP de janela fixa não apresente resultados satisfatórios.

Uma revisão de aprendizagem profunda para modelagem de séries temporais é apresentada por Långkvist, Karlsson e Loutfi (2014). O principal problema na modelagem de séries de tempo é considerar o tempo de forma estática, assim os autores sugerem o uso de técnicas que façam a captura das dependências tanto de longo quanto de curto prazo, sendo o modelo LSTM uma das técnicas possíveis e recomendadas, já que é capaz de memorizar mais de cem períodos.

Nenni, Giustiniano e Pirolo (2013) observam que as características da indústria da moda, vistas como dificultadoras da previsão em si, tais como curtas coleções, grau de incerteza e falta de histórico de dados, fizeram com que, ao longo do tempo, grandes empresas do setor focassem mais nas suas respectivas cadeias de suprimentos em vez de focar única e exclusivamente no processo de previsão de vendas. Gradativamente a realidade aceita pelas pessoas que atuam no setor, seja trabalhando diretamente ou realizando pesquisas, é que a demanda por produtos de moda não pode ser prevista de forma satisfatória e que os mercados da moda são complexos sistemas abertos que frequentemente demonstram altos níveis de “caos”.

Assim como Nenni, Giustiniano e Pirolo, o estudo de Liu *et al.* (2013) revisa a literatura e apresenta conjunto de *papers* sobre a previsão de venda para o varejo de moda. Os autores afirmam que, comparado a outros tipos de varejo, a previsão de vendas na moda se torna tarefa árdua pela alta volatilidade da demanda, afetada pela mudança constante do gosto do consumidor e pelo ciclo de vida curto dos produtos. As vendas são impactadas pela sazonalidade, pelas tendências de moda e por uma série de outras variáveis e os varejistas de moda devem carregar muitos SKUs distintos, tendo em vista variações de cor e tamanho com histórico de vendas limitado, o que torna o processo de previsão de vendas no setor desafiador, passando a exigir ferramentas sofisticadas e flexíveis. Liu *et al.* (2013) agrupam os métodos de previsão em três grandes grupos: estatísticos, baseados em inteligência artificial e híbridos.

Os métodos estatísticos têm seu uso popularizado no setor pela sua simplicidade e velocidade de desempenho. Porém, seus resultados não podem ser considerados satisfatórios, pois escolher o método estatístico mais adequado não é uma tarefa simples, o desempenho em termos de acurácia é inferior aos métodos baseados em inteligência computacional e, por fim, eles não conseguem lidar bem com questões como tendência, sazonalidade e padrões irregulares.

Com o avanço da tecnologia computacional, surgiram métodos baseados em inteligência artificial que, em comparação aos estatísticos, têm desempenho superior em termos de acurácia. Os métodos mais utilizados são ANN, *fuzzy*, ENN, ELM e EELM. Apesar da performance superior, em termos de precisão, os métodos de inteligência artificial demandam quantidade de tempo expressiva para que a previsão seja gerada.

Os métodos híbridos são usados por aproveitarem os principais pontos fortes de diversos métodos. Podem ser considerados, inclusive, mais eficientes quando comparados com os métodos estatísticos e os de inteligência artificial puros. Nos últimos anos, houve crescimento das pesquisas considerando métodos híbridos e as principais combinações encontradas foram estabelecidas com as seguintes técnicas: *fuzzy*, RNN e ELM com técnicas estatísticas, *grey model* (GM) etc.

Em termos de aplicação na indústria da moda, Liu *et al.* (2013) avaliam que sobre o aspecto de horizonte da previsão, a maior parte dos trabalhos tem como objetivo o médio e o longo prazo, existindo uma carência de previsão de curto prazo, o que faria mais sentido, considerando as especificidades do setor. As pesquisas podem ser divididas em dois tipos de previsão: de produtos existentes ou de produtos novos, sendo o último tipo de previsão o mais complexo, dominado pelos métodos de inteligência computacional. No âmbito da velocidade de geração da previsão, métodos estatísticos são muito mais rápidos que os AI. Apesar disto, surgiram alternativas, como o ELM, e mesmo métodos híbridos, tendo o modelo *fuzzy* destaque.

Em relação a pesquisas futuras, os autores propõem alguns temas interessantes, como: execução das previsões, considerando dados transversais e longitudinais, ainda pouco explorados pela indústria, sendo que o maior foco consiste no uso somente dos dados série de tempo em si; levando em conta que a cor é um elemento crítico para o planejamento de demanda desse tipo de produto, existem poucos trabalhos focados nesse aspecto ou até mesmo considerando outros atributos, como nível de moda ou qualidade etc.; pesquisas sobre previsão de vendas sobre feriados importantes e/ou datas comemorativas é um tema pouco explorado e que vem apresentando um impacto significativo sobre as vendas das companhias do setor.

Tendo como base a indústria de moda *fast fashion*, Choi *et al.* (2013) propõem como novo algoritmo o *fast fashion forecasting* (3F), cujo objetivo é realizar previsões de vendas com baixo volume de dados e para um curto espaço de tempo. O 3F é um

algoritmo híbrido que combina dois métodos de inteligência computacional: ELM e GM. Os autores testaram o algoritmo e alcançaram níveis aceitáveis de precisão com base no MAPE. Além disso, constataram que o algoritmo é mais recomendado em conjuntos de dados com alta tendência ou grande variância do ciclo sazonal.

No trabalho de Beheshti-Kashi *et al.* (2015) foi possível revisar os métodos tidos como o estado da arte para previsão de vendas de produtos de moda e novos produtos. Assim como foi visto em Thomassey (2010, 2014), os autores mencionam as características críticas do setor, que afetam a tarefa de previsão: ciclos de vida curto de produtos, alta variabilidade dos produtos e a própria incerteza da demanda.

Em termos de técnicas, as abordagens existentes são: técnicas estatísticas, como suavização exponencial, modelos de regressão e ARIMA; técnicas mais avançadas, que são as redes neurais ou de mineração de dados; e técnicas híbridas.

Em relação às técnicas avançadas, Thomassey (2014) destaca que as especificidades da indústria da moda e a própria falta de dados impossibilitam a adoção em massa desses métodos pelas empresas que atuam no setor.

Diante desse fato, Beheshti-Kashi *et al.* (2014) apresentam abordagem distinta para a previsão de vendas de novos produtos de moda que tem como base pesquisas e conteúdo gerado em redes sociais pelos consumidores finais. Em suma, os autores afirmam que a maior parte dos trabalhos na área de previsão de vendas para moda propõe modelos complexos para obter previsões precisas. Contudo, as empresas do setor ainda enfrentam desafios na adaptação desses modelos complexos e na sua própria implementação em produção. Portanto, um aspecto interessante que pode ser foco de pesquisas futuras, é explorar modelos que sejam, simultaneamente, precisos e efetivamente aplicáveis ao dia a dia das empresas de moda.

A previsão de demanda é vista como um processo desafiador que tende a apresentar um alto nível de imprecisão, o que acaba afetando a própria gestão dos estoques. Assim, é fundamental o desenvolvimento de sistemas de gestão de risco para a gestão de estoques de produtos de moda.

Choi e Ren (2016) discutem como empresas focadas em *fast fashion* podem desenvolver sistemas inteligentes de previsão de demanda que minimizem o risco de estoque inadequado, permitindo uma gestão mais eficiente. Aspecto interessante é a revisão da operação de *fast fashion*, do processo de previsão de demanda em si e dos modelos de controle de estoques, que levam o fator risco em consideração. Os

autores destacam que tanto a previsão de demanda como a gestão dos estoques são processos altamente interligados que suportam o modelo de *fast fashion*.

Considerando as peculiaridades da indústria de moda rápida, Brahmadeep e Thomassey (2016) propõem um sistema de previsão para novos produtos em duas etapas. A primeira é a realização de uma previsão de longo prazo, com o intuito de prever as primeiras semanas de vendas. Nessa etapa, existem poucos dados, então é sugerido usar uma rede neural ou uma árvore de decisão com base no histórico de dados de produtos similares. Na segunda etapa, há a previsão de curto prazo, que é um ajuste com base nos dados reais existentes. Nessa etapa, existe, mais uma vez, uma limitação tanto de prazo como de volume de dados, sendo recomendado, então, o uso do algoritmo ELM, que ajustará as previsões de longo prazo, a partir da terceira semana, de acordo com os dados reais das duas últimas semanas. De forma conclusiva, o método de duas etapas apresentou níveis satisfatórios de precisão, baseados nos resultados do RMSE, e foram bastante complementares entre si, considerando o objetivo de previsão de vendas para novos produtos de moda.

A despeito desta Tese ter como foco de análise o setor de moda, o setor de tecnologia também pode ser considerado referência para verificação dos métodos de previsão utilizados, levando em conta a presença de características similares dos seus produtos, como ciclo de vida curto, baixa quantidade de histórico de dados, mudança constante nos hábitos de consumo, obsolescência etc. Sendo assim, o trabalho de Chen e Lu (2017) mostra a importância de os varejistas de computadores terem previsões de vendas precisas para gerenciamento eficaz do marketing e dos seus estoques. Eles aplicaram os algoritmos K-means, SOM e GHSOM para o agrupamento dos produtos e SVR e ELM, duas técnicas de aprendizado de máquina, para construir seis modelos de previsão de vendas, considerando cada agrupamento de produtos. Os resultados apontaram o GHSOM-ELM como o mais robusto e com menor erro de previsão para os três grupos de produtos nos seis grupos contemplados, sendo o método recomendado para previsão no setor de computadores.

Complementando os trabalhos vistos até o momento, que mostram as peculiaridades da indústria da moda e os principais métodos de previsão utilizados, o estudo de Ren, Chan e Ram (2017) tem como diferencial aplicar uma análise AHP, abordando diversos critérios para avaliação dos métodos de previsão em um caso

prático. Os métodos foram ARIMA, GM, ELM e *pure panel data* (PPD), um modelo de regressão que, além da série de tempo em si, é capaz de incorporar na previsão a demanda cruzada entre múltiplos produtos e os respectivos preços de venda.

A definição dos pesos da análise AHP foi estabelecida em uma pesquisa com um grupo heterogêneo de pessoas, contemplando os seguintes aspectos: tipo de produto, cargo na empresa e mercado-alvo. Os critérios contemplados foram: acurácia, velocidade, estabilidade, facilidade de uso e os requisitos de dados.

A ordenação da análise AHP permitiu verificar que o método com o melhor desempenho foi PDD, seguido pelo ARIMA, ELM e GM. Além disso, o estudo constatou algumas descobertas relevantes, tais como: em geral, a acurácia é o principal critério de um sistema de previsão de demanda, sendo que o PDD é o sistema mais apropriado para produtos de moda; há uma correlação positiva significativa entre as importâncias percebidas nos critérios de desempenho, no entanto, essa relação é insignificante entre a importância percebida da “acurácia” e da “facilidade de uso”; a importância percebida de critérios de desempenho de previsão é diferente entre os gerentes e os operadores para as empresas que vendem produtos básicos ou muito modernos; para as empresas que vendem produtos com maior nível de moda, os operadores percebem que os critérios “velocidade” e “requisitos dados” são relativamente menos importantes para a realização da previsão, no entanto, os gerentes valorizam muito mais esses dois critérios ao avaliar o desempenho da previsão de demanda de um determinado método; essa mesma diferença de relevância de critérios entre operadores e gestores também é percebida para o mercado de alta renda, o que não ocorre no mercado de massa.

A previsão de demanda é um desafio de longa data na indústria da moda. É o que diz Luce (2018) ao ressaltar aspectos que tornam a tarefa ainda mais difícil, como as curtas estações e a mudança de comportamento do consumidor, que resultam em um baixo volume de dados, sendo que, em geral, algoritmos de aprendizagem profunda demandam grande volume de dados. Assim, o autor sugere os seguintes algoritmos para pequenos conjuntos de dados: *transfer learning*, LSTM e Prophet.

Assim como Luce (2018), Loureiro, Miguéis e Silva (2018) avaliaram métodos de aprendizagem profunda para realizar previsões de vendas de novos produtos, o que implica, necessariamente, ausência de histórico. Os autores afirmam que métodos de previsão mais precisos tornam a gestão do estoque mais eficiente dentro

das corporações e, conseqüentemente, gera maior lucratividade. No mais, a manutenção de níveis de estoques adequados aumenta o nível de serviço entregue ao consumidor, elevando a sua satisfação e lealdade frente aos concorrentes.

Os autores foram inovadores ao propor para o setor um modelo que contemplasse, simultaneamente, os seguintes pontos: características físicas dos produtos, preço e características representativas do domínio de conhecimento dos gestores. As conclusões observadas por Loureiro, Miguéis e Silva (2018) evidenciam que o uso de técnicas de *deep neural networking* (DNN) e de aprendizado por máquina são adequadas para realizar previsões de vendas no setor de varejo de moda, mesmo quando não há o histórico de informações. A rede neural profunda foi o método com maior precisão, considerando as diversas métricas calculadas. Os métodos estabelecidos para comparação foram: árvores de decisão, RF, SVR, rede neural e regressão linear. Embora geralmente sugeridas como apropriadas para a análise de grandes bancos de dados, as RNs profundas também podem funcionar de forma adequada quando aplicadas a conjuntos de dados menores. Apesar da superioridade da DNN em termos de acurácia, ela não foi considerada a técnica mais adequada para os autores, pois seu processo de treinamento é mais complexo e ela oferece baixa capacidade de interpretação. Nesse sentido, eles recomendam a RF por oferecer tanto capacidade preditiva como maior facilidade de interpretação do modelo em si.

Ren, Hui e Choi (2018) versam sobre métodos de inteligência artificial na previsão de vendas de itens de moda no contexto de *big data*. Aspecto interessante ressaltado pelos autores é que os itens de moda são caracterizados, sob o ponto de vista do tipo de dado, como sendo amostras pequenas e de multidimensões, o que leva a previsões menos precisas e menos eficientes devido à maldição da dimensionalidade. Adicionalmente, estruturas de *big data* intensificam esse problema.

Catal *et al.* (2019) realizaram um estudo com o objetivo de prever as vendas semanais para produtos do Walmart com vários algoritmos de aprendizado por máquina e, posteriormente, fazer a comparação de desempenho entre eles. Sem uma previsão precisa, as companhias podem sofrer com problemas de fluxo de caixa, afirmam os autores, que usaram os seguintes algoritmos de aprendizagem de máquina: regressão linear, regressão bayesiana, rede neural, RF e *boosted decision tree*. Para efeito de comparação, foram aplicadas as seguintes técnicas de séries de tempo: ARIMA, ARIMA sazonal, suavização exponencial, suavização exponencial

sazonal, método *Naive*, método da média e método *drift*. Os autores constataram que os métodos de aprendizagem de máquina tiveram um desempenho superior aos métodos de séries temporais, em termos de precisão. Além disso, o método com o melhor desempenho foi o *boosted decision tree*. Para trabalhos futuros, é sugerida a aplicação de métodos híbridos e de métodos de *deep learning*.

A previsão de demanda para produtos de moda continua sendo uma tarefa desafiadora tanto do ponto de vista acadêmico como para o setor em si e Ren, Chan e Siqin (2020) argumentam que a chegada da era do *big data* permite uma revolução no processo de previsão para esses produtos, impactando negativamente os métodos tradicionais de previsão de demanda e planejamento de estoques. Os autores fazem uma revisão do processo de previsão na indústria e reforçam a importância do contexto de *big data*, assim como visto em Ren, Hui e Choi (2018), tendo a opinião do consumidor disponível em vários pontos de contato e as informações de mídias sociais como aspectos fundamentais para aprimoramento dos métodos de previsão.

Ren, Chan e Siqin (2020) destacam aspectos que podem contribuir para a revolução do processo de previsão de demanda de artigos de moda, os quais são: considerando a grande volumetria de dados disponível, o grande desafio passa a ser reduzir o nível de ruído presente nos dados para torná-los efetivamente úteis; antigamente, existia pouca informação para que o consumidor pudesse realizar as suas escolhas e a demanda dependia das tendências de moda estabelecidas pela própria indústria, enquanto nos dias de hoje, a informação de moda está mais acessível aos clientes, sendo fundamental que as empresas consigam capturar as suas demandas para aperfeiçoarem os seus respectivos métodos de previsão; por último, os métodos tradicionais de previsão de demanda não são adequados para o ambiente atual, considerando a sua baixa acurácia e sua incapacidade de reagir às rápidas mudanças dos gostos dos consumidores, gerando a necessidade de adoção de tecnologias avançadas para melhorar o desempenho da previsão de demanda.

Em resumo, Ren, Chan e Siqin (2020) afirmam que, ao coletar cada interação junto aos clientes e, assim, conseguir obter as suas preferências, as empresas podem obter previsões mais precisas. Ademais, os métodos de previsão e de recomendação podem sugerir produtos dos quais os clientes não precisariam, mas pelos quais possivelmente pagariam, aumentando a receita das companhias do setor.

Um estudo prático de previsão de vendas para uma multinacional asiática de sapato foi conduzido por Liu, Ming e Hu (2020). A previsão foi feita no nível de SKU, mas não de loja, pois os autores consideraram que, dessa forma, melhorariam a eficiência tanto da gestão dos estoques quanto da logística. Com o objetivo principal de evitar sobreajuste ou subajuste, o método escolhido foi MDL-NN. Essa técnica é um tipo de rede neural criado por Small e Tse (2002), que usa um critério de informação teórico, denominado como comprimento mínimo (MDL), para estabelecer o número de neurônios necessário na camada intermediária. O MDL é definido pela compressão de dados e representa a melhor descrição compacta de série temporal.

Para efeito de comparação, foram utilizados quatro métodos de referência: GM, SVM, *Markov regime switching* (MS) e ELM. Como resultado, o MDL-NN apresentou o menor erro de previsão frente aos outros métodos, sendo recomendado para esse tipo de problema. Além das métricas convencionais, como MAPE, MASE e SMAPE, os autores propuseram uma métrica denominada RMAPE para endereçar o problema da presença de valores zerados na série, que impacta o cálculo do MAPE. O cálculo obtido com o novo indicador foi consistente com os anteriores. Por fim, os autores afirmam que as previsões geradas pelos diversos algoritmos considerados produzem números decimais e não inteiros, mas que, na realidade, sempre são considerados como números inteiros. Esse arredondamento leva a resultados distintos de níveis de precisão, o que pode impactar o método de previsão a ser escolhido.

Último trabalho a ser evidenciado é a pesquisa de Masini, Medeiros e Mendes (2021) sobre os avanços na aplicação de métodos de aprendizagem e em estatística de alta dimensão para séries temporais. O estudo cobriu tanto métodos lineares como não lineares, sendo que a maior parte dos métodos apresentados foi utilizada nesta Tese. A evidência empírica mostrou que modelos de aprendizado por máquina não lineares combinados com grandes conjuntos de dados podem ser extremamente úteis para previsões econômicas e financeiras. Salienta-se que a parte prática desta pesquisa foi desenvolvida com base em um conjunto pequeno de dados, o que não necessariamente exclui a utilidade e o desempenho superior que pode ser obtido através da aplicação de métodos de aprendizagem de máquina não lineares.

Diante da análise de todos esses trabalhos, pode-se afirmar que esta Tese contribui de forma relevante, tendo em vista o baixo volume de trabalhos produzidos para o setor de moda com todos os objetivos aqui contemplados, abarcando os

questionamentos do impacto da sazonalidade, das variáveis externas e do processo de seleção de variáveis para previsão de vendas com métodos de aprendizado por máquina. Ao analisar esses aspectos no âmbito nacional, a inovação é ainda maior.

Os trabalhos evidenciados no referencial bibliográfico foram fundamentais para trazer um histórico e um panorama das técnicas utilizadas pelo setor de moda para produzir suas previsões, assim como foram bastante esclarecedores em apresentar todas as peculiaridades intrínsecas ao setor, que motivam o estudo e o desenvolvimento de modelos próprios para atender tais especificidades.

Após a revisão bibliográfica, que serve como fio condutor para os próximos passos do trabalho, será exposta detalhadamente a metodologia utilizada para a realização da previsão de vendas das séries temporais desta Tese.

5. Metodologia

A metodologia aplicada para o desenvolvimento desta Tese passa pela definição do problema, especificação dos métodos de previsão a serem implantados, apresentação do conjunto de dados usados e variáveis criadas, indicação do conjunto dos hiperparâmetros testados e divulgação das formas de avaliação implementadas, visando responder os principais questionamentos propostos.

O método científico empregado foi o dedutivo, um método racionalista (GIL, 1994). Essa metodologia leva em conta a razão como única maneira de se determinar o conhecimento verdadeiro. Logo, é utilizado um raciocínio descendente, da análise mais geral para particular, até que se possa chegar a uma dada conclusão.

No que diz respeito ao tipo de classificação de pesquisa, este trabalho apresenta vários tipos de classificações, sendo que cada um deles atende a um ponto de vista específico. Diante dessas diversas possibilidades, Gil (1994) destaca que um estudo pode ser categorizado por meio das seguintes características: sua natureza, sua abordagem, seu objetivo e os procedimentos técnicos aplicados.

Sob o ponto de vista da natureza, é uma pesquisa aplicada por produzir conhecimentos para aplicações práticas. Especificamente, o objetivo central é contribuir para gerar previsões de vendas com desempenho superior para empresas varejistas de vestuário, o que inclui o Lojão do Brás, empresa que concedeu os dados.

Em termos da abordagem empregada, pode ser classificada como uma pesquisa quantitativa, tendo em vista que todo o projeto é quantificável e foi empregado um conjunto de métodos matemáticos e estatísticos para testar e fundamentar de forma quantitativa o relacionamento entre a variável resposta e as variáveis explicativas.

Na perspectiva do objetivo de pesquisa, é vista como uma pesquisa descritiva, pois busca verificar vários aspectos que impactam a realização da previsão de curto prazo das vendas para o varejo de moda.

Na esfera do tipo de procedimentos técnicos usados, este trabalho pode ser considerado uma pesquisa bibliográfica, pois foi elaborado, em grande parte, com base em material já publicado, como também uma pesquisa experimental, já que foram aplicadas diversas técnicas de previsão, visando configurá-las e testá-las com o objetivo de responder aos problemas de pesquisas presentes nesta Tese.

Finalizado o posicionamento deste trabalho em relação aos diversos tipos de classificação propostos por Gil (1994), o próximo passo é realizar uma definição do problema de estudo, do conjunto de dados envolvidos, dos métodos aplicados e da forma de avaliação e apresentação dos resultados obtidos.

5.1. Definição do problema de previsão

Nesta seção é especificado o escopo do problema de previsão analisado nesta Tese. Assim, temos setor, variável a ser prevista, horizonte de previsão, requisitos para previsão no varejo de moda e objetivo, os quais são mais detalhados a seguir.

Setor:

O trabalho tem como foco o setor de varejo de moda, no qual a empresa Lojão do Brás, que concedeu os dados para a pesquisa, atua. A companhia possui um *mix* bastante diversificado de produtos, que contempla itens dos seguintes departamentos: moda masculina, moda feminina, moda infantil, moda íntima, moda bebê, moda praia, acessórios e moda lar. Apesar da presença de uma gama ampla de produtos, apenas as categorias mais relevantes em termos de faturamento são analisadas.

Variável a ser prevista:

A variável a ser prevista são as vendas semanais – em termos de quantidade e em unidades – de quatro categorias de produtos: camiseta de homem, camiseta de menino, blusa de mulher e blusa de menina, sendo considerados somente os itens básicos. As categorias foram escolhidas por sua representatividade no faturamento da empresa, correspondendo a, aproximadamente, 35% das vendas totais do departamento de moda, dentre as 70 categorias existentes. Salienta-se que o departamento de moda corresponde a 80% do faturamento total da rede de lojas.

Horizonte de previsão:

A previsão a ser realizada é no horizonte de curto prazo, a próxima semana, indo ao encontro da carência de métodos de previsão de curto prazo notada por Liu *et al.* (2013). A granularidade semanal foi escolhida, pois é nessa frequência que a empresa estudada faz suas projeções e, coincidentemente, foi a forma verificada nos trabalhos de Thomassey, com diversos artigos sobre o assunto abordado nesta Tese.

Requisitos para previsão no varejo de moda:

Conforme observado anteriormente, Thomassey (2014) destacou que algumas definições devem ser verificadas previamente para que o processo de previsão de vendas seja feito para o varejo de moda. Sendo assim, a previsão de vendas neste trabalho é realizada considerando os seguintes aspectos:

- **Horizonte de previsão:** curto prazo com foco em reposição na fase *in-season*;
- **Ciclo de vida de produto:** itens básicos que vendem ao longo de todo ano, tendo em vista a aplicação majoritária de métodos de previsão baseados em aprendizagem de máquina, que demandam um alto volume de dados;
- **Nível de agregação de vendas:** a previsão é realizada por categoria de produto, que pode ser considerado um nível de agregação intermediário;
- **Sazonalidade:** são previstos produtos não sazonais vendidos ao longo do ano, diferentemente de produtos específicos para o inverno e para o verão;
- **Variáveis externas:** são consideradas as variáveis macroeconômicas, as variáveis climáticas e as variáveis de datas comemorativas.

Não é foco desta pesquisa realizar a previsão de produtos do grupo *fashion*, cuja previsão é muito difícil de ser executada, segundo Thomassey (2010).

Objetivo:

Este trabalho é um ensaio com múltiplos objetivos dentro do contexto de previsão de vendas de curto prazo para o varejo de moda. Certamente, o principal propósito desta Tese é realizar a recomendação do modelo mais adequado para as previsões de vendas no varejo de moda, o que é feito por meio do uso da análise AHP e com base no trabalho de Ren, Chan e Ram (2017). Além disso, outros objetivos são:

- A aplicação de tratamentos sazonais impacta a realização das previsões?
- O emprego de técnicas de seleção de variáveis ajuda a aumentar a precisão das estimativas? Qual técnica apresenta o melhor desempenho?
- A utilização de seleção de variáveis reduz o custo computacional das previsões?
- Quais variáveis exógenas e variáveis criadas tiveram mais importância no processo de previsão de vendas?
- Qual modelo, diante dos 79 métodos testados, apresentou o melhor desempenho em termos de precisão?

5.2. Definição dos métodos de previsão

Neste ensaio são avaliados, no total, 16 algoritmos distintos. Além dos algoritmos, são aplicados tratamentos sazonais, tratamentos de seleção de variáveis e métodos *ensembles*. A combinação das diversas opções de tratamentos com os algoritmos disponíveis gerou 79 modelos possíveis de previsão de vendas.

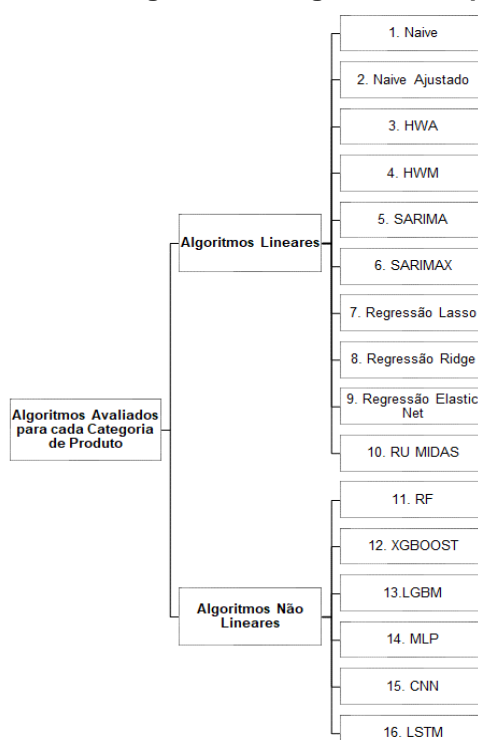
No âmbito da taxonomia e, sob uma perspectiva mais restrita, poderia se afirmar que cada modelo de previsão estabelecido, em que houve a mudança de um dos seus hiperparâmetros, se configura como um modelo de previsão distinto.

Buscou-se agrupar os diversos modelos de previsão implementados sob a ótica dos principais algoritmos aplicados, mesmo que parte deles apresente, estruturalmente, poucas distinções entre si por terem como origem um mesmo método base. Esse é o caso dos modelos de regressão, que poderiam ser agrupados na classe de regressão; dos modelos Naives e SARIMAs, que poderiam ser reunidos na

categoria modelos Box-Jenkins; dos modelos baseados em árvores de decisão, em que o LGBM é uma melhoria do XGBOOST e que poderiam ser classificados como árvores de decisão; e dos modelos baseados em redes neurais, sendo que a rede LSTM pode ser vista como uma evolução da MLP, abrangendo o fator da recorrência.

Na Figura 1.22 é possível verificar os algoritmos aplicados, os quais foram subdivididos em algoritmos lineares e não lineares.

Figura 1.22 – Categorias dos algoritmos de previsão



Fonte: elaboração própria.

Uma outra forma mais harmônica de agrupamento dos métodos é a separação inicial entre métodos lineares e não lineares e, posteriormente, subdividir os métodos lineares em métodos de regressão e em métodos Box-Jenkins. Em relação aos não lineares, a sugestão é associar os métodos disponíveis em duas principais categorias: métodos baseados em árvores de decisão e métodos baseados em redes neurais.

Em relação aos métodos baseados em redes neurais, tanto a LSTM como a CNN são arquiteturas de redes mais utilizadas, respectivamente, para resolver problemas de reconhecimento de linguagem/escrita e para classificação de imagens. Apesar disso e de forma prática, existem vários trabalhos e competições no Kaggle em que ambos os algoritmos foram aplicados para previsões de séries temporais.

No âmbito acadêmico, existem vários trabalhos evidenciando a aplicação da LSTM, inclusive a sua superioridade em relação a outros métodos de previsão em termos de acurácia, como pode ser observado em Abbasimehr, Shabani e Yousefi (2020). Por isso, optou-se por sua consideração.

A aplicação da CNN é encontrada em uma volumetria muito menor quando comparada com a MLP e à própria LSTM, contudo optou-se por sua inclusão, dada sua menor complexidade diante da LSTM. Em trabalhos em que a CNN foi implementada, destaca-se artigo de Jiang, Kelly e Xiu (2020), em que codificam os dados de mercado como imagem e tentam realizar a previsão do retorno. Os autores constataram que os padrões preditivos identificados pela CNN foram altamente robustos a variações na especificação do modelo e no controle de um grande conjunto de variáveis preditivas.

Tabela 1.5 – Quadro de modelos aplicados

Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Tratamento Sazonal	Tratamento de Seleção de Variáveis/Previsão	Tipo Ensemble
NAIVE	Linear	Naive	Não	Não	Não
NAIVEA	Linear	Naive	Não	Não	Não
Regressão Elastic Net	Linear	Regressão	Não	Não	Não
Regressão Lasso	Linear	Regressão	Não	Não	Não
Regressão Ridge	Linear	Regressão	Não	Não	Não
RUMIDAS	Linear	Regressão	Não	Não	Não
SARIMA	Linear	Regressão	Não	Não	Não
SARIMAX	Linear	Regressão	Não	Não	Não
HWA	Linear	Suavização Exponencial	Não	Não	Não
HWM	Linear	Suavização Exponencial	Não	Não	Não
Regressão Ridge	Linear	Regressão	Não	Não	Ensemble Stacking
Regressão Ridge	Linear	Regressão	Não	Rfa	Ensemble Stacking
MÉDIA SIMPLES PREVISÕES	Linear	Média	Não	Não	Ensemble Blending
CNN	Não Linear	Rede Neural	Diferença	Não	Não
MLP	Não Linear	Rede Neural	Diferença	Não	Não
CNN	Não Linear	Rede Neural	Não	Não	Não
LSTM	Não Linear	Rede Neural	Não	Não	Não
MLP	Não Linear	Rede Neural	Não	Não	Não
CNN	Não Linear	Rede Neural	Polyfit	Não	Não
LSTM	Não Linear	Rede Neural	Polyfit	Não	Não
MLP	Não Linear	Rede Neural	Polyfit	Não	Não
CNN	Não Linear	Rede Neural	Diferença	Pvalue	Não
MLP	Não Linear	Rede Neural	Diferença	Pvalue	Não
CNN	Não Linear	Rede Neural	Não	Pvalue	Não
LSTM	Não Linear	Rede Neural	Não	Pvalue	Não
MLP	Não Linear	Rede Neural	Não	Pvalue	Não
CNN	Não Linear	Rede Neural	Polyfit	Pvalue	Não
LSTM	Não Linear	Rede Neural	Polyfit	Pvalue	Não
MLP	Não Linear	Rede Neural	Polyfit	Pvalue	Não
MLP	Não Linear	Rede Neural	Diferença	Rfa	Não
LSTM	Não Linear	Rede Neural	Não	Rfa	Não
MLP	Não Linear	Rede Neural	Não	Rfa	Não
LSTM	Não Linear	Rede Neural	Polyfit	Rfa	Não
MLP	Não Linear	Rede Neural	Polyfit	Rfa	Não
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking
LGBM	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Stacking
RF	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Stacking
XGBOOST	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Stacking
CNN	Não Linear	Rede Neural	Não	Não	Ensemble Stacking
MLP	Não Linear	Rede Neural	Não	Não	Ensemble Stacking
MLP	Não Linear	Rede Neural	Não	Rfa	Ensemble Stacking
LGBM	Não Linear	Árvore de Decisão	Diferença	Não	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Diferença	Não	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Polyfit	Não	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Polyfit	Não	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Diferença	Pvalue	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Diferença	Pvalue	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Não	Pvalue	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Não	Pvalue	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Polyfit	Pvalue	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Polyfit	Pvalue	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Diferença	Rfa	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Diferença	Rfa	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Polyfit	Rfa	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Polyfit	Rfa	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Diferença	Shap	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Diferença	Shap	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Não	Shap	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Não	Shap	Ensemble Boosting
LGBM	Não Linear	Árvore de Decisão	Polyfit	Shap	Ensemble Boosting
XGBOOST	Não Linear	Árvore de Decisão	Polyfit	Shap	Ensemble Boosting
RF	Não Linear	Árvore de Decisão	Diferença	Não	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Polyfit	Não	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Diferença	Pvalue	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Não	Pvalue	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Polyfit	Pvalue	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Diferença	Rfa	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Não	Rfa	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Polyfit	Rfa	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Diferença	Shap	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Não	Shap	Ensemble Bagging
RF	Não Linear	Árvore de Decisão	Polyfit	Shap	Ensemble Bagging

Fonte: elaborado pelo autor.

Ainda a respeito da CNN, nesta Tese foi aplicada sua arquitetura convencional. Todavia, já existe uma arquitetura mais adequada para lidar com séries temporais chamada rede neural convolucional temporal (TCN), proposta por Lea *et al.* (2016). Suas principais vantagens são a possibilidade de uso do paralelismo computacional e maior estabilidade dos gradientes, que muitas vezes explodem ou desaparecem no caso das RNN. A validação dos resultados do uso da TCN pode gerar estudo futuro.

Outro aspecto é que para os métodos *Naive*, *Naive* ajustado, *Holt Winters* com sazonalidade aditiva (HWA), *Holt Winters* com sazonalidade multiplicativa (HWM) e SARIMA apenas foi usada a série de dados pura. Para o SARIMAX, além da série da quantidade de cada categoria, foram consideradas também as variáveis exógenas. Para os demais métodos, foram contemplados os dados de cada série, as variáveis exógenas e as variáveis criadas, detalhadas na próxima seção.

Para os algoritmos de regressão, considerados de forma individual e não de forma combinada, não foram aplicados métodos de tratamento de sazonalidade e nem de tratamento de seleção de variáveis, já que os métodos selecionados têm como objetivo fazer uma seleção das variáveis mais importantes através dos tratamentos de regularização contemplados.

Para os algoritmos que compõem o grupo de métodos não lineares, foi aplicado dois possíveis métodos de tratamento sazonal e três possíveis métodos de seleção de variáveis, o que está especificado a seguir. Considerando, esses tratamentos, juntamente com os métodos *ensembles* foram desenvolvidos 79 modelos para cada uma das séries de produtos estudadas, os quais podem ser observados na Figura 1.5.

Além de todos esses modelos, foram testadas diversas configurações de hiperparâmetros possíveis para os métodos não lineares, visando encontrar o modelo mais adequado em termos de precisão para cada série. O processo de definição dos hiperparâmetros foi feito antes da implementação dos métodos de seleção de variáveis e com base em um método Bayesiano para reduzir o custo computacional.

5.3. Base de dados e variáveis

Nesta seção do trabalho são abordadas as bases de dados usadas, suas respectivas fontes e a divisão temporal feita nos dados para treinamento e testes dos modelos, além das variáveis criadas através da variável original, denominado *feature engineering*, já explicado anteriormente.

Em termos de bases de dados, o estudo utiliza bases primárias e secundárias. Dados de natureza primária, capturados na sua forma original, são representados pelas quantidades e pelos preços médios das quatro categorias de produtos selecionadas, cujos preços foram corrigidos pelo Índice de Preços ao Consumidor (IPCA). Sobre os dados de natureza secundárias, os quais são coletados e

manipulados por outras pessoas com outros fins, algumas variáveis exógenas foram contempladas com o intuito de contribuir para geração de previsões mais precisas. As variáveis exógenas podem ser divididas nas classes: variáveis exógenas climáticas, variáveis exógenas macroeconômicas e variáveis exógenas de eventos.

A escolha das variáveis exógenas macroeconômicas tem como base o estudo da SBVC (2020) e o trabalho de Moraes e Silva (2015). No primeiro, vê-se que quatro elementos são fundamentais para o crescimento e desenvolvimento do setor de varejo: renda, emprego, crédito e confiança. No trabalho de Moraes e Silva, nota-se que as vendas no varejo são condicionadas por três fatores macroeconômicos: taxa de juros, taxa de desemprego e renda média real. Com base na importância desses fatores, foram selecionadas bases de dados equivalentes a estas informações.

As variáveis exógenas climáticas têm como representantes: a temperatura máxima registrada, a mínima registrada e o nível de precipitação registrado. Para o desenvolvimento dos modelos de previsão foram usadas as informações climáticas registradas – e não as previsões –, da empresa Somar. Entretanto, quando o modelo for colocado em produção, deverão ser utilizadas as previsões de temperatura e de nível de precipitação que a empresa disponibiliza com 15 dias de antecedência.

As variáveis exógenas de eventos foram criadas como binárias para controlar o efeito dos principais eventos do calendário varejista, que geram picos de venda difíceis de prever. Os principais eventos definidos foram: Dia das Mães, Dia dos Namorados, Dia dos Pais, Dia das Crianças, *Black Friday* e Natal. Para esses eventos, a binária foi incluída para duas semanas, iniciando uma semana antes do evento.

Dada a importância do mês de dezembro para o faturamento, atingindo cerca de 25% do total do ano, foi considerada uma variável binária para todas as semanas do mês. Também foram consideradas binárias para algumas iniciativas do governo para promoção do consumo, como: redução da alíquota do IPI da linha branca de móveis de mar/2010 a dez/2014; redução da alíquota do IPI da linha branca de dez/2011 a dez/14; liberação do saque do FGTS de mar/17 a julho/17; e segunda liberação do saque do FGTS de set/2019 a mar/2020. Foi colocada uma variável para o evento da greve dos caminhoneiros em 21/05/2018. No período que durou essas iniciativas governamentais, foi incluído o valor unitário nas respectivas variáveis.

A Tabela 1.6 mostra um quadro das variáveis exógenas com o resumo, explicitando sua fonte, frequência e forma de divulgação. A maior parte das variáveis

foi extraída do banco de dados de séries temporais do Banco Central. As variáveis exógenas de eventos não foram colocadas na tabela, pois elas são apontamentos gerados pela própria autora, não sendo extraídas de nenhuma base secundária.

Tabela 1.6 – Quadro das variáveis exógenas

Variáveis Exógenas										
Item	Tipo	Descrição	Unidade	Site	Fonte	Frequência	Código Fonte	Descrição Fonte	Data Início	Divulgação
1	Climática	Temperatura mínima registrada	°C		Somar	Diária			01/01/2000	Diária
2	Climática	Temperatura máxima registrada	°C		Somar	Diária			01/01/2000	Diária
3	Climática	Nível de precipitação registrada	mm		Somar	Diária			01/01/2000	Diária
4	Macroeconômica	Índice de Confiança do Consumidor (ICC)	Índice	BCB	Fecomercio	Mensal	4393	Índice de Confiança do Consumidor	31/03/1999	2 meses após
5	Macroeconômica	Taxa de desemprego da RMSP	Participação %	IPEA DATA	Fundação Sistema Estadual de Análise de Dados, Pesquisa de Emprego e Desemprego (Seade/PED)	Mensal		Taxa de desemprego na RMSP	01/12/1984	2 meses após
6	Macroeconômica	Massa salarial ampliada disponível real	R\$	BCB	BCB-Depec	Mensal	22080	Massa Salarial Ampliada Disponível - PNADC (acumulada em 12 meses)	01/12/2004	2 meses após
7	Macroeconômica	Crédito para Pessoa Física % PIB	Participação %	BCB	BCB-DSTAT	Mensal	20624	Saldo da carteira de crédito a pessoas físicas em relação ao PIB	01/03/2007	2 meses após
8	Macroeconômica	Índice em volume da Pesquisa Mensal do Comércio para o Vestuário em SP	Índice	IBGE	IBGE	Mensal	3418	Índice de volume de vendas no comércio varejista (índice de base fixa 2014=100)	01/01/2000	2 meses após

Fonte: elaboração própria.

Com exceção das variáveis de temperatura e de nível de precipitação, as demais variáveis exógenas apresentam um atraso de dois meses, em média, no momento de sua divulgação. Esse atraso foi contemplado nos modelos de previsão com o objetivo de realizar um experimento em condições extremamente próximas ao real. Exemplificando, para previsão das vendas das semanas de janeiro, foram usadas as informações de índice de confiança do consumidor de novembro do ano anterior.

Fundamental frisar que a maior parte das variáveis exógenas tem periodicidade mensal e não semanal. Assim, foi aplicada uma interpolação polinomial de terceiro grau para transformar a frequência das informações de mensal para semanal. Outra estratégia seria a repetição do mesmo valor para as diversas semanas compreendidas em um dado mês. O uso dessa abordagem não seria adequado para as redes neurais, por isso optou-se pela primeira alternativa. Essa transformação não foi aplicada para o método RU-MIDAS, já que ele é capaz de tratar dados com frequências distintas.

Além das variáveis mencionadas acima e nas páginas anteriores, foram criadas variáveis baseadas na variável de quantidade vendida e na data mediante o uso de técnicas *feature engineering*. Através da variável de data, foi criada a variável ano, mês e semana. Por meio da variável de quantidade vendida, foram geradas as variáveis: quantidade vendida defasada, diferença da quantidade vendida defasada, diferença da quantidade vendida percentual, média móvel da quantidade vendida,

desvio padrão móvel da quantidade vendida e distância da defasagem em relação à média. A escolha foi feita a partir dos estudos de Brownlee (2020a) e Filho (2019a).

A seguir, é apresentada de forma detalhada cada uma das variáveis contempladas no trabalho, iniciando pelas variáveis internas.

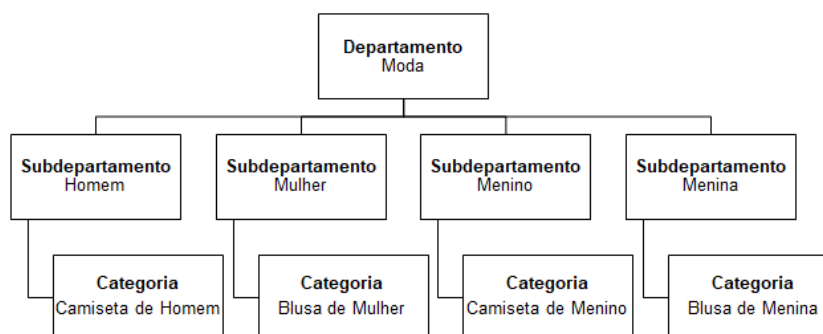
Vendas semanais por categoria de produto em quantidades

O objetivo deste trabalho é realizar a previsão de vendas em quantidades das principais categorias de produtos vendidas na empresa Lojão do Brás. Verificou-se que quatro categorias de produto correspondem por 35% das vendas totais do departamento de moda, o qual representa 80% das vendas da empresa.

Para o Lojão do Brás, uma categoria de produto é a soma da venda de diversos SKUs, o menor nível de cadastramento de um produto dentro da empresa. Especificamente, o nível de agrupamento a ser previsto é obtido através da junção do subdepartamento com uma categoria, determinando a categoria de produtos.

As quatro categorias de produtos foco desta Tese são: camiseta de homem, blusa de mulher, camiseta de menino e blusa de menina. Na Figura 1.23 é evidenciada a estrutura hierárquica de cadastramento de produtos na companhia estudada.

Figura 1.23 – Hierarquia de cadastro dos produtos do Lojão do Brás



Fonte: elaboração própria.

Aspecto fundamental a ser colocado é que as vendas são previstas na forma de diferença, já que a série de vendas não é estacionária.

Preço médio categoria de produto

O Lojão do Brás tem como foco os públicos C e D, sendo o preço uma variável que afeta o nível de consumo de seus clientes. Assim, a inclusão da variável tem como objetivo capturar impactos nas quantidades vendidas como consequência de

alterações nos preços, devido a mudanças nos seus respectivos custos, promoções, liquidações etc. Como a empresa possui o histórico de preço dos seus produtos na granularidade diária, foi realizada uma média semanal para inclusão dessa informação no modelo na periodicidade requisitada. Para que a informação fosse comparável, foi feita a correção dos preços pelo IPCA.

Em termos de variáveis internas criadas, foram constituídas as seguintes:

Defasagens

As defasagens representam as vendas em semanas passadas. Nesse sentido, foram incorporados ao modelo as vendas das 52 semanas anteriores em relação a cada semana existente no conjunto de dados.

Diferença defasadas

A *feature* diferença é parecida com a defasagem. Porém, ao invés de usar o valor anterior em si, aplica-se a diferença entre dois períodos. Assim também foram incorporadas ao modelo sete diferenças defasadas.

Diferença percentual defasada

Essa variável é a diferença percentual dividida pelo valor real da quantidade vendida no período imediatamente anterior. Foram contempladas sete defasagens.

Média Móvel

Foi considerada a média móvel da primeira defasagem com a janela de 2 a 12 períodos. Não foi aplicado sobre o dado real, pois incorreria no problema de vazamento de informação, mas sim na primeira defasagem.

Desvio Padrão Móvel

Foi considerado o desvio padrão móvel da primeira defasagem com a janela de 2 a 16 períodos. Assim como na variável média móvel, não foi calculado sobre o dado real, pois haveria vazamento de informação, o que acabaria interferindo na previsão.

Distância da defasagem em relação à média

O objetivo dessa variável é constatar o quão distante a primeira defasagem está da média móvel com a janela variando de 2 a 16 períodos.

As variáveis criadas são altamente correlacionadas entre si, podendo levar a problemas de multicolinearidade. Por isso, é fundamental a aplicação de métodos de seleção de variáveis para minimizar os problemas de multicolinearidade, aumentar a parcimônia, reduzir o custo computacional e gerar previsões mais precisas.

Como já colocado, também foram criadas as variáveis de ano, mês e semana, fundamentado na data de início de cada observação pertencente às séries estudadas. Conceitos de ano, mês e semana são conhecidos, não necessitando esclarecimentos.

Em termos de variáveis exógenas, temos:

Temperatura e nível de precipitação médios registrados

O Lojão do Brás adquire informação meteorológica junto a Somar Meteorologia e possui as condições registradas em termos de temperatura e nível de precipitação na região onde atua. Para os modelos que foram previstos, como os dados recebidos são diários, foi feito o cálculo da média semanal tanto para temperatura como para o nível de precipitação. A temperatura afeta, principalmente, as vendas de produtos de inverno. Já o nível de precipitação afeta o volume das vendas em geral, devido a empresa ter foco em lojas de rua e não de *shopping*, como outros varejistas.

Calendário varejista (binárias)

Foi criada uma binária para cada data relevante do calendário varejista, tendo em vista a importância de controlar seus efeitos, que afetam com distintas intensidades a quantidade vendida: Dia das Mães, Dia dos Namorados, Dia dos Pais, Dia das Crianças, *Black Friday*, que tem ganhado força nos últimos anos, e Natal.

Índice de confiança do consumidor

É utilizado o índice da Federação do Comércio de Bens, Serviços e Turismo do Estado de São Paulo (FecomercioSP) de confiança do consumidor com periodicidade mensal. Sabe-se que quando o consumidor está satisfeito e otimista em relação ao futuro, ele tende a gastar mais e é esse aspecto que se almeja incluir no modelo.

Crédito disponível para pessoas físicas como participação do PIB

O nível de crédito disponível na economia é um dos principais fatores para determinar o nível de consumo das famílias. O indicador é divulgado na frequência mensal e, quando ocorre um aumento do valor agregado de um bem, ocorre um aumento de necessidade de crédito por parte dos consumidores para conseguir

adquiri-lo. No varejo de vestuário, as principais redes de lojas do Brasil, como C&A, Marisa, Pernambucanas, Riachuelo e Renner, possuem seus cartões *private label* exatamente para aumentar o volume de crédito disponível dos seus clientes.

Massa salarial ampliado disponível real

O indicador macroeconômico reflete, de forma conjunta, o aumento do número de empregos e o nível de renda, evidenciando a soma de todos os salários pagos. Ele é de suma importância, pois corresponde ao total que poderia ser gasto para consumo, afetando, assim, a atividade varejista como um todo. A informação é disponibilizada pelo Banco Central e é divulgada na frequência mensal.

Taxa de desemprego na região metropolitana de SP

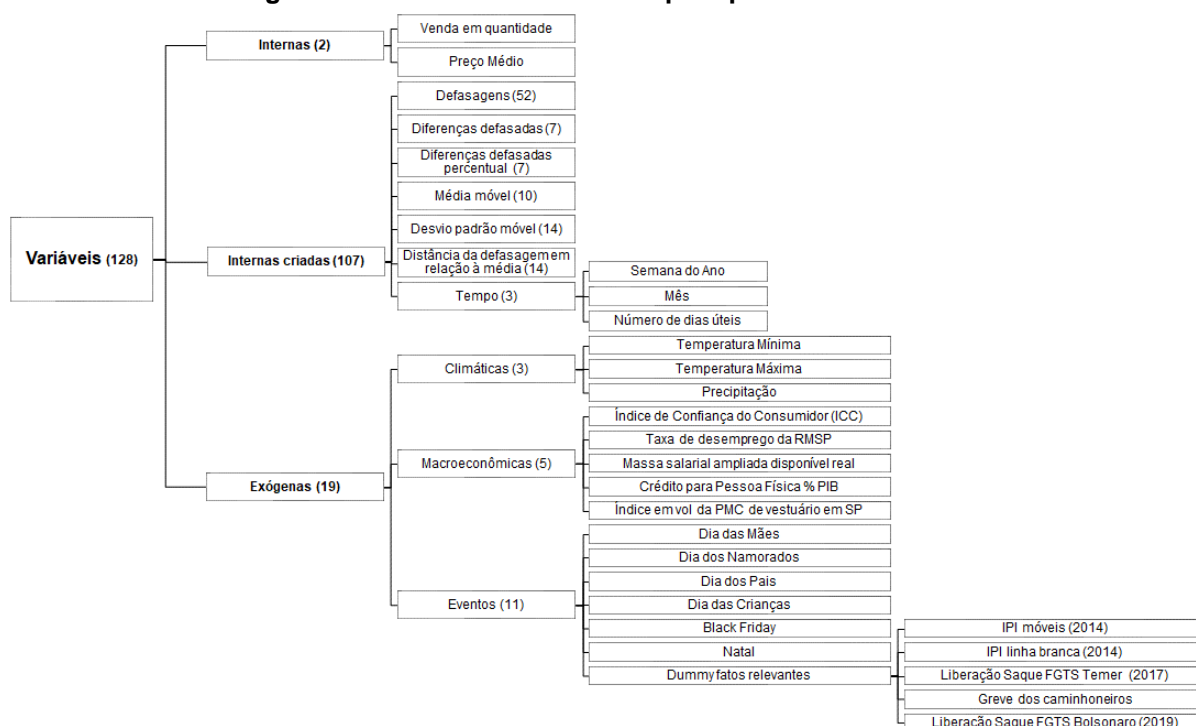
Um maior nível de desemprego, certamente, afeta o nível de consumo. Nesse sentido, a Tese leva em conta a taxa de desocupação medida pela Pesquisa de Emprego e Desemprego (PED), da Fundação Sistema Estadual de Análise de Dados (SEADE) de São Paulo, com periodicidade de divulgação mensal.

Índice em volume da pesquisa mensal do comércio para categoria de vestuário no estado de São Paulo

O IBGE divulga mensalmente a pesquisa sobre o comércio, que funciona como um termômetro de como está evoluindo a economia para esse setor. A pesquisa é publicada de forma detalhada para cada um dos segmentos existentes, os quais são estabelecidos por meio da classificação nacional de atividade econômica (CNAE).

O objetivo da inclusão da variável no modelo de previsão é incluir e/ou controlar o efeito do setor, tendo em vista que melhorias no indicador podem refletir aumentos nas quantidades comercializadas. Foi selecionado especificamente o índice de volume de vendas para comércio varejista de tecidos, vestuário e calçados no estado de São Paulo, pois está mais diretamente relacionado com o tipo de atividade exercido pelo Lojão do Brás. O indicador compara o nível de volume de venda do mês com a média mensal obtida no ano de 2014 e apresenta frequência de divulgação mensal.

A Figura 1.24 apresenta um quadro-resumo com todas as variáveis contempladas nos modelos de previsão que são implementados.

Figura 1.24 – Variáveis utilizadas para previsão de vendas

Fonte: elaboração própria.

Em relação às variáveis apresentadas, aspecto que merece ser salientado é sobre a estacionariedade das variáveis utilizadas, independente se a variável é endógena ou exógena. Tal aspecto é um pré-requisito para a construção dos modelos Box-Jenkins, como o SARIMA e o SARIMAX, mas não é uma condição necessária para os modelos não lineares. Apesar disso, aplicou-se a técnica de diferenciação para todas as variáveis não estacionárias, a partir do resultado do teste de raiz unitária feito por meio de três testes: ADF, KPSS e DFGLS. Na Tabela 1.7, um quadro-resumo mostra as variáveis que foram consideradas estacionárias e não estacionárias. Não foram consideradas as variáveis binárias, pois elas são estacionárias.

Tabela 1.7 – Quadro-resumo variáveis estacionárias e não estacionárias

Variável	Teste ADF	Teste KPSS	Teste DFGLS	# Qtd Não Estacionário	# Qtd Estacionário	Resultado Final
Crédito disponível para pessoas físicas como % do PIB	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Índice de confiança do consumidor	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Índice em volume da PMC para categoria de vestuário em SP	Não Estacionário	Estacionário	Não Estacionário	2	1	Não Estacionário
Massa salarial ampliada disponível real	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Nível de precipitação	Estacionário	Estacionário	Estacionário	0	3	Estacionário
Preço médio blusa de menina	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Preço médio blusa de mulher	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Preço médio camiseta de homem	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Preço médio camiseta de menino	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Quantidade vendida blusa de menina	Não Estacionário	Não Estacionário	Estacionário	2	1	Não Estacionário
Quantidade vendida blusa de mulher	Estacionário	Não Estacionário	Não Estacionário	2	1	Não Estacionário
Quantidade vendida camiseta de homem	Não Estacionário	Estacionário	Não Estacionário	2	1	Não Estacionário
Quantidade vendida camiseta de menino	Não Estacionário	Não Estacionário	Estacionário	2	1	Não Estacionário
Taxa de desemprego na região metropolitana de SP	Não Estacionário	Não Estacionário	Não Estacionário	3	0	Não Estacionário
Temperatura máxima	Estacionário	Estacionário	Estacionário	0	3	Estacionário
Temperatura mínima	Estacionário	Estacionário	Estacionário	0	3	Estacionário

Fonte: elaboração própria.

Além da aplicação do cálculo da diferença para tornar as variáveis estacionárias, foi implementada a normalização em todas as variáveis, com exceção das variáveis binárias. A normalização é um tratamento recomendado, sobretudo, para aplicação de algoritmos baseados em redes neurais. É fundamental destacar que a normalização foi aplicada de forma separada entre o conjunto de dados de treino e de teste para evitar o problema de vazamento de informação, que compromete os resultados gerados.

As quatro categorias de produtos analisadas possuem séries de vendas semanais, compreendendo as semanas de janeiro de 2007 a dezembro de 2019, e representam a quantidade de unidades vendidas para cada semana.

Para a formulação do modelo de previsão mais apropriado é preciso separar parte dos dados para treinar o modelo e a outra parte para testá-lo. Essa divisão é conhecida como *insample* (treino) e *outsample* (teste). O período de treinamento compreendeu as semanas de janeiro de 2007 a dezembro de 2017, o que representa 576 observações. Para verificação efetiva do desempenho, em termos de acurácia, foram analisados os resultados obtidos no período de teste, o qual compreendeu as semanas dos anos de 2018 e 2019, que significa 104 observações.

Ademais, é importante destacar que não foi utilizado período para validação e que foi utilizada a técnica de janelamento expansível, isto é, conforme o tempo vai passando, os dados de treino aumentam. A forma de avaliação dos resultados está mostrada a seguir

5.4. Forma de avaliação das previsões

A avaliação das previsões geradas no período teste será avaliada a partir do cálculo das métricas de precisão visto no capítulo 3.5, as quais são: MAE, RMSE, MAEP, RMSEP, RMSLE e o MAPE, sendo que as duas últimas são métricas padronizadas.

Para a aplicação do teste MCS, que busca verificar o poder preditivo dos métodos de previsão aplicados, utilizou-se um nível de confiança de 95% e 1000 amostras *bootstrap*.

5.5. Tratamentos sazonais aplicados

Não é unanimidade na academia que métodos não lineares, sobretudo os baseados em redes neurais, são capazes de apreender o padrão sazonal presente em uma série temporal. Assim, optou-se por testar dois tipos de tratamento sazonal: (1) dessazonalização por diferença e (2) dessazonalização polinomial e, assim, verificar o impacto positivo na realização das previsões. O objetivo principal da dessazonalização é limpar o padrão sazonal de uma determinada série de dados, tornando, a princípio, mais fácil de entender o comportamento da série em foco.

A dessazonalização por diferença consiste em calcular a diferença semanal e não somente a primeira diferença da série de dados. Um ponto negativo dessa metodologia é a perda de informação, já que são eliminados 52 pontos de dados.

A modelagem da sazonalidade é obtida em um processo de decomposição. O componente sazonal em uma determinada série temporal é, provavelmente, uma onda senoidal com período e amplitude geralmente fixos. Tal componente pode ser aproximado usando um método de ajuste de curva, o qual é, geralmente, polinomial e, no estudo em questão, foi definido como sendo um polinômio de terceiro grau.

Além dos tratamentos de sazonalidade citados, foram feitos testes preliminares com o tratamento da sazonalidade por meio de binárias juntamente com os métodos de regressão lasso e do LGBM. Apesar do maior poder de interpretação desse modelo, em que é possível entender com mais clareza a importância de cada semana para a previsão considerada, as estimativas geradas não se mostraram mais precisas do que as estimativas produzidas pelos demais modelos contemplados, o que fez com que essa forma de tratamento sazonal fosse preterida em relação às demais.

5.6. Tratamento de seleção de variáveis

A aplicação de métodos de seleção de variáveis busca reduzir o custo computacional da modelagem e aprimorar o nível de precisão do modelo. Neste estudo foram empregados os seguintes métodos: (1) seleção das variáveis, considerando o nível de significância de 5%; (2) aplicação do método SFS, também denominada RFA (*recursive feature Addition*), que consiste em uma pesquisa iterativa, na qual são adicionadas variáveis por meio da avaliação de alguma medida de

desempenho, que nesse caso foi o RMSLE; e (3) SHAP, um método que calcula a contribuição marginal de cada variável para gerar a previsão do modelo.

Em termos de categorização, o primeiro método é categorizado como de filtragem e os demais são do tipo *wrappers*. Para todos os métodos de seleção de variáveis utilizados, definiu-se previamente que fossem selecionadas, no máximo, 30 variáveis para compor cada modelo. O valor é arbitrário e pode ser testado. A decisão por esse valor foi com base na regra de Pareto, a qual afirma que para muitos eventos, aproximadamente, 80% dos efeitos vêm de 20% das causas. Como temos 128 variáveis, 20% corresponderiam a cerca de 26 variáveis.

5.7. Especificação dos hiperparâmetros dos métodos de previsão

Além da aplicação dos tratamentos de sazonalidade e de seleção de variáveis para os métodos que têm hiperparâmetros, presentes tanto nos métodos lineares como nos não lineares, foram testadas alternativas de configuração de modelos.

Os métodos não lineares apresentaram uma maior quantidade de hiperparâmetros e nem todos os hiperparâmetros possíveis de serem ajustados foram contemplados. Cada configuração sugerida pode ser vista como um modelo independente, a ser testado para se alcançar o melhor desempenho possível. Em um processo experimental, os métodos usados são considerados heurísticos, tendo em vista que são algoritmos exploratórios que levam à resolução do problema em questão por meio de aproximações, que não necessariamente chegam à solução exata e definitiva, mas sim à melhor possível. Ou seja, não há uma forma exata para que hiperparâmetros sejam estabelecidos e, assim, obtenha-se um modelo ideal. É tarefa do analista testar um conjunto de possibilidades para alcançar a solução mais eficiente.

Todos os métodos usam, pelo menos, a própria variável a ser prevista na sua forma defasada como variável de *input* para a realização da previsão em si. Adicionalmente, os métodos de regressão e os não lineares contemplam as variáveis exógenas e as demais variáveis criadas para compor os seus respectivos modelos.

Os métodos lineares SARIMA e SARIMAX utilizam a modelagem Box-Jenkins para identificar os termos autorregressivos e de médias móveis necessários para cada um dos modelos, mais especificamente, essa informação é obtida pela análise da FAC e da FACP. Posteriormente, através do critério de AIC é estabelecido o modelo mais

parcimonioso. Para cada uma das categorias de produtos foram criados 17 modelos SARIMA e 17 modelos SARIMAX com o intuito de constatar a melhor configuração.

Os métodos de regressão têm como hiperparâmetro apenas o *alpha*, que é um parâmetro que determina a força da regularização a ser aplicada.

Os métodos não lineares possuem uma série de hiperparâmetros que precisam ser ajustados. Esses hiperparâmetros são muitos distintos entre os modelos de árvores de decisão e os modelos embasados em redes neurais.

Os principais hiperparâmetros de métodos baseados em árvores de decisão:

- **Profundidade máxima (max_depth):** refere-se à profundidade máxima que uma árvore pode ter;
- **Número máximo de features (max_features):** é o número máximo de atributos a serem contemplados durante a divisão;
- **Amostragem mínima para folha (min_sample_leaf):** é o número de amostras no nível folha (terminal);
- **Amostragem mínima para divisão (min_samples_split):** é o número mínimo de amostras para considerar um nó para divisão;
- **Taxa de aprendizado (learning_rate):** visa controlar o peso de novas árvores adicionadas ao modelo;
- **N_estimators:** número de árvores contempladas.

No caso dos métodos baseados em redes neurais, os hiperparâmetros mais importantes são:

- **Número de camadas escondidas:** número de camadas que lidam com a não linearidade;
- **Número de neurônios em cada camada escondida:** especificação da quantidade de neurônios em cada camada escondida;
- **Taxa de aprendizado:** é a taxa que determina a velocidade de aprendizado;
- **Dropout:** é o nível de desligamento da rede neural, funciona como uma técnica de regularização;
- **Função de ativação:** é a responsável por gerar a saída y do neurônio através da ponderação dos valores dos pesos e das entradas;

- **Otimizador:** define qual o método de otimização, o que afeta o tempo de treinamento do algoritmo, sendo o gradiente descendente um dos representantes mais conhecidos.

Considerando todos os hiperparâmetros envolvidos e as variações para cada um deles e para cada uma das séries de categorias de produto estudadas, foram desenvolvidas 2052 alternativas possíveis de modelos de previsão. A Tabela 1.8 traz um quadro-resumo especificando as alternativas testadas por tipo de algoritmo.

Salienta-se que não foram testadas todas as possibilidades de variações de hiperparâmetros estabelecidos. Foi aplicada uma otimização bayesiana com base em um processo gaussiano, através do pacote Skopt existente para o Python. Esse tipo de otimização foi empregado, uma vez que o custo computacional de avaliação de todas as alternativas, inviabilizaria a realização deste estudo. A ideia por trás da otimização Bayesiana é aproximar a função usando um processo gaussiano, ou seja, assume-se que os valores da função seguem uma gaussiana multivariada. A covariância dos valores da função é fornecida por um *kernel* GP entre os parâmetros. Logo, a definição do próximo hiperparâmetro a ser avaliado pode ser feita pela função de aquisição em vez da priori gaussiana, que é muito mais rápido de ser avaliada.

Tabela 1.8 – Alternativas de modelos para cada algoritmo de previsão

Método	Configurações de Hiperparâmetros testados
RF	Avaliadas 30 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Max_depth: 3-20; • Max_features: auto, sqrt; • Min_sample_leaf: 1-4; • Min_samples_split: 2-10; • Criterion: mse; • n_estimators: 300.
XGBOOST	Avaliadas 30 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • learning_rate: 0.002-0.01; • max_depth: 8,15,24; • objective: reg:squarederror; • eval_metric: rmse.
LGBM	Avaliadas 30 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • learning_rate: 0.002-0.01; • max_depth: 8,15,24; • objective: regression; • eval_metric: rmse.
MLP	Avaliadas 30 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Número de camadas escondidas: 1,2; • Número de camadas de saída: 1; • Neurônios na camada escondida : 52,104,156 • Taxa de aprendizado: 0.002-0.01; • Dropout: 0.0001- 0.05; • Funções de ativação da camada escondida: relu,elu,sigmoid,selu; • Otimizadores: ADAM, RMSProp e Adagrad.
CNN	Avaliadas 30 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Número de camadas escondidas: 1,2; • Número de camadas de saída: 1; • Neurônios na camada escondida : 52,104,156 • Taxa de aprendizado: 0.002-0.01; • Funções de ativação da camada escondida: relu,elu,sigmoid,selu; • Otimizadores: ADAM, RMSProp e Adagrad.
LSTM	Avaliadas alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Número de camadas escondidas: 1,2; • Número de camadas de saída: 1; • Neurônios na camada escondida : 52,104,156 • Taxa de aprendizado: 0.002-0.01; • Dropout: 0.0001- 0.05; • Funções de ativação da camada escondida: relu,elu,sigmoid,selu; • Otimizadores: ADAM, RMSProp e Adagrad.

Fonte: elaboração própria.

Último ponto que merece ser salientado é que, além das alternativas de modelos geradas com base nos hiperparâmetros selecionados, após a definição da melhor configuração, obtida por uma estratégia bayesiana que minimizasse a métrica RMSLE, foi realizado um procedimento de validação, que consiste na replicação de

30 arquiteturas do modelo com a configuração que apresentou o menor RMSLE, com o intuito de conferir os respectivos resultados encontrados, que são médios.

Tal procedimento é fundamental quando se utiliza modelos afetados por aleatoriedade, como é o caso dos modelos baseados em árvore de decisão do tipo *boosting* e das redes neurais. Ele consiste na mudança do valor da semente que é responsável pela geração de números aleatórios. Para as RNs, a mudança da semente implica na inicialização distinta dos pesos iniciais, o que pode levar a rede neural a encontrar um ponto de convergência diferente. Já no caso dos modelos baseados em árvores de decisão do tipo *boosting*, o efeito da aleatoriedade é menor, mas ela acontece quando o GBDT faz uma amostra dos dados ou das variáveis para criar a próxima árvore, conforme Filho (2019). Esse efeito não se observa na *random forest*, pois o processo de criação de árvore é independente.

5.8. Pacotes computacionais

Todas as previsões foram realizadas com o Python. As bibliotecas utilizadas para que todo o trabalho fosse produzido, considerando os tratamentos e os métodos de previsão foram: (1) Sklearn para cálculo das métricas de precisão, tratamento de normalização dos dados, divisão do período de treino e teste e para o algoritmo Random Forest; (2) Keras para os métodos de redes neurais aplicados neste estudo; (3) Lighgbm para o algoritmo Lightgbm; (4) Xgboost para o método de previsão Xgboost; (5) Skopt para a otimização bayesiana aplicação para a seleção dos hiperparâmetros e (6) SHAP, que é a biblioteca completa SHapley Additive exPlanations, que utilizamos como um dos tratamentos empregados para seleção de variáveis.

5.9. Análise exploratória das séries temporais

Esta parte do trabalho traz a análise exploratória de cada uma das séries de produtos modeladas, as quais são variáveis quantitativas discretas. Além disso, também será realizada a mesma análise para as principais variáveis exógenas. O objetivo é realizar a descrição das séries através das suas medidas descritivas.

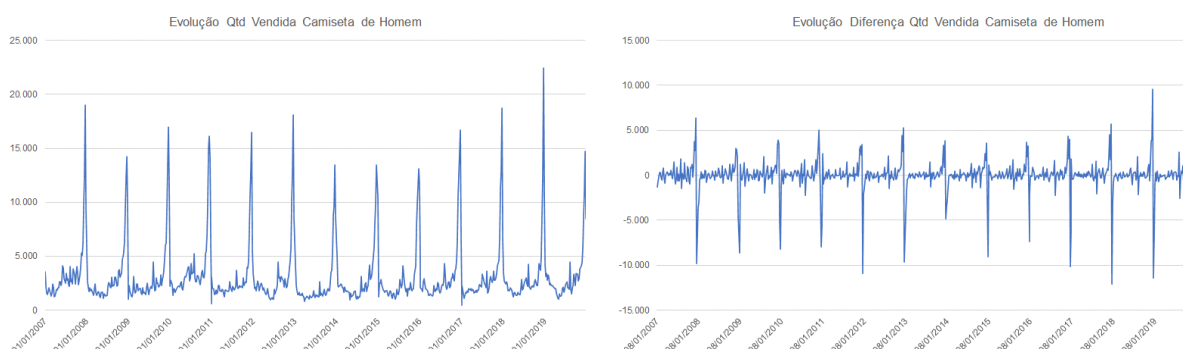
5.9.1. Análise exploratória das séries de quantidade vendida por categoria de produto

A seguir cada uma das séries em quantidade vendida em unidades das categorias de produto selecionadas serão analisadas.

Camiseta de Homem

Na Figura 1.25 pode-se observar a evolução das séries de quantidade vendida e da diferença da quantidade vendida de camiseta de homem ao longo do período de 2007 a 2019. É perceptível pela análise gráfica que a série de quantidade vendida não é estacionária, por isto procedeu-se com a sua diferenciação.

Figura 1.25 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série camiseta de homem



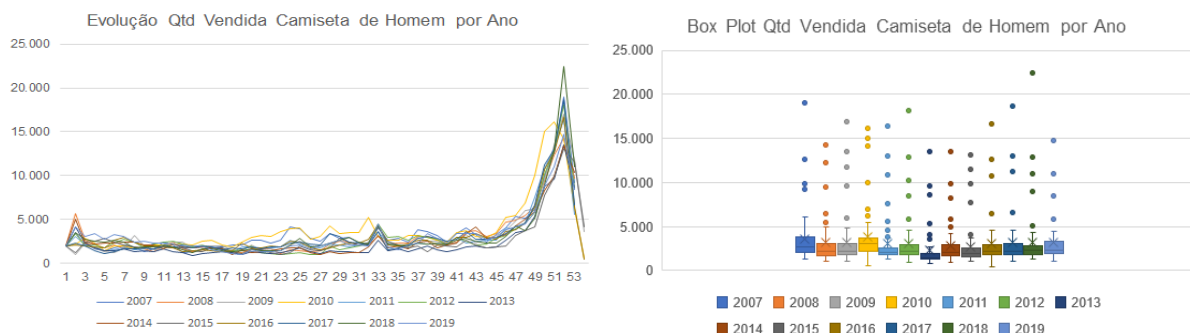
Fonte: elaboração própria.

Considerando a série em questão, verifica-se a presença de um forte padrão sazonal, com picos de vendas maiores nos meses de novembro e dezembro, devido a *Black Friday* e ao Natal, respectivamente, e em agosto, graças ao Dia dos Pais.

A partir da análise das medidas descritivas, nota-se que a distribuição dos dados é bastante assimétrica à esquerda, com a média superior à mediana. O padrão sazonal é tão intenso que as vendas nas últimas semanas do ano poderiam ser tratadas como dados anômalos, correspondendo a cerca de 11% das observações. Essa quantidade amostral com valores muito elevados acaba aumentando significativamente o valor da média em relação ao valor da mediana. Isso também ajuda a explicar a alta variância dos dados e o próprio coeficiente de variação.

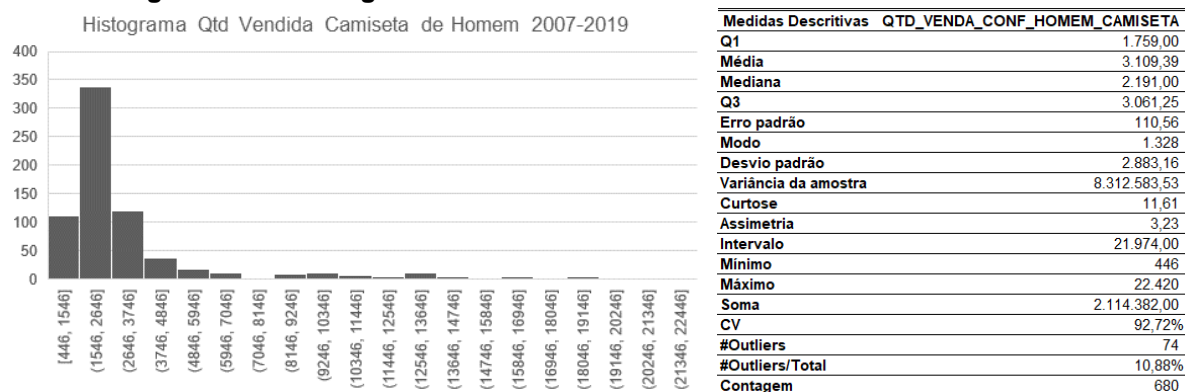
As Figuras 1.26 e 1.27 mostram os gráficos destacando a questão sazonal, o histograma e as medidas descritivas da série de camiseta de homem.

Figura 1.26 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série camiseta de homem



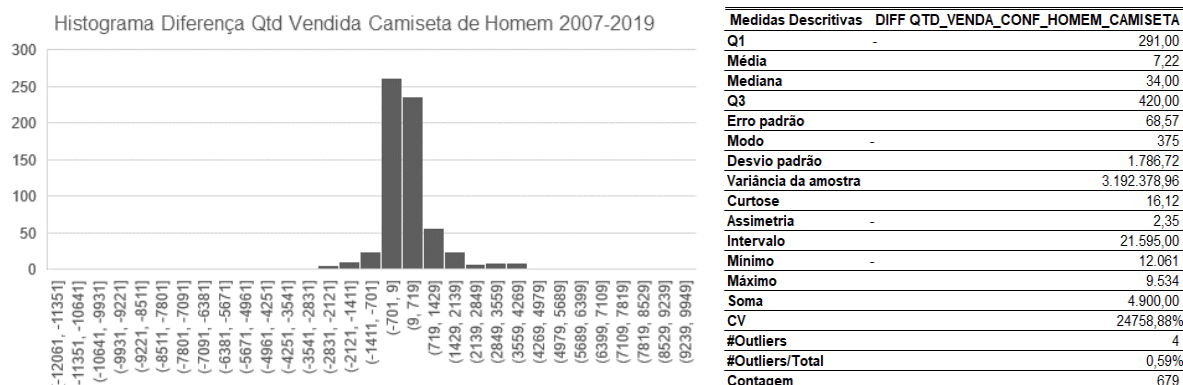
Fonte: elaboração própria.

Figura 1.27 – Histograma e medidas descritivas da série camiseta de homem



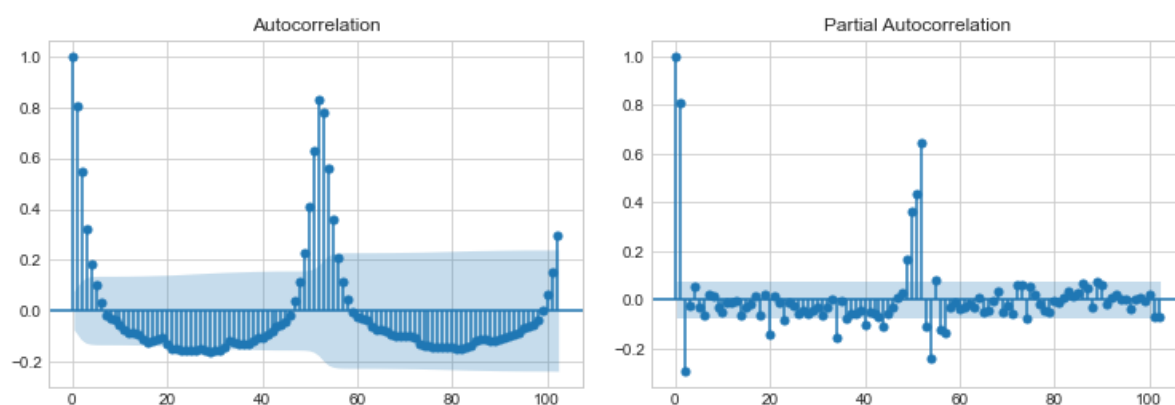
Fonte: elaboração própria.

Tendo em vista a questão da presença da não estacionariedade, na Figura 1.28 é evidenciada o histograma da série diferenciada, assim como as suas medidas descritivas. Neste sentido, verifica-se uma assimetria à direita, um coeficiente de variação superior quando comparado com a série original, mas com uma variância inferior. Ademais, a série de camiseta de homem apresentou o menor coeficiente de variação, considerando todas as séries de produtos diferenciadas.

Figura 1.28 – Histograma e medidas descritivas da série camiseta de homem em diferença

Fonte: elaboração própria.

Para os modelos Box Jenkins, é importante a análise da FAC e da FACP para definição dos termos autorregressivos e dos termos de médias móveis. Na Figura 1.29 pode-se verificar ambos os gráficos para a série de camiseta de homem. Assim, com base na FACP e na FAC, a sugestão seria a inclusão de 2 termos autorregressivos e 4 termos de médias móveis.

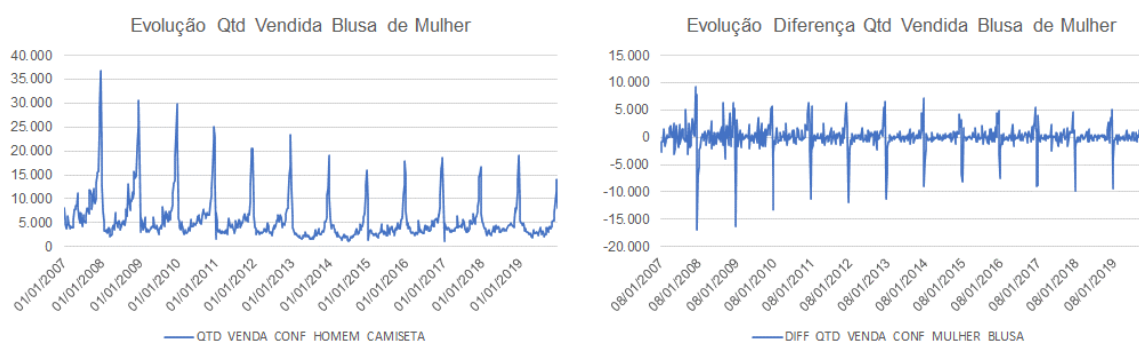
Figura 1.29 – FAC e FACP da série camiseta de homem

Fonte: elaboração própria.

Blusa de Mulher

Na Figura 1.30 é evidenciada a evolução das séries de quantidade vendida e da diferença da quantidade vendida de blusa de mulher para o intervalo de 2007 a 2019. Assim como foi destacado na série de camiseta de homem, a série de vendas de blusa de mulher também não é estacionária.

Figura 1.30 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série blusa de mulher



Fonte: elaboração própria.

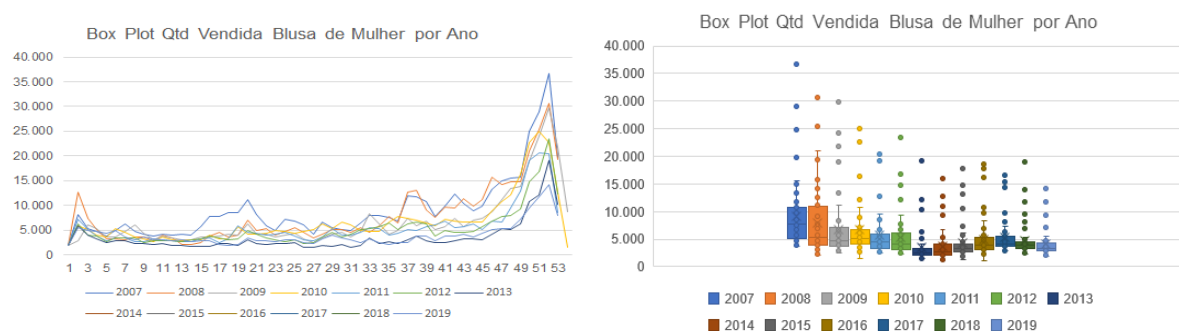
A série de mulher apresenta um comportamento um pouco diferente da série de camiseta de homem, já que a importância das vendas do final de ano, nos meses de novembro e dezembro, tem uma relevância menor. Ademais, o mês de maio, e não o mês de agosto, é um mês relevante para o seu faturamento.

Assim como ocorre na série de camiseta de homem, constata-se uma assimetria à esquerda, com a média sendo superior à mediana. Das 680 observações, cerca de 11% também poderiam ser classificados como dados anômalos, os quais ajudam a justificar o alto coeficiente de variação da série, que é de aproximadamente 81%.

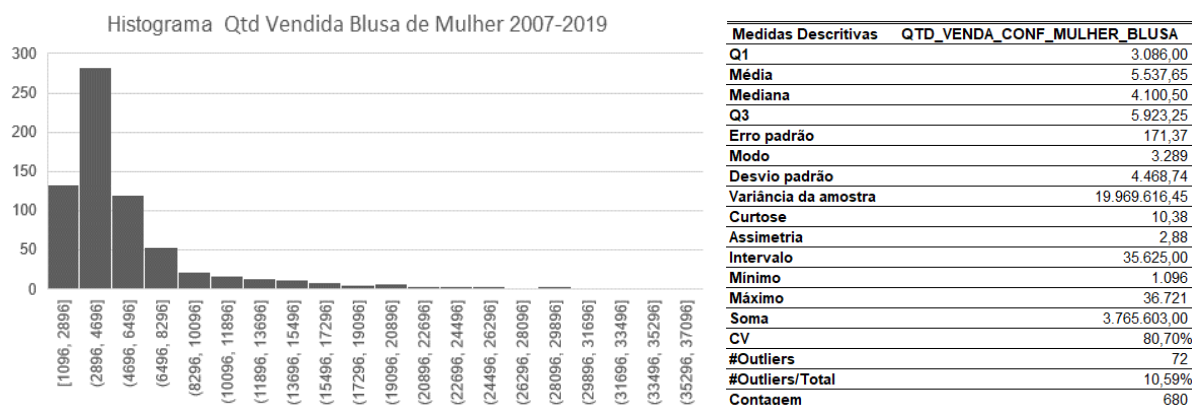
Interessante destacar que os valores dos dados aberrantes são muito superiores até 2012 perante os valores verificados nos demais anos.

As Figuras 1.31 e 1.32 mostram os gráficos destacando a questão sazonal, o histograma e as medidas descritivas da série de blusa de mulher.

Figura 1.31 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série blusa de mulher

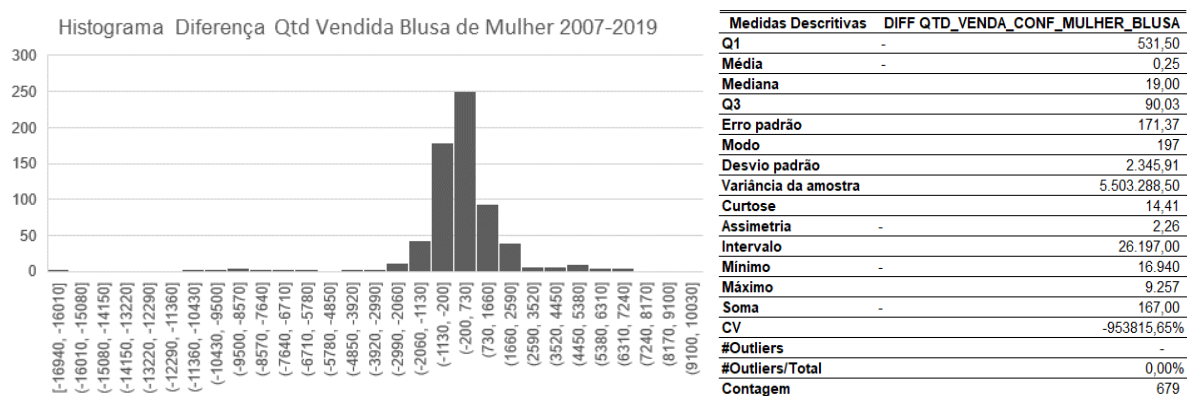


Fonte: elaboração própria.

Figura 1.32 – Histograma e medidas descritivas da série blusa de mulher

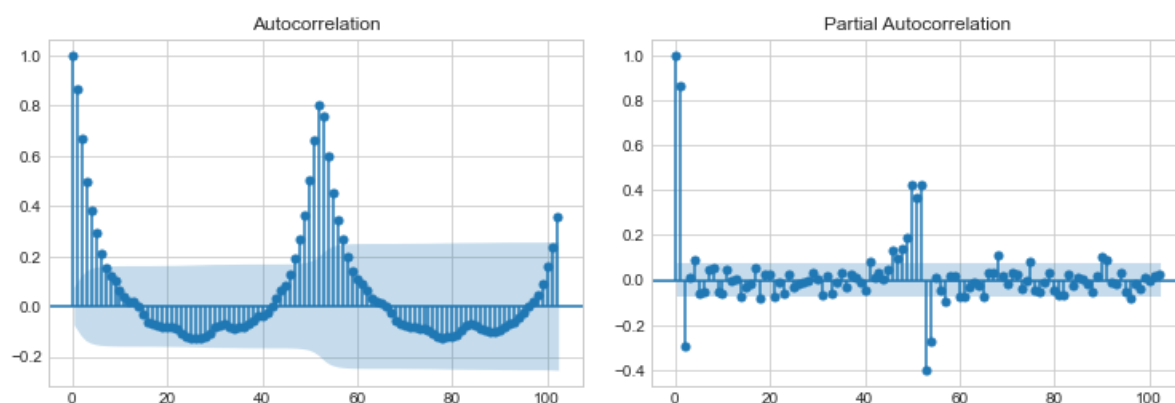
Fonte: elaboração própria.

Diante da presença da não estacionariedade da série de blusa de mulher, na Figura 1.33 é evidenciado o histograma da série diferenciada, bem como as suas medidas descritivas. Desta forma, constata-se uma assimetria à direita, um coeficiente de variação muito superior quando comparado com a série original e com o maior valor quando comparado com as demais categorias de produto.

Figura 1.33 – Histograma e medidas descritivas da série blusa de mulher em diferença

Fonte: elaboração própria.

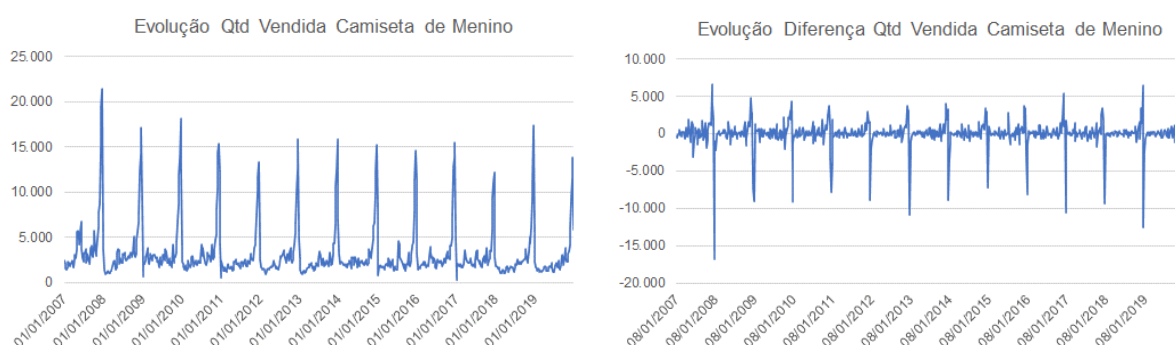
Para a construção dos modelos SARIMAs, é importante a análise da FAC e da FACP para definição dos termos autorregressivos e dos termos de médias móveis. Então, na Figura 1.34 é apresentado os dois gráficos, e com base na análise de ambos se recomenda a inclusão de 2 termos autorregressivos e 6 termos de médias móveis.

Figura 1.34 – FAC e FACP da série blusa de mulher

Fonte: elaboração própria.

Camiseta de Menino

A partir da Figura 1.35 pode-se verificar a evolução das séries de quantidade vendida e da diferença da quantidade vendida de camiseta de menino para o período de 2007 a 2019. Mais uma vez, constata-se um padrão não estacionário, como já foi observado nas demais séries analisadas até o momento.

Figura 1.35 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série camiseta de menino

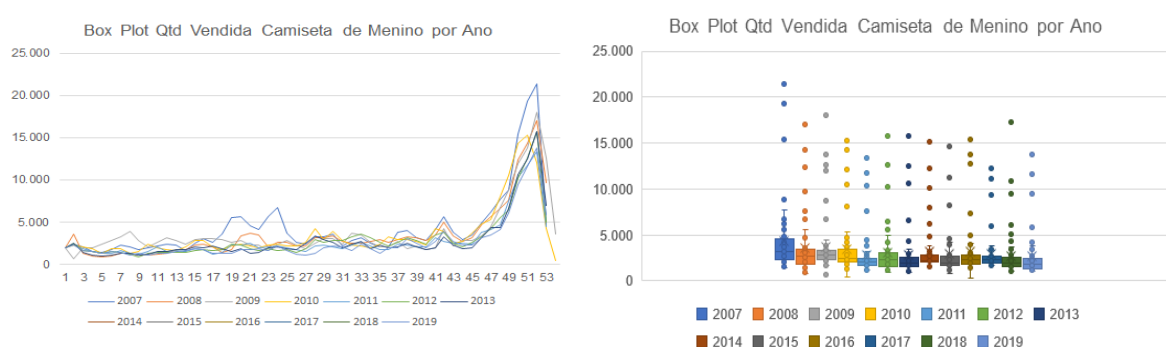
Fonte: elaboração própria.

Sobre a série de camiseta de menino, constata-se padrão similar ao da série de camiseta de homem, com maior peso para as semanas dos meses de novembro e dezembro. No entanto, aqui o pico de vendas é verificado no Dia das Crianças.

A distribuição assimétrica à esquerda também é vista nesse conjunto de dados, com o valor da média muito próximo ao valor do terceiro quartil e bem superior à mediana. Certamente, a presença de cerca de 10% de observações dados anômalos colaboram para justificar esse fato em relação às medidas quartílicas e também colaboram para explicar o alto valor de coeficiente de variação.

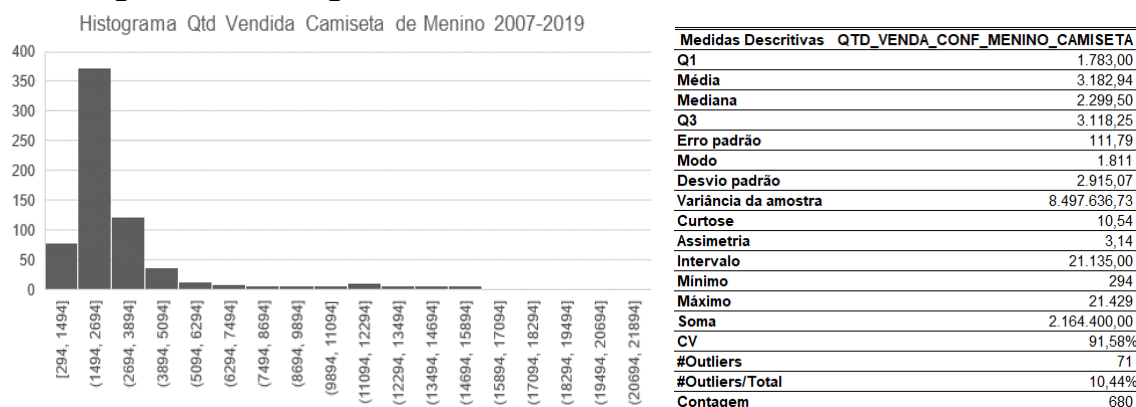
As Figuras 1.36 e 1.37 trazem os gráficos destacando a questão sazonal, o histograma e as medidas descritivas da série de blusa de mulher.

Figura 1.36 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série camiseta de menino



Fonte: elaboração própria.

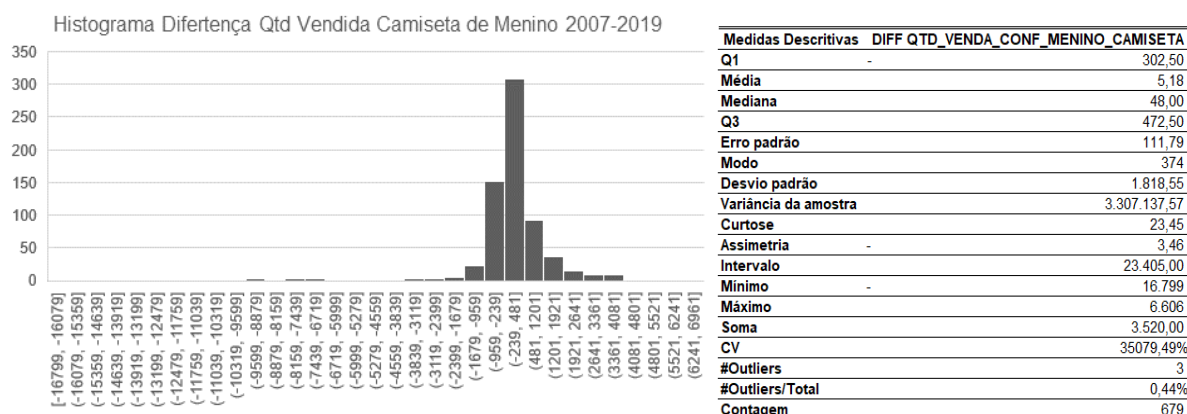
Figura 1.37 – Histograma e medidas descritivas da série camiseta de menino



Fonte: elaboração própria.

Considerando a presença de uma série não estacionária, é fundamental realizar a análise das medidas descritivas na série com a aplicação do tratamento da diferenciação para torná-la estacionária. Assim, na Figura 1.38 pode-se verificar o histograma da série diferenciada, como também as suas medidas descritivas. Neste sentido, é identificada uma assimetria à direita, um coeficiente de variação superior quando comparado com a série original, no entanto, com uma dispersão menor.

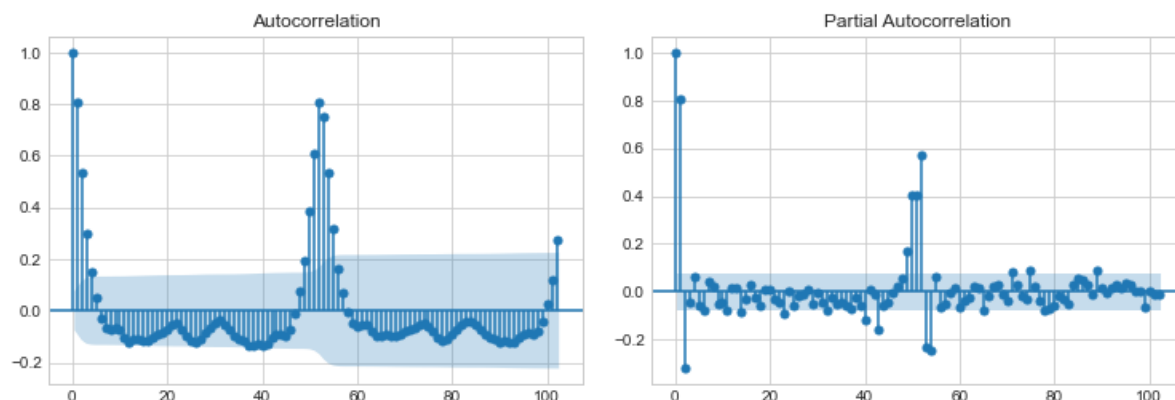
Figura 1.38 – Histograma e medidas descritivas da série camiseta de menino em diferença



Fonte: elaboração própria.

Baseada na análise da FAC e da FACP, cujos gráficos são observados na Figura 1.39, que é uma etapa importante para especificação dos modelos Box Jenkins, recomenda-se a inclusão de 2 termos autorregressivos e 4 termos de médias móveis, como ocorreu na série de camiseta de homem.

Figura 1.39 – FAC e FACP da série camiseta de menino

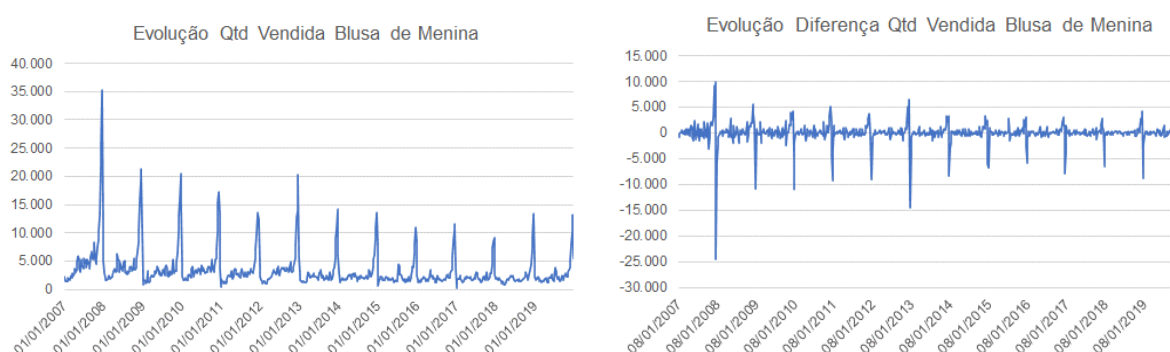


Fonte: elaboração própria.

Blusa de Menina

Na Figura 1.40 é possível observar a evolução das séries de quantidade vendida e da diferença da quantidade vendida de blusa de menina ao longo dos anos de 2007 a 2019. Tal como ocorreu para todas as séries de produtos analisadas, a série de blusa de menina também apresenta um padrão não estacionário.

Figura 1.40 – Gráficos da evolução da quantidade vendida e da quantidade vendida em diferença da série blusa de menina



Fonte: elaboração própria.

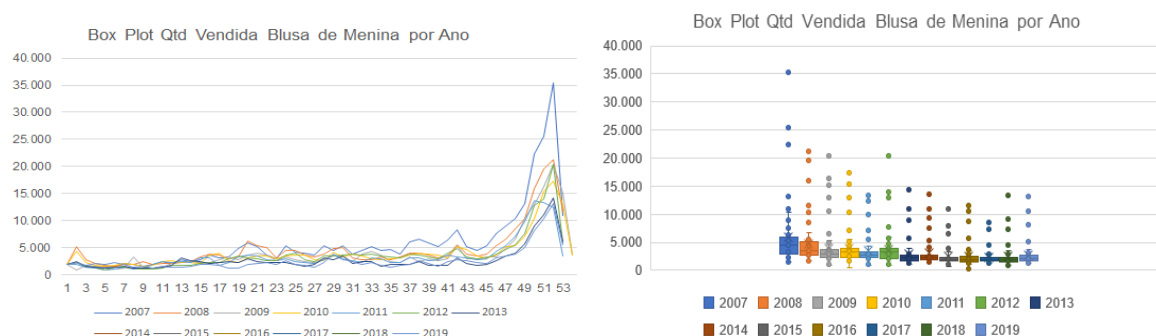
A série de blusa de menina, assim como as demais, tem seus picos de vendas nos meses de novembro e dezembro, devido, respectivamente, ao período da *Black Friday*, que se intensificou nos últimos quatro anos, e do Natal. No entanto, diferentemente do que ocorre com a série de camiseta de menino, foi observado um incremento inferior das vendas no Dia das Crianças.

A análise das medidas descritivas dessa série permite verificar a maior assimetria entre as séries contempladas neste estudo, sendo o valor da média extremamente próximo do valor do terceiro quartil.

A série possui a maior quantidade de dados que podem ser considerados como dados anômalos, correspondendo a 14% do total da contagem. Diante desse número de observações com valores muito superiores, é esperado um grande desvio padrão, assim como um alto coeficiente de variação, que também é o mais alto quando comparado com as demais categorias de produtos, sendo de 97%. Assim como observado na série de blusa de mulher, é perceptível que os dados anômalos apresentaram valores absolutos muito maiores até 2010, reduzindo de forma significativa posteriormente.

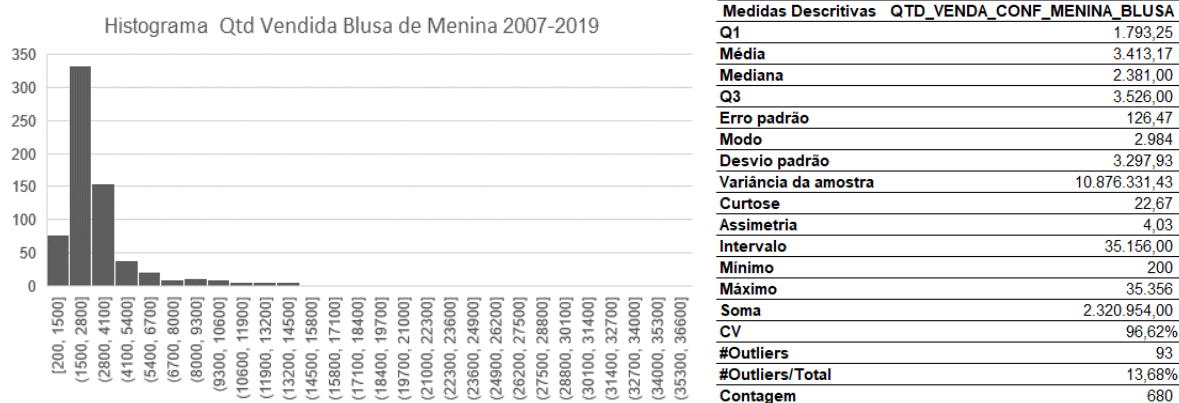
As Figuras 1.41 e 1.42 evidenciam os gráficos destacando a questão sazonal, o histograma e as medidas descritivas da série em foco.

Figura 1.41 – Gráfico da evolução da quantidade vendida e Boxplot por ano da série blusa de Menina



Fonte: elaboração própria.

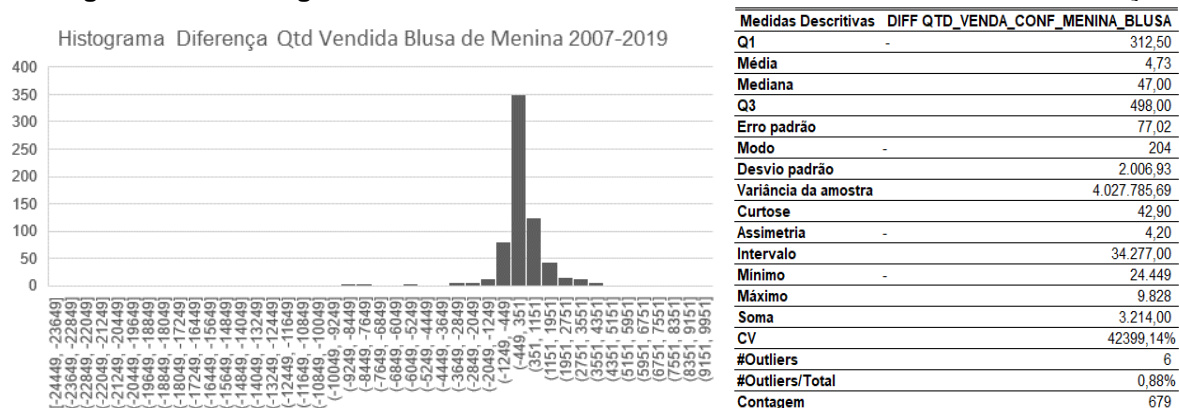
Figura 1.42 – Histograma e medidas descritivas da série blusa de Menina



Fonte: elaboração própria.

Perante a presença de um padrão não estacionário, é imprescindível realizar a diferenciação da série e depois rever a análise descritiva. Desta forma, na Figura 1.43 é apresentado o histograma da série diferenciada e as medidas descritivas do padrão estacionário. Logo, é identificada uma assimetria à direita, um coeficiente de variação bem superior quando comparado com a série original, mas com uma variância menor.

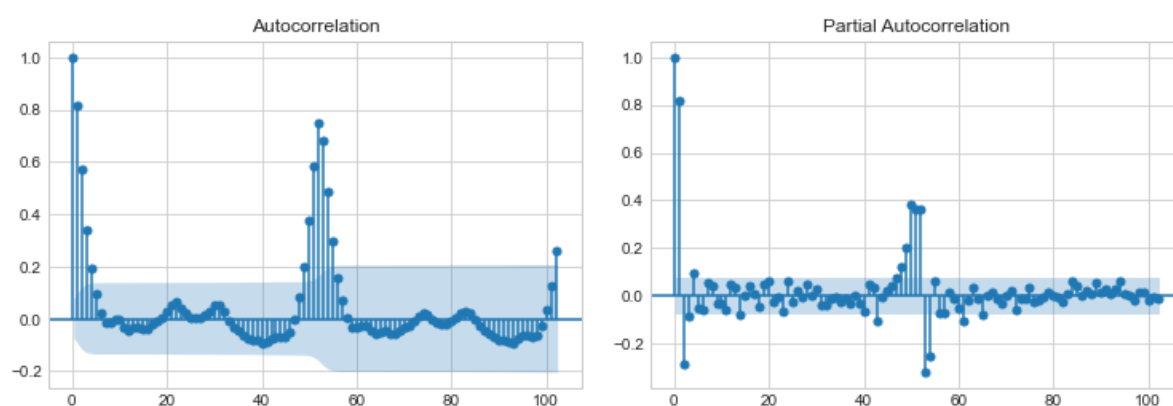
Figura 1.43 – Histograma e medidas descritivas da série blusa de menina em diferença



Fonte: elaboração própria.

Considerando a verificação da FAC e da FACP, cujos gráficos são observados na Figura 1.44, que é uma atividade relevante para especificação dos modelos SARIMAs, a sugestão é a inclusão de 2 termos autorregressivos e 4 termos de médias móveis, como ocorreu na série de camiseta de menino.

Figura 1.44 – FAC e FACP da série blusa de menina



Fonte: elaboração própria.

Com base na avaliação de todas as séries contempladas neste estudo, foi possível averiguar um forte padrão sazonal, sendo que os maiores picos de venda estão concentrados nas semanas dos meses de novembro e de dezembro, que é o período que contém tanto a *Black Friday* como o Natal.

Ademais, foi possível notar um volume médio de 11,4% de observações como dados anômalos, que nos modelos de aprendizagem de máquina geralmente são tratados de forma separada, sendo muitas vezes expurgados dos modelos de previsão ou recebendo uma abordagem específica para minimizar o seu impacto. Neste estudo não foi aplicado nenhum tratamento distinto para os dados aberrantes encontrados e um dos objetivos dos modelos gerados será apreender o comportamento da sazonalidade existente.

Conceitualmente, considerando a definição de dados anômalos, como àqueles que são inferiores à 1,5x ao valor do primeiro quartil ou 1,5x superior ao valor do terceiro quartil, teríamos 11,4% dos dados que poderiam ser consideradas outliers. No entanto, esta classificação não foi adotada, pois tais observações ocorrem sempre

no mesmo período do ano, especificamente no fim do ano, refletindo muito mais um padrão sazonal.

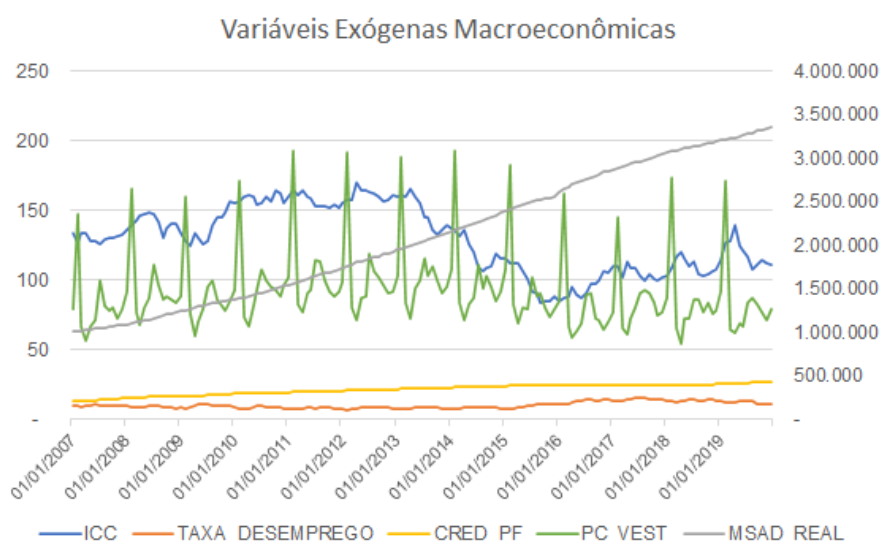
Em relação à análise dos gráficos FAC e FACP foi possível constatar que para as séries de quantidade vendida de camiseta de homem, camiseta de menino e blusa de menina foi feita a mesma recomendação de especificação de modelo com a indicação para a inclusão de 2 termos autorregressivos e 4 termos de médias móveis. No caso da série de blusa de mulher, foi também sugerido a consideração de 2 termos autorregressivos, porém com 6 médias móveis.

5.9.2. Análise exploratória principais variáveis exógenas

Nesta secção será realizada a análise exploratória das principais variáveis exploratórias deste estudo. Mais especificamente, será realizada a análise descritiva tanto das variáveis exógenas macroeconômicas como das variáveis exógenas climáticas.

Na Figura 1.45 é possível verificar a evolução de 2007 a 2019 das variáveis macroeconômicas: índice de confiança ao consumidor (ICC), taxa de desemprego da RMSP, massa salarial ampliada disponível real, crédito para pessoa física como participação do PIB e índice em volume da pesquisa mensal do comércio para categoria de vestuário em São Paulo. Dado o nível superior da massa salarial ampliada disponível real, esta variável foi colocada no eixo secundário.

A análise do gráfico da Figura 1.45 permite constatar um padrão sazonal do índice em volume da pesquisa mensal do comércio para categoria de vestuário em São Paulo série, que não é observado nas demais variáveis macroeconômicas. Além disso, é possível verificar claramente uma piora do quadro econômico entre 2015 e 2017, com a redução do ICC, queda do volume da PMC para a categoria de vestuário em SP e aumento do desemprego.

Figura 1.45 – Gráficos da evolução das variáveis exógenas macroeconômicas

Fonte: elaboração própria.

Na Tabela 1.9 pode-se observar as principais medidas descritivas para as séries das variáveis exógenas macroeconômicas. Neste sentido, verifica-se que as séries ICC e crédito disponível a PF como % PIB são as séries com os menores coeficientes de variação, indicando um patamar de variância menor. Por outro lado, as séries massa salarial ampliada disponível real e índice em volume da pesquisa mensal do comércio para categoria de vestuário em São Paulo apresentaram os maiores coeficientes de variação, evidenciando uma alta dispersão. Adicionalmente, a série da PMC segmentada para vestuário foi a série com maior assimetria e curtose.

Tabela 1.9 – Medidas descritivas das variáveis exógenas macroeconômicas

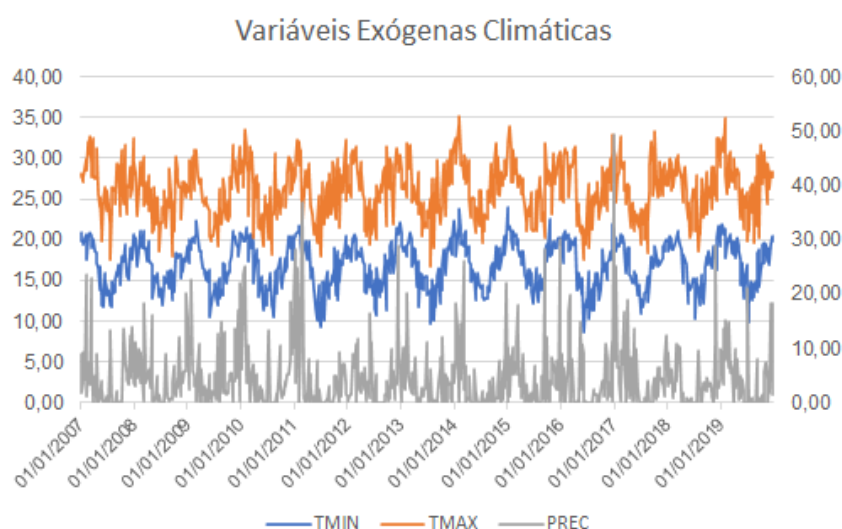
Medidas Descritivas	ICC	TAXA_DESEMPREGO	MSAD_REAL	CRED_PF	PC_VEST
Q1	109,53	8,80	1.416.636,70	18,85	77,08
Média	129,74	10,58	2.107.545,65	21,73	92,32
Mediana	130,59	9,70	2.040.209,60	22,68	86,45
Q3	153,20	12,43	2.804.702,03	25,03	96,45
Erro padrão	1,90	0,19	59.518,64	0,30	2,24
Modo	161	9		14	100
Desvio padrão	23,76	2,35	743.387,52	3,74	27,96
Variância da amostra	564,41	5,54	552.625.000.349,84	13,95	781,98
Curtose	- 1,17 -	0,84 -	1,34 -	0,95	4,51
Assimetria	- 0,12 -	0,67	0,14 -	0,53	2,10
Intervalo	85,63	9,00	2.359.390,40	13,34	138,30
Mínimo	85	7	1.010.515	14	55
Máximo	170	16	3.369.906	27	193
Soma	20.239,09	1.649,70	328.777.121,90	3.390,48	14.401,80
CV	18,31%	22,26%	35,27%	17,19%	30,29%
Contagem	156	156	156	156	156

Fonte: elaboração própria.

Em relação às medidas quartilícas, nota-se que a distribuição dos dados é bastante assimétrica à esquerda, com a média superior à mediana para as variáveis taxa de desemprego na RMSP e para índice em volume da PMC para vestuário em SP. Já no caso da série ICC e massa salarial ampliada disponível real constata-se uma assimetria à direita, com a média inferior à mediana.

Na Figura 1.46 foi evidenciada a evolução de 2007 a 2019 das variáveis climáticas: temperatura máxima, temperatura mínima e nível de precipitação. Verifica-se claramente um padrão sazonal em todas as variáveis climáticas, pois ocorrem temperaturas maiores nas semanas do verão e menores nas semanas do inverno e, em termos de nível de precipitação, os maiores volumes são verificados nas semanas de dezembro, janeiro e fevereiro, que correspondem exatamente ao período do verão.

Figura 1.46 – Gráficos da evolução das variáveis exógenas climáticas



Fonte: elaboração própria.

Na Tabela 1.10 foram sumarizadas as principais medidas descritivas para as séries climáticas. Assim, o maior coeficiente de variação ficou para a série de nível de precipitação, indicando uma dispersão maior. Já a série de temperatura mínima apresentou o menor coeficiente de variação, gerando uma dispersão inferior. Vale destacar também que a série de volume de precipitação apresentou a maior curtose e a maior assimetria.

Sobre a distribuição dos dados, pode-se observar que as duas séries de temperaturas são assimétricas à direita, já que suas médias são menores que suas

respectivas medianas. Em contrapartida, a série de volume de chuvas mostrou uma assimetria à esquerda.

Tabela 1.10 – Medidas descritivas das variáveis exógenas climáticas

Medidas Descritivas	TMIN	TMAX	PREC
Q1	14,64	23,78	0,11
Média	16,88	26,15	4,41
Mediana	17,16	26,32	2,44
Q3	19,25	28,62	6,35
Erro padrão	0,11	0,13	0,23
Desvio padrão	2,86	3,36	5,89
Variância da amostra	8,18	11,29	34,67
Curtose	- 0,66 -	0,46	8,76
Assimetria	- 0,31 -	0,13	2,44
Intervalo	15,23	18,40	49,20
Mínimo	9	17	-
Máximo	24	35	49
Soma	11.478,46	17.783,08	2.998,31
CV	16,94%	12,85%	133,54%
Contagem	680	680	680

Fonte: elaboração própria.

REFERÊNCIAS

ABBASIMEHR, Hossein; SHABANI, Mostafa; YOUSEFI, Mohsen. An optimized model using LSTM network for demand forecasting. **Computers & Industrial Engineering**, v. 143, 2020.

AKSOY, Asli; OZTURK, Nursel; SUCKY, Eric. A decision support system for demand forecasting in the clothing industry. **International Journal of Clothing Science and Technology**, v. 24, n. 4, p. 221-236, 2012.

ALMEIDA, Fernando Carvalho de; PASSARI, Antônio Fabrizio Lima. Previsão de vendas no varejo por meio de redes neurais. **Revista da Administração**, v. 41, n. 3, p. 257-272, 2006.

ALON, Ilan; QI, Min; SADOWSKI, Robert J. Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. **Journal of Retailing and Consumer Services**, v. 8, n. 3, p. 147-156, 2001.

ARMSTRONG, J. **Sales Forecasting**. Marketing Department – University of Pennsylvania. 16 p. 20 jul. 2008. Disponível em SSRN: <<http://ssrn.com/abstract=1164602>>. Acesso em: 10 nov. 2017.

AZEVEDO, Ana; SANTOS, M. F. KDD, SEMMA and CRISP-DM: A parallel overview. **MCCSIS'08** – IADIS Multi Conf. Comput. Sci. Inf. Syst. Proc. Informatics, n. jun., p. 182-185, 2008.

BATISTA, Valquiria Constancio. **Previsão de vendas**: um estudo no varejo de vestuário. Dissertação (Mestrado Profissional em Engenharia de Produção). Universidade Federal do Rio Grande do Sul, Porto Alegre. 74 p. 2017.

BERMAN, Berry; EVANS, Joel R. **Retail management**: a strategic approach. 7th ed. Upper Saddle River: Prentice Hall, 1998.

BEHESHTI-KASHI, Samaneh *et al.* A survey on retail sales forecasting and prediction in fashion markets. **Systems science & Control engineering: an open access journal**, v. 3, n. 1, p. 154-161, 2015.

BRAHMADEEP; THOMASSEY, Sébastien. Intelligent demand forecasting systems for fast fashion. In: CHOI, Tsan-Ming (ed.). **Information systems for the fashion and apparel industry**, Woodhead publishing series in textiles, p. 145-161, 2016.

BREIMAN, Leo. Bagging predictors. **Machine learning**, v. 24, p. 123-140, 1996. Disponível em: <<https://doi.org/10.1023/A:1018054314350>> Acesso em: 28 maio 2021.

BROWNLEE, Jason. **A gentle introduction to XGBoost for applied machine learning**. Machine Learning Mastery, 17 ago. 2016. Disponível em: <machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/> Acesso em: 2 jan. 2021.

BROWLEE, Jason. **How to identify and remove seasonality from time series data with python**. Machine Learning Mastery, 23 dez. 2016. Disponível em: <machinelearningmastery.com/time-series-seasonality-with-python/> Acesso em: 10 jan. 2021.

BROWLEE, Jason. When to Use MLP, CNN, and RNN Neural Networks. Machine Learning Mastery, 23 jul. 2018. Disponível em: <machinelearningmastery.com/when-to-use-mlp-cnn-and-rnn-neural-networks/> Acesso em: 10 mar. 2021.

BROWNLEE, Jason. **Data preparation for machine learning**. Data cleaning, feature selection and data transforms in python. Machine Learning Mastery: e-book, 2020a.

BROWLEE, Jason. **Hyperparameter optimization with random search and grid search**. Machine Learning Mastery, 14 set. 2020b. Disponível em: <machinelearningmastery.com/hyperparameter-optimization-with-random-search-and-grid-search/>. Acesso em: 2 jan. 2021.

BROWLEE, Jason. **Why use ensemble learning?** Machine Learning Mastery, 26 out. 2020c. Disponível em: <machinelearningmastery.com/why-use-ensemble-learning/>. Acesso em: 14 fev. 2021.

BUENO, R. L. S. **Econometria de séries temporais**. 2. ed. São Paulo: Cengage Learning, 2008.

CATAL, Cagatay *et al.* Benchmarking of regression algorithms and time series analysis techniques for sales forecasting. **Balkan Journal of Electrical and Computer Engineering**. v. 7, p. 20-26, 2019.

CELIA, Frank *et al.* **Forecasting women's apparel sales using mathematical modeling**. International Journal of Clothing Science and Technology, v. 15, n. 2, p. 107-125, abr. 2003.

CHAPMAN, Pete *et al.* **CRISP-DM 1.0 – Step-by-step data mining guide**. SPSS [1996?]. Disponível em: <the-modeling-agency.com/crisp-dm.pdf> Acesso em: 27 abr. 2020.

CHEN, I-Fei; LU, Chi-Jie. Sales forecasting by combining clustering and machine-learning techniques for computer retailing. **Neural Computing and Applications**, v. 28, p. 2633-2647, 2017.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: a scalable tree boosting system. **KDD'16: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**, p. 785-794, ago. 2016.

CHOI, Tsan-Ming *et al.* Fast fashion sales forecasting with limited data and time. **Decision Support Systems**, v. 59, n. 1, 2013.

CHOI, Tsan-Ming; HUI, Chi-Leung; YU, Young. **Intelligent fashion forecasting systems: models and applications**. Nova York: Springer, 2014.

CHOI, Tsan-Ming; REN, Shuyun. Intelligent demand forecasting supported risk management systems for fast fashion inventory management. **Information systems for the fashion and apparel industry**, p. 263-271, 2016.

CHU, Ching-Wu; ZHANG, Guoqiang Peter. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. **International Journal of Production Economics**, v. 86, n. 3, p. 217-231, 2003.

CHUNG, Junyoung *et al.* Empirical evaluation of gated recurrent neural networks on sequence modeling. **NIPS 2014 Deep Learning and Representation Learning Workshop**, dez. 2014.

COSTA, Helder Gomes; MOLL, Roberto Nunes. Emprego do método de análise hierárquica (AHP) na seleção de variedades para o plantio de cana-de-açúcar. **Gestão e Produção**, v. 6, n. 3, p. 243-256, dez. 1999. Disponível em: <scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X1999000300009&lng=en&nrm=iso> Acesso em: 23 jan. 2021.

DESARDA, Akash. **Getting data ready for modelling**: feature engineering, feature selection, dimension reduction (part two). Towards Data Science, 13 dez. 2018. Disponível em: <towardsdatascience.com/getting-data-ready-for-modelling-feature-engineering-feature-selection-dimension-reduction-39dfa267b95a> Acesso em: 31 ago. 2020.

DESHPANDE, Adit. **A beginner's guide to understanding convolutional neural networks**. Engineering at Forward, 2019. Disponível em: <adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/> Acesso em: 3 jan. 2021.

DEVAY, Andre. **Modelos de Predição | Regressão de Ridge e Lasso**. Medium, 29 set. 2019. Disponível em: <<https://medium.com/turing-talks/turing-talks-20-regress%C3%A3o-de-ridge-e-lasso-a0fc467b5629>> Acesso em: 4 jan. 2020.

DIAS, George Paulus Pereira. **Proposta de processo de previsão de vendas para bens de consumo**. Anais do XIX ENEGEP, CD-ROM, Rio de Janeiro, 1999.

DIETTERICH, Thomas G. Ensemble Methods in Machine Learning. **Lecture Notes in Computer Science book series**, v. 1857, p. 1-15, 2000.

DORFFNER, Georg. **Neural networks for time series processing neural network world**. Computer Science, 1996. Disponível em: <<http://img.clubic.com/06675738-photo-machine-learning.jpg>>. Acesso em: 10 fev. 2018.

DU, Wei; LEUNG, Sunney; KWONG, C. K. A multiobjective optimization-based neural network model for short-term replenishment forecasting in fashion industry. **Neurocomputing**. v. 151, p. 342-353, mar. 2015.

EBIT. **Webshoppers**. 42. ed. 2020. Disponível em: <company.ebit.com.br/webshoppers/webshoppersfree> Acesso em: 23 jan. 2021.

ENDERS, Walter. **Applied econometric time series**. 3. ed. Nova York: Wiley, 2009.

FAVA, Vera Lúcia. Análise de séries de tempo. In: VASCONCELLOS, Marco Antonio Sandoval; ALVES, Denisard (coord.). **Manual de Econometria**. Cap. 9. São Paulo: Atlas, 2000.

FAYYAD U., PIATETSKY-SHAPIRO G., SMYTH P. **From data mining to knowledge discovery in databases**. AI Mag., vol. 17, no. 3, pp. 37–53, 1996.

FELISONI DE ANGELO, Claudio *et al.* Séries temporais e redes neurais: uma análise comparativa de técnicas na previsão de vendas do varejo brasileiro. **Brazilian Business Review** (portuguese edition), v. 8, n. 2, p.1-21, 2011.

FILHO, Mario. **Guia de machine learning para time series**. 2019a.

FILHO, Mario. **O manual prático de data science**. E-book, 2019b.

FORONI, Claudia; MASSIMILIANO, Marcellino. A survey of econometric methods for mixed frequency data. **Economics Working Papers ECO 2013/02**, Department of Economics, European University Institute, 48 p. 2013.

FORONI, Claudia; GUÉRIN, Pierre; MASSIMILIANO, Marcellino. Using low frequency information for predicting high frequency variables. **International Journal of Forecasting**, v. 34, p. 774-787, 2018.

FREUND, Yoav; SCHAPIRE, Robert E. Decision-theoretic generalization of on-line learning and an application to boosting. **Computer and System Sciences**, v. 55, n. 1, p. 119-139, 1997.

GARCIA, Eduardo *et al.* **Gestão de Estoques**: otimizando a logística e a cadeia de suprimentos. Rio de Janeiro: E-Papers Serviços Editoriais, 2006.

GERS, Felix A.; SCHMIDHUBER, Jürgen; CUMMINS, Fred. Learning to Forget: Continual Prediction with LSTM. **Neural Computation**, v. 12, n. 10, p. 2451-2471, F Italiana Reti Neuroniche (SIREN), 12th Italian workshop, Salerno, Itália, 2001.

GIBBONS, Jean Dickinson; CHAKRABORTI, Subhabrata. **Nonparametric statistical inference**. 4. ed. Nova York: Marcel Decker, 2003.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 4. ed. São Paulo: Atlas, 1994.

GHYSELS, Eric; SANTA-CLARA, Pedro; VALKANOV, Rossen. **The MIDAS touch**: mixed data sampling regression models. Mimeo, Chapel Hill, NC. 2004.

GHYSELS, Eric; WRIGHT, Jonathan H. Forecasting professional forecasters. **Finance and Economics Discussion Series 2006-10**, Board of Governors of the Federal Reserve System (U.S.), Washington, DC, 2006.

GOMES, Pedro A. B. *et al.* Avaliação de diferentes arquiteturas de redes neurais recorrentes na previsão do nível de atividade de abelhas. **XXXVII Congresso da Sociedade Brasileira de Computação**, São Paulo, 2017.

GREFF, Klaus *et al.* LSTM: A search space odyssey. **IEEE transactions on neural networks and learning systems**, v. 28, 2015.

GUJARATI, Damodar N. **Econometria Básica**. 4. ed. Rio de Janeiro: Editora Campus, 2006.

GUYON, Isabelle; ELISSEEFF, André. An introduction to variable and feature selection. **Jornal of Machine Learning Research**, v. 3, p. 1157-1182, 2003.

HAYKIN, Simon S. Redes neurais: princípios e práticas. Porto Alegre: Bookman, 2001.

HANSEN, Peter R.; LUNDE, Asger; NASON, James M. The model confidence set. **Econometrica**, v. 79, n. 2, p. 453-497, 2011.

HINES, Tony; BRUCE, Margaret. **Fashion marketing**: contemporary issues. 2. ed. Oxford: Butterworth-Heinemann, 2007.

HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735-1780, 1997.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting**: principles and practise. 1. ed. Melbourne: OTexts, 2013. Disponível em: <<https://otexts.com/fpp2/>> Acesso em: 25 jan. 2021.

HYNDMAN, Rob; KOEHLER, Anne. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, p. 679-688, 2006.

IANNONE, Raffaele *et al.* Merchandise and replenishment planning optimisation for fashion retail. **International Journal of Engineering Business Management**, v. 5, p. 1-14, jan. 2013.

IANNONE, Raffaele *et al.* Modeling fashion retail supply chain through causal loop diagram. **IFAC-PapersOnLine**, v. 48, n. 3, p. 1290-1295, dez. 2015.

IEMI. **O público consumidor de vestuário**. 7 fev. 2018. Disponível em:<iemi.com.br/o-publico-consumidor-de-vestuario/> Acesso em: 28 dez. 2020.

IPAR, Maria da Conceição Silva; ALMEIDA, Luciene Nascimento de. Gestão do varejo de moda feminina: um estudo dos fatores críticos observados em lojas da cidade do Rio de Janeiro. **Revista de Design Inovação e Gestão Estratégica – REDIGE**, v. 1, n. 1, 2010.

Jl, Shouwen *et al.* An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise. **Mathematical Problems in Engineering**, v. 2019, 15 p. 2019.

JIANG, Jingwen; KELLY, Bryan T.; XIU, Dacheng. (Re-)imag(in)ing price trends. **Chicago Booth Research Paper**, n. 21, v. 1, 66 p. 4 jan. 2021. Disponível em: <ssrn.com/abstract=3756587> Acesso em: 10 abr. 2021.

KE, Guolin *et al.* **LightGBM: A Highly Efficient Gradient Boosting Decision Tree**. 31st Conference on Neural Information Processing Systems 30, NIPS 2017, Long Beach, Califórnia, 2017.

KIM, Hyun-jung; SHIN, Kyung-shik. A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. **Applied Soft Computing**, v. 7, n. 2, p. 569-576, 2007.

KRAJEWSKI, Lee; RITZMAN, Larry; MALHOTRA, Manoj. **Administração de produção e operações**. 8. ed. São Paulo: Pearson Prentice Hall, 2009.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. Nova York: Springer, 2013.

LÄNGKVIST, Martin; KARLSSON, Lars; LOUTFI, Amy. A Review of Unsupervised Learning and Deep Learning for Time-Series Modeling. **Pattern Recognition Letters**, v. 42, p. 11-24, 2014.

LEA, Colin *et al.* **Temporal Convolutional Networks: A Unified Approach to Action Segmentation**. Cornell University, Ithaca, 29 ago. 2016. Disponível em: <arxiv.org/abs/1608.08242> Acesso em: 10 abr. 2021.

LEE, Hau L.; PADMANABHAN, V.; WHANG, Seungjin. Information distortion in a supply chain: the bullwhip effect. **Management Science**, v. 43, n. 4, p. 546-558, 1997.

LINDBERG, Erik; ZACKRISSON, Uno. Deciding about the uncertain: the use of forecasts as an aid to decision-making. **Scandinavian Journal of Management**, v. 7, n. 4, p. 271-283. 1991.

LITTLE, J.D.C.. Integrated measures of sales, merchandising and distribution. **International Journal of Research in Marketing**, v.15, 473–485.1998

LIU, Peipei; MING, Wei; HU, Bin. Sales forecasting in rapid market changes using a minimum description length neural network. **Neural Computing and Applications**, v. 33, p. 937-948, 2020.

LIU, Na *et al.* Sales forecasting for fashion retailing service industry: a review. **Mathematical Problems in Engineering**, v. 2013, p. 1-9. 2013.

LOUREIRO, A.L.D.; MIGUÉIS, V.L.; SILVA, Lucas F.M. Exploring the use of deep neural networks for sales forecasting in fashion retail. **Decision Support Systems**, v. 114, p. 81-93, out. 2018.

LUCE, Leanne. Deep Learning and Demand Forecasting. In: **Artificial Intelligence for Fashion**. Berkeley: Apress. p. 155-166, 2018.

MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven C.; HYNDMAN, Rob J. **Forecasting: methods and applications**. 3. ed. Nova York: John Wiley, 1998.

MANDOT, Pushkar. **What is LightGBM, how to implement it? How to fine tune the parameters?** Medium, 17 ago. 2017. Disponível em: <medium.com/@pushkarmandot/https-medium-com-pushkarmandot-what-is-lightgbm-how-to-implement-it-how-to-fine-tune-the-parameters-60347819b7fc> Acesso em: 3 jan. 2021.

MARCANO-CEDENO, Alexis *et al.* Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. **IECON 2010** – 36th Annual Conference on IEEE Industrial Electronics Society, p. 2845-2850, 2010.

MARQUARDT, Donald W.; SNEE, Rondald D. Ridge regression in practice. **The American Statistician**, v. 29, n. 1, p. 3-20, 1975.

MASINI, Ricardo; MEDEIROS, Marcelo; MENDES, Eduardo. **Machine Learning Advances for Time Series Forecasting**. Cornell University, Ithaca, 23 dez. 2020. Disponível em: <arxiv.org/abs/2012.12802> Acesso em: 15 mar. 2021.

MEHRJOO, Marzieh. **Supply chain risk assessment for perishable products applying system dynamics methodology: a case of fast fashion apparel industry**. Tese (Doutorado), University of Windsor, Windsor. 161 p. 2014.

MENEZES JUNIOR, José Maria Pires de. **Redes Neurais dinâmicas para predição e modelagem não-linear de séries temporais**. Dissertação (Mestrado em Engenharia de Teleinformática). Centro de Tecnologia, Universidade Federal do Ceará, Fortaleza. 136 p. 2006.

MERCADO E CONSUMO. **Varejistas de moda estão otimistas com o fim de ano**. 17 dez. 2019. Disponível em: <mercadoeconsumo.com.br/2019/12/17/varejistas-de-moda-estao-otimistas-com-o-fim-de-an> Acesso em: 28 dez. 2020.

MORAES, Gustavo Inácio de; SILVA, Thiago Leal Flores da. Determinantes macroeconômicos das vendas no varejo no período 2004-2011. **Revista de Economia Contemporânea**, v. 19, p. 30-48, 2015.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blücher, 2006.

NELSON, Michael *et al.* Time series forecasting using neural networks: should the data be deseasonalized first? **Journal of forecasting**, v. 18, n. 5, p. 359-367, 1999.

NENNI, Maria; GIUSTINIANO, Luca; PIROLO, Luca. Demand forecasting in the fashion industry: a review. **International Journal of Engineering Business Management**, v. 5, jan. 2013.

NI, Yanrong; FAN, Feiya. A two-stage dynamic sales forecasting model for the fashion retail. **Expert Systems with Applications**, v. 38, n.3, p. 1529-1536, mar. 2011.

NIELSEN, Aileen. **Practical time series analysis: prediction with statistics and machine learning**. Sebastopol: O'Reilly Media, 2019.

NUNNARI, Giuseppe; NUNNARI, Valeria. Forecasting monthly sales retail time series: a case study. **IEEE**, 19th Conference on Business Informatics (CBI), Thessaloniki, p. 1-6, 2017.

PARENTE, Juracy. **Varejo no Brasil**. São Paulo: Atlas, 2000.

PAVLYSHENKO, Bohdan. Machine-learning models for sales time series forecasting. **Data**, v. 4, n. 15, 2019.

PASQUOTTO, Jorge Luís Durgante. **Previsão de redes temporais no varejo brasileiro**: uma investigação comparativa da aplicação de redes neurais recorrentes de Elman. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo. 200 p. 2010.

PASSARI, Antonio. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais**. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo. 150 p. 2003.

PELLEGRINI, F.; FOGLIATTO, F. Estudo Comparativo entre os Modelos de Winters e de Box-Jenkins para Previsão de Demanda Sazonal. **Produto & Produção**, v. 4, n. especial, p. 72-85. 2000.

REN, Shuyun; CHAN, Hau, Ling; RAM, Pratibha. A comparative study on fashion demand forecasting models with multiple sources of uncertainty. **Annals of Operations Research**, v. 257, n. 1, p. 335-355, out. 2017.

REN, Shuyun; HUI, Patrick; CHOI, Tsun-ming. AI-based fashion sales forecasting methods in big data era. In: THOMASSEY S., ZENG X. (eds). **Artificial Intelligence for Fashion Industry in the Big Data Era**. Springer Series in Fashion Business. Singapura: Springer. p. 9-26, 2018.

REN, Shuyun; CHAN, Hau Ling; SIQIN, Tana. Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. **Annals of Operations Research**, v. 291, n. 1 p. 761-777, 2020

RENNER. **Apresentação Renner Day 2018**. 2018. Disponível em: <mz-filemanager.s3.amazonaws.com/13154776-9416-4fce-8c46-3e54d45b03a3/file_manager/85f89d55-4279-4d9f-bee5-9431034d158c/apresentacao_renner_day_2018.pdf>. Acesso em: 20 set. 2020.

REN, Shuyun; CHAN, Hau Ling; RAM, Pratibha. A Comparative Study on Fashion Demand Forecasting Models with Multiple Sources of Uncertainty. **Annals of Operations Research**, v. 257, n. 1, p. 335-355, out. 2017.

RINK, David R.; SWAN, John E. Product life cycle research: a literature review. **Journal of Business Research**, v. 7, n. 3, p. 219-242, 1979.

SAATY, R. W. The analytic hierarchy process – What it is and how it is used. **Mathematical Modelling**, v. 9, n. 3-5, p. 161-176, 1987.

SAATY, Thomas L. **Método de análise hierárquica**. Tradução e revisão por Wainer da Silveira e Silva. São Paulo: McGraw-Hill, 1991.

SAATY, Thomas L. How to make a decision: the analytic hierarchy process. **Interfaces**, v. 24, p. 19-43, 1994. Disponível em: <<https://doi.org/10.1287/inte.24.6.19>> Acesso em: 12 jan. 2021.

SANTANA, Felipe. **Árvores de decisão (projeto passo a passo)**. Minerando Dados, 17 jan. 2020. Disponível em: <minerandodados.com.br/arvores-de-decisao-conceitos-e-aplicacoes> Acesso em: 2 jan. 2021.

SARADHI, Rashmi Priyanka; NARESH, Nelaturi. An advanced sales forecasting system using XGBoost algorithm. **International Journal of Electrical Electronics Computers & Mechanical Engineering**, v. 8, n. 7, set. 2018.

SAYYAD, Rehan Ahmed. **How to use convolutional neural networks for time series classification**. Medium, 13 jun. 2020. Disponível em: <medium.com/@Rehan_Sayyad/how-to-use-convolutional-neural-networks-for-time-series-classification-80575131a474> Acesso em: 3 jan. 2021.

SBVC. **O papel do varejo na economia**. 2020a. Disponível em: <http://www.sbvc.com.br/wpcontent/uploads/2015/04/O-Papel-do-Varejo-na-economia_apresenta%C3%A7%C3%A3ooficial1.pdf>. Acesso em: 10 jan. 2020.

SBVC. Varejo passa por transformação inédita: 5 anos em 5 meses. **SBVC na mídia**, São Paulo, 30 abr. 2020b. Disponível em: <sbvc.com.br/varejo-transformacao-inedita-2020/?utm_source=SBVC&utm_campaign=242fa472f3-SBVC+20200501&utm_medium=email&utm_term=0_b9f2f0b5b0-242fa472f3-101659889> Acesso em: 2 jan. 2020.

SEN, Alper. The US fashion industry: a supply chain review. **International Journal of Production Economics**, v. 114, n. 2, p. 571-593, 2008.

SMALL, Michael; TSE, Chi. Minimum description length neural networks for time series prediction. **Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics**, v. 66, n. 6, 2002.

SZOZDA, Natalia. Analogous forecasting of products with a short life cycle. **Decision Making in Manufacturing and Services**, v. 4, n. 1-2, p. 71-85, 2010.

SWALIM. A. **CatBoost vs. LightGBM vs. XGBoost**. Towards Data Science, 13 mar. 2018. Disponível em: <<https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db>> Acesso em: 2 jan. 2021.

THOMASSEY, Sebastien; HAPPIETTE, Michel; CASTELAIN, Jean Marie. An automatic textile sales forecast using fuzzy treatment of explanatory variables. **Journal of Textile and Apparel, Technology and Management**, v. 2, n. IV, p. 1-15, 2002.

THOMASSEY, Sebastien; HAPPIETTE, Michel; CASTELAIN, Jean Marie. A short and mean-term automatic forecasting system – application to textile logistics. **European Journal of Operational Research**, v. 161, n. 1, p. 275-285, 10 p., 16 fev. 2005.

THOMASSEY, Sébastien. Sales forecasts in clothing industry: the key success factor of the supply chain management. **International Journal of Production Economics**, v. 128, n. 2, p. 470-483, 2010.

THOMASSEY, Sébastien. Sales forecasting in apparel and fashion industry: a review. In: CHOI, Tsan-Ming; HUI, Chi-Leung; YU, Young (ed.). **Intelligent fashion forecasting systems: models and applications**. Cap. 2, p. 9-27. Nova York: Springer, 2014.

TIBSHIRANI, Robert. Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society, Series B (Methodological)**, vol. 58, n. 1, p. 267-288, 1996. Disponível em: <jstor.org/stable/2346178> Acesso em: 20 set. 2020.

VANDEPUT, Nicolas. **Forecast KPI: RMSE, MAE, MAPE & Bias**. Towards Data Science, 5 jul. 2019. Disponível em: <towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d> Acesso em: 4 set. de 2020.

VELLASCO, Marley Maria Bernardes Rebuzzi. **Redes neurais artificiais**. ICA – Laboratório de Inteligência Computacional Aplicada, PUC-RJ, 2007. Disponível em: <docplayer.com.br/44905591-Redes-neurais-artificiais-marley-maria-bernardes-rebuzzi-vellasco.html> Acesso em: 1 ago. 2020.

WERNER, Liane. **Um modelo composto para realizar previsão de demanda através da integração de previsões e do ajuste baseado na opinião**. Tese (Doutorado), Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, Porto Alegre. 166 p. 2005.

WHITE, H. Reality check for data snooping. **Econometrica**, v. 68, n. 5, p. 1097-1126, set. 2000.

WINKLHOFER, Heidi; DIAMANTOPOULOS, Adamantios; WITT, Stephen F. Forecasting practice: a review of the empirical literature and an agenda for future research. **International Journal of Forecasting**, v. 12, p. 193-221, 1996.

WONG, Wai Keung; GUO, Zheng Xiao. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. **International Journal of Production Economics**, v. 128, n. 2, p. 614-624, 2010.

XIA, Min *et al.* Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. **Knowledge-Based Systems**, v. 36, p. 253-259, dez. 2012.

ZHANG, Peter G.; PATUWO, Eddy; HU, Michael Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v. 14, p. 35-62, 1998.

ZHENG, Alice. **Evaluating machine learning models**: a beginner's guide to key concepts and pitfalls. Sebastopol: O'Reilly Media, 2015.

ZHENG, Alice; CASARI, Amanda. **Feature engineering for machine learning**: principles and techniques for data scientists. Sebastopol: O'Reilly Media, 2018.

ZHOU, Zhi-Hua. **Ensemble Methods**: foundations and algorithms. Nova York: Chapman and Hall/CRC, 2012.

ZOU, Hui; HASTIE, Trevor. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society Series B**, v. 67, p. 768-768, 2005.

II. Análise do tratamento sazonal nos métodos de previsão de vendas baseados em aprendizagem de máquina

RESUMO

A geração de previsões de vendas mais assertivas é um aspecto crítico para a maior parte das organizações, pois permite tornar o processo de planejamento mais eficiente, impactando positivamente nos resultados financeiros das empresas por meio de redução dos níveis de estoque, diminuição do número de remarcações, minimização de ruptura e aumento das vendas.

Inúmeros segmentos do varejo são susceptíveis a questões sazonais, entre eles o segmento de moda. Assim, a sazonalidade é um componente importante para as séries temporais desse setor e se apresenta como uma das dificuldades do processo de previsão.

O presente trabalho empregou duas técnicas de dessazonalização em seis métodos de previsão de vendas não lineares e de aprendizagem de máquina, visando verificar o efeito do tratamento sazonal para a obtenção de previsões mais precisas.

Assim, no caso dos métodos de previsão baseados em árvores de decisão, não houve ganhos de previsão com a aplicação dos tratamentos. Por outro lado, observou-se que a capacidade de aprendizagem das redes neurais, mais especificamente, a CNN e a LSTM, tende a melhorar quando os componentes sazonais são extraídos das séries temporais antes da sua modelagem, levando à geração de previsões mais precisas.

Palavras-chave: Previsão de vendas. Sazonalidade. Métodos de aprendizagem de máquina.

ABSTRACT

Generating more assertive sales forecasts is a critical aspect for most organizations, as it makes the planning process more efficient, positively impacting the financial results of companies by reducing inventory levels, reducing the number of markdowns, minimizing disruption and increasing sales.

Numerous retail segments are susceptible to seasonal issues, including the fashion segment. Thus, seasonality is an important component for the time series of this sector and presents itself as one of the difficulties in the forecasting process.

The present work used two seasonality techniques in six nonlinear sales forecasting and machine learning methods, aiming to verify the effect of seasonal treatment to obtain more accurate forecasts.

Thus, in the case of forecasting methods based on decision trees, there were no forecast gains with the application of treatments. On the other hand, it was observed that the learning capacity of neural networks, more specifically, CNN and LSTM, tends to improve when seasonal components are extracted from the time series before their modeling, leading to the generation of more accurate forecasts.

Keywords: Sales forecasting. Seasonality. Machine learning methods.

1. Introdução

*“If a man will begin with certainties,
he shall end in doubts,
but if he will content to begin with doubts,
he shall end in certainties.”*

Francis Bacon

Entende-se como sazonalidade um padrão que ocorre todo ano sempre no mesmo período. Esse é um tema importante para o varejo de forma geral, não só para o de moda, considerando o calendário de datas comemorativas que o varejo lança mão para fomentar seu faturamento. Assim, temos o Dia da Mãe, o Dia do Namorado, o Dia dos Pais, o Dia das Crianças, a *Black Friday* e o Natal. Essas datas comemorativas podem ser vistas como um padrão sazonal, já que se repetem ao longo dos anos e geram picos de vendas em determinadas categorias de produtos.

Para entender a relevância dessas datas no *e-commerce*, vale acompanhar o desempenho da *Black Friday* no Brasil que, em 2020, faturou R\$ 5,1 bilhões de reais entre 26 e 27 de novembro, segundo a Mercado e Consumo (2020), representando cerca de 8% do faturamento total de 2019 (ECOMMERCE BRASIL, 2020). Na China, o Dia de Solteiro, em 11 de novembro, é um marco nas vendas e, em 2020, bateu o próprio recorde nas vendas on-line atingindo cerca de US\$ 115 bilhões em faturamento, conforme Kharpal (2020). O *e-commerce* chinês tem uma penetração de 20% do total das vendas do seu varejo (RENNER, 2018).

Diante do peso que tais datas representam dentro do faturamento, torna-se relevante ajustar os métodos de previsão para identificar esses padrões sazonais. Ademais, pesquisas sobre previsão de vendas para feriados importantes e/ou datas comemorativas é tema pouco explorado pela academia, segundo Liu *et al.* (2013).

Em um ambiente em constante flutuação e com uma série de incertezas, as empresas de vestuário devem aperfeiçoar a gestão de suas cadeias de abastecimento com sistemas de previsão de vendas adequados, recomenda Thomassey (2002). Frank *et al.* (2003) reforçam que um sistema de previsão é determinante para evitar problemas com escassez ou excesso de estoque, que podem ser originados devido redução de vendas, perda de clientes e paradas de produção.

Nesse contexto, surge o problema a ser explorado neste capítulo: verificar se o tratamento da sazonalidade, tornando uma determinada série de dados

dessazonalizadas, contribui para gerar previsões mais precisas com a implementação de métodos de previsão de aprendizagem de máquina.

Nelson *et al.* (1999) e Zhang e Qi (2005) destacam que as redes neurais não são capazes de capturar o padrão sazonal e, assim, uma abordagem que contempla um pré-processamento dessazonalizando os dados é fundamental para garantir que as previsões geradas sejam mais precisas. Por outro lado, Adhikari e Agrawal (2012) afirmam ser possível configurar uma rede neural de forma adequada para que ela seja capaz de identificar a sazonalidade presente na série de dados.

A fim de verificar o impacto do tratamento sazonal na geração das estimativas deste trabalho, foram aplicados dois tipos de tratamentos sazonais, diferenciação e ajuste por uma curva polinomial, comparados com o uso do dado bruto sem ajuste para quatro categorias de produtos e para seis métodos de previsão: CNN, LGBM, LSTM, MLP, RF e XGBOOST.

Em termos de avaliação, foram calculadas seis medidas de precisão: RMSE, RMSLE, MAE, MAPE, MAEP e RSMEP, sendo que as duas últimas são, respectivamente, MAE e RMSE ponderados de forma a punir, mais severamente, as subestimativas. Ademais, foi realizado o teste MCS para comprovar a capacidade efetiva de predição dos modelos de previsão contemplados neste estudo.

O uso do tratamento sazonal se mostrou pouco eficiente para os métodos de previsão baseados em árvores de decisão (LGBM, RF e XGBOOST), pois não houve redução do erro de previsão a partir das métricas avaliadas. No caso dos modelos embasados em redes neurais, especificamente CNN e LSTM, verificou-se que o emprego dos tratamentos sazonais, a diferenciação para CNN e o ajuste de curva para LSTM contribuíram para gerar previsões mais precisas.

Os algoritmos com as previsões mais precisas foram obtidos por meio dos modelos baseados em árvores de decisão, indo ao encontro do resultado do teste MCS que, para todas as séries e para as diversas métricas de erros consideradas, revelou que somente tais modelos evidenciaram uma capacidade preditiva superior.

O capítulo está estruturado em quatro seções, sendo essa a primeira. A segunda seção revela os procedimentos metodológicos empregados no desenvolvimento da pesquisa, enquanto a terceira expõe os resultados obtidos. Por fim, a quarta seção traz as principais conclusões obtidas.

2. Tipos de tratamento sazonais

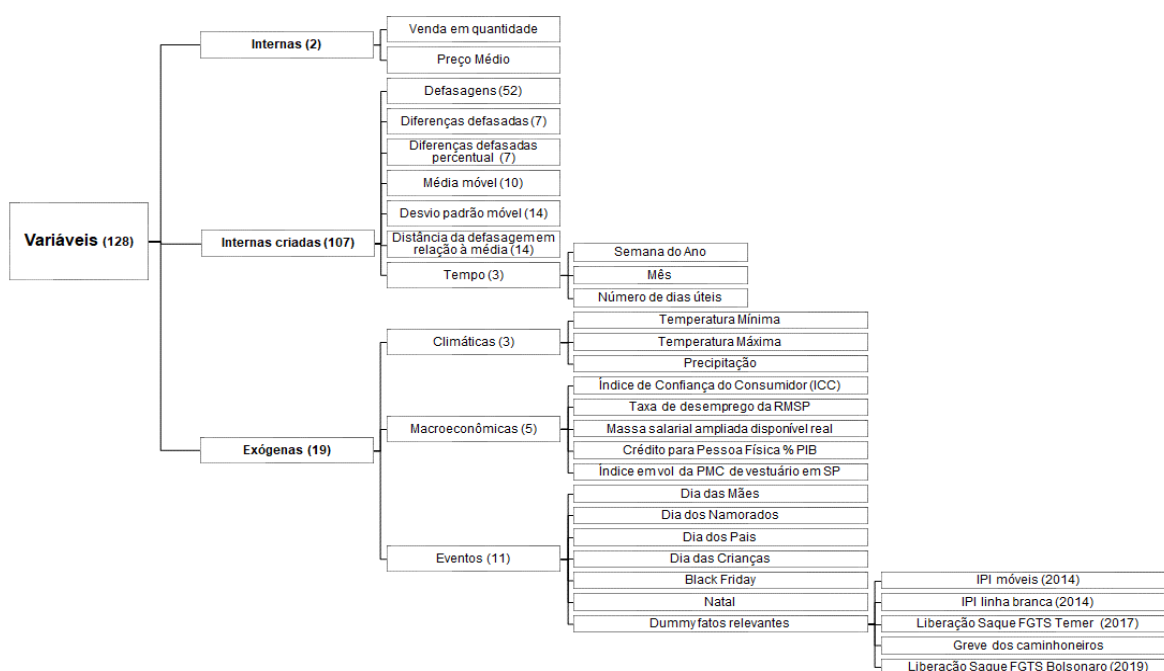
Este capítulo busca constatar a importância do uso de técnicas de dessazonalização para a previsão de vendas no setor de moda por meio de métodos não lineares baseados em inteligência computacional. A importância aqui diz respeito ao quanto é possível serem realizadas estimativas mais precisas por meio de uso de dados dessazonalizados. Em outras palavras, busca-se através da eliminação do ruído da sazonalidade alcançar previsões com maior acurácia.

A previsão é realizada em quantidade no horizonte semanal para quatro categorias de produtos: blusa de mulher, blusa de menina, camiseta de homem e camiseta de menino. Na análise descritiva, apresentada no capítulo anterior, foi possível constatar o forte padrão sazonal presente em todas as séries, podendo, inclusive, as vendas das semanas do final do ano serem consideradas como observações dados anômalos.

Em termos de base de dados, são utilizados os dados de vendas de cada uma das séries de 2007 a 2019. Para a realização do treinamento dos algoritmos de previsão de vendas, são contemplados os dados de 2007 a 2017. Já para validação desses mesmos algoritmos, são considerados os dados de 2018 e 2019.

A Figura 2.1 mostra as variáveis consideradas e suas respectivas tipologias.

Figura 2.1 – Variáveis utilizadas para previsão de vendas



Fonte: elaboração própria.

As 128 variáveis evidenciadas na seção de metodologia do capítulo introdutório estão incluídas nos modelos de previsão deste capítulo. Assim, são consideradas as duas variáveis internas, as 107 variáveis criadas através de *feature engineering* e as 19 variáveis exógenas, as quais podem ser subdivididas em variáveis exógenas climáticas, variáveis exógenas macroeconômicas e variáveis exógenas de eventos.

Os métodos de previsão de vendas para a construção das estimativas deste capítulo são baseados em inteligência computacional, contemplando métodos de árvores de decisão (RF, LGBM e XGBOOST) e de redes neurais (MLP, CNN e LSTM).

Antes da aplicação dos algoritmos citados, as variáveis que não eram estacionárias foram diferenciadas para eliminar a tendência e normalizadas para garantir uma convergência mais rápida dos modelos baseados em redes neurais.

Como já observado, não há consenso sobre a capacidade dos métodos não lineares aprenderem o padrão presente nos dados. Tal discussão está mais ligada aos métodos embasados em redes neurais. Objetivo principal da dessazonalização é eliminar o aspecto sazonal de uma determinada série de dados, tornando, a princípio, mais fácil prever o comportamento presente na série estudada. São analisados, aqui, dois tipos de tratamento sazonal para verificar o impacto do uso dos tratamentos sazonais: (1) dessazonalização por diferença e (2) dessazonalização polinomial.

Dessazonalização por diferença é a aplicação da diferença semanal sobre a série de dados, o que, obrigatoriamente, gera a perda de 52 observações. A diferenciação consiste em tirar a primeira diferença da variável em questão y_t : Δy_t . Abaixo, tem-se, respectivamente, a equação da primeira e da n-ésima diferença.

$$\Delta y_t = y_t - y_{t-1} \quad \Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}$$

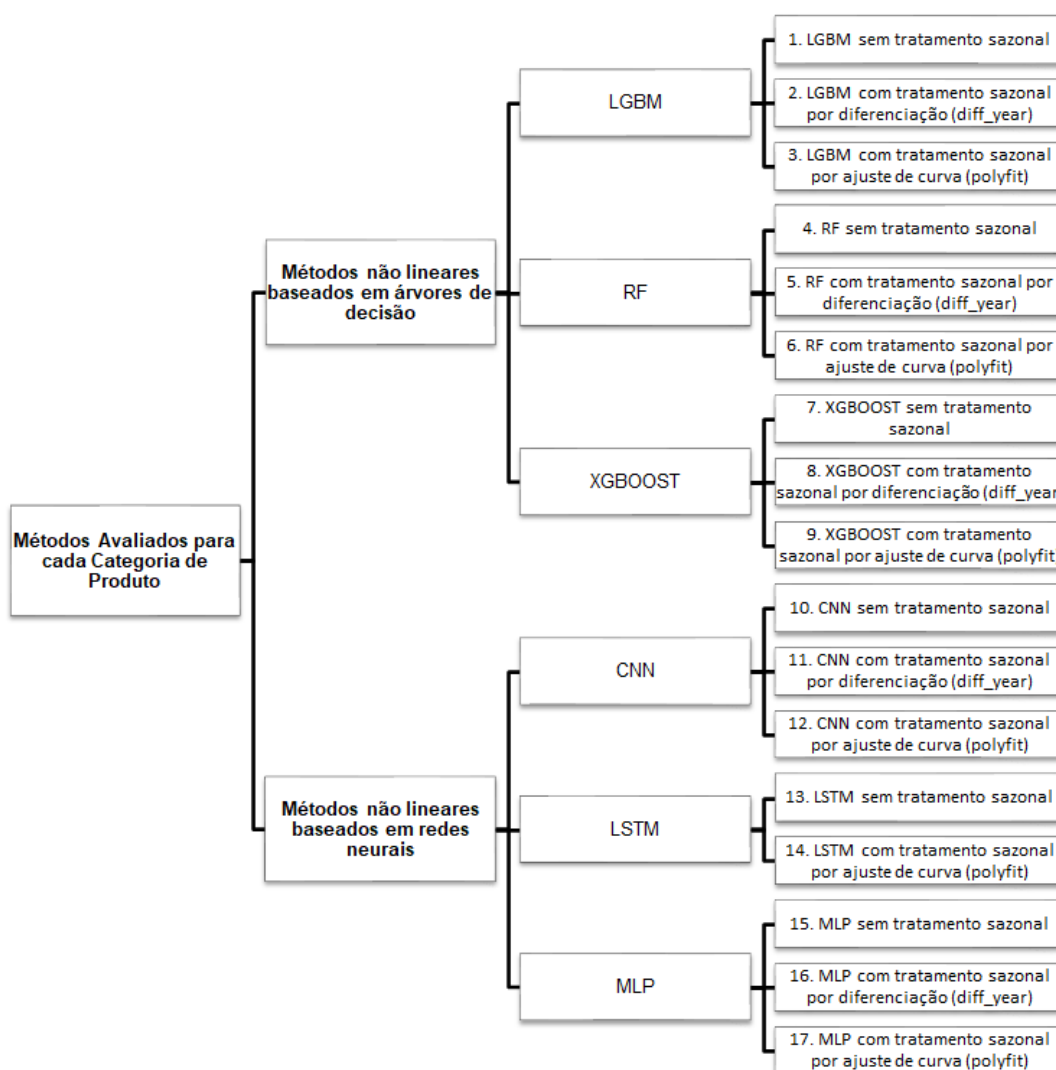
O outro método aplicado é o da modelagem da sazonalidade, obtido através de um processo de decomposição. O componente sazonal em uma determinada série temporal é, provavelmente, uma onda senoidal com período e amplitude, em geral, fixos. Tal componente pode ser aproximado usando um método de ajuste de curva, o qual é, geralmente, polinomial e, no estudo em questão, definido como sendo um polinômio de terceiro grau. Especificamente, usa-se a curva para realizar a estimativa da previsão. A subtração entre o valor estimado e o valor real determina o componente

sazonal. Afirma-se que, em certa medida, o processo realizado é similar ao processo de decomposição de uma série de dados em tendência, sazonalidade e erro.

Para avaliar o efeito do uso dos tratamentos sazonais, foram calculadas as métricas de acurácia: RMSE, RMSLE, MAE, MAEP e RMSEP, já apresentadas na parte de métricas de avaliação. As métricas visam constatar o nível de precisão das previsões geradas e, pela média, são agrupadas por tipo de tratamento sazonal, tipo de algoritmo e algoritmo em si. Os resultados são verificados por categoria de produto e, de novo, a média evidencia o efeito médio obtido por todas as séries estudadas.

A Figura 2.2 apresenta o conjunto de algoritmos de previsão juntamente com os tratamentos sazonais empregados, sendo que *diff_year* remete ao tratamento por diferenciação semanal e *polyfit* diz respeito ao ajuste pela curva.

Figura 2.2 – Métodos de previsão e tratamentos sazonais utilizados



Fonte: elaboração própria.

Adicionalmente, é submetido o teste MCS com o objetivo de verificar, efetivamente, se os tratamentos sazonais empregados são capazes de gerar previsões mais precisas. Esse é o teste formal que deve ser feito para poder comprovar a eficiência do uso de tratamentos de dessazonalização.

Salienta-se que não foi possível aplicar o tratamento de dessazonalização por diferenciação e o uso conjunto do algoritmo LSTM pela pequena quantidade de dados. O teste feito não foi capaz de gerar as previsões. Logo, são analisados os resultados obtidos da previsão de 17 alternativas. Além disso, para este estudo, não são aplicadas técnicas de seleção de variáveis, o que será visto no próximo capítulo. A não consideração desse aspecto, neste momento, tem como intuito verificar de forma isolada o efeito do uso dos tratamentos sazonais.

3. O Tratamento Sazonal faz diferença?

Nesta seção do capítulo são mostrados de forma detalhada os resultados obtidos por cada modelo e feita uma discussão comparativa entre eles. Assim, o primeiro ponto, evidenciado na Tabela 2.1, é o resultado médio em termos das métricas de acurácia obtido pelo emprego do tratamento sazonal, sem a especificação do tipo do tratamento sazonal e do tipo de algoritmo.

A partir da análise da Tabela 2.1, verifica-se que, em média, o emprego do tratamento sazonal não conseguiu aumentar a acurácia obtida nos indicadores RMSLE, MAE, MAPE e MAEP. Apenas trouxe um ganho marginal nos indicadores do RMSE e RMSEP, sendo que o último penaliza mais as subestimativas e ambos são muito afetados pela presença de dados anômalos na série. Quando se analisa o resultado apresentado por categoria de produto, constata-se que, para a série de camiseta de homem e de blusa de mulher, na média, o emprego das técnicas de dessazonalização não trouxeram ganhos de precisão. Já quando se verifica tal aspecto para as séries de camiseta de menino e blusa de menina, para a parte dos indicadores, o uso do tratamento sazonal trouxe uma melhora de acurácia, sendo que em ambos foi constatado nos indicadores RMSE e RMSEP.

Sob a perspectiva do tipo de tratamento sazonal empregado na Tabela 2.2, verifica-se um desempenho distinto entre as medidas de acurácia. Na média, entre as

séries estudadas para os indicadores RMSE e RMSEP, as previsões se mostraram mais precisas com o uso do tratamento por diferenciação semanal. Já para os indicadores RMSLE, MAE e MAPE, as melhores previsões foram verificadas sem o emprego de nenhum tratamento sazonal. No caso MAEP, deve ser empregada a diferenciação por ajuste de curva.

Tabela 2.1 – Resultado das métricas de erro como uso do tratamento sazonal

Categoria de Produto	Métrica	Tratamento Sazonal	
		Não	Sim
Média	RMSE	1.157,85	1.112,81
	RMSLE	29,63	44,45
	MAE	584,69	635,33
	MAPE	17,15	20,24
	MAEP	290,91	303,59
	RMSEP	851,47	837,42
Camiseta de Homem	RMSE	1.249,88	1.347,07
	RMSLE	21,12	23,49
	MAE	592,84	656,96
	MAPE	16,05	18,71
	MAEP	305,41	311,71
	RMSEP	856,48	1.031,44
Blusa de Mulher	RMSE	1.095,88	1.308,82
	RMSLE	31,80	31,02
	MAE	679,22	829,19
	MAPE	15,08	19,25
	MAEP	339,17	342,15
	RMSEP	782,71	1.027,10
Camiseta de Menino	RMSE	1.281,01	956,34
	RMSLE	33,51	50,55
	MAE	563,66	512,61
	MAPE	19,04	20,48
	MAEP	272,50	241,76
	RMSEP	995,96	743,19
Blusa de Menina	RMSE	1.004,65	839,02
	RMSLE	32,08	72,73
	MAE	503,04	542,56
	MAPE	18,42	22,52
	MAEP	246,57	318,73
	RMSEP	770,73	547,96

Fonte: elaboração própria.

Na avaliação do resultado alcançado em cada série de produto, nota-se comportamentos diferentes entre eles. Para a série de camiseta de homem, as estimativas mais precisas foram obtidas sem o uso de tratamento sazonal. No caso da série de blusa de mulher, o uso do tratamento sazonal com ajuste de curva teve previsões mais precisas com base no MAPE e no MAEP e o uso do tratamento sazonal por diferenciação gerou previsões mais assertivas com base no RMSLE. Na série de camiseta de menino, para a maior parte dos indicadores foram obtidas estimativas mais precisas por meio do uso do tratamento sazonal por diferenciação.

A série de blusa de menina apresentou o comportamento menos conclusivo, não sendo possível afirmar claramente qual tipo de tratamento sazonal é mais eficiente.

Tabela 2.2 – Resultado das métricas de erro por tipo de tratamento sazonal

Categoria de Produto	Tipo de Tratamento Sazonal			
	Métrica	No	Diff_year	Polyfit
Média	RMSE	1.157,85	1.020,13	1.190,04
	RMSLE	29,63	46,80	42,48
	MAE	584,69	682,24	596,24
	MAPE	17,15	23,05	17,90
	MAEP	290,91	328,66	282,69
	RMSEP	851,47	774,90	889,53
Camiseta de Homem	RMSE	1.249,88	1.205,65	1.464,91
	RMSLE	21,12	25,00	22,24
	MAE	592,84	672,33	644,16
	MAPE	16,05	20,95	16,85
	MAEP	305,41	303,84	318,27
	RMSEP	856,48	995,39	1.061,49
Blusa de Mulher	RMSE	1.095,88	1.372,36	1.255,86
	RMSLE	31,80	29,56	32,23
	MAE	679,22	979,30	704,09
	MAPE	15,08	24,45	14,92
	MAEP	339,17	352,50	333,54
	RMSEP	782,71	1.145,70	928,26
Camiseta de Menino	RMSE	1.281,01	724,30	1.149,71
	RMSLE	33,51	49,15	51,71
	MAE	563,66	479,25	540,40
	MAPE	19,04	21,54	19,60
	MAEP	272,50	235,60	246,89
	RMSEP	995,96	551,65	902,81
Blusa de Menina	RMSE	1.004,65	778,21	889,70
	RMSLE	32,08	83,50	63,76
	MAE	503,04	598,08	496,29
	MAPE	18,42	25,26	20,24
	MAEP	246,57	422,72	232,07
	RMSEP	770,73	406,84	665,55

Fonte: elaboração própria.

A próxima análise é sobre o impacto do tratamento sazonal no âmbito do tipo de algoritmo empregado. Mais especificamente, são mostrados os resultados obtidos agrupados e na média tendo em vista o tipo de algoritmo e o tipo de método sazonal. Assim, com base na Tabela 2.3, pode-se afirmar pelos valores negritados que para a maior parte dos indicadores a não utilização de tratamentos sazonais gerou previsões mais precisas no caso dos algoritmos de árvore de decisão, sendo que os dados em vermelho são os menores valores obtidos no geral e os em preto consideram somente a comparação do emprego do tratamento sazonal dentro do tipo de algoritmo. Por outro lado, no caso dos algoritmos baseados em redes neurais, o emprego de

tratamentos sazonais gerou menores níveis de erros em termos de RMSE, RMSEP e MAEP, o que não foi observado no caso do RMSLE, MAE e do MAPE.

Tabela 2.3 – Resultado métricas de erro por tipo de algoritmo e por tipo de tratamento sazonal

Categoria de Produto	Métrica	Tipo de Algoritmo + Tratamento Sazonal			
		Árvore de Decisão-Não	Árvore de Decisão-Sim	Rede Neural-Não	Rede Neural-Sim
Média	RMSE	786,92	912,16	1.528,79	1.353,59
	RMSLE	28,52	27,88	30,74	64,33
	MAE	439,20	550,29	730,17	737,37
	MAPE	14,43	17,74	19,86	23,25
	MAEP	215,89	259,26	365,93	356,78
	RMSEP	557,23	679,19	1.145,70	1.027,30
Camiseta de Homem	RMSE	877,62	1.126,92	1.622,14	1.611,24
	RMSLE	16,08	20,65	26,17	26,91
	MAE	419,71	556,01	765,96	778,11
	MAPE	12,22	16,61	19,89	21,23
	MAEP	222,70	260,21	388,11	373,51
	RMSEP	558,25	860,37	1.154,70	1.236,74
Blusa de Mulher	RMSE	811,28	1.122,43	1.380,49	1.532,48
	RMSLE	20,48	24,78	43,11	38,50
	MAE	563,25	766,95	795,18	903,86
	MAPE	13,24	18,40	16,92	20,28
	MAEP	279,44	310,75	398,90	379,84
	RMSEP	551,72	874,59	1.013,69	1.210,11
Camiseta de Menino	RMSE	827,37	753,21	1.734,65	1.200,10
	RMSLE	39,62	30,77	27,41	74,27
	MAE	405,32	415,18	721,99	629,52
	MAPE	16,05	16,89	22,03	24,79
	MAEP	188,94	184,05	356,07	311,01
	RMSEP	638,94	590,94	1.352,98	925,89
Blusa de Menina	RMSE	631,41	646,10	1.377,88	1.070,52
	RMSLE	37,91	35,31	26,26	117,64
	MAE	368,53	463,03	637,54	637,99
	MAPE	16,23	19,05	20,61	26,70
	MAEP	172,48	282,03	320,65	362,76
	RMSEP	480,01	390,86	1.061,44	736,48

Fonte: elaboração própria.

Considerando essa abordagem para análise por categoria de produto, pode-se afirmar que as árvores de decisão sem tratamento sazonal levaram a previsões mais precisas, tendo a maior parte dos indicadores para as séries de camiseta de homem e blusa de mulher. Para ambas as séries, o uso do tratamento sazonal não contribuiu para gerar previsões mais precisas no caso dos algoritmos de redes neurais.

No caso da série de menino, fica evidente que o emprego de tratamento sazonais leva a previsões mais precisas tanto para modelos com base em árvores de decisão como para modelos baseados em redes neurais. Já na série de menina, constata-se que as previsões mais precisas, com base na maior parte dos indicadores, foram obtidas sem o uso de tratamento sazonal, porém no caso do RMSLE foram obtidas previsões superiores com aplicação de tratamentos sazonais.

Após a análise agrupada por tipo de algoritmo e tratamento sazonal, é feita uma análise detalhada para cada algoritmo aplicado. Na Tabela 2.4 apresenta o impacto do tratamento sazonal sob essa perspectiva. Os números em vermelho representam

os menores valores para cada métrica e para cada série, já os dados em negrito na cor preta são os menores valores encontrados para cada algoritmo e série de produto.

Tabela 2.4 – Resultado métricas de erro por algoritmo e por tipo de tratamento sazonal

		CNN			LSTM			MLP			LGBM			RF			XGBOOST		
Categoria de Produto	Métrica	No	Diff_year	Polyfit	No	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	
Média	RMSE	1.673,70	936.84	1.646,14	1.679,71	1.643.25	1.232.96	1.248,44	1.293,27	804.98	933,28	977,38	750.86	1.037,73	768,87	804.92	944,37	811,36	
	RMSLE	25,00	23.09	59,16	25.27	57,95	41.95	135,93	45,53	35,55	25.14	51,42	17.54	24,96	19,54	32,47	24,90	21.30	
	MAE	771,65	578.22	735,43	778,27	732.61	640.59	945,64	694,94	468.19	594,78	503,35	419.84	672,04	450,19	429.58	620,51	460,89	
	MAPE	19,83	18.39	19,76	20,10	20,19	19.65	36,22	21,69	15.55	19,42	15,67	13.73	20,63	14,94	14.03	20,60	15,18	
	MAEP	388,85	255.35	357,32	393,70	351.80	315.25	496,60	322,84	225.20	295,95	236,82	205.94	325,32	212,29	216.53	270,10	215.09	
	RMSEP	1.274,44	739.26	1.267,61	1.277.99	1.278.32	884.67	897,45	953,87	578.29	712,66	715,14	521.02	782,00	542,80	572.39	743,11	579,41	
Camiseta de Homem	RMSE	1.981,48	1.276.69	1.842,10	1.973,98	1.854.09	910.95	1.229,53	1.853,82	897.16	1.243,37	1.181,84	978,06	1.115,03	958.60	757.63	1.163,63	1.099,02	
	RMSLE	27,79	21.24	25,55	27,36	25.61	23.36	36,54	25,59	16.21	20,76	17,15	17.02	21,82	18,65	15.00	24,63	20,87	
	MAE	889,44	623.54	805,39	882,45	805.88	526.00	850,03	805,68	419.49	623,87	462,16	454.45	612,31	486,40	385.18	651,89	499,44	
	MAPE	21,13	17.35	19,76	21,45	19.78	17.09	29,48	19,76	12,12	17,81	11.83	12.94	18,90	14,44	11.60	21,19	15,52	
	MAEP	451,81	294.13	399,87	433,12	397.68	279.40	377,86	398,01	225.01	284,95	240,34	234.22	282,31	239,73	208.88	279,96	234,00	
	RMSEP	1.457,23	1.063.91	1.362,53	1.463,05	1.383.22	543.80	991,14	1.382,89	578.22	1.040,21	784,26	653.28	919,42	660,34	443.26	962,28	795,67	
Blusa de Mulher	RMSE	1.527,88	1.241.19	1.525,05	1.540.53	1.573.73	1.073.05	1.748,25	1.574,19	913.02	1.123,30	1.114,66	811.64	1.438,05	838,08	709.19	1.311,04	909,46	
	RMSLE	20,07	26.29	50,32	20.38	35,02	88.88	45,87	35.02	27.02	24,07	30,30	18,17	25,50	17.11	16.25	26,07	25,61	
	MAE	817,23	925,03	783.70	824.27	802,11	744.04	1.206,30	802,18	632.42	824,36	659,86	555.17	988,62	573,15	502.17	952,18	603,55	
	MAPE	16,15	24,98	15.75	16,17	16,34	18,42	28,02	16.34	14,65	21,60	14.30	13.01	22,94	13,07	12.07	24,73	13,75	
	MAEP	404,66	293.35	377,30	418.85	385,20	373.20	458,12	385,22	305.91	308,34	311,18	272,57	388,57	263.44	259.85	314,10	278,87	
	RMSEP	1.163,47	1.054.60	1.140,76	1.163.76	1.207,96	713.83	1.438,75	1.208,47	631.13	927,33	792,92	557.22	1.197,67	584,76	466.83	1.110,16	634,69	
Camiseta de Menino	RMSE	1.773,15	649.83	1.782,31	1.779.49	1.788.87	1.651.32	946,82	832.67	828,19	640.07	940,35	760,18	690.62	784,32	893,72	694.16	769,73	
	RMSLE	26,80	18.75	80,65	27.46	80,65	27.97	163,46	27.86	22.58	19.49	79,17	17.73	22,78	23,24	78.53	21,30	18.67	
	MAE	732,27	345.86	720,12	747,36	719.51	686.36	808,41	553.70	433,05	356.67	465,90	353.79	481,44	394,79	429,11	403,87	388.40	
	MAPE	21,60	14.13	22,36	21.94	22,42	22.54	42,25	22,78	17.76	14,62	17,91	14.43	19,51	16,89	15,97	17,20	15.42	
	MAEP	367,36	145.95	348,54	384.59	345,09	316.25	487,35	228.12	192,08	167.69	209,11	165.85	206,58	183,05	208.88	170,43	167.46	
	RMSEP	1.385,15	534.34	1.422,00	1.383.74	1.431,80	1.290,03	615.68	625,65	688.03	517.26	759,71	571,31	533.78	595,52	677,48	557.19	582,20	
Blusa de Menina	RMSE	1.412,29	579.65	1.435,12	1.424.84	1.356.30	1.296.52	1.069,17	912.39	581,53	726.40	672,66	453.54	907,20	494,48	859,16	608,64	467.25	
	RMSLE	25,33	26.06	80,10	25.86	90,51	27.59	297,86	93,67	76.39	36.23	79,04	17.24	29,75	19,17	20,11	27,60	20.06	
	MAE	647,67	418.44	632,52	659,00	602.94	605.96	917,82	618,20	387.79	574,23	425,47	315.94	605,80	346,43	401,86	474,10	352.16	
	MAPE	20,43	17.09	21,16	20.86	22,21	20.54	45,14	27,88	17.67	23,64	18,64	14.53	21,16	15,55	16,48	19,28	16.01	
	MAEP	331,59	287.96	303,55	338,23	279.23	292,15	663,05	280,02	177.80	422,81	186,66	151.13	423,82	162,94	188,52	315,93	180.02	
	RMSEP	1.091,90	304.19	1.145,17	1.101,43	1.090.30	991,00	544.23	598,48	435.77	365.85	523,67	302.26	477,13	330,59	702,01	342,80	305.10	

Fonte: elaboração própria.

Para o algoritmo CNN e, na média entre as séries estudadas, o tratamento sazonal e por diferenciação gerou previsões mais precisas na maior parte das métricas de erro. Já para o LSTM, ocorreu um empate entre os resultados obtidos, entre o uso e o não uso do tratamento sazonal, não sendo possível afirmar que o uso do tratamento sazonal trouxe ganhos. Diferentemente do que era esperado, no caso do MLP, majoritariamente, o tratamento sazonal não gerou estimativas mais precisas.

Com relação aos algoritmos de árvores de decisão, LGBM, RF e XGBOOST, em média, o uso de tratamentos sazonais pioraram as estimativas produzidas em termos de precisão.

Na análise específica por série de produto, diferentemente do observado acima, para a série de camiseta de menino, as melhores previsões foram obtidas pelo emprego de tratamento sazonal na maior parte dos algoritmos. Assim, para o algoritmo CNN as melhores previsões foram através do uso da diferenciação semanal; no MLP e no XGBOOST, pelo uso do tratamento sazonal por ajuste de curva; apenas no RF e no LSTM, obteve-se previsões mais precisas sem o uso do tratamento sazonal.

Na série de blusa de menina, no caso da CNN e do LGBM, as estimativas mais precisas foram com o emprego do tratamento da sazonalidade por diferenciação. Para o XGBOOST, o melhor resultado foi alcançado com o uso do tratamento sazonal por ajuste de curva, assim como no caso do LSTM. Já para o MLP e para o RF, não é recomendado o uso de tratamento sazonais como pré-processamento no processo de previsões, pois tal iniciativa não tem impactos no desempenho das métricas de erro.

Tanto a série de camiseta de homem como a série de blusa de mulher alcançaram os menores erros de previsão sem o uso de tratamentos sazonais para a maior parte dos algoritmos.

Apresentada a análise detalhada em relação às diversas métricas de precisão e por meio de diversas perspectivas sobre o efeito do tratamento sazonal para as quatro categorias de produtos encontradas, é fundamental destacar o resultado obtido no teste MCS, que visa apresentar quais são os modelos com maior poder preditivo.

Apesar dos resultados vistos até agora, na Tabela 2.5 é possível constatar que, para a maior parte das séries estudadas e para cada uma das métricas de erro calculadas, os modelos de árvores de decisão sem tratamento sazonal apresentam maior poder preditivo, sendo considerado um nível de confiança de 95% no teste MCS aplicado. Ademais, somente para a série de camiseta de menino e para a métrica MAPE foi verificado três modelos com capacidade preditivas equivalentes.

Tabela 2.5 – Resultado do teste MCS por algoritmo e por tipo de tratamento sazonal

		CNN			LSTM			MLP			LGBM			RF			XGBOOST		
Categoria de Produto	Métrica	No	Diff_year	Polyfit	No	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	No	Diff_year	Polyfit	
Camiseta de Homem	RMSE	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
	RMSLE	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
	MAE	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
	MAPE	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
	MAEP	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
	RMSEP	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
Blusa de Mulher	RMSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
	MAE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
	MAPE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
	MAEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
Camiseta de Menino	RMSE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAPE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAEP	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
Blusa de Menina	RMSE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAPE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	MAEP	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	

Fonte: elaboração própria.

Análise mais detalhada do teste MCS permite observar que para a série de camiseta de homem e para as métricas de RMSE, MAE, MAPE, MAEP e RMSEP, o

modelo com maior capacidade preditiva foi o LGBM sem tratamento sazonal. Somente para o RMSLE o melhor modelo foi o LGBM com o tratamento sazonal por diferença.

Para a série de blusa de mulher, uniformemente para todas as métricas de erro, o algoritmo XGBOOST apresentou a maior capacidade preditiva. Essa uniformidade também foi observada na série de blusa de menina, no entanto o método de previsão RF sem tratamento sazonal foi considerado o com maior poder preditivo.

Para a série de camiseta de menino, assim como observado na série de menina, houve uma supremacia do método RF sem tratamento sazonal. Contudo, para o indicador MAEP três modelos apresentaram a mesma capacidade de predição, os quais foram: CNN com tratamento sazonal por diferença, RF sem tratamento sazonal e RF com tratamento sazonal por ajuste de curva.

Portanto, pela análise do teste do MCS, nota-se que o tratamento sazonal não foi capaz de gerar modelos com maior capacidade preditiva para a maior parte das métricas de todas as séries estudadas, nem mesmo para os modelos baseados em redes neurais, com exceção da análise tendo como métrica o MAEP e apenas para a série de camiseta de menino.

4. Conclusões

O objetivo deste capítulo foi identificar se a aplicação do tratamento sazonal para ajustar as séries de produtos estudados melhoraria o desempenho das estimativas em termos de precisão, quando comparada com as previsões sem esse tipo de ajuste. Como dito, a sazonalidade é um aspecto forte e afeta o processo de previsão das empresas do setor de moda, o que pode ser constatado no caso prático estudado. Adicionalmente, as datas comemorativas têm ganhado um peso maior no faturamento das empresas varejistas em geral.

O tratamento sazonal foi feito por meio de duas técnicas: diferenciação semanal e ajuste da curva, através de um polinômio de terceiro grau. Após aplicadas ambas as técnicas, as séries de dados passaram a estar dessazonalizadas e com menos ruídos, o que, supostamente, geraria previsões mais precisas.

As técnicas de dessazonalização foram analisadas sob a perspectiva de seis métodos de previsão não lineares, sendo três baseados em árvores de decisão

(LGBM, RF e XGBOOST), e três em redes neurais (CNN, MLP e LSTM), e por meio de seis métricas de precisão (RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP).

O tratamento sazonal não contribuiu para gerar previsões mais precisas, o que foi quase uma unanimidade para os métodos de árvores de decisão. Ademais, para a maior parte das séries estudadas, os modelos baseados em árvores de decisão foram os com maior precisão diante das diversas métricas contempladas.

Nos métodos de previsão construídos por meio de redes neurais, verificou-se a importância da aplicação do ajuste sazonal, sobretudo para a CNN e para o LSTM, não sendo possível afirmar o mesmo para a MLP. Os resultados mostram a capacidade de aprendizagem de parte dos métodos com base em redes neurais de melhorar quando os componentes sazonais são extraídos das séries temporais antes da modelagem, já que diminui a quantidade e oscilações advindas de aspectos previamente conhecidos e computáveis da série histórica.

Em termos do tipo de tratamento sazonal, o tratamento por diferenciação gerou previsões mais precisas em frequência superior ao método por ajuste de curva.

O ganho em termos de precisão da aplicação do tratamento sazonal foi mais notado na série de camiseta de menino. Para quase todos os algoritmos, com exceção do LSTM e do RF, as estimativas com menores erros para as diversas métricas calculadas vieram do uso de dados dessazonalizados. Na série de blusa de menina, o uso de dados dessazonalizado também contribuiu para previsões mais precisas em modelos de árvores de decisão, o que não foi observado nas demais categorias.

Em relação ao teste MCS, não foi possível constatar que o uso de dados ajustados sazonalmente seja capaz de gerar previsões mais assertivas, apesar dos resultados observados anteriormente referentes ao nível de acurácia absoluta. Assim, os modelos com maior capacidade preditiva foram os baseados em árvores de decisão sem a aplicação de tratamento da sazonalidade.

Conclui-se, de forma geral, que o tratamento sazonal não é recomendado para modelos de árvores de decisão, já que não trouxe ganhos de desempenho em termos de precisão, devendo ser aplicado somente nos modelos baseados em redes neurais, especificamente para algoritmo CNN e LSTM. No caso do método de previsão CNN, recomenda-se o uso do tratamento por diferenciação, e do LSTM, do tratamento por ajuste de curva, tendo em vista os tratamentos de ajuste sazonal testados neste capítulo.

Apesar da recomendação do uso do tratamento sazonal para parte dos algoritmos apoiados em redes neurais e considerando o teste MCS, somente os modelos baseados em árvores de decisão apresentaram maior capacidade preditiva.

Uma recomendação futura é a aplicação de outras técnicas de ajuste sazonal, com o intuito de verificar seu impacto em termos de acurácia no processo de previsão.

REFERÊNCIAS

ADHIKARI, Ratnadip; AGRAWAL, R. Forecasting strong seasonal time series with artificial neural networks. **Journal of scientific and industrial research**, v. 71, p. 657-666, 2012.

E-COMMERCE BRASIL. **E-commerce brasileiro faturou R\$ 61,9 bilhões em 2019, 16,3% acima de 2018**. 19 fev. 2020. Disponível em: <ecommercebrasil.com.br/noticias/e-commerce-brasileiro-faturou-r-619-bilhoes-em-2019-163-acima-de-2018/> Acesso em: 6 fev. 2021.

FRANK, Celia *et al.* Forecasting women's apparel sales using mathematical modeling. **International Journal of Clothing Science & Technology**, v. 15, n. 2, p.107-25, 2003.

KHARPAL, Arjun. **Alibaba, JD set new records to rack up record \$115 billion of sales on Singles Day as regulations loom**. CNBC, 12 nov. 2020. Disponível em: <cnbc.com/2020/11/12/singles-day-2020-alibaba-and-jd-rack-up-record-115-billion-of-sales.html> Acesso em: 28 dez. 2020.

LIU, Na *et al.* Sales Forecasting for Fashion Retailing Service Industry: A Review. **Mathematical Problems in Engineering**, v. 4, p. 1-9, 2013.

MERCADO E CONSUMO. **Vendas do e-commerce na Black Friday 2020 crescem 31% e passam de R\$ 5 bi**. 29 nov. 2020. Disponível em: <mercadoeconsumo.com.br/2020/11/28/vendas-do-e-commerce-na-black-friday-2020-crescem-31-e-passam-de-r-5-bi/> Acesso em: 28 dez. 2020.

NELSON, Michael *et al.* Time series forecasting using neural networks: should the data be deseasonalized first? **Journal of forecasting**, v. 18, n. 5, p. 359-367, 1999.

RENNER. **Apresentação Renner Day 2018**. 2018. Disponível em: <mz-filemanager.s3.amazonaws.com/13154776-9416-4fce-8c46-3e54d45b03a3/file_manager/85f89d55-4279-4d9f-bee5-9431034d158c/apresentacao_renner_day_2018.pdf> Acesso em: 20 set. 2020.

THOMASSEY, Sebastien; HAPPIETTE, Michel; CASTELAIN, Jean Marie. An automatic textile sales forecast using fuzzy treatment of explanatory variables. **Journal of Textile and Apparel, Technology and Management**, v. 2, n. IV, p. 1-15, 2002.

ZHANG, Peter; QI, Min. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, v. 160, n. 2, p. 501-514, 2005.

III. Seleção de variáveis nos métodos de previsão de vendas baseados em aprendizagem de máquina

RESUMO

A previsão de vendas é um processo crucial para as operações varejistas de forma geral. Para o setor de moda, o processo de previsão de vendas se torna ainda mais crítico por conta da alta volatilidade da demanda dos seus produtos, do fato dos produtos apresentarem um ciclo de vida extremamente encurtado e da venda no setor ser afetada por múltiplos aspectos.

Diante desse contexto, a geração de previsões cada vez mais precisas ajuda os varejistas a manterem níveis adequados de produtos, em que não há falta e nem sobra estoque, sendo possível prestar um serviço de qualidade aos seus consumidores e sem a necessidade de números altos de promoções e remarcações, que acabam reduzindo a margem de lucro das empresas desse setor.

O objetivo deste trabalho é entender o impacto do emprego de técnicas de seleção de variáveis, tendo em vista o crescimento exponencial de informação disponível para realizar as previsões dos produtos do setor em foco. Para isso, foram avaliadas o uso de três técnicas de seleção de variáveis para seis algoritmos de previsão, os quais foram implementados para quatro categorias de produtos.

A partir da análise das métricas de acurácia, verificou-se que o uso de técnicas de seleção de variáveis colaborou para geração de previsões mais precisas em 13%. Os modelos com maior capacidade preditiva foram os baseados em árvores de decisão e com técnicas de seleção de variáveis, o que foi comprovado pelo teste MCS. No entanto, dado que a técnica predominante foi a RFA, houve um expressivo aumento do custo computacional. Com base na verificação da frequência das variáveis utilizadas para cada modelo, constatou-se a baixa frequência das variáveis exógenas, evidenciando uma maior relevância das variáveis criadas, sobretudo as defasagens e as médias móveis.

Palavras-Chaves: Previsão de vendas. Seleção de variáveis. Métodos de aprendizagem de máquina.

ABSTRACT

Sales forecasting is a crucial process for overall retail operations. For the fashion industry, the sales forecasting process becomes even more critical due to the high volatility of demand for its products, the fact that products have an extremely short life cycle and sales in the sector are affected by multiple aspects.

In this context, the generation of increasingly accurate forecasts helps retailers to maintain adequate levels of products, in which there is no shortage or surplus stock, making it possible to provide a quality service to its consumers and without the need for high numbers of promotions and markdowns, which end up reducing the profit margin of companies in this sector.

The objective of this work is to understand the impact of employing variable selection techniques, considering the exponential growth of information available to forecast the products in the sector in question. For this, we evaluated the use of three variable selection techniques for six prediction algorithms, which were implemented for four product categories.

From the analysis of accuracy metrics, it was found that the use of variable selection techniques contributed to generating more accurate forecasts in 13%. The models with greater predictive capacity were those based on decision trees and variable selection techniques, which was proven by the MCS test. However, given that the predominant technique was RFA, there was a significant increase in the computational cost. Based on the verification of the frequency of the variables used for each model, the low frequency of exogenous variables was found, showing a greater relevance of the created variables, especially lags and moving averages.

Keywords: Sales forecasting. Feature selection. Machine learning methods.

1. Introdução

“What we have is a data glut.”

Vernor Vinge

O rápido avanço da internet e da capacidade computacional fez com que o volume de dados de diversos tipos (texto, foto, voz e vídeo etc.) crescesse substancialmente. A alta dimensionalidade dos dados é um desafio enorme para a sua análise e a tomada de decisão. Assim, a seleção de variáveis é crucial para a eficiência não só do processamento dos dados, mas para a aprendizagem estatística.

A seleção de variáveis pode ser definida como o processo de obtenção de um subconjunto de um conjunto de variáveis original de acordo com um determinado critério, sendo que atualmente existem diversos métodos de seleção de variáveis. O objetivo principal do uso desse tipo de tratamento, visto como um pré-processamento, é remover variáveis redundantes e irrelevantes para o processo de previsão. Conforme Cai *et al.* (2018), empregar técnicas de seleção de variáveis traz como benefícios: melhora a precisão do aprendizado, reduz o tempo de aprendizado e simplifica os resultados da aprendizagem, tornando o modelo mais compreensível.

A previsão de séries temporais consiste em prever valores desconhecidos de uma série com base, sobretudo, em valores anteriores conhecidos, segundo Verleysen e François (2005). Devido ao caráter não aleatório da série temporal, existem dependências entre as coordenadas dos regressores, os valores anteriores da série. Portanto, essa é a típica situação em que a redução de dimensão deveria ser possível, levando a melhores desempenhos de previsão.

O uso de técnicas de seleção de variáveis pode ser aplicado a qualquer problema de previsão, inclusive para as séries temporais. Assim, o foco deste capítulo é avaliar o uso de técnicas de seleção de variáveis para a realização da previsão de vendas na indústria da moda, onde a previsão é crucial e colabora significativamente para os resultados de uma empresa por conta das especificidades do setor.

O impacto do uso de métodos de seleção de variáveis é mensurado em termos de aumento dos níveis de acurácia obtidos, considerando a aplicação de técnicas de seleção de variáveis, quando comparado com o não uso de tais métodos e em termos de redução de custo computacional. Ademais, são apresentadas as variáveis mais relevantes para as séries de produtos estudadas do Lojão do Brás.

Para verificar o impacto do uso de técnicas de seleção de variáveis, são aplicadas três técnicas: a filtragem por meio do p-valor das variáveis com nível de significância inferior a 5%, a técnica *recursive feature Addition* (RFA) e a técnica *Shapley additive explanations* (SHAP), sendo as duas últimas iterativas. As técnicas são comparadas com o uso do conjunto completo de variáveis para a previsão de vendas das quatro categorias de produtos mais relevantes para a empresa, considerando seis métodos de previsão: CNN, LGBM, LSTM, MLP, RF e XGBOOST.

A avaliação da precisão foi feita através de seis métricas, RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP, que evidenciam o nível de erro das previsões obtidas, em que, quanto menor o valor, melhores são as estimativas geradas. Especificamente, o MAEP e RMSEP são medidas ponderadas, respectivamente, do MAE e do RMSEP, que penalizam mais as previsões menores do que o dado real, já que a ruptura é um problema mais impactante financeiramente para a empresa, o que é intensificado em um cenário econômico de baixas taxas de juros. Adicionalmente, foi realizado o teste MCS para comprovar, formalmente, a capacidade preditiva das combinações entre algoritmos de previsão e técnicas de seleção de variáveis presentes neste estudo.

O uso de técnicas de seleção de variáveis gerou estimativas com níveis de erro 13% menores, o que comprova a importância da sua aplicação como um pré-processamento dos dados a serem previstos. Houve melhora dos níveis de precisão com o uso de técnicas de seleção de variáveis tanto para os modelos baseados em árvores de decisão quanto para os modelos baseados em redes neurais.

Em termos médios, considerando as diversas métricas de erro e as quatro categorias de produtos estudadas, as estimativas mais precisas foram obtidas por meio do algoritmo RF e do uso do tratamento de seleção de variáveis RFA.

O fato de o uso das técnicas de seleção de variáveis melhorar os níveis de precisão das previsões geradas vai ao encontro do estudo de Pajares, Benítez e Palmero (2008), que colocam a seleção de variáveis como estágio importante para a modelagem de séries temporais, pois permite produzir modelos mais simples, enquanto mantém o nível de desempenho ou, em alguns casos, torna-os melhor. Além disso, minimiza o problema da maldição da dimensionalidade.

O teste MCS confirmou os resultados da análise das métricas de precisão em valores absolutos, evidenciando que os modelos baseados em árvores de decisão e

com o emprego predominante da técnica de seleção de variáveis RFA foram os com as maiores capacidades preditivas.

Em relação ao custo computacional, na média para todos os algoritmos e para todas as categorias de produtos estudadas, a não utilização de técnica de seleção de variáveis permitiu um treinamento mais rápido. Sendo que o método de seleção de variáveis, o RFA, que garantiu uma acurácia superior, teve um custo computacional cerca de 3,5 vezes superior em relação às demais técnicas de seleção de variáveis.

Por fim, sobre a questão da importância de variáveis inseridas nos modelos através de uma análise de frequência, verificou-se que as variáveis criadas por *feature engineering* foram mais relevantes, considerando a seleção das 30 variáveis mais importantes, tendo como destaques as defasagens e as médias móveis. Somente cerca de 4% das variáveis selecionadas são variáveis exógenas, sendo as mais relevantes: temperatura máxima e a binária de Natal.

A estrutura deste capítulo tem quatro seções, sendo essa a primeira delas. A segunda seção apresenta os procedimentos metodológicos adotados, enquanto a terceira mostra os resultados verificados. A quarta e última seção traz as conclusões.

2. Métodos de seleção de variáveis

O emprego de técnicas de seleção de variáveis visa reduzir o custo computacional e até mesmo melhorar o nível de acurácia da previsão obtida.

O objetivo deste estudo é constatar a importância do emprego de técnicas de seleção de variáveis no processo de previsão de vendas para o varejo de moda em um contexto do uso de múltiplas variáveis e do uso de métodos não lineares de previsão. Para medir essa importância, são analisadas as métricas de precisão e o tempo de processamento como indicador do custo computacional envolvido.

A seleção de variáveis pode ser vista como um método de redução de variância, que busca um equilíbrio entre obter os benefícios da minimização da variância, resultante da redução na dimensionalidade, e diminuir o impacto do viés aumentado, devido à eliminação de algumas variáveis irrelevantes, de acordo com Munson e Caruana (2009). Estudo feito pelos autores mostrou que o uso de técnicas de seleção de variáveis permite determinar o conjunto que representa a melhor compensação entre o viés de ter poucas variáveis e a variação de ter muitos recursos.

A previsão é feita em termos de quantidade para quatro categorias de produtos comercializados no Lojão do Brás: camiseta de homem, blusa de mulher, camiseta de menino e blusa de menina. Descritas no capítulo de introdução, as séries foram selecionadas pela relevância, em termos de faturamento, para a empresa que concedeu os dados de pesquisa. O horizonte da previsão é de curto prazo, sendo de uma semana.

A base de dados coletada contempla o período de 2007 a 2019, sendo que os dados entre 2007 e 2017 são utilizados para treinar os modelos de previsão de vendas e os dados entre 2018 e 2019 são utilizados para a validação desses algoritmos.

No total, são 128 variáveis utilizadas, sendo que 107 foram criadas pelo processo de *feature engineering*. Essas variáveis foram denominadas “criadas”, pois foi aplicada alguma função matemática ou de data/tempo para que elas fossem geradas. Especificamente, foram criadas as seguintes variáveis:

- **Defasagens:** é o uso de um valor anterior da variável-alvo como *feature* para prever o valor atual. Foram consideradas 52 defasagens;
- **Diferenças defasadas:** é a diferença entre o valor observado entre dois períodos defasados para evitar que tenhamos o problema de vazamento de informação, pois a variável-alvo será prevista em diferença. Foram consideradas sete diferenças defasadas;
- **Diferenças defasadas percentual:** é a diferença defasada dividida pelo valor da observação no período mais antigo, o que permite gerar um valor percentual. Foram consideradas mais sete atributos desse tipo;
- **Média móvel:** foi aplicada a média sobre algumas janelas de período, sendo considerado de 2 a 12 períodos, gerando mais 10 variáveis para uso nos modelos. Importante considerar que a média foi aplicada sobre a variável correspondente à primeira defasagem para evitar problemas de uso de dados do presente para realização da previsão.
- **Desvio padrão móvel:** aplicação da função desvio padrão para as janelas de período de 2 a 16 observações, sendo considerada como variável base a primeira defasagem. Com isso, ao modelo foram acrescentadas 14 variáveis;

- **Distância da defasagem em relação à média:** é a divisão da primeira defasagem em relação à média móvel para janelas de dois a 16 períodos, o que agregou mais 14 variáveis para os modelos de previsão;
- **Semana do ano:** especificação da semana do ano;
- **Mês:** especificação do mês da semana;
- **Número de dias úteis:** quantidade de dias úteis da semana, considerando que isso afeta o volume de vendas.

Ademais, foram consideradas 19 variáveis exógenas, as quais foram subdivididas em climáticas, macroeconômicas e de eventos. As variáveis climáticas são: os valores reais de temperatura mínima, de temperatura máxima e de volume de precipitação para a cidade de São Paulo, agrupados por meio da função média para obter os valores semanais, já que as informações são disponibilizadas diariamente pela Somar Meteorologia, prestador de serviço do Lojão do Brás.

As variáveis macroeconômicas são: índice de confiança do consumidor, taxa de desemprego na região metropolitana de São Paulo, massa salarial ampliada disponível real, crédito para pessoa física, como participação do PIB, e índice em volume da pesquisa mensal do comércio para categoria de vestuário em São Paulo. A escolha dessas variáveis foi pela influência no varejo, segundo a SBVC (2020).

Por fim, temos as 11 variáveis de eventos, que visam controlar o impacto de datas comuns no calendário varejista, como Dia das Mães, Dia dos Namorados, Dia dos Pais, Dia das Crianças, *Black Friday* e Natal, e eventos oriundos de ações governamentais diretas, como isenção de IPI sobre móveis, linha branca e liberação do saque do FGTS, além da greve dos caminhoneiros.

Diante desse extenso número de variáveis, o primeiro questionamento que surge é a necessidade de contemplar todas elas nos métodos de previsão. Sendo a resposta negativa, quais métodos de seleção de variáveis poderiam ser aplicados?

Em regressão, o conceito de multicolinearidade diz respeito à presença de regressores altamente correlacionados entre si, e não apenas correlacionados à variável resposta. Isso implica na presença de variáveis redundantes, tornando mais difícil verificar a importância real de cada uma para o modelo. Como consequência, a multicolinearidade aumenta os erros padrão dos coeficientes, o que, por sua vez,

significa que os coeficientes, para algumas variáveis independentes, podem se tornar estatisticamente insignificantes, impactando o nível de acurácia do modelo.

Apesar da multicolinearidade ser um ponto mais discutido nos modelos de regressão dado o impacto sobre os coeficientes e considerando que existe um foco maior na interpretação do modelo em si, no caso dos demais métodos de previsão de aprendizagem de máquina, passando pelos modelos baseados em árvores e em redes neurais, a aplicação de técnicas de regularização contribui para amenizar esse aspecto intrinsecamente e, adicionalmente, o foco maior desses modelos, geralmente, é a previsão em si e não o aspecto causal, que é o mais relevante nas regressões.

A seleção de um subconjunto de variáveis é importante, primeiramente, para que os atributos efetivamente relevantes sejam considerados, evitando redundâncias. Além disso, a sua aplicação gera os seguintes benefícios: minimiza o impacto da maldição da dimensionalidade, que demanda um número exponencial de amostra com o aumento dos atributos contemplados no modelo; reduz o tempo de treinamento; e minimiza o problema de sobreajuste, que aparece em conjunto com o aumento do número de variáveis, permitindo, então, uma maior generalização do modelo.

O primeiro capítulo apresentou várias técnicas de seleção de variáveis. Para este capítulo foram escolhidas três técnicas, sendo uma de filtragem, que avalia o p-valor calculado em relação a um nível de significância para a escolha de um conjunto de variáveis, e duas técnicas de seleção do tipo *wrapper*: *recursive feature Addition* (RFA) e *Shapley additive explanations* (SHAP). As técnicas *wrapper* avaliam diversos subconjuntos de variáveis de entrada e selecionam o subconjunto com melhor desempenho, a partir de uma métrica previamente definida. Essas diversas avaliações iterativas têm como consequência um maior custo computacional quando comparada com a aplicação de técnicas de seleção de filtragem e com a aplicação de técnicas *embedded*, como ocorre nos modelos baseados em árvores de decisão.

A aplicação da técnica de seleção por p-valor primeiramente descarta as variáveis altamente correlacionadas, o que significa eliminar as variáveis com índice com coeficiente de Spearman superior a 0.9 e, depois, iterativamente vai eliminando as variáveis com o nível descritivo do teste superior ao nível de significância de 0.05 até que todo o conjunto de variáveis selecionadas respeite tal nível de significância.

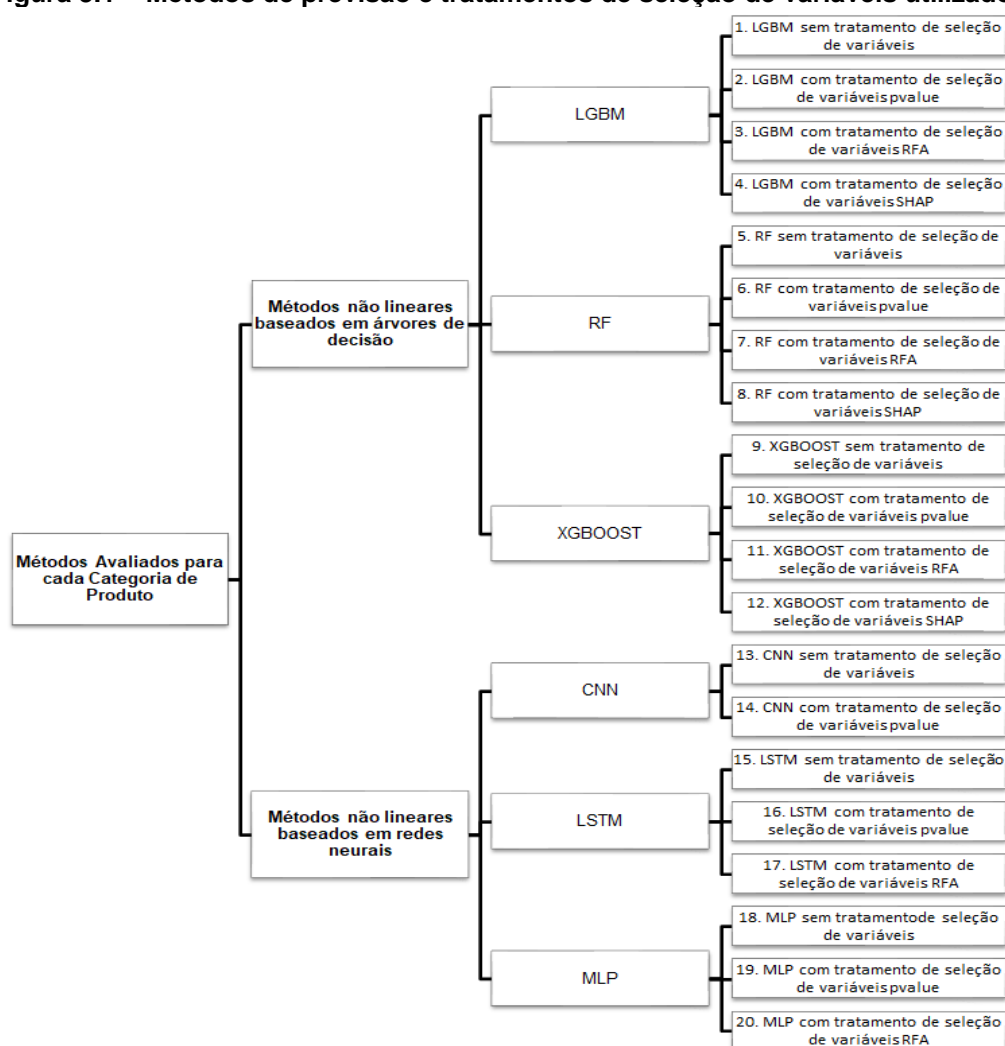
O método RFA consiste em um processo de adição de variáveis. O primeiro passo calcula a previsão com o modelo contendo apenas uma variável. Em seguida,

seleciona-se o modelo com a variável que apresentou a estimativa mais precisa. Posteriormente, o processo é repetido iterativamente até ser selecionado o número de variáveis desejadas, respeitando o critério que a adição de uma nova variável deve necessariamente reduzir o erro de previsão gerado. Portanto, o conjunto de variáveis selecionadas completo deve produzir um erro menor quando comparado a outro conjunto de variáveis que contenha mais variáveis. Para este trabalho, fixou-se um número máximo k de 30 variáveis.

O terceiro método é o SHAP, embasado na teoria dos jogos, que quantifica a contribuição marginal que cada variável traz para o jogo, sendo o jogo definido como cada observação. Como visto, o SHAP busca a interpretabilidade local de um modelo preditivo. A vantagem do método é possuir rápida implementação para modelos baseados em árvores de decisão, enquanto sua desvantagem é ignorar a dependência existente entre as variáveis. Na implementação desse método para o Python, ela apenas funciona para aplicação junto a modelos de árvores de decisão.

Especificados os métodos de seleção, é importante apresentar os algoritmos de previsão usados. A Figura 3.1 traz o conjunto de modelos de previsão e os tratamentos de seleção de variáveis empregados. No total, são feitas 20 combinações.

Figura 3.1 – Métodos de previsão e tratamentos de seleção de variáveis utilizados



Fonte: elaboração própria.

A avaliação é somente sobre métodos não lineares, subdivididos em modelos baseados em árvores de decisão (LGBM, RF e XGBOOST) e em redes neurais (CNN, MLP e LSTM). Não foi possível implementar o tratamento RFA no algoritmo CNN. Em todas as tentativas, as previsões geradas foram infinitas.

A avaliação da acurácia obtida através do uso das técnicas de seleção de variáveis é feita por meio da comparação das seguintes métricas: RMSE, RMSLE, MAE, MAEP e RMSEP. Elas estão formalmente descritas no primeiro capítulo. Ademais, é evidenciado o tempo médio de treinamento de cada modelo e com cada técnica com o objetivo de demonstrar o custo computacional envolvido. O custo computacional em termos de tempo foi calculado considerando um computador com as seguintes especificações em termos de processamento, memória e capacidade de armazenamento: processador Intel Core i7-7500U CPU 2.70GHz 2.90GHz e 32GB de memória RAM e um disco SSD de 1 TB.

A partir do cálculo das métricas, verifica-se o nível de precisão alcançado com cada técnica e sem a mesma, permitindo uma efetiva comparação. A análise é feita por diversas perspectivas: agrupada em relação à presença de tratamento de seleção de variável, por tipo de tratamento aplicado, por tipo de algoritmo e por algoritmo em si, sendo utilizada a média dos resultados para geração dos subtotais de cada agrupamento. Esses cálculos também são disponibilizados por cada categoria de produto e, por fim, é evidenciada a média de todas as categorias estudadas.

Além da comparação das métricas de acurácia para avaliação do uso das técnicas de seleção no processo de previsão, também é feito o teste MCS para estabelecer, formalmente, qual combinação entre algoritmo e técnica de seleção possui o maior poder preditivo, já que não necessariamente melhores níveis de precisão garantem o maior poder preditivo.

A última análise seleciona as 30 variáveis mais importantes para os modelos através da distribuição de frequência das diversas variáveis contempladas neste estudo. Ademais, tendo em vista a presença das variáveis exógenas e das variáveis criadas e com base na distribuição de frequência, é possível examinar qual tipo de variável foi mais importante para as previsões construídas e, mais do que isso, afirmar se a inclusão de variáveis exógenas contribui para previsões mais assertivas.

3. Vale a pena selecionar variáveis?

Nesta parte do capítulo são apresentados os resultados obtidos de forma minuciosa da aplicação dos tratamentos de seleção de variáveis para a realização da previsão de vendas das categorias de produtos estipuladas.

O primeiro resultado a ser evidenciado diz respeito ao resultado médio em termos de acurácia para cada uma das métricas RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP, do uso ou não de algum tratamento de seleção de variáveis, sem o detalhamento do tipo de tratamento do tipo de algoritmo.

Na Tabela 3.1 verifica-se que, independente, da categoria de produto, na média e para todos os indicadores de precisão, foram obtidas previsões com menor nível de erro com a aplicação de um método de seleção de variáveis. Na média geral, as estimativas com a implementação de técnicas de seleção de variáveis foram cerca de 13% mais precisas. Nas categorias de produtos infantis o ganho foi ainda maior: em

torno de 15%. Assim, é evidente a vantagem do emprego de um conjunto reduzido de variáveis em termos de acurácia *versus* a possibilidade do uso do conjunto completo.

Tabela 3.1 – Resultado das métricas de erro como uso de tratamento de seleção de variáveis

Tratamento de Seleção de Variáveis			
Categoria de Produto	Métrica	Não	Sim
Média	RMSE	1.157,85	1.015,56
	RMSLE	29,63	22,43
	MAE	584,69	523,13
	MAPE	17,15	15,62
	MAEP	290,91	256,76
	RMSEP	851,47	746,68
Camiseta de Homem	RMSE	1.249,88	1.083,29
	RMSLE	21,12	18,78
	MAE	592,84	520,83
	MAPE	16,05	14,39
	MAEP	305,41	269,14
	RMSEP	856,48	732,34
Blusa de Mulher	RMSE	1.095,88	1.045,05
	RMSLE	31,80	21,12
	MAE	679,22	639,07
	MAPE	15,08	14,04
	MAEP	339,17	311,41
	RMSEP	782,71	766,00
Camiseta de Menino	RMSE	1.281,01	1.142,62
	RMSLE	33,51	21,67
	MAE	563,66	502,80
	MAPE	19,04	17,28
	MAEP	272,50	241,28
	RMSEP	995,96	884,07
Blusa de Menina	RMSE	1.004,65	791,28
	RMSLE	32,08	28,17
	MAE	503,04	429,82
	MAPE	18,42	16,76
	MAEP	246,57	205,20
	RMSEP	770,73	604,33

Fonte: elaboração própria.

A Tabela 3.2 mostra desempenhos distintos entre os métodos de seleção de variáveis e entre as categorias de produtos. Na média de categorias de produtos, os maiores níveis de precisão para as métricas RMSE, MAE, MAEP e RMSEP foram obtidos com o tratamento SHAP. Já para as métricas percentuais, os melhores resultados foram com o uso do RFA.

Tabela 3.2 – Resultado das métricas de erro por tipo de tratamento de seleção de variáveis

Tipo de Tratamento Seleção de Variáveis					
Categoria de Produto	Métrica	no	pvalue	rfa	shap
Média	RMSE	1.157,85	1.203,35	926,69	788,10
	RMSLE	29,63	25,83	17,64	23,63
	MAE	584,69	607,71	472,72	437,99
	MAPE	17,15	17,57	14,06	14,31
	MAEP	290,91	301,65	230,21	211,23
	RMSEP	851,47	890,76	679,66	570,24
Camiseta de Homem	RMSE	1.249,88	1.250,68	992,36	900,04
	RMSLE	21,12	21,63	16,68	16,59
	MAE	592,84	609,11	471,37	426,73
	MAPE	16,05	16,48	12,95	12,60
	MAEP	305,41	312,41	245,65	221,74
	RMSEP	856,48	862,96	656,57	597,36
Blusa de Mulher	RMSE	1.095,88	1.233,43	950,14	826,48
	RMSLE	31,80	25,89	15,63	20,70
	MAE	679,22	736,30	564,91	568,18
	MAPE	15,08	15,70	12,55	13,21
	MAEP	339,17	370,28	262,71	274,85
	RMSEP	782,71	898,44	719,25	579,07
Camiseta de Menino	RMSE	1.281,01	1.334,00	1.105,13	822,37
	RMSLE	33,51	23,92	20,23	19,55
	MAE	563,66	576,82	474,86	401,34
	MAPE	19,04	19,19	15,98	15,63
	MAEP	272,50	277,79	232,26	183,33
	RMSEP	995,96	1.035,37	849,24	639,52
Blusa de Menina	RMSE	1.004,65	995,28	659,13	603,52
	RMSLE	32,08	31,88	18,03	37,67
	MAE	503,04	508,62	379,75	355,69
	MAPE	18,42	18,90	14,76	15,82
	MAEP	246,57	246,12	180,21	165,00
	RMSEP	770,73	766,30	493,57	465,00

Fonte: elaboração própria.

Na análise detalhada por categoria de produto, na camiseta de homem e na camiseta de menino, para todos os indicadores de acurácia, as melhores estimativas foram obtidas através do uso do tratamento SHAP. Para a série de blusa de mulher, apenas para RMSE e RMSEP, as previsões mais precisas foram com a aplicação do tratamento de seleção de variáveis SHAP. Para os demais indicadores, o uso do tratamento RFA garantiu previsões com menor nível de erro. Para a série de blusa de menina, os resultados foram equivalentes aos médios, em que, para a maior parte das métricas de precisão, o melhor resultado foi com o emprego do SHAP; somente para as métricas percentuais o melhor desempenho foi oriundo do uso do tratamento RFA.

A próxima análise demonstra o uso do tratamento de seleção de variáveis sobre determinado tipo de algoritmo. Assim, são mostrados os resultados agrupados pela média por tipo de algoritmo, ou seja, se é baseado em árvores de decisão ou em redes neurais, e pela presença de tratamento de seleção de variáveis.

Com base na Tabela 3.3, em que valores negritados em preto são os melhores modelos em termos de precisão para cada tipo de algoritmo e valores negritados em

vermelhos evidenciam os mais precisos no geral para cada série de produto, atesta-se que, na média, tanto para os modelos de árvores de decisão como para os de redes neurais, o tratamento de seleção de variáveis produziu estimativas mais precisas.

Tabela 3.3 – Resultado das métricas de erro por tipo de algoritmo e por tipo de tratamento de seleção de variáveis

Tipo de Algoritmo + Tratamento Seleção de Variáveis					
Categoria de Produto	Métrica	Árvore de Decisão-Não	Árvore de Decisão-Sim	Rede Neural-Não	Rede Neural-Sim
MÉDIA	RMSE	804,92	754,40	1.528,79	1.528,03
	RMSLE	32,47	21,81	30,74	24,43
	MAE	429,58	424,77	730,17	725,16
	MAPE	14,03	13,86	19,86	19,49
	MAEP	216,53	206,50	365,93	359,38
	RMSEP	572,39	534,89	1.145,70	1.161,05
HOMEM	RMSE	757,63	833,37	1.622,14	1.690,94
	RMSLE	15,00	15,59	26,17	26,65
	MAE	385,18	409,67	765,96	789,33
	MAPE	11,60	12,14	19,89	19,81
	MAEP	208,88	213,13	388,11	404,57
	RMSEP	443,26	530,95	1.154,70	1.223,04
MULHER	RMSE	709,19	788,31	1.380,49	1.528,43
	RMSLE	16,25	21,80	43,11	20,28
	MAE	502,17	546,41	795,18	819,56
	MAPE	12,07	12,92	16,92	16,39
	MAEP	259,85	266,34	398,90	398,81
	RMSEP	466,83	550,05	1.013,69	1.170,23
MENINO	RMSE	893,72	817,57	1.734,65	1.743,98
	RMSLE	78,53	18,98	27,41	27,07
	MAE	429,11	386,62	721,99	725,66
	MAPE	15,97	14,94	22,03	22,06
	MAEP	208,88	178,57	356,07	360,56
	RMSEP	677,48	623,95	1.352,98	1.364,10
MENINA	RMSE	859,16	578,36	1.377,88	1.148,75
	RMSLE	20,11	30,86	26,26	23,74
	MAE	401,86	356,37	637,54	566,08
	MAPE	16,48	15,42	20,61	19,70
	MAEP	188,52	167,94	320,65	273,59
	RMSEP	702,01	434,63	1.061,44	886,84

Fonte: elaboração própria.

Cada série de produto produziu resultados diferentes. A série de camiseta de homem teve resultados superiores em termos de acurácia tanto para modelos baseados em árvores de decisão como para redes neurais sem o uso de tratamento de seleção de variáveis. No caso da série de blusa de mulher, os modelos de árvores de decisão também apresentaram desempenhos superiores sem o tratamento de seleção de variáveis, o que não foi verificado nos modelos com base em redes neurais, em que majoritariamente o uso de um conjunto menor de variáveis gerou estimativas mais precisas. Nas séries de produtos infantis, o uso do tratamento de seleção de variáveis produziu previsões mais precisas para os modelos de árvores de decisão. Para os algoritmos de redes neurais, no caso da série de blusa de menina, a

aplicação de tratamento de seleção de variáveis gerou um desempenho superior frente ao não uso, o que não ocorreu na série de camiseta de menino.

A Tabela 3.4 mostra os resultados do uso das técnicas seleção de variáveis para cada tipo de algoritmo e para cada série de produto estudada. Mais uma vez, os valores negritados em vermelho representam os menores níveis de erro para cada série de produto, já os negritados em preto representam os menores níveis de erros por tipo de algoritmo.

Tabela 3.4 – Resultado das métricas de erro por algoritmo e por tratamento de seleção de variáveis

Categoria de Produto	Métrica	CNN		LSTM		MLP		LGBM		RF		XGBOOST	
		No	Pvalue	No	Pvalue	No	Pvalue	No	Pvalue	No	Pvalue	No	Pvalue
Média	RMSE	1.673,70	1.679,03	1.679,71	1.677,54	1.421,65	1.232,96	1.333,88	1.316,13	804,98	901,98	652,10	811,25
	RMSLE	25,00	25,26	25,27	25,51	23,27	41,95	23,70	20,09	35,55	42,57	14,53	35,75
	MAE	771,65	777,01	778,27	784,04	683,05	640,59	656,53	600,31	488,19	511,30	365,45	462,23
	MAPE	19,83	19,92	20,10	20,71	18,49	19,65	18,84	15,99	15,55	16,42	11,72	15,16
	MAEP	388,85	396,28	393,70	386,04	328,66	315,25	326,54	298,63	225,20	241,76	179,15	220,73
	RMSEP	1.274,44	1.274,86	1.277,99	1.279,68	1.086,90	884,67	1.002,78	995,33	578,29	651,74	451,89	587,43
Camiseta de Homem	RMSE	1.981,48	1.985,90	1.973,98	1.968,03	1.975,55	910,95	834,30	901,95	897,16	944,77	761,38	912,33
	RMSLE	27,79	27,90	27,36	27,30	27,53	23,36	23,87	16,06	16,21	16,53	12,17	16,27
	MAE	889,44	891,24	882,45	876,55	878,18	526,00	511,33	447,30	419,49	456,68	344,36	421,21
	MAPE	21,13	21,11	21,45	20,84	20,49	17,09	16,78	12,95	12,12	12,84	9,39	12,10
	MAEP	451,81	455,30	433,12	439,48	454,03	279,40	269,46	231,51	225,01	232,26	185,60	222,01
	RMSEP	1.457,23	1.458,92	1.463,05	1.453,14	1.449,12	543,80	530,99	582,00	578,22	639,51	452,20	603,41
Blusa de Mulher	RMSE	1.527,88	1.537,95	1.540,53	1.542,49	1.491,02	1.073,05	1.542,24	1.422,21	913,02	1.027,76	602,14	941,77
	RMSLE	20,07	20,34	20,38	20,47	19,90	88,88	20,40	18,30	27,02	53,55	13,02	28,99
	MAE	817,23	822,53	824,27	829,53	801,21	744,04	824,99	750,98	632,42	722,93	415,52	642,57
	MAPE	16,15	16,13	16,17	16,47	16,78	18,42	16,19	14,77	14,65	17,00	10,09	14,82
	MAEP	404,66	418,01	418,85	413,14	344,90	373,20	419,19	367,44	305,91	342,88	208,23	310,02
	RMSEP	1.163,47	1.161,98	1.163,76	1.168,60	1.185,01	713,83	1.165,33	1.092,71	631,13	714,95	415,85	646,31
Camiseta de Menino	RMSE	1.773,15	1.774,96	1.779,49	1.776,16	1.711,71	1.651,32	1.713,10	1.662,70	828,19	1.040,40	684,99	808,85
	RMSLE	26,80	27,19	27,46	27,55	26,96	27,97	26,57	24,25	22,58	23,59	15,35	21,70
	MAE	732,27	740,72	747,36	751,60	704,46	686,36	705,85	657,09	433,05	471,79	329,96	412,83
	MAPE	21,60	21,76	21,94	22,83	21,42	22,54	22,25	19,15	17,76	17,67	12,53	16,71
	MAEP	367,36	378,11	384,59	365,62	358,47	316,25	340,05	326,58	192,08	210,86	150,71	182,49
	RMSEP	1.385,15	1.382,63	1.383,74	1.389,34	1.340,48	1.290,03	1.343,96	1.305,05	698,03	805,52	517,99	653,26
Blusa de Menina	RMSE	1.412,29	1.417,30	1.424,84	1.423,46	508,32	1.296,52	1.245,89	1.277,68	581,53	595,00	559,90	582,03
	RMSLE	25,33	25,61	25,96	26,73	18,68	27,59	23,93	21,76	76,39	76,63	17,59	76,07
	MAE	647,67	653,57	659,00	678,46	348,34	605,96	583,94	545,88	387,79	393,81	371,96	372,31
	MAPE	20,43	20,69	20,86	22,70	15,26	20,54	20,14	17,07	17,67	18,17	14,88	17,01
	MAEP	331,59	333,72	338,23	325,93	157,24	292,15	277,47	266,99	177,80	181,06	172,06	168,40
	RMSEP	1.091,90	1.095,92	1.101,43	1.107,64	372,97	991,00	970,83	1.001,55	435,77	446,96	421,52	446,72

Fonte: elaboração própria.

Para a média dos tipos de produtos estudados e para todos os indicadores de precisão, as previsões mais precisas foram com o uso do algoritmo RF e do tratamento de seleção de variáveis RFA. Por tipo de algoritmo, o RFA também foi a técnica dominante para a geração das estimativas mais precisas, com exceção da CNN, em que a recomendação é a não aplicação de um tratamento de seleção de variáveis.

Cada categoria de produto apresentou comportamento diferente. Na série de camiseta de homem, os menores níveis de erros foram obtidos com a combinação do algoritmo XGBOOST e da técnica de seleção de variáveis RFA para a maior parte das métricas, com exceção do RMSLE e do MAPE, em que o melhor desempenho se deu com o LGBM em detrimento do XGBOOST, mas mantendo a RFA. Interessante destacar o predomínio da RFA como melhor técnica, mas, no caso do modelo LSTM, a recomendação foi o uso do tratamento de filtragem do p-valor.

No caso de blusa de mulher, o resultado foi muito similar ao obtido pela média das categorias de produto, em que a técnica de seleção de variáveis com melhor desempenho foi a RFA, sendo que a combinação que resultou no melhor nível de erro foi com o emprego do algoritmo LGBM e com a técnica RFA.

A série de camiseta de menino teve um comportamento similar à média das categorias de produtos com o tratamento RFA, destacando-se como técnica capaz de gerar os maiores níveis de precisão para os diversos algoritmos estudados, não sendo recomendada apenas para a CNN. Em termos de algoritmo, o melhor modelo foi obtido através da implementação do algoritmo RF com a RFA, sendo que somente para métrica RMSLE o algoritmo recomendado foi o LGBM.

Para a série de blusa de menina, constatou-se uma apuração distinta da observada nas demais categorias de produto, já que o modelo com maior nível de precisão, para a maior parte dos indicadores, foi gerado por meio do uso do algoritmo RF e do tratamento de seleção de variáveis SHAP. Para os demais algoritmos, o uso da técnica RFA gerou as estimativas mais precisas.

A Tabela 3.5 mostra o resultado do teste MCS, que visa comparar a capacidade preditiva de todos os modelos levados em consideração.

Tabela 3.5 – Resultado de teste MCS por algoritmo e por tipo de tratamento de seleção de variáveis

		CNN		LSTM		MLP		LGBM				RF				XGBOOST						
Categoria de Produto	Métrica	No	Pvalue	No	Pvalue	RFA	No	Pvalue	RFA	No	Pvalue	RFA	SHAP	No	Pvalue	RFA	SHAP	No	Pvalue	RFA	SHAP	
Camiseta de Homem	RMSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	RMSLE	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	MAE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	MAPE	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
	MAEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Blusa de Mulher	RMSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	MAE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	MAPE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	MAEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Camiseta de Menino	RMSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	MAE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	MAPE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	MAEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Blusa de Menina	RMSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	RMSLE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	MAE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	MAPE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	MAEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	RMSEP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Fonte: elaboração própria.

Considerando um nível de 95% de confiança para todas as séries de produtos, os modelos com as maiores capacidades preditivas foram os baseados em árvores

de decisão e com o emprego predominante da RFA. Porém, analisando cada série de produto, constatou-se algoritmos distintos como resultado do modelo mais apropriado.

Apesar do teste MCS selecionar um conjunto dos melhores modelos em termos de capacidade preditiva, na maior parte dos casos verificou-se que um único modelo foi eleito como o mais eficiente para cada métrica de precisão e para cada categoria de produto, com exceção do indicador MAPE para camiseta de homem, em que foram selecionados dois modelos.

Para a série de camiseta de homem e para as métricas de erro RMSE, MAEP e RMSEP, o modelo com maior capacidade preditiva foi o XGBOOST com a técnica RFA. Para a métrica MAPE, houve empate, em termos de capacidade preditiva, entre os algoritmos XGBOOST e LGBM com o tratamento de seleção de variáveis RFA. Para o RMSLE, o melhor modelo é resultado da combinação do LGBM com a RFA.

No caso da categoria blusa de mulher, de forma unânime para todas as métricas de erros contempladas neste estudo, os modelos com maior capacidade preditiva são oriundos do uso XGBOOST com a técnica de seleção de variáveis RFA.

Para a série camiseta de menino, os modelos com maior capacidade preditiva e para todos os indicadores de precisão considerados resultaram da combinação do algoritmo RF com o tratamento de seleção de variáveis RFA.

Resultado diferente da média foi encontrado na categoria blusa de menina, em que o modelo que gerou as estimativas mais precisas foi o algoritmo RF junto com a técnica de seleção de variáveis SHAP.

Ao comparar o resultado do teste MCS com o resultado obtido em termos dos níveis de precisão, nota-se uma sobreposição importante e que não necessariamente ocorre. Assim, constata-se que os modelos com maiores níveis de precisão em termos absolutos, tendo em vista as métricas de acurácia escolhidas, são em sua maior parte os mesmos com maior capacidade preditiva de acordo com o resultado do teste MCS para cada categoria de produto. Tal fato reforça que os resultados obtidos são robustos. Vale destacar que somente para a métrica MAPE de blusa de menina e para as métricas RMSE, RMSEP e RMSLE de blusa de mulher esta sobreposição não ocorreu.

Além da questão da melhora em termos de acurácia, que pôde ser verificada até o momento com o uso de técnicas de seleção de variáveis para modelos de aprendizagem de máquina, outro objetivo proposto pelo capítulo foi validar o impacto

em termos de custo computacional, o qual é mostrado na Tabela 3.6. A tabela quantifica o tempo gasto em segundos para treinar cada algoritmo com cada tipo de tratamento de seleção de variáveis para cada série de produto. Os valores negritados em vermelho representam os menores tempos de treinamento por série e os valores negritados em pretos, os menores tempos obtidos por algoritmo. Considerando as técnicas de seleção de variáveis aplicadas, constata-se que o algoritmo mais rápido na média foi o LGBM sem a aplicação de técnica de seleção de variáveis.

Tabela 3.6 – Análise do custo computacional – tempo em segundos

Tempo de Treinamento em segundos						
Algoritmo	Tratamento de Seleção de Variáveis	MÉDIA	HOMEM	MULHER	MENINO	MENINA
CNN	no	2.525	3.224	1.163	3.471	2.240
CNN	pvalue	2.848	3.582	1.162	3.425	3.225
LSTM	no	1.683	2.233	1.164	1.658	1.679
LSTM	pvalue	1.456	1.406	1.169	1.398	1.851
LSTM	rfa	3.150	9.083	1.185	1.128	1.204
MLP	no	1.436	1.713	714	2.153	1.166
MLP	pvalue	1.554	1.739	1.165	1.650	1.661
MLP	rfa	14.032	13.316	1.093	23.072	18.647
LGBM	no	432	425	631	365	308
LGBM	pvalue	483	415	715	370	433
LGBM	rfa	3.682	4.084	416	7.014	3.213
LGBM	shap	443	417	646	376	332
RF	no	1.325	1.561	557	1.825	1.356
RF	pvalue	1.434	1.823	556	1.752	1.606
RF	rfa	5.618	10.757	459	4.678	6.577
RF	shap	1.456	1.809	573	1.931	1.512
XGBOOST	no	907	961	467	1.197	1.005
XGBOOST	pvalue	985	1.113	624	1.215	987
XGBOOST	rfa	2.882	3.964	444	3.807	3.313
XGBOOST	shap	911	953	518	1.197	976

Fonte: elaboração própria.

Na média para todos os algoritmos, a não utilização de técnica de seleção de variáveis permitiu um treinamento mais rápido. O custo computacional do RFA é, cerca de, 3,5 vezes superior quando comparado a outros métodos de seleção de variáveis. Ademais, o aumento do tempo médio para treinamento dos modelos levando em conta a aplicação das técnicas p-valor e SHAP foi de, aproximadamente, 8%, ao comparar com o treinamento sem o uso de técnicas de seleção de variáveis.

Para todas as séries de produtos e para os algoritmos LSTM, MLP e RF, o não uso de técnicas de seleção de variáveis permitiu um treinamento mais rápido dos modelos. No caso do CNN, foi pela seleção p-valor para as séries de blusa de mulher,

camiseta de menino e blusa de menina, o que não houve com a série de homem, em que o não uso de um conjunto restrito de variáveis gerou um treinamento mais veloz.

A aplicação do método LGBM apresentou resultados dispares entre as séries de produtos presentes neste estudo, pois para as séries infantis o não uso de técnicas de seleção de variáveis garantiu um custo computacional menor. Por outro lado, para a série de camiseta de homem, foi conquistado com o uso do método p-valor e para a série de blusa de mulher, com o uso do RFA, um resultado inesperado. Por último, para o algoritmo XGBOOST, o treinamento mais rápido para a maior parte das séries foi feito com o emprego da técnica SHAP.

Independentemente do tipo de técnica de seleção de variáveis aplicado, o algoritmo LGBM foi o menos custoso computacionalmente, demandando somente, cerca de, 50% do tempo, quando comparado, em média, com os demais algoritmos considerados. Inicialmente era esperado que o uso de técnicas de seleção de variáveis reduzisse o custo computacional, o que foi mensurado, nesta pesquisa, por meio da contagem do tempo em segundos de treinamento dos modelos. No entanto, conforme verificado, na maior parte dos casos, a implementação de técnicas de seleção de variáveis aumentou o tempo de treinamento, de forma marginal, no caso das técnicas de seleção de variáveis de filtragem pelo valor do p-valor e pelo SHAP, mas, significativamente, pelo uso da RFA.

Com objetivo de examinar quais variáveis foram consideradas mais vezes, tendo em vista o corte estabelecido das 30 variáveis mais importantes, foi examinada a frequência das variáveis para cada uma das séries de produtos. Além disso, também foi feita uma verificação por tipo de variável com o intuito de constatar qual tipo de variável foi mais relevante para os modelos de previsão produzidos.

Tabela 3.7 – Análise de frequência das variáveis por série de produto

HOMEM			MULHER			MENINO			MENINA		
Variáveis	Tipo	Frequência	Variáveis	Tipo	Frequência	Variáveis	Tipo	Frequência	Variáveis	Tipo	Frequência
SEMANA	Criada	100,00%	MEDIA_MOVEL_QTD_3_SEMANAS	Criada	78,57%	SEMANA	Criada	100,00%	QTD_ATUAL_52	Criada	92,86%
QTD_ATUAL_32	Criada	85,71%	QTD_ATUAL_5	Criada	78,57%	QTD_ATUAL_46	Criada	78,57%	QTD_ATUAL_3	Criada	78,57%
QTD_ATUAL_5	Criada	85,71%	QTD_ATUAL_52	Criada	78,57%	DIFF_QTD_ATUAL_1	Criada	71,43%	STD_MOVEL_QTD_2_SEMANAS	Criada	78,57%
MEDIA_MOVEL_QTD_3_SEMANAS	Criada	78,57%	TMAX	Exógena	78,57%	MEDIA_MOVEL_QTD_3_SEMANAS	Criada	71,43%	DU	Criada	71,43%
QTD_ATUAL_1	Criada	78,57%	QTD_ATUAL_16	Criada	71,43%	MEDIA_MOVEL_QTD_4_SEMANAS	Criada	71,43%	MEDIA_MOVEL_QTD_4_SEMANAS	Criada	71,43%
QTD_ATUAL_2	Criada	78,57%	QTD_ATUAL_19	Criada	71,43%	QTD_ATUAL_30	Criada	71,43%	MEDIA_MOVEL_QTD_7_SEMANAS	Criada	71,43%
QTD_ATUAL_12	Criada	78,57%	QTD_ATUAL_22	Criada	71,43%	QTD_ATUAL_4	Criada	71,43%	QTD_ATUAL_27	Criada	71,43%
QTD_ATUAL_22	Criada	78,57%	QTD_ATUAL_39	Criada	71,43%	QTD_ATUAL_37	Criada	71,43%	QTD_ATUAL_28	Criada	71,43%
QTD_ATUAL_4	Criada	78,57%	QTD_ATUAL_49	Criada	71,43%	QTD_ATUAL_40	Criada	71,43%	QTD_ATUAL_37	Criada	71,43%
QTD_ATUAL_48	Criada	78,57%	STD_MOVEL_QTD_2_SEMANAS	Criada	71,43%	QTD_ATUAL_5	Criada	71,43%	QTD_ATUAL_41	Criada	71,43%
QTD_ATUAL_49	Criada	78,57%	DIFF_QTD_ATUAL_1	Criada	64,29%	QTD_ATUAL_41	Criada	71,43%	QTD_ATUAL_46	Criada	71,43%
TMAX	Exógena	78,57%	MEDIA_MOVEL_QTD_4_SEMANAS	Criada	64,29%	QTD_ATUAL_49	Criada	71,43%	MEDIA_MOVEL_QTD_2_SEMANAS	Criada	64,29%
MEDIA_MOVEL_QTD_4_SEMANAS	Criada	71,43%	QTD_ATUAL_1	Criada	64,29%	QTD_ATUAL_6	Criada	71,43%	MEDIA_MOVEL_QTD_9_SEMANAS	Criada	64,29%
MEDIA_MOVEL_QTD_5_SEMANAS	Criada	71,43%	QTD_ATUAL_20	Criada	64,29%	QTD_ATUAL_51	Criada	71,43%	QTD_ATUAL_2	Criada	64,29%
QTD_ATUAL_16	Criada	71,43%	QTD_ATUAL_3	Criada	64,29%	DNATAL	Exógena	64,29%	QTD_ATUAL_11	Criada	64,29%
QTD_ATUAL_20	Criada	71,43%	QTD_ATUAL_33	Criada	64,29%	MEDIA_MOVEL_QTD_5_SEMANAS	Criada	64,29%	QTD_ATUAL_17	Criada	64,29%
QTD_ATUAL_3	Criada	71,43%	QTD_ATUAL_44	Criada	64,29%	QTD_ATUAL_3	Criada	64,29%	QTD_ATUAL_22	Criada	64,29%
QTD_ATUAL_37	Criada	71,43%	QTD_ATUAL_45	Criada	64,29%	QTD_ATUAL_22	Criada	64,29%	QTD_ATUAL_4	Criada	64,29%
QTD_ATUAL_39	Criada	71,43%	QTD_ATUAL_46	Criada	64,29%	QTD_ATUAL_27	Criada	64,29%	QTD_ATUAL_31	Criada	64,29%
QTD_ATUAL_8	Criada	71,43%	STD_MOVEL_QTD_15_SEMANAS	Criada	64,29%	QTD_ATUAL_28	Criada	64,29%	QTD_ATUAL_35	Criada	64,29%
QTD_ATUAL_51	Criada	71,43%	DNATAL	Exógena	57,14%	QTD_ATUAL_31	Criada	64,29%	QTD_ATUAL_42	Criada	64,29%
STD_MOVEL_QTD_2_SEMANAS	Criada	71,43%	DU	Criada	57,14%	QTD_ATUAL_7	Criada	64,29%	QTD_ATUAL_45	Criada	64,29%
DIFF_QTD_ATUAL_1	Criada	64,29%	QTD_ATUAL_12	Criada	57,14%	STD_MOVEL_QTD_3_SEMANAS	Criada	64,29%	QTD_ATUAL_49	Criada	64,29%
QTD_ATUAL_26	Criada	64,29%	QTD_ATUAL_14	Criada	57,14%	DIFF_QTD_ATUAL_PERC_1	Criada	57,14%	DCRIANCAS	Exógena	57,14%
QTD_ATUAL_30	Criada	64,29%	QTD_ATUAL_21	Criada	57,14%	MSAD_REAL	Exógena	57,14%	DIFF_QTD_ATUAL_1	Criada	57,14%
QTD_ATUAL_31	Criada	64,29%	QTD_ATUAL_28	Criada	57,14%	QTD_ATUAL_12	Criada	57,14%	PRECO_MEDIO	Exógena	57,14%
QTD_ATUAL_35	Criada	64,29%	QTD_ATUAL_30	Criada	57,14%	QTD_ATUAL_19	Criada	57,14%	QTD_ATUAL_14	Criada	57,14%
QTD_ATUAL_36	Criada	64,29%	QTD_ATUAL_4	Criada	57,14%	QTD_ATUAL_25	Criada	57,14%	QTD_ATUAL_16	Criada	57,14%
QTD_ATUAL_40	Criada	64,29%	QTD_ATUAL_40	Criada	57,14%	QTD_ATUAL_29	Criada	57,14%	QTD_ATUAL_18	Criada	57,14%
QTD_ATUAL_45	Criada	64,29%	QTD_ATUAL_41	Criada	57,14%	QTD_ATUAL_32	Criada	57,14%	QTD_ATUAL_19	Criada	57,14%

Fonte: elaboração própria.

A Tabela 3.7 mostra que, para todas as categorias de produtos estudadas, as variáveis mais presentes foram as criadas, sendo elencadas duas variáveis exógenas por categoria de produto, no máximo. Entre as variáveis exógenas mais frequentes, foram observadas as seguintes: temperatura máxima, massa salarial real disponível, binária de Natal e binária do Dia das Crianças. A baixa frequência, cerca de 4%, reflete uma maior importância das variáveis criadas. Especificamente as defasagens e as médias móveis são as mais usadas para construção das estimativas.

A Tabela 3.8 mostra as variáveis que mais se repetem entre as diversas séries de produtos. Entre as 30 variáveis selecionadas, seis aparecem para todos os tipos de produtos, as quais são: a primeira diferença defasada, a média móvel das quatro últimas semanas, a defasagem 3, a defasagem 4, a defasagem 22 e a defasagem 49.

Tabela 3.8 – Ranking das 30 variáveis mais frequentes para todas as séries de produtos

Variavel	Tipo	Homem	Mulher	Menino	Menina	Total
DIFF_QTD_ATUAL_1	Criada	1	1	1	1	4
MEDIA_MOVEL_QTD_4_SEMANAS	Criada	1	1	1	1	4
QTD_ATUAL_3	Criada	1	1	1	1	4
QTD_ATUAL_22	Criada	1	1	1	1	4
QTD_ATUAL_4	Criada	1	1	1	1	4
QTD_ATUAL_49	Criada	1	1	1	1	4
MEDIA_MOVEL_QTD_3_SEMANAS	Criada	1	1	1	0	3
QTD_ATUAL_12	Criada	1	1	1	0	3
QTD_ATUAL_16	Criada	1	1	0	1	3
QTD_ATUAL_19	Criada	0	1	1	1	3
QTD_ATUAL_28	Criada	0	1	1	1	3
QTD_ATUAL_30	Criada	1	1	1	0	3
QTD_ATUAL_31	Criada	1	0	1	1	3
QTD_ATUAL_37	Criada	1	0	1	1	3
QTD_ATUAL_40	Criada	1	1	1	0	3
QTD_ATUAL_5	Criada	1	1	1	0	3
QTD_ATUAL_41	Criada	0	1	1	1	3
QTD_ATUAL_45	Criada	1	1	0	1	3
QTD_ATUAL_46	Criada	0	1	1	1	3
STD_MOVEL_QTD_2_SEMANAS	Criada	1	1	0	1	3
DNATAL	Exógena	0	1	1	0	2
DU	Criada	0	1	0	1	2
MEDIA_MOVEL_QTD_5_SEMANAS	Criada	1	0	1	0	2
QTD_ATUAL_1	Criada	1	1	0	0	2
QTD_ATUAL_2	Criada	1	0	0	1	2
QTD_ATUAL_14	Criada	0	1	0	1	2
QTD_ATUAL_20	Criada	1	1	0	0	2
QTD_ATUAL_27	Criada	0	0	1	1	2
QTD_ATUAL_32	Criada	1	0	1	0	2
QTD_ATUAL_35	Criada	1	0	0	1	2
QTD_ATUAL_39	Criada	1	1	0	0	2
QTD_ATUAL_6	Criada	1	0	1	0	2
QTD_ATUAL_51	Criada	1	0	1	0	2
QTD_ATUAL_52	Criada	0	1	0	1	2
SEMANA	Criada	1	0	1	0	2
TMAX	Exógena	1	1	0	0	2

Fonte: elaboração própria.

4. Conclusões

Este capítulo teve como principal problema de estudo a avaliação da relevância do uso de técnicas de seleção de variáveis para a previsão de vendas em quantidade para as quatro categorias de produtos mais importantes em termos de faturamento para a rede Lojão do Brás. Ressalta-se, mais uma vez, a importância do processo de previsão de vendas para o setor de moda, dadas suas especificidades. Ademais, o aumento exponencial de informação disponível, por conta da internet e da capacidade computacional, torna o processo de seleção de variáveis fundamental.

Como já colocado, o emprego de técnicas de seleção de variáveis tem vários benefícios, pois reduz o custo computacional, melhora a acurácia dos modelos produzidos, torna os modelos mais simples e fáceis de serem explicados, minimiza o problema da maldição da dimensionalidade e, assim, mitiga a chance de sobreajuste.

Com base na aplicação de três técnicas de seleção de variáveis – filtragem do p-valor das variáveis, RFA e SHAP –, constatou-se a importância do uso de tais

técnicas frente à possibilidade do uso do conjunto completo de 128 variáveis disponíveis, entre variáveis exógenas e variáveis criadas. A importância neste trabalho foi medida em termos de melhora de acurácia e redução de custo computacional. Tais técnicas de seleção de variáveis foram aplicadas sobre seis métodos de previsão, os quais foram: LGBM, RF, XGBOOST, CNN, MLP e LSTM.

Apoiado nos resultados obtidos, verificou-se que, em média, o uso das técnicas de seleção de variáveis para as quatro categorias de produtos estudadas foi capaz de produzir estimativas com níveis menores de erros, com base nas métricas RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP, em torno de 13%.

Em termos de tipo de técnicas, as melhores foram as *wrappers* RFA e SHAP, predominantemente e em média diante dos algoritmos de previsão propostos.

A análise por tipo de algoritmo de previsão, considerando tanto os modelos baseados em árvores de decisão, assim como os modelos baseados em redes neurais, permitiu verificar que a aplicação de técnicas de seleção de variáveis gerou menores níveis de erro de previsão para as duas categorias de algoritmos.

Na avaliação por tipo de algoritmo, de tratamento empregado e por série de produto, constatou-se que, majoritariamente, os modelos mais precisos foram aqueles que combinam um algoritmo de árvore de decisão, destacando-se o RF e o XGBOOST, juntamente com, na maior parte dos casos, o emprego da técnica de seleção de variáveis RFA. Uma ressalva deve ser feita para a série de blusa de menina, na qual a aplicação da técnica de seleção de variáveis SHAP garantiu melhores resultados.

O resultado do teste MCS, com nível de confiança de 95%, que visa verificar o conjunto de modelos com maior capacidade preditiva, confirmou a maior parte dos resultados da análise dos valores absolutos das métricas de precisão, reiterando que os modelos com maiores capacidades de predição para as séries estudadas foram os métodos de previsão baseados em árvores de decisão com a aplicação da técnica de seleção de variáveis RFA para todas as séries, com exceção da série de blusa de mulher, em que a técnica de seleção de variáveis com maior capacidade preditiva foi a SHAP.

De forma detalhada, no MCS, para as séries de produto camiseta de homem e blusa de mulher, os melhores resultados foram obtidos com o uso do algoritmo XGBOOST e da técnica de seleção de variáveis RFA. Para a série de camiseta de

menino, os melhores resultados foram obtidos com a combinação do método RF com a técnica RFA. Por último, para a série de blusa de menina, o melhor desempenho foi atribuído ao uso do algoritmo RF com a técnica de seleção de variáveis SHAP.

No que diz respeito à avaliação do custo computacional, medido pelo tempo de treinamento dos modelos de previsão em segundos, as técnicas de seleção de variáveis, sobretudo a RFA, ao invés de reduzir o custo computacional, acabou por aumentá-lo significativamente, o que ocorreu para todas as séries de produtos. A RFA apresentou um tempo de treinamento, aproximadamente, 3,5 vezes maior em relação às demais técnicas de seleção, sendo que essas foram cerca de 8%, em média, mais demoradas no processo de realização do treinamento dos modelos, quando comparadas com os casos de não aplicação de técnicas de seleção de variáveis.

Vale destacar os resultados obtidos na análise de frequência das variáveis mais presentes nos modelos em que foram empregadas as técnicas de seleção de variáveis. Nesse sentido, verificou-se para todas as séries uma alta frequência das variáveis criadas em detrimento das variáveis exógenas, considerando o subconjunto de 30 variáveis selecionadas. Em termos de variáveis exógenas, as mais frequentes são: a temperatura máxima e a binária de Natal. Em relação às variáveis criadas, seis variáveis apareceram na maior parte dos modelos para todas as séries de produtos estudadas, as quais são: a primeira diferença defasada, a média móvel das quatro últimas semanas, a defasagem 3, a defasagem 4, a defasagem 22 e a defasagem 49.

Diante disso, fica evidente que um conjunto menor de variáveis, composto sobretudo por variáveis criadas, em que as variáveis exógenas poderiam ser removidas dos modelos de previsão, é capaz de gerar modelos de previsão, baseados em aprendizagem de máquina, com maiores níveis de precisão, porém com maior custo computacional. Conclui-se, então, que é recomendado o uso de técnicas de seleção de variáveis para o processo de previsão de vendas no varejo de moda.

REFERÊNCIAS

CAI, Jie *et al.* Feature selection in machine learning: a new perspective. **Neurocomputing**, v. 300, p. 70-79, 26 jul. 2018.

MUNSON, Arthur; CARUANA, Rich. On feature selection, bias-variance, and bagging. In: BUNTINE, Wray et al. (ed.). **Machine learning and knowledge discovery in database**. Lecture Notes in Computer Science, v. 5782, p. 144-159. Berlim: Springer, 2009.

PAJARES, Rubén García; BENÍTEZ, Jose M.; PALMERO, Gregorio Sáinz. Feature selection for time series forecasting: a case study. **2008 Eighth International Conference on Hybrid Intelligent Systems**, IEEE, Barcelona, 10 a 12 set. 2008.

SBVC. **O papel do varejo na economia**. 2020. Disponível em: <http://www.sbvc.com.br/wpcontent/uploads/2015/04/O-Papel-do-Varejonaeconomia_apresenta%C3%A7%C3%A3ooficial1.pdf> Acesso em: 10 jan. 2020.

VERLEYSEN, Michael; FRANÇOIS, Damien. The curse of dimensionality in data mining and time series prediction. **Lecture Notes in Computer Science**, v. 3512, p. 758-770, 2005.

IV. Métodos de combinação de previsão: uma análise comparativa para o varejo de moda

RESUMO

O varejo de moda é um dos setores mais relevantes para o varejo brasileiro e, nos últimos anos, tem sido impactado pelo aumento da competitividade, pela maior inserção em cadeias globais e pelo aumento do nível de incerteza. As peculiaridades e a grande extensão da cadeia de produção do setor do varejo de moda, tornam o processo de previsão de vendas uma questão mais complexa que deve ser tratada com a devida importância pelas empresas que atuam no setor. A geração de previsões com maior nível de acurácia é capaz de garantir uma cadeia de abastecimento mais eficiente, produzir maiores resultados financeiros e aumentar o nível de atendimento da demanda dos consumidores para as organizações.

Este trabalho tem dois objetivos fundamentais: verificar o mais preciso e o melhor modelo de previsão de vendas para o varejo de moda, o que foi feito através de uma análise AHP, em que multicritérios são contemplados. Para ambos os objetivos, foi utilizado um estudo de caso, em que foram analisados os resultados obtidos para 79 modelos de previsão, com foco maior nos modelos de combinação de previsão para a geração das estimativas de vendas de quatro categorias de produtos.

Com base nos cálculos das seis métricas de acurácia consideradas, constatou-se por meio da análise de frequência que o algoritmo XGBOOST, modelo *ensemble* do tipo *boosting*, acompanhado do tratamento sazonal por ajuste de curva e do tratamento de seleção de variáveis RFA, foi o mais preciso em 37,5% dos casos.

Em média, o modelo de previsão de vendas mais preciso superou em 82,9% o pior modelo de previsão considerado e foi 48,7% mais assertivo que o modelo médio, considerando as seis métricas e as quatro categorias de produtos contempladas.

Através da análise AHP, o modelo de previsão *Naive* ajustado foi considerado o melhor entre os demais modelos de previsão de vendas no varejo de moda e, de forma unânime, entre as quatro categorias de produtos estudadas.

Palavras-chave: Previsão de vendas. Métodos de previsão por combinação. Métodos de aprendizagem de máquina. Análise AHP.

ABSTRACT

Fashion retail is one of the most relevant sectors for Brazilian retail and, in recent years, it has been impacted by increased competitiveness, greater involvement in global chains and the increased level of uncertainty. The peculiarities and the large extension of the fashion retail sector's production chain make the sales forecasting process a more complex issue that must be treated with due importance by companies operating in the sector. The generation of forecasts with a higher level of accuracy is capable of guaranteeing a more efficient supply chain, producing greater financial results and increasing the level of meeting consumer demand for organizations.

This work has two fundamental objectives: to verify the most accurate and the best sales forecast model for fashion retail, which was done through an AHP analysis, in which multi-criteria are considered. For both purposes, a case study was used, in which the results obtained for 79 forecasting models were analyzed, with a greater focus on forecasting combination models for the generation of sales estimates for four product categories.

Based on the calculations of the six accuracy metrics considered, it was found through frequency analysis that the XGBOOST algorithm, a boosting ensemble model, accompanied by the seasonal treatment by curve fitting and the RFA variable selection treatment, was the more accurate in 37.5% of cases.

On average, the most accurate sales forecast model surpassed the worst forecast model considered by 82.9% and was 48.7% more assertive than the average model, considering the six metrics and four product categories covered.

Through the AHP analysis, the adjusted Naive forecast model was considered the best among the other fashion retail sales forecast models and, unanimously, among the four studied product categories.

Keywords: Sales forecasting. Ensemble learning. Machine learning methods. AHP analysis.

1. Introdução

“The goal of forecasting is not to predict the future but to tell you what you need to know to take meaningful action in the present.”

Paul Saffo

A indústria de moda, que engloba o setor de varejo de moda e outras indústrias, passou a lidar com um ambiente cada vez mais competitivo, com um maior nível de incertezas e clientes mais exigentes. A fim de se manterem competitivas, as empresas do setor devem melhorar a gestão da sua cadeia de abastecimento, o que leva ao investimento em sistema de informação, competências logísticas e, em especial, sistemas de previsão com alta acurácia (THOMASSEY, 2010).

Algumas limitações do setor tornam o processo de previsão de vendas complexo, como a alta volatilidade da demanda, um conjunto amplo de variáveis exógenas afeta o comportamento das vendas, o aspecto sazonal é crítico e deve ser considerado, a previsão muitas vezes deve ser realizada no nível de item de cor e tamanho, dificultando a realização pela falta de histórico de dados e, por fim, a indústria de moda é efêmera, gerando um volume crescente de novos produtos de ciclo curto e, conseqüentemente, com baixa quantidade de dados para realimentar os sistemas de previsão.

Característica extremamente importante da indústria da moda é que ela possui uma cadeia de abastecimento longa, composta por diversos fornecedores, sendo muitos deles globais, o que intensifica o efeito chicote, quando ele ocorre. Esse fato reforça ainda mais a necessidade da implantação de sistemas de previsão precisos.

Em termos de relevância, o comércio varejista de vestuário e complementos, equivalente ao nomeado como varejo de moda, teve um faturamento de R\$ 124,5 bilhões, em 2018, o que representa cerca de 6,8% do total apurado pelo comércio varejista, de acordo com a Pesquisa Anual do Comércio (PAC), do IBGE.

Diante do contexto e do setor apresentados, este trabalho possui dois objetivos. Primeiro, demonstrar qual o método de previsão de vendas com maior acurácia para o setor de varejo de moda, tendo em vista a diversidade disponível que se multiplicou exponencialmente com o desenvolvimento computacional. Os modelos *ensembles*, que podem ser sumariamente definidos como combinações de previsões geradas por

modelos de previsão individuais, são o foco deste estudo, os quais são comparados com outros modelos de previsão.

Em sua pesquisa *survey*, Liu *et al.* (2013) destacam que os métodos de previsão de vendas para o varejo de moda podem ser agrupados em métodos estatísticos, métodos de inteligência computacional e métodos híbridos. Perante essa abordagem, afirma-se que este trabalho considera métodos em todas as categorias, sendo que a maior concentração está nos métodos de inteligência computacional e nos híbridos. No caso dos métodos *ensembles*, os *ensembles bagging* e *boosting* são considerados como de inteligência computacional e os demais, como híbridos.

No total são avaliados 79 métodos de previsão para as quatro principais categorias de produtos da rede de varejo de moda Lojão do Brás. Os modelos de previsão de vendas selecionados contemplam variações em termos de tratamento sazonal aplicado, de tratamento de seleção de variáveis utilizados, em termos de algoritmo e em termos de tipo de *ensemble*.

O uso de métodos de combinação de previsões, tanto para tarefas de regressão, como é o caso deste trabalho, como para tarefas de classificação, tem seu início na década de 1960, a partir de trabalhos de Nilsson (1965), Ablow e Kaylor (1965) e Bates e Granger (1969). Outra contribuição relevante é o artigo de Granger e Ramanathan (1984). A partir dos anos 1990, houve crescimento da pesquisa na área, fato diretamente ligado ao desenvolvimento computacional, e, nessa década, foram produzidos os trabalhos iniciais para os principais tipos de métodos combinados. Para o tipo *bagging*, se destaca o estudo de Breiman (1996a); para o *boosting*, os trabalhos de Schapire (1990) e Freund e Schapire (1997); e para o tipo *stacking*, os trabalhos de Wolpert (1992) e Breiman (1996b).

Em termos práticos, em uma série de competições do Kaggle, geralmente o modelo vencedor aplicou algum modelo *ensemble* para produzir as suas previsões.

Ao se falar de modelos *ensembles*, salienta-se o *puzzle* da combinação das previsões, que coloca que, ao se levar em conta um número fixo de previsões ($N < \infty$), combiná-los usando pesos iguais ($1 / N$) gera um resultado superior em termos de precisão por meio do MSE, do que ao usar “pesos ideais” construídos para superar qualquer outra combinação, segundo Issler e Lima (2008) e Claeskens *et al.* (2014).

Outro aspecto relevante em termos de análise de acurácia, é que nem sempre é simples superar o modelo de previsão *Naive*, que considera o último valor observado

como valor previsto. Tal fato é bastante comum em previsões de taxas de juros da economia e para taxas de câmbio, em que o modelo *Random Walk* só é batido em previsões de longo prazo, conforme observado nos trabalhos de Meese e Rogoff (1983), Yoon (1998) e Butter e Jansen (2013).

Para apurar a acurácia dos métodos de previsão escolhidos, foram calculadas seis métricas de precisão: RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP. Tais indicadores apresentam o nível de erro das previsões obtidas, sendo que quanto menor o valor, melhores são as estimativas geradas. MAEP e RMSEP são variações de MAE e RMSE, respectivamente. Conceitualmente, MAEP e RMSEP são medidas ponderadas, respectivamente, de MAE e RMSE, que penalizam mais as previsões menores que o dado real, já que a falta de estoque, desencadeada por uma previsão a menor, é um problema mais impactante financeiramente para a empresa do que a previsão a maior, fato acentuado em um contexto econômico de baixas taxas de juros.

O teste MCS também foi realizado para comprovar, formalmente, a capacidade preditiva de todos os modelos de previsão contemplados neste estudo.

Os resultados das métricas de precisão para as quatro categorias de produtos avaliadas permitiu constatar que foram encontrados resultados distintos por métrica e por categoria de produto, no que diz respeito ao modelo de previsão com menor erro. Apesar disto, mediante a análise da frequência dos resultados obtidos junto a cada métrica de acurácia e a cada categoria de produto, constata-se que, para o setor de moda, o modelo de previsão recomendado em termos de maior assertividade seria um XGBOOST, que é um *ensemble* do tipo *boosting*, juntamente com a aplicação do tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA.

O modelo *Naive* apresentou desempenho superior em relação a vários outros modelos de previsão, o que vai ao encontro de diversos estudos da academia, tanto na área econômica como na financeira, sobre a dificuldade de superar esse modelo, podendo ser considerado efetivamente como uma referência representativa.

Em relação ao resultado do teste MCS, através da verificação por frequência, mais uma vez entre as métricas de precisão e para as quatro categorias de produto, constatou-se que o modelo com maior capacidade preditiva para o setor de moda foi o RF, que é um *ensemble* do tipo *bagging*, sem tratamento sazonal e com tratamento de seleção de variáveis por RFA. Comparando o resultado do teste MCS com o resultado obtido em termos dos níveis de precisão absolutos, verifica-se uma

divergência entre eles, em termos do algoritmo, do tratamento sazonal e do tipo de método de combinação de previsão.

O segundo objetivo deste trabalho é realizar uma análise AHP, a qual contempla múltiplos critérios para ajudar no processo de tomada de decisões dentro das organizações. Elaborada por Saaty, na década de 1980, o objetivo por trás da implementação desta análise é realizar uma recomendação final de método de previsão de vendas para o varejo de moda, o que não poderia ser feito somente considerando o aspecto da acurácia, já que, na realidade das organizações, vários critérios são levados em conta para estabelecer qual melhor método de previsão. Dessa forma, foram selecionados os mesmos critérios e os mesmos pesos atribuídos na avaliação realizada no artigo de Ren, Chan e Ram (2017), que também foi aplicado para o setor de moda. Especificamente, os critérios são acurácia, velocidade, requerimento de dados, estabilidade e facilidade de uso.

Com base na metodologia AHP, o melhor modelo de previsão foi o *Naive* ajustado, de forma unânime entre as quatro categorias estudadas, em que basicamente é considerada a previsão como sendo o valor observado na última semana, sendo que, no caso das semanas do mês de dezembro, são utilizadas as previsões das semanas desse mês pertencentes ao ano anterior.

Desconsiderando os modelos *Naives*, a recomendação de modelo final para o setor de varejo de moda para a previsão de vendas seria o modelo *ensemble* de média simples, o qual contempla todos os modelos de previsão usados neste trabalho.

A estrutura deste capítulo é composta de cinco seções, sendo essa a primeira delas. A segunda seção apresenta os métodos de combinação de previsão. A terceira mostra conceitualmente a análise AHP, enquanto a quarta apresenta os resultados obtidos de forma completa para cada um dos objetivos deste capítulo. Na quinta e última seção são demonstradas as principais conclusões alcançadas.

2. Métodos de combinação de previsão

Nesta parte do artigo são mostrados os métodos de combinação de previsão de vendas usados para a realização das estimativas no caso de estudo deste trabalho, juntamente com a metodologia empregada. Importante destacar que os métodos de

combinação de previsão são apenas um dos tipos de técnicas de previsão aplicados neste trabalho.

O foco deste estudo é o uso de métodos *ensembles*, que conceitualmente realizam a combinação de previsões, as quais podem ser geradas por diversos algoritmos, assim como as combinações em si podem ser feitas de diversas maneiras. Deve-se ter mente que a principal suposição para a aplicação dos métodos *ensemble* em detrimento do uso dos métodos de forma individual está no fato de que a combinação de várias previsões leva a um resultado superior em termos de acurácia, o que somente é obtido pela inclusão de modelos diversos entre si.

Os métodos *ensembles* são um dos tipos de máquina de comitê. Haykin (2001) define que uma máquina de comitê é formada pela combinação de algoritmos especialistas, que funde o conhecimento adquirido por cada especialista, o que gera uma solução melhor do que aquela gerada de forma isolada por cada especialista. Os métodos *ensemble* empregam a combinação de um conjunto de modelos ou classificadores redundantes, sendo que cada modelo determina uma solução completa para a tarefa em questão, segundo Hansen e Salamon (1990). O outro tipo de máquina de comitê são os sistemas modulares, em que é feita a decomposição do problema para que, a partir da combinação da solução obtida em cada subtarefa, seja encontrada uma solução mais eficiente para o problema.

O uso de máquinas de comitê e, respectivamente, de métodos *ensemble*, que são o foco deste capítulo, tem como principal objetivo evitar a perda de informação, que ocorre quando é feita a opção por um determinado estimador em detrimento de outros. Adicionalmente, os estimadores combinados, segundo Dietterich (2002), possuem como características maior poder de generalização e maior robustez. Zhang e Ma (2012) observam o aumento de robustez através do uso de *ensembles* pela redução da variância das previsões geradas, o que acaba levando a um aumento da acurácia das previsões produzidas. Em termos de desvantagens, pode-se afirmar que os modelos *ensembles* são mais complexos de interpretar, apresentam maior custo computacional por contemplarem diversos modelos e a atividade de seleção dos modelos que farão parte do modelo combinado não é uma tarefa simples, ou seja, a geração da diversidade pode ser vista, praticamente, como uma tarefa artesanal.

O problema que este capítulo busca verificar é o desempenho da aplicação dos métodos *ensembles* em comparação aos métodos individuais para a previsão de

vendas no varejo de moda em termos de acurácia. Outro objetivo é especificar o método de previsão mais adequado para o setor, o que é feito através de uma análise AHP, em que são considerados vários aspectos, e não somente a precisão. Para essa análise, o trabalho de Ren, Chan e Ram (2017) é utilizado como base.

Para verificação do desempenho dos métodos *ensembles* em termos de precisão, é feita a previsão semanal em quantidade para as quatro categorias de produtos mais relevantes para o faturamento da rede Lojão do Brás, as quais são: camiseta de homem, blusa de mulher, camiseta de menino e blusa de menina. Todas essas séries foram apresentadas e descritas na primeira etapa deste trabalho. As previsões são geradas no curtíssimo prazo, sempre para a próxima semana.

O conjunto de dados utilizado compreende os anos de 2007 a 2019. Para a realização do treinamento dos modelos, foram usados os dados de 2007 a 2017 e, para a realização dos testes, os dados compreendidos entre 2018 e 2019.

São consideradas 128 variáveis, sendo a maioria criada por *feature engineering*, o que significa a aplicação de alguma função matemática ou de data. Foram criadas as seguintes variáveis: 52 defasagens, considera a quantidade vendida em um período anterior; sete diferenças defasadas, diferença entre o valor observado entre dois períodos no tempo; sete diferenças percentual defasadas, última variável apresentada de forma percentual; 10 médias móveis, considera de dois a 12 períodos; 14 desvios padrões móveis, leva em conta o intervalo de duas a 16 observações no seu cálculo; a semana do ano; o mês e o número de dias úteis existentes na semana.

Além disso, foram acrescentadas 19 variáveis exógenas com o intuito de levar em conta outros aspectos e, assim, gerar previsões mais precisas. Desse modo, foram utilizadas variáveis climáticas, representadas por temperatura máxima, temperatura mínima e volume de precipitação; e variáveis macroeconômicas, como índice de confiança do consumidor, taxa de desemprego na região metropolitana de São Paulo, massa salarial ampliada disponível real, crédito para pessoa física e índice em volume da pesquisa mensal do comércio para categoria de vestuário em São Paulo.

Também foram utilizadas 11 variáveis de eventos, cujo objetivo é absorver o impacto de datas comuns no calendário varejista, como Dia das Mães, Dia dos Namorados, Dia dos Pais, Dia das Crianças, *Black Friday* e Natal, além de eventos criados por ação governamental, como isenção de IPI sobre móveis, linha branca e liberação do saque do FGTS, e ainda a greve dos caminhoneiros.

Antes de especificar os métodos *ensembles* usados, é essencial frisar que foram feitas variações de tais métodos conforme o tratamento sazonal aplicado, sendo os mesmos analisados no segundo capítulo desta Tese: o tratamento da sazonalidade por diferenciação e o tratamento da sazonalidade por ajuste de uma curva polinomial. Ademais, também foram contemplados, junto aos métodos *ensembles* aplicados, os tratamentos de seleção de variáveis avaliados no terceiro capítulo deste trabalho, os quais são: seleção por p-valor, seleção por RFA e seleção por SHAP.

Na teoria de aprendizagem de métodos de previsão por combinação, modelos que podem compor um modelo mais complexo são chamados de modelos bases ou aprendizes fracos. O objetivo é unir aprendizes fracos para gerar um aprendiz forte, o que leva a uma performance superior por meio da redução do viés e/ou da variância. A combinação dos modelos pode ser feita de forma homogênea, em que é usado o mesmo tipo de algoritmo, ou de forma heterogênea, com algoritmos distintos.

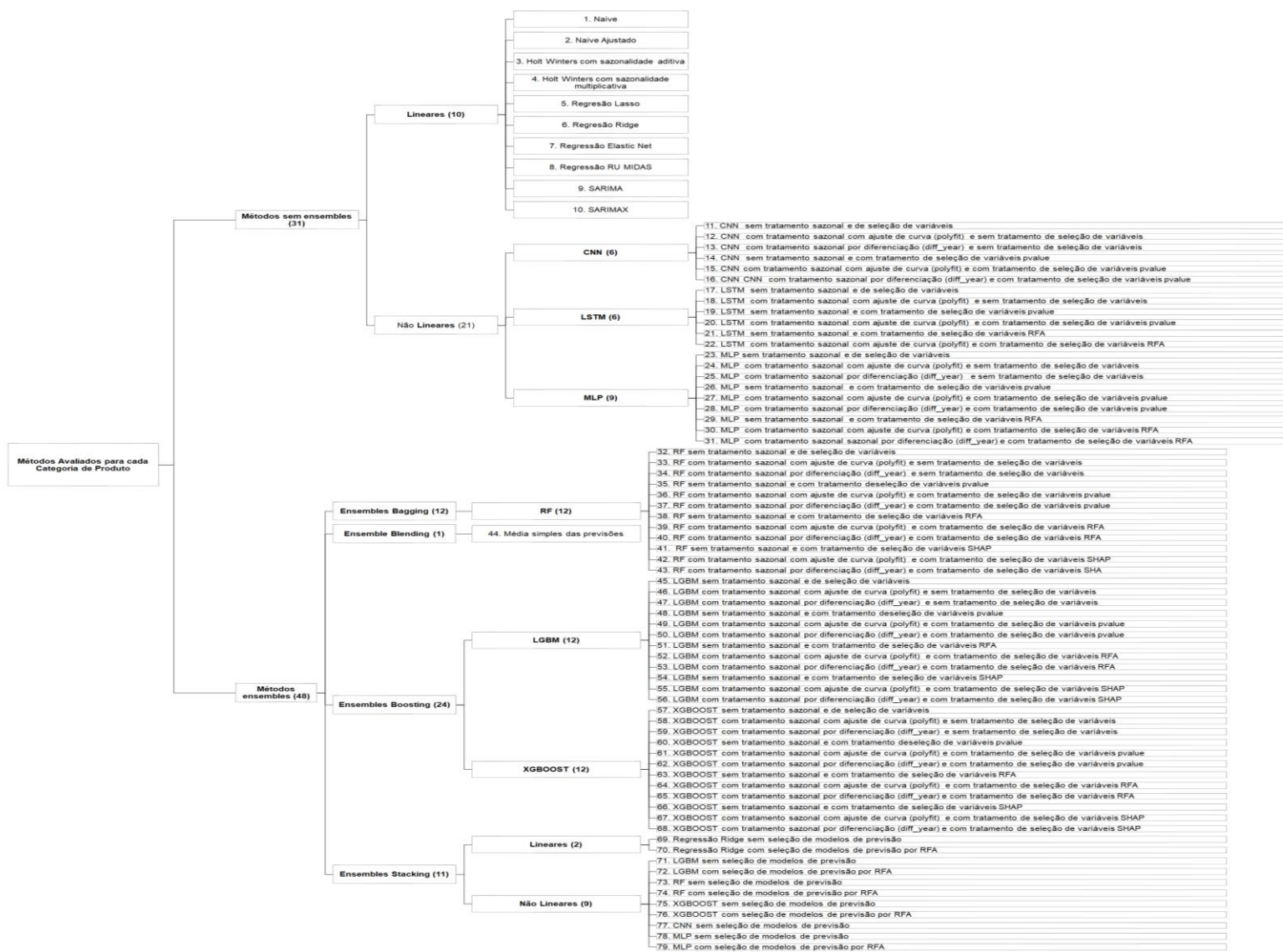
Nesta Tese empregou-se quatro métodos *ensembles*. O primeiro deles é o *blending*, que é a aplicação da média simples ou ponderada, sendo aqui aplicada a simples, das várias previsões geradas pelos diversos algoritmos. O segundo é o *bagging*, com aprendizes homogêneos fracos e reposição amostral. O aprendizado é feito de forma independente entre os aprendizes e sua combinação é através de um processo determinístico. O terceiro é o *boosting*, que também considera aprendizes homogêneos, porém o aprendizado é de forma sequencial e adaptativa, em que os erros recebem uma penalização maior para que, iterativamente, os modelos adicionados gerem um desempenho cada vez maior. O quarto é o *stacking*, que lida com modelos bases heterogêneos. O aprendizado se dá de forma independente e a combinação é pelo treinamento de uma meta modelo, que produz uma previsão com base nas previsões dos modelos fracos.

Os algoritmos de cada uma das abordagens consideradas de *ensembles* são:

- **Blending:** aplicação da média simples para as diversas previsões geradas;
- **Bagging:** aqui representado pelo algoritmo *random forest* (RF);
- **Boosting:** aqui representado pelo algoritmo LightGBM e XGBoost;
- **Stacking:** foi implementado com os algoritmos regressão Ridge, CNN, MLP, *random forest* (RF), LightGBM e XGBoost. Nesse caso, não foi considerado a variação do tratamento sazonal, tendo em vista a existência de métodos que não tinham essa opção e, adicionalmente, optou-se por

usar somente a técnica de seleção de variáveis RFA, a que teve o maior ganho de acurácia, consoante à conclusão do terceiro capítulo deste ensaio.

Figura 4.1 – Métodos de previsão *ensembles* e não *ensembles* aplicados



Fonte: elaboração própria.

Para comparar os resultados dos métodos *ensembles*, foram selecionados 31 modelos, sendo 10 lineares e 21 não lineares, os quais têm como representantes modelos baseados em redes neurais. No total, foram considerados 79 modelos de previsão, levando em conta uma diversidade em termos de tratamento sazonal, tratamento de seleção de variáveis, tipo de *ensemble* e tipo de algoritmo para fazer a comparação dos níveis de precisão obtidos. Cada um dos modelos está na Figura 4.1.

O desempenho das previsões geradas por cada um dos 79 modelos foi comparado pelo critério da precisão, a qual foi mensurada através dos seguintes indicadores: RMSE, RMSLE, MAE, MAEP e RMSEP, sendo que os mesmos estão descritos no capítulo introdutório.

Com base no cálculo de cada uma destas métricas, é possível verificar o nível de precisão alcançado por cada tipo de modelo e, conseqüentemente, possibilita realizar uma série de análises e comparações. Especificamente, torna possível constatar quais modelos *ensembles* geram modelos preditivos mais precisos em relação aos modelos não combinados, que tipo de modelo *ensemble* é mais efetivo e que tipo de algoritmo tem maior acurácia. Em resumo, permite apurar qual o modelo é mais preciso para cada uma das categorias de produtos estudadas e para a média das mesmas, sendo possível concluir qual o modelo de previsão com maior acurácia para o varejo de moda.

Além do cálculo da precisão para cada um dos modelos elencados para participar deste estudo comparativo de modelos de previsão para o varejo de moda e da comparação entre eles, também é realizado o teste MCS para comprovar, formalmente, qual o modelo de previsão possui o maior poder preditivo, tendo em vista que não necessariamente melhores níveis de acurácia garantem o maior poder preditivo.

3. Análise AHP

A análise AHP foi selecionada para este trabalho, pois a decisão do modelo de previsão mais adequado leva em conta vários critérios, e não somente o aspecto da precisão, e a análise AHP é um dos métodos de tomada de decisão multicritério mais aplicado nas organizações.

O método AHP, desenvolvido por Saaty (1980), define um objetivo, estrutura o problema de forma hierárquica, através de critérios que podem ser tanto quantitativos

como qualitativos, realiza a comparação paritária com base em uma escala – e não por *ranking* – entre as alternativas existentes, sendo geradas as matrizes de prioridade, calcula os índices de prioridade e a consistência e, finalmente, avalia as opções consideradas de acordo com as prioridades estabelecidas.

A análise AHP presente neste capítulo é feita com base no trabalho de Ren, Chan e Ram (2017). Assim, são considerados os mesmos critérios e pesos estipulados no artigo dos autores para realizar a análise dos 79 modelos de previsão presentes neste estudo. Ademais, o objetivo é o mesmo: responder qual é o melhor modelo.

Os critérios contemplados no trabalho de Ren, Chan e Ram (2017) são: (1) acurácia, (2) velocidade, no sentido de saber qual método é o mais rápido, gerando um menor custo computacional, (3) requerimento de dados, que está relacionado ao volume demandado para a realização das previsões, (4) estabilidade, que diz respeito à robustez dos resultados gerados, que é maior nos modelos estatísticos quando comparado com os modelos de aprendizagem de máquina, e (5) facilidade de uso e outros parâmetros, que diz respeito à facilidade de uso e implementação dos modelos, sendo que, mais uma vez, os modelos de aprendizagem de máquinas são mais complexos que os modelos estatísticos lineares.

A Tabela 4.1 mostra a escala de avaliação para a realização da comparação paritária entre os modelos de previsão contemplados, enquanto a Tabela 4.2 traz o peso atribuído a cada critério especificado, de acordo com o resultado da pesquisa realizada pelos autores com 123 profissionais com experiência profissional no setor de moda.

Alicerçada na escala fundamental, é possível realizar a comparação paritária entre os modelos para cada um dos critérios e desenvolver a matriz de prioridade, a qual posteriormente é multiplicada pelos pesos da tabela 4.2 e, assim, é determinada a avaliação resultante entre as alternativas de modelos de previsão elencadas. Essa avaliação é feita para cada uma das categorias de produtos analisadas nesta Tese.

Tabela 4.1 – Escala fundamental

Importância	Definição
1	Altamente sem importância
2	Sem importância
3	Neutro
4	Importante
5	Muito Importante

Fonte: Ren, Chan e Ram (2017).

Tabela 4.2 – Pesos dos critérios para avaliação das previsões

Critério	Acurácia	Velocidade	Requerimentos de dados	Estabilidade	Facilidade de uso
Peso	0,213	0,188	0,194	0,208	0,196

Fonte: Ren, Chan e Ram (2017).

A matriz prioridade para os critérios: requerimentos de dados, estabilidade e facilidade de uso foi igual entre as diferentes categorias de produtos. Já a acurácia, constatada pela métrica RMSLE, e a velocidade, calculada pelo tempo de treinamento dos modelos, foram distintas entre os produtos averiguados.

4. Qual o modelo de previsão é o mais adequado para o setor de moda?

Nesta seção são mostrados os resultados alcançados, em termos de precisão, da aplicação de métodos de combinação para a previsão de vendas no varejo de moda. Além disso, levando em conta a apuração dos resultados da análise AHP feita, é possível indicar o método mais adequado para a previsão de vendas no setor-alvo.

A Tabela 4.3 mostra o resultado médio em termos de acurácia para cada uma das métricas RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP, do uso ou não de um método de previsão *ensemble*. Pode-se afirmar que a aplicação dos métodos *ensembles* foi capaz de gerar, de forma unânime, para todas as categorias de produtos e para todos os indicadores de precisão, um resultado superior quando comparado com métodos de previsão individuais. Na média, o uso dos métodos de combinação de previsão gerou previsões, cerca de, 36% mais que precisas do que o uso de métodos não combinados. Deve-se ressaltar que a categoria de produto camiseta de menino foi a que apresentou o maior ganho de precisão, que foi de, aproximadamente, 45%.

Na análise do tipo de *ensemble*, que produziu a maior precisão entre as diversas abordagens contempladas neste estudo, *bagging*, *blending*, *boosting* e *stacking*, foi possível constatar que, em média para todas as métricas de precisão incluídas e para todas as categorias de produto, o tipo *blending*, que é a média simples das diversas previsões individuais, gerou o melhor resultado em termos de acurácia. É importante ter em mente que esse é um resultado médio e, como será visto na próxima análise, alguns algoritmos específicos conseguiram superar tal performance. Apesar disso, essa

conclusão vai ao encontro do *puzzle* de combinação de previsões, que observa que a combinação por média simples gera resultados superiores do que combinações ponderadas mais sofisticadas, conforme Smith e Wallis (2009).

Tabela 4.3 – Resultado das métricas de erro como uso de métodos *ensembles*

Ensembles			
Categoria de Produto	Métrica	Não	Sim
Média	RMSE	1.419,20	854,11
	RMSLE	41,71	25,07
	MAE	757,59	519,58
	MAPE	22,81	16,81
	MAEP	394,39	228,05
	RMSEP	1.021,57	645,86
Camiseta de Homem	RMSE	1.585,86	900,76
	RMSLE	27,36	17,90
	MAE	777,15	465,62
	MAPE	21,28	13,92
	MAEP	380,32	233,50
	RMSEP	1.144,69	650,79
Blusa de Mulher	RMSE	1.477,00	1.125,46
	RMSLE	31,56	24,75
	MAE	892,64	745,38
	MAPE	19,63	17,53
	MAEP	438,31	286,39
	RMSEP	1.094,99	890,51
Camiseta de Menino	RMSE	1.447,80	730,21
	RMSLE	51,12	26,75
	MAE	716,03	417,17
	MAPE	25,78	17,58
	MAEP	396,51	169,80
	RMSEP	1.043,13	578,07
Blusa de Menina	RMSE	1.166,14	660,00
	RMSLE	56,78	30,87
	MAE	644,54	450,14
	MAPE	24,56	18,21
	MAEP	362,41	222,51
	RMSEP	803,47	464,05

Fonte: elaboração própria.

A Tabela 4.4 traz os resultados obtidos por tipo de método *ensemble* de forma detalhada, sendo constatada a superioridade do desempenho do *ensemble blending*.

Tabela 4.4 – Resultado das métricas de erro por tipo de método *ensemble*

Categoria de Produto	Métrica	Tipos de Ensembles				
		Não	Ensemble Bagging	Ensemble Blending	Ensemble Boosting	Ensemble Stacking
Média	RMSE	1.426,69	835,11	507,52	850,18	914,91
	RMSLE	42,34	19,95	14,77	29,64	21,61
	MAE	765,65	500,39	350,27	494,60	610,39
	MAPE	23,12	15,99	11,81	16,00	19,92
	MAEP	401,32	243,02	173,99	235,73	199,88
	RMSEP	1.025,38	602,32	365,58	628,00	757,80
Camiseta de Homem	RMSE	1.587,21	1.013,15	519,96	1.020,42	551,70
	RMSLE	27,97	18,57	14,92	18,37	16,40
	MAE	784,04	510,06	329,59	492,37	371,15
	MAPE	21,58	15,04	11,56	14,31	12,07
	MAEP	386,34	251,16	152,05	239,01	209,59
	RMSEP	1.144,17	733,18	395,42	746,33	375,68
Blusa de Mulher	RMSE	1.486,06	960,80	573,62	967,31	1.700,30
	RMSLE	31,81	19,24	13,55	28,20	24,26
	MAE	897,50	670,64	424,83	665,18	1.031,03
	MAPE	19,73	15,88	11,11	16,09	23,05
	MAEP	442,32	292,17	169,27	286,38	290,75
	RMSEP	1.102,30	733,51	461,34	726,29	1.459,08
Camiseta de Menino	RMSE	1.456,28	752,51	435,70	787,13	608,48
	RMSLE	52,23	20,26	15,15	31,96	23,51
	MAE	728,36	402,68	308,72	401,19	477,71
	MAPE	26,32	16,40	12,68	15,84	23,09
	MAEP	407,19	182,72	164,55	181,18	131,34
	RMSEP	1.044,82	575,77	307,99	616,41	521,48
Blusa de Menina	RMSE	1.177,20	613,96	500,78	625,87	799,18
	RMSLE	57,34	21,74	15,46	40,02	22,28
	MAE	652,68	418,20	337,92	419,67	561,67
	MAPE	24,85	16,63	11,91	17,75	21,50
	MAEP	369,41	246,01	210,09	236,33	167,83
	RMSEP	810,23	366,81	297,56	422,95	674,95

Fonte: elaboração própria.

Aspecto que vai ao encontro do resultado obtido no terceiro capítulo deste ensaio é que a aplicação de tratamento de seleção de variáveis para os *ensembles stacking*, mais especificamente, o RFA, trouxe ganhos de acurácia para praticamente todas as métricas e categorias de produtos, sendo que o ganho médio foi de 8,9% de precisão.

A próxima análise do desempenho da aplicação de métodos *ensembles* apresenta o desempenho por algoritmo aplicado. Dessa forma, na Tabela 4.5, pode-se observar resultados muito distintos por categoria de produto e por métrica de precisão. O primeiro aspecto a ser destacado é que o algoritmo XGBOOST, aplicado tanto como o tipo *boosting* como o tipo *stacking*, foi o que apareceu com mais frequência entre os melhores resultados de precisão, mais especificamente, em cerca de 55% dos casos, sendo que em 37,5% dos casos os menores erros foram obtidos pela forma *boosting*. Essa descoberta é bastante interessante, tendo em vista o crescimento do uso do tipo *stacking*, que não gerou os melhores níveis de erro nesse trabalho.

Dois aspectos relevantes são: (1) a única categoria em que o melhor algoritmo para a maior parte das métricas não foi um método de combinação de previsão, e sim o uso do algoritmo individual MLP, foi a de camiseta de menino, e (2) no caso da categoria blusa de menina, para todos os indicadores de precisão o melhor algoritmo foi o XGBOOST aplicado por meio de um *boosting*.

Em termos de média dos resultados das categorias de produtos analisadas, para cada métrica de precisão, um determinado algoritmo teve um melhor desempenho.

Sendo assim, tanto para RMSE como para RMSEP, os melhores resultados de precisão foram obtidos pela média das diversas previsões. Para o RMSLE, isso se deu com o LGBM *boosting*; para o MAE, os menores erros foram atingidos com o XGBOOST *boosting*; para o MAPE, a maior precisão ocorreu com a aplicação do RF *bagging*; e para o MAEP, o melhor desempenho foi alcançado com o XGBOOST *stacking*.

No contexto somente de acurácia, o melhor algoritmo, juntamente com os melhores tratamentos de sazonalidade e de seleção de variáveis, considerando o critério de frequência entre as seis métricas de precisão analisadas para cada uma das categorias de produto, foi:

- **Camiseta de homem:** LGBM sem tratamento sazonal e com tratamento de seleção de variáveis RFA;
- **Blusa de mulher:** XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA;

Tabela 4.5 – Resultado das métricas de erro por tipo de algoritmo

MÉDIA										HOMEM										MULHER										MENINO										MENINA									
Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Tratamento de Seleção de Variáveis/Previsão		Tratamento Sazonal	Ensemble	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP							
NAIVE	Linear	Naive	Não	Não	Não	Não	1.684,21	25,45	781,45	20,03	400,09	1.277,75	1.989,25	28,03	893,43	21,11	458,73	1.459,61	1.542,24	20,40	824,99	16,19	419,19	1.165,33	1.779,66	27,46	747,42	21,96	383,62	1.384,09	1.425,69	25,90	659,96	20,89	338,83	1.101,97	1.582,44	25,83	648,46	20,47	330,24	1.229,56	1.131,52	23,89	538,60	19,16	284,43	860,94	
NAIVEA	Linear	Naive	Não	Não	Não	Não	1.459,24	24,22	714,60	19,01	351,02	1.186,67	1.995,22	27,43	865,90	20,61	408,80	1.510,64	1.487,78	19,72	805,45	15,81	380,61	1.145,56	1.582,44	25,83	648,46	20,47	330,24	1.229,56	1.131,52	23,89	538,60	19,16	284,43	860,94													
Regressão Elastic Net	Linear	Regressão	Não	Não	Não	Não	1.465,60	22,86	667,14	17,71	343,31	1.114,02	1.663,38	24,39	749,94	18,45	385,82	1.204,40	1.235,96	17,38	680,57	14,05	348,92	937,09	1.582,44	25,83	648,46	20,47	330,24	1.229,56	1.131,52	23,89	538,60	19,16	284,43	860,94													
Regressão Lasso	Linear	Regressão	Não	Não	Não	Não	868,85	26,76	561,56	19,38	282,53	603,38	1.026,51	26,09	616,77	20,08	310,53	690,45	905,62	22,14	693,87	16,51	343,34	637,74	862,48	27,33	458,09	18,95	233,39	656,47	650,76	31,49	476,61	21,98	248,84	428,87													
Regressão Ridge	Linear	Regressão	Não	Não	Não	Não	924,43	29,38	621,76	21,84	300,78	648,62	996,64	24,04	573,25	18,54	293,26	663,50	1.051,68	25,05	819,67	19,99	368,18	784,49	893,11	29,41	514,02	19,74	254,27	635,78	756,29	39,02	580,11	27,08	280,40	510,64													
RUMIDAS	Linear	Regressão	Não	Não	Não	Não	1.012,32	37,74	693,27	24,19	426,54	675,25	1.071,33	36,39	687,38	22,87	426,49	694,90	1.061,42	23,38	709,48	16,24	381,49	786,87	1.077,66	55,27	784,91	33,37	518,82	676,46	838,87	35,91	591,32	24,26	379,38	542,77													
SARIMA	Linear	Regressão	Não	Não	Não	Não	2.668,33	106,33	1.848,61	64,71	1.195,80	1.560,28	2.807,80	44,98	1.432,62	40,81	620,07	1.745,51	2.281,29	47,27	1.574,53	32,97	882,51	1.599,83	3.065,96	170,30	2.260,87	94,61	1.691,59	1.534,10	2.718,24	162,79	2.126,42	90,44	1.589,03	1.361,68													
SARIMAX	Linear	Regressão	Não	Não	Não	Não	1.571,01	79,22	1.368,64	54,04	813,00	969,21	1.454,61	45,10	1.272,24	52,99	349,62	1.225,69	1.915,07	86,09	1.719,71	47,83	1.235,83	1.017,55	2.924,72	151,87	1.840,99	84,37	1.380,74	1.102,36	889,64	33,81	641,62	30,97	285,80	621,22													
HWA	Linear	Suavização Exponencial	Não	Não	Não	Não	1.890,87	37,06	920,55	24,00	570,59	1.030,60	2.189,48	38,02	1.007,25	23,79	603,45	1.233,42	1.723,80	30,66	1.065,70	22,82	640,57	1.006,45	1.940,76	36,70	764,90	21,06	519,49	983,40	1.708,53	42,87	844,34	28,31	518,85	899,13													
HWM	Linear	Suavização Exponencial	Não	Não	Não	Não	1.899,30	37,20	925,54	24,04	575,69	1.034,44	2.190,07	38,04	1.007,77	23,80	603,74	1.233,98	1.721,14	30,56	1.062,07	22,75	637,60	1.005,00	1.977,19	37,38	787,59	21,25	543,06	999,14	1.708,81	42,82	844,75	28,34	518,88	899,66													
CNN	Não Linear	Rede Neural	Não	Não	Polyfit	Não	1.646,14	59,16	736,43	19,76	357,32	1.287,61	1.842,10	25,55	895,69	17,76	399,87	1.362,53	1.525,05	50,32	775,75	37,30	373,30	1.140,76	1.782,31	80,65	727,27	22,36	348,54	1.422,00	1.435,10	82,10	632,52	21,16	303,55	1.145,17													
CNN	Não Linear	Rede Neural	Não	Não	Polyfit	Não	1.673,70	25,00	771,65	19,83	398,85	1.274,44	1.981,48	27,79	889,44	21,13	451,81	1.457,23	1.527,88	20,07	817,23	16,15	404,66	1.163,47	1.773,15	26,80	732,27	21,60	387,36	1.385,15	1.412,29	25,33	647,67	20,43	331,59	1.091,90													
CNN	Não Linear	Rede Neural	Não	Diferença	Polyfit	Não	936,84	23,09	578,22	18,39	255,35	736,26	1.276,69	21,24	623,54	17,35	294,13	1.063,91	1.241,19	26,29	925,03	24,98	293,35	1.054,60	649,83	18,75	345,86	14,13	145,95	534,34	579,65	36,06	418,44	17,09	287,96	304,19													
CNN	Não Linear	Rede Neural	P-Value	Polyfit	Não	Não	1.656,42	55,15	736,68	19,88	358,79	1.286,07	1.847,65	25,58	804,47	19,73	399,38	1.373,85	1.785,15	80,65	719,09	22,39	346,16	1.427,15	1.785,15	80,65	719,09	22,39	346,16	1.427,15	1.435,53	80,10	632,50	21,18	302,71	1.146,08													
CNN	Não Linear	Rede Neural	P-Value	Não	Não	Não	1.679,03	25,26	777,01	19,92	396,28	1.274,86	1.985,90	27,90	891,24	21,11	455,30	1.458,92	1.537,25	50,32	822,53	16,13	418,01	1.161,98	1.774,96	27,19	740,72	21,76	378,11	1.382,63	1.417,30	25,81	653,57	20,69	333,72	1.095,92													
CNN	Não Linear	Rede Neural	P-Value	Diferença	Não	Não	932,16	22,94	575,35	18,29	253,34	736,02	1.265,11	21,06	616,34	17,13	292,42	1.053,38	1.243,54	26,34	926,96	25,04	292,96	1.057,05	648,44	18,72	346,66	14,15	145,95	534,34	579,65	36,06	418,44	17,09	287,96	304,19													
LSTM	Não Linear	Rede Neural	Não	Polyfit	Não	Não	1.643,25	55,95	732,61	20,19	351,80	1.278,32	1.854,09	25,61	805,88	19,78	397,68	1.383,22	1.573,73	35,02	802,11	16,34	385,20	1.207,96	1.788,87	80,65	719,51	22,42	345,09	1.431,80	1.356,30	90,51	602,44	22,21	279,23	1.090,30													
LSTM	Não Linear	Rede Neural	Não	Não	Não	Não	1.679,71	27,17	778,27	20,10	393,70	1.277,99	1.678,93	27,66	882,45	21,45	433,12	1.463,05	1.540,53	20,38	824,27	16,17	418,25	1.163,76	1.779,49	27,46	747,36	21,94	384,59	1.383,74	1.424,84	25,86	659,00	20,86	338,23	1.101,43													
LSTM	Não Linear	Rede Neural	P-Value	Polyfit	Não	Não	1.639,43	54,96	743,10	20,80	367,34	1.241,04	1.859,12	25,64	806,54	19,61	404,82	1.375,77	1.508,85	30,03	827,55	16,33	426,95	1.036,41	1.753,09	79,04	705,12	21,97	338,18	1.403,16	1.436,64	80,11	633,18	21,31	299,38	1.148,82													
LSTM	Não Linear	Rede Neural	P-Value	Não	Não	Não	1.677,54	25,51	784,04	20,71	386,04	1.279,68	1.968,03	27,30	876,55	20,84	439,48	1.453,14	1.542,49	20,47	829,53	16,47	413,14	1.168,60	1.776,16	27,55	751,60	22,83	365,62	1.389,34	1.423,46	26,73	678,46	22,70	325,93	1.107,64													
LSTM	Não Linear	Rede Neural	Rta	Polyfit	Não	Não	1.211,53	36,92	601,67	17,07	309,26	893,07	1.273,20	20,07	575,36	15,83	291,60	950,06	1.541,24	24,47	798,02	15,97	400,77	1.174,71	1.208,89	23,64	529,45	18,41	247,31	966,84	824,29	79,51	503,88	18,29	297,38	480,67													
LSTM	Não Linear	Rede Neural	Rta	Não	Não	Não	1.421,65	23,27	683,05	18,49	328,66	1.086,90	1.975,55	27,53	878,18	20,49	454,03	1.449,12	1.544,02	19,90	801,21	16,78	344,90	1.185,01	1.711,71	26,96	704,46	21,42	358,47	1.340,48	808,32	16,68	348,34	15,26	157,24	372,87													
MLP	Não Linear	Rede Neural	Não	Polyfit	Não	Não	1.293,27	45,53	694,94	21,69	322,84	953,87	1.853,82	25,59	805,68	19,76	398,01	1.382,89	1.574,19	35,02	802,18	16,34	385,22	1.208,47	832,67	27,86	553,70	22,78	228,12	625,85	91,39	618,20	27,88	280,02	598,48														
MLP	Não Linear	Rede Neural	Não	Não	Não	Não	1.232,96	41,95	640,59	19,65	315,25	884,67	1.910,95	23,36	526,00	17,09	279,40	1.362,93	1.073,05	88,88	744,04	18,42	373,30	1.715,83	1.651,32	27,97	688,36	22,54	316,25	1.290,03	1.296,52	27,59	605,98	20,54	292,15	991,00													
MLP	Não Linear	Rede Neural	Diferença	Não	Não	Não	1.248,50	19,62	461,04	15,50	208,00	527,47	1.556,97	21,43	491,51	15,29	253,17	969,54	1.748,25	35,87	673,96	14,38	316,95	941,30	646,82	163,46	350,41	42,25	487,35	615,86	531,81	17,13	391,24	14,63	154,25	544,29													
MLP	Não Linear	Rede Neural	P-Value	Não	Não	Não	961,21	28,78	543,94	17,52	282,35	651,00	776,43	23,84	410,95	13,46	226,35	439,82	1.574,19	35,02	802,18	16,34	385,22	1.208,47	722,74	22,71	447,82	18,47	209,07	502,08	771,50	33,55	514,79	22,41	298,75	453,64													
MLP	Não Linear	Rede Neural	P-Value																																														

- **Camiseta de menino:** MLP com tratamento sazonal por diferença e com tratamento de seleção de variáveis RFA;
- **Blusa de menina:** XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA.

Levando em conta que as métricas de precisão RMSLE e MAPE são as mais interpretáveis, já que são métricas padronizadas, obteve-se um resultado médio a partir dos cálculos das métricas acima para os modelos mais precisos para as quatro categorias de produtos de 12,76% de RMSLE e de 10,07% de MAPE, o que pode ser considerado um bom resultado.

Fundamentado nos resultados para cada indicador de precisão e para cada categoria de produto, o modelo de previsão, na média, mais preciso gerou uma acurácia 82,9% maior do que o pior modelo de previsão existente e 48,7% superior em relação à média de todos os modelos de previsão contemplados nessa análise.

Diante do exposto até o momento, caso fosse necessário realizar a recomendação de um único modelo de previsão de vendas para o varejo de moda, levando em conta somente o aspecto da precisão, o modelo recomendado seria um XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA, já que pela análise de frequência esse modelo de previsão apresentou a melhor performance em 37,5% dos casos. A Tabela 4.6 mostra a análise de frequência por modelo de previsão.

Tabela 4.6 – Análise de frequência dos algoritmos mais precisos

Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Tratamento de Seleção de Variáveis/Previsão	Tratamento Sazonal	Ensemble	n(i)	f(i)
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Boosting	2	8,3%
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	3	12,5%
MÉDIA SIMPLES PREVISÕES	Linear	Média	Não	Não	Ensemble Blending	3	12,5%
MLP	Não Linear	Rede Neural	Rfa	Diferença	Não	3	12,5%
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	4	16,7%
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Polyfit	Ensemble Boosting	9	37,5%
Total						24	100%

Fonte: elaboração própria.

Apesar do *ensemble* de média simples não ser o modelo de previsão com maior acurácia, nota-se que apresentou um ótimo resultado e pode ser considerado um *benchmarking* a ser superado. A constatação vai ao encontro de trabalhos na área financeira de alocação de portfólio com intuito de maximização de retorno e minimização de risco, em que a regra mais simples 1/N é difícil de ser batida e é uma referência relevante, segundo DeMiguel, Garlappi e Uppal (2005). Em trabalho posterior, os autores constatarem que a diferença de desempenho entre alocações sofisticadas e

alocação *naive* 1/N é estatisticamente insignificante (DEMIGUEL; GARLAPPI; UPPAL, 2009).

Apresentados os resultados em termos de precisão do uso de modelos *ensembles* versus a sua não utilização por um conjunto de métricas de erro e por categoria de produto, é essencial mostrar o resultado do teste MCS para cada um dos algoritmos contemplados, cujo objetivo é validar a capacidade preditiva de todos os modelos levados em consideração. O mesmo pode ser conferido na Tabela 4.7.

Considerando um nível de 95% de confiança para a maior parte das séries de produtos estudadas, o *ensemble bagging* aqui representado pela RF, apresentou, majoritariamente, o melhor desempenho, o que foi acompanhado pelo não uso de tratamento sazonal e pelo uso de RFA. No entanto, analisando cada série de produto foi possível verificar algoritmos distintos com melhor capacidade preditiva para cada métrica de precisão. Apesar do teste MCS permitir selecionar um conjunto dos melhores modelos em termos de capacidade preditiva, um único modelo foi escolhido como mais eficiente para cada métrica de precisão e para cada categoria de produto, com exceção do indicador MAE para camiseta de menino, em que foram selecionados dois modelos.

No caso da série de produto camiseta de homem, para todas as métricas, com exceção do RMSLE, o modelo com a maior capacidade preditiva foi o RF *bagging* sem tratamento sazonal e com tratamento de seleção de variáveis RFA. Especificamente para a métrica RMSLE, o modelo com maior capacidade preditiva foi o LGBM *stacking* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

No que diz respeito à categoria de blusa de mulher, para todos os indicadores de precisão, o modelo com maior poder preditivo foi o LGBM *boosting* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

Já no caso da camiseta de menino, ocorreu o comportamento mais diferenciado entre todas as séries de produtos estudadas, pois foram selecionados alguns modelos que não foram eleitos para as demais séries de produtos. Sendo assim, para RMSE, MAPE, MAEP e RMSEP, o modelo recomendado é o RF *bagging* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA. Já para a métrica RMSLE, o modelo com maior capacidade preditiva foi o LGBM *boosting* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA. Por fim, para o MAE ocorreu um empate na capacidade preditiva entre dois modelos, os quais são: RF *bagging* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA, e LGBM *boosting* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

Fonte: elaboração própria

						HOMEM						MULHER						MENINO						MENINA						
Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Seleção de Variáveis/Previsão	Tratamento Sazonal	Ensemble	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	RMSE	RMSLE	MAE	MAPE	MAEP	RMSEP	
CNN	Não Linear	Rede Neural	Não	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Não	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Não	Diferença	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Pvalue	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Pvalue	Diferença	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HWA	Linear	Suavização Exponencial	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HWM	Linear	Suavização Exponencial	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Não	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Pvalue	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Rfa	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LSTM	Não Linear	Rede Neural	Rfa	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Não	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Não	Diferença	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Pvalue	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Pvalue	Diferença	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Rfa	Polyfit	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Rfa	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Rfa	Diferença	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
NAIVE	Linear	Naive	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
NAIVEA	Linear	Naive	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Regressão Elastic Net	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Regressão Lasso	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Regressão Ridge	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RUMIDAS	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SARIMA	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
SARIMAX	Linear	Regressão	Não	Não	Não	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CNN	Não Linear	Rede Neural	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
MLP	Não Linear	Rede Neural	Rfa	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Regressão Ridge	Linear	Rede Neural	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Regressão Ridge	Linear	Rede Neural	Rfa	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
RF	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Não	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Não	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Pvalue	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Pvalue	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Pvalue	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Rfa	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Boosting	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0
LGBM	Não Linear	Árvore de Decisão	Rfa	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Shap	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Shap	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
LGBM	Não Linear	Árvore de Decisão	Shap	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Não	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Diferença	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Shap	Polyfit	Ensemble Boosting	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
XGBOOST	Não Linear	Árvore de Decisão	Shap	Não	Ensemble Boosting																									

Para a categoria de produto blusa de menina, para todas as métricas com exceção do RMSLE, o modelo com maior capacidade preditiva foi o RF *bagging* sem tratamento sazonal e sem tratamento de seleção de variáveis. Para a métrica RMSLE, o modelo recomendado foi o XGBOOST *boosting* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

Conclui-se pela análise do teste MCS, através da verificação por frequência, que o modelo com maior capacidade preditiva para o setor de moda foi o RF *bagging* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

Comparando o resultado do teste MCS com o obtido em termos dos níveis de precisão de maneira absoluta, nota-se que não houve sobreposição dos modelos de previsão. Ou seja, a análise dos níveis de precisão evidenciou como modelo de previsão mais preciso, com base em uma análise de frequência, o XGBOOST *boosting* com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA. Já o modelo que mais se repetiu em termos de capacidade preditiva foi o RF *bagging* sem tratamento sazonal e com tratamento de seleção de variáveis por RFA. Sendo assim, ocorreu divergência tanto no tipo de método de combinação como no tratamento sazonal, existindo a aderência somente em relação ao método de seleção de variáveis.

O próximo passo é apresentar a apuração da análise AHP. Entende-se que esse é um resultado mais completo para recomendação da técnica de previsão de vendas mais apropriada para o setor de moda por contemplar cinco critérios, e não somente a acurácia.

Com base na Tabela 4.1 foi possível calcular o vetor prioridade para cada um dos critérios e para cada uma das séries de produtos estudadas. A comparação paritária entre os modelos foi feita a partir de avaliação subjetiva da autora desta Tese para os seguintes critérios: requerimentos de dados, estabilidade e facilidade de uso. Já para a acurácia e velocidade usou-se os resultados calculados.

Na Tabela 4.7 verifica-se cada um dos vetores prioridades para cada uma das técnicas de previsão consideradas, sendo que o seu cálculo foi feito através de uma média geométrica e, posteriormente, foi aplicado o processo de normalização. Importante destacar que o vetor prioridade dos critérios requerimentos de dados, estabilidade e facilidade de uso foram iguais entre as categorias de produtos contempladas, já que a mudança de produto não gerou impactos distintos nesses critérios, somente na acurácia e na velocidade.

Na análise do vetor prioridade para a estabilidade constata-se que os melhores modelos são os lineares *Naïves*, seguidos pelos baseados em suavização exponencial e, logo na sequência, pelos baseados em regressão. Os próximos modelos da lista seriam aqueles baseados em árvores de decisão, ficando, por último, os baseados em redes neurais, sendo o algoritmo LSTM o menos estável.

No caso do vetor prioridade de requisitos de dados, o comportamento é similar ao visto no vetor de prioridade de estabilidade, porém os modelos de árvores de decisão ficaram equivalentes aos modelos de regressão. Ademais, os modelos de regressão se equivaleram aos *Naïve*. Por outro lado, os modelos SARIMA e os de suavização exponencial apresentaram uma necessidade um pouco maior de dados, sendo então considerados um pouco piores que os modelos *Naïve* e os de regressão e equivalentes aos modelos de árvores de decisão. Mais uma vez, os piores modelos foram os de redes neurais e o algoritmo LSTM ficou na última posição entre os modelos analisados.

Sobre o vetor prioridade de facilidade de uso, verificou-se que os modelos *Naïve*, seguidos pelos de regressão, com exceção dos SARIMA, foram os que apresentaram a maior facilidade de uso. Na sequência, vêm os modelos de árvores de decisão, seguidos pelos SARIMA e, por fim, os modelos com a aplicação de redes neurais, e mais uma vez, a técnica LSTM teve o pior desempenho frente às demais contempladas.

Para o vetor prioridade velocidade, apesar de ter um vetor diferente para a categoria de produto, os valores calculados foram muito próximos, não sendo necessária uma análise apartada para cada categoria de produto. Desse modo, os melhores modelos foram os *Naïves*, seguidos pelos de suavização exponencial, de regressão e, por último, pelos RU-MIDAS e o de média simples no contexto de *ensembles*. Após esse primeiro bloco de modelos, os próximos são, predominantemente, modelos de árvores de decisão sem o uso do tratamento de seleção de variáveis RFA. Por fim, existe o bloco de modelos baseados em árvores de decisão com tratamento de seleção de variáveis por RFA e, na última posição, estão os modelos baseados em redes neurais.

Diante do exposto, fica evidente que os modelos com a aplicação de algoritmos de redes neurais e o uso do tratamento de seleção de variáveis por RFA afetam significativamente a velocidade de tratamento dos algoritmos considerados. Ademais, o uso do tratamento sazonal teve um impacto bem menor quando comparado com o uso do tratamento de seleção de variáveis.

Tabela 4.8 – Vetor prioridade por critério e por tipo de produto

HOMEM										MULHER					MENINO					MENINA					
Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Tratamento de Seleção de Variáveis/Previsão	Tratamento Sazonal	Ensemble	Acurácia	Velocidade	Requerimentos de dados	Estabilidade	Facilidade de uso	Acurácia	Velocidade	Requerimento s de dados	Estabilidade	Facilidade de uso	Acurácia	Velocidade	Requerimento s de dados	Estabilidade	Facilidade de uso	Acurácia	Velocidade	Requerimento s de dados	Estabilidade	Facilidade de uso
NAIVE	Linear	Naive	Não	Não	Não	0.430	1.000	1.000	1.000	1.000	0.551	1.000	1.000	1.000	1.000	0.478	1.000	1.000	1.000	1.000	0.550	1.000	1.000	1.000	1.000
NAIVEA	Linear	Naive	Não	Não	Não	0.439	1.000	1.000	1.000	1.000	0.570	1.000	1.000	1.000	1.000	0.506	1.000	1.000	1.000	1.000	0.597	1.000	1.000	1.000	1.000
Regressão Elastic Net	Linear	Regressão	Não	Não	Não	0.495	0.009	1.000	0.798	0.798	0.648	0.060	1.000	0.798	0.798	0.507	0.061	1.000	0.798	0.798	0.597	0.065	1.000	0.798	0.798
Regressão Lasso	Linear	Regressão	Não	Não	Não	0.462	0.063	1.000	0.798	0.798	0.507	0.067	1.000	0.798	0.798	0.478	0.067	1.000	0.798	0.798	0.451	0.059	1.000	0.798	0.798
Regressão Ridge	Linear	Regressão	Não	Não	Não	0.502	0.065	1.000	0.798	0.798	0.448	0.064	1.000	0.798	0.798	0.444	0.066	1.000	0.798	0.798	0.363	0.068	1.000	0.798	0.798
RUMIDAS	Linear	Regressão	Não	Não	Não	0.330	0.060	0.798	0.798	0.596	0.480	0.060	0.798	0.798	0.596	0.234	0.060	0.798	0.798	0.596	0.395	0.060	0.798	0.798	0.596
SARIMA	Linear	Regressão	Não	Não	Não	0.266	0.008	0.596	0.798	0.596	0.235	0.005	0.596	0.798	0.596	0.075	0.007	0.596	0.798	0.596	0.086	0.005	0.596	0.798	0.596
SARIMAX	Linear	Regressão	Não	Não	Não	0.265	0.004	0.596	0.798	0.596	0.128	0.004	0.596	0.798	0.596	0.084	0.005	0.596	0.798	0.596	0.420	0.004	0.596	0.798	0.596
HWA	Linear	Suavização Exponencial	Não	Não	Não	0.316	0.711	0.798	1.000	0.798	0.365	0.711	0.798	1.000	0.798	0.355	0.711	0.798	1.000	0.798	0.330	0.711	0.798	1.000	0.798
HWM	Linear	Suavização Exponencial	Não	Não	Não	0.315	0.711	0.798	1.000	0.798	0.366	0.711	0.798	1.000	0.798	0.348	0.711	0.798	1.000	0.798	0.331	0.711	0.798	1.000	0.798
CNN	Não Linear	Rede Neural	Não	Polyfit	Não	0.472	0.001	0.395	0.395	0.196	0.221	0.002	0.395	0.395	0.196	0.160	0.001	0.395	0.395	0.196	0.175	0.003	0.395	0.395	0.196
CNN	Não Linear	Rede Neural	Não	Polyfit	Não	0.433	0.001	0.395	0.395	0.196	0.560	0.001	0.395	0.395	0.196	0.488	0.001	0.395	0.395	0.196	0.563	0.002	0.395	0.395	0.196
CNN	Não Linear	Rede Neural	Não	Diferença	Não	0.569	0.001	0.395	0.395	0.196	0.426	0.002	0.395	0.395	0.196	0.701	0.002	0.395	0.395	0.196	0.547	0.002	0.395	0.395	0.196
CNN	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0.471	0.001	0.395	0.395	0.196	0.326	0.002	0.395	0.395	0.196	0.160	0.002	0.395	0.395	0.196	0.175	0.002	0.395	0.395	0.196
CNN	Não Linear	Rede Neural	Pvalue	Não	Não	0.432	0.001	0.395	0.395	0.196	0.553	0.002	0.395	0.395	0.196	0.481	0.001	0.395	0.395	0.196	0.556	0.001	0.395	0.395	0.196
CNN	Não Linear	Rede Neural	Pvalue	Diferença	Não	0.574	0.001	0.395	0.395	0.196	0.425	0.002	0.395	0.395	0.196	0.702	0.002	0.395	0.395	0.196	0.556	0.002	0.395	0.395	0.196
LSTM	Não Linear	Rede Neural	Não	Polyfit	Não	0.471	0.005	0.196	0.196	0.196	0.319	0.004	0.196	0.196	0.196	0.160	0.004	0.196	0.196	0.196	0.155	0.005	0.196	0.196	0.196
LSTM	Não Linear	Rede Neural	Não	Não	Não	0.440	0.002	0.196	0.196	0.196	0.552	0.005	0.196	0.196	0.196	0.476	0.003	0.196	0.196	0.196	0.551	0.003	0.196	0.196	0.196
LSTM	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0.470	0.006	0.196	0.196	0.196	0.319	0.005	0.196	0.196	0.196	0.163	0.004	0.196	0.196	0.196	0.175	0.006	0.196	0.196	0.196
LSTM	Não Linear	Rede Neural	Pvalue	Não	Não	0.441	0.003	0.196	0.196	0.196	0.549	0.006	0.196	0.196	0.196	0.474	0.003	0.196	0.196	0.196	0.533	0.003	0.196	0.196	0.196
LSTM	Não Linear	Rede Neural	Rta	Polyfit	Não	0.603	0.003	0.196	0.196	0.196	0.458	0.000	0.196	0.196	0.196	0.554	0.004	0.196	0.196	0.196	0.177	0.000	0.196	0.196	0.196
LSTM	Não Linear	Rede Neural	Rta	Não	Não	0.438	0.001	0.196	0.196	0.196	0.565	0.006	0.196	0.196	0.196	0.485	0.004	0.196	0.196	0.196	0.766	0.004	0.196	0.196	0.196
MLP	Não Linear	Rede Neural	Não	Polyfit	Não	0.395	0.003	0.395	0.395	0.395	0.319	0.003	0.395	0.395	0.395	0.469	0.003	0.395	0.395	0.395	0.150	0.004	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Não	Não	Não	0.395	0.017	0.395	0.395	0.395	0.395	0.024	0.395	0.395	0.395	0.467	0.002	0.395	0.395	0.395	0.516	0.004	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Não	Diferença	Não	0.328	0.003	0.395	0.395	0.395	0.243	0.003	0.395	0.395	0.395	0.078	0.002	0.395	0.395	0.395	0.046	0.004	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0.506	0.003	0.395	0.395	0.395	0.319	0.003	0.395	0.395	0.395	0.577	0.003	0.395	0.395	0.395	0.423	0.003	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Pvalue	Não	Não	0.506	0.003	0.395	0.395	0.395	0.551	0.003	0.395	0.395	0.395	0.492	0.003	0.395	0.395	0.395	0.596	0.003	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Pvalue	Diferença	Não	0.310	0.003	0.395	0.395	0.395	0.339	0.003	0.395	0.395	0.395	0.067	0.002	0.395	0.395	0.395	0.111	0.003	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Rta	Polyfit	Não	0.471	0.001	0.395	0.395	0.395	0.319	0.001	0.395	0.395	0.395	0.719	0.000	0.395	0.395	0.395	0.175	0.001	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Rta	Não	Não	0.755	0.000	0.395	0.395	0.395	0.615	0.000	0.395	0.395	0.395	0.540	0.000	0.395	0.395	0.395	0.656	0.000	0.395	0.395	0.395
MLP	Não Linear	Rede Neural	Rta	Diferença	Não	0.657	0.001	0.395	0.395	0.395	0.592	0.000	0.395	0.395	0.395	1.000	0.000	0.395	0.395	0.395	0.538	0.000	0.395	0.395	0.395
MÉDIA SIMPLES PREVISÕES	Linear	Média	Não	Não	Ensemble Blending	0.814	0.053	1.000	1.000	1.000	0.834	0.122	1.000	1.000	1.000	0.869	0.021	1.000	1.000	1.000	0.928	0.057	1.000	1.000	1.000
Regressão Ridge	Linear	Regressão	Não	Não	Ensemble Stacking	0.672	0.132	1.000	1.000	0.798	0.450	0.128	1.000	1.000	0.798	0.514	0.132	1.000	1.000	0.798	0.597	0.139	1.000	1.000	0.798
Regressão Ridge	Linear	Regressão	Não	Não	Ensemble Stacking	0.965	0.106	1.000	1.000	0.798	0.495	0.097	1.000	1.000	0.798	0.570	0.084	1.000	1.000	0.798	0.719	0.064	1.000	1.000	0.798
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0.794	0.025	0.798	0.596	0.596	0.329	0.035	0.798	0.596	0.596	0.466	0.024	0.798	0.596	0.596	0.470	0.034	0.798	0.596	0.596
LGBM	Não Linear	Árvore de Decisão	Rta	Não	Ensemble Stacking	0.971	0.002	0.798	0.596	0.596	0.384	0.006	0.798	0.596	0.596	0.519	0.004	0.798	0.596	0.596	0.597	0.006	0.798	0.596	0.596
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0.599	0.012	0.798	0.596	0.596	0.439	0.006	0.798	0.596	0.596	0.497	0.008	0.798	0.596	0.596	0.588	0.016	0.798	0.596	0.596
RF	Não Linear	Árvore de Decisão	Rta	Não	Ensemble Stacking	0.907	0.001	0.798	0.596	0.596	0.514	0.001	0.798	0.596	0.596	0.617	0.003	0.798	0.596	0.596	0.750	0.004	0.798	0.596	0.596
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0.559	0.011	0.798	0.596	0.596	0.555	0.011	0.798	0.596	0.596	0.674	0.011	0.798	0.596	0.596	0.836	0.014	0.798	0.596	0.596
XGBOOST	Não Linear	Árvore de Decisão	Rta	Não	Ensemble Stacking	0.738	0.004	0.798	0.596	0.596	1.000	0.003	0.798	0.596</											

Em relação ao vetor prioridade de acurácia, calculado a partir do valor observado do indicador RMSLE, constatou-se comportamentos distintos entre as categorias de produtos. Considerando a quantidade de modelos existentes, optou-se por detalhar os resultados sobre os dez melhores e piores modelos para cada tipo de produto.

Assim, para a categoria de produto camiseta de homem os modelos mais precisos foram os com algoritmos de árvores de decisão e com tratamento de seleção de variáveis por RFA. Já os piores modelos foram os SARIMAS, seguidos pelos MLPs com tratamento sazonal por diferença e pelos modelos de suavização exponencial.

No caso da categoria de produto de blusa de mulher, o cálculo do vetor prioridade de acurácia revelou que os melhores modelos foram o XGBOOST *stacking* com tratamento de seleção de variáveis por RFA e sem tratamento sazonal, seguido pelo XGBOOST com tratamento de seleção de variáveis por RFA e com tratamento sazonal por ajuste de curva. Em relação aos piores modelos, o pior foi o MLP sem tratamento sazonal e sem tratamento de seleção de variáveis, seguido pelos modelos LGBM com tratamento sazonal por ajuste de curva e pelo modelo SARIMAX.

Para a camiseta de menino, constatou-se resultados bem diferentes dos observados nas demais categorias de produto, pois o melhor modelo de previsão foi obtido por meio do emprego do algoritmo MLP com tratamento de seleção de variáveis por RFA e com tratamento sazonal por diferença. Na sequência, entre os melhores modelos, apareceu o de média simples das previsões geradas com outros algoritmos de forma individual, seguidos pelos de previsão com o uso de árvores de decisão com tratamento de seleção de variáveis RFA. Na parte dos piores modelos estão os com aplicação do algoritmo MLP com tratamento sazonal por diferenças com outros tratamentos de seleção de variáveis, seguidos pelos SARIMAs e pelos LSTM e CNN com tratamento sazonal por ajuste de curva.

No que diz respeito à categoria de blusa de menina, o vetor prioridade de acurácia revelou que os melhores modelos foram, nessa ordem, o XGBOOST com tratamento de seleção de variáveis com RFA e com tratamento sazonal por ajuste de curva; o de média simples das previsões geradas por todos os métodos contemplados neste estudo; e, por último, como destaque positivo, o modelo RF com tratamento de seleção de variáveis com RFA e sem tratamento sazonal. No conjunto dos piores modelos, o MLP com tratamento sazonal por diferença, em companhia do SARIMA e dos modelos MLP, LSTM e CNN com tratamento sazonal por ajuste de curva, apresentam desempenhos ruins em termos de acurácia.

Após o cálculo do vetor prioridade para cada critério e para cada tipo de produto através da multiplicação com pesos da Tabela 4.2, que é igual para todas as categorias de produto, é possível elencar qual o modelo mais apropriado para cada categoria de produto considerando os cinco critérios escolhidos. Assim, como os pesos dos critérios são muito próximos, pode-se afirmar, previamente, que os modelos de previsão que tiverem um resultado de avaliação melhor no maior número de critérios, têm mais chance de serem selecionados como modelo recomendado. Logo, não importa ter uma excelente avaliação de acurácia e uma avaliação muito ruim nos demais fatores de análise, pois tal modelo não será eleito como melhor solução de previsão de vendas para o varejo de moda.

Diante dessa observação, verifica-se por meio da Tabela 4.9, que os dez melhores modelos para todas as séries de produtos, com exceção para a categoria de produto de blusa de menina, que teve uma mudança no último colocado, foram, respectivamente, mas não necessariamente nessa ordem: *Naive ajustado*, *Naive*, modelo *ensemble* de média simples de todas previsões geradas, modelo *ensemble stacking* com regressão Ridge com tratamento por seleção de variáveis e sem tratamento sazonal, modelo *ensemble stacking* com regressão Ridge sem tratamento por seleção de variáveis por RFA e sem tratamento sazonal, HWM, HWA, regressão Ridge, regressão Lasso e regressão Elastic Net. No caso da categoria de produto de blusa de menina, a regressão Ridge foi substituída pelo modelo XGBOOST com tratamento de seleção de variáveis por RFA e com tratamento sazonal por ajuste curva.

Perante o que foi observado até o momento, apesar da acurácia pior dos modelos *Naives*, a sua facilidade uso, em conjunto com a sua baixa requisição de dados, a sua estabilidade e a sua alta velocidade, fazem desse modelo uma opção difícil de ser superada pela ampla gama de modelos contemplados neste estudo. Depois dos *Naives*, os melhores modelos de previsão foram *ensembles* de média simples e *stacking* de regressão Ridge com tratamento de seleção de variáveis por RFA e sem tratamento de seleção de variáveis, nessa ordem.

Com base no constatado até o momento e desconsiderando os modelos *Naives*, a recomendação de modelo final para o setor de varejo de moda para a realização da previsão de vendas é o modelo de combinação de previsões de vendas através da média simples, contemplando todas as opções em termos de algoritmos, tratamentos de sazonalidade e tratamentos de seleção de variáveis aplicados neste estudo.

Tabela 4.9 – Resultado da análise AHP por categoria de produto

Algoritmo	Tipo de Algoritmo	Algoritmo baseado	Tratamento de Seleção de Variáveis/Previsão	Tratamento Sazonal	Ensemble	HOMEM	MULHER	MENINO	MENINA
NAIVEA	Linear	Regressão	Não	Não	Não	0,88	0,91	0,89	0,91
NAIVE	Linear	Regressão	Não	Não	Não	0,88	0,90	0,89	0,90
MÉDIA SIMPLES PREVISÕES	Linear	Média	Não	Não	Ensemble Blending	0,78	0,80	0,79	0,81
Regressão Ridge	Linear	Regressão	Rta	Não	Ensemble Stacking	0,78	0,68	0,70	0,72
Regressão Ridge	Linear	Regressão	Não	Não	Ensemble Stacking	0,73	0,68	0,69	0,71
HWA	Linear	Suavização Exponencial	Não	Não	Não	0,72	0,73	0,73	0,72
HWM	Linear	Suavização Exponencial	Não	Não	Não	0,72	0,73	0,73	0,72
Regressão Ridge	Linear	Regressão	Não	Não	Não	0,64	0,62	0,62	0,61
Regressão Lasso	Linear	Regressão	Não	Não	Não	0,63	0,64	0,63	0,62
Regressão Elastic Net	Linear	Regressão	Não	Não	Não	0,62	0,67	0,64	0,66
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Boosting	0,61	0,58	0,58	0,57
LGBM	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	0,60	0,48	0,51	0,52
RF	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Bagging	0,59	0,58	0,57	0,59
RF	Não Linear	Árvore de Decisão	Rfa	Não	Ensemble Stacking	0,59	0,51	0,53	0,56
XGBOOST	Não Linear	Árvore de Decisão	Rta	Não	Ensemble Boosting	0,57	0,57	0,55	0,58
LGBM	Não Linear	Árvore de Decisão	Rta	Diferença	Ensemble Boosting	0,57	0,53	0,56	0,49
RF	Não Linear	Árvore de Decisão	Rta	Polyfit	Ensemble Bagging	0,57	0,57	0,57	0,57
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0,57	0,47	0,50	0,50
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting	0,57	0,54	0,43	0,55
XGBOOST	Não Linear	Árvore de Decisão	Shap	Não	Ensemble Boosting	0,56	0,55	0,55	0,55
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Não	Ensemble Boosting	0,56	0,50	0,54	0,54
XGBOOST	Não Linear	Árvore de Decisão	Rfa	Polyfit	Ensemble Boosting	0,56	0,58	0,55	0,61
LGBM	Não Linear	Árvore de Decisão	Não	Não	Ensemble Boosting	0,56	0,49	0,52	0,44
LGBM	Não Linear	Árvore de Decisão	Shap	Não	Ensemble Boosting	0,56	0,48	0,53	0,44
LGBM	Não Linear	Árvore de Decisão	Pvalue	Não	Ensemble Boosting	0,55	0,44	0,52	0,44
XGBOOST	Não Linear	Árvore de Decisão	Rta	Não	Ensemble Stacking	0,55	0,61	0,57	0,58
LGBM	Não Linear	Árvore de Decisão	Rta	Polyfit	Ensemble Boosting	0,55	0,54	0,43	0,43
LGBM	Não Linear	Árvore de Decisão	Não	Polyfit	Ensemble Boosting	0,55	0,48	0,43	0,44
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Bagging	0,55	0,53	0,55	0,57
RF	Não Linear	Árvore de Decisão	Shap	Polyfit	Ensemble Bagging	0,54	0,55	0,53	0,56
RF	Não Linear	Árvore de Decisão	Shap	Não	Ensemble Bagging	0,54	0,53	0,55	0,57
LGBM	Não Linear	Árvore de Decisão	Shap	Polyfit	Ensemble Boosting	0,54	0,42	0,43	0,44
RF	Não Linear	Árvore de Decisão	Pvalue	Não	Ensemble Bagging	0,54	0,53	0,54	0,57
RF	Não Linear	Árvore de Decisão	Não	Polyfit	Ensemble Bagging	0,54	0,54	0,52	0,56
RF	Não Linear	Árvore de Decisão	Pvalue	Polyfit	Ensemble Bagging	0,53	0,54	0,54	0,56
XGBOOST	Não Linear	Árvore de Decisão	Rta	Diferença	Ensemble Boosting	0,53	0,51	0,55	0,53
XGBOOST	Não Linear	Árvore de Decisão	Shap	Polyfit	Ensemble Boosting	0,53	0,55	0,54	0,55
RF	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0,53	0,49	0,50	0,52
RF	Não Linear	Árvore de Decisão	Rta	Diferença	Ensemble Bagging	0,52	0,52	0,55	0,49
LGBM	Não Linear	Árvore de Decisão	Não	Diferença	Ensemble Boosting	0,52	0,50	0,54	0,48
LGBM	Não Linear	Árvore de Decisão	Shap	Diferença	Ensemble Boosting	0,52	0,50	0,54	0,48
LGBM	Não Linear	Árvore de Decisão	Pvalue	Diferença	Ensemble Boosting	0,52	0,50	0,54	0,48
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Polyfit	Ensemble Boosting	0,52	0,53	0,54	0,55
XGBOOST	Não Linear	Árvore de Decisão	Não	Polyfit	Ensemble Boosting	0,52	0,49	0,55	0,55
RUMIDAS	Linear	Regressão	Não	Não	Não	0,52	0,55	0,50	0,53
LGBM	Não Linear	Árvore de Decisão	Pvalue	Polyfit	Ensemble Boosting	0,52	0,42	0,43	0,44
XGBOOST	Não Linear	Árvore de Decisão	Shap	Diferença	Ensemble Boosting	0,52	0,49	0,53	0,50
XGBOOST	Não Linear	Árvore de Decisão	Não	Não	Ensemble Stacking	0,52	0,52	0,54	0,58
RF	Não Linear	Árvore de Decisão	Shap	Diferença	Ensemble Bagging	0,52	0,49	0,51	0,50
RF	Não Linear	Árvore de Decisão	Não	Diferença	Ensemble Bagging	0,51	0,49	0,52	0,50
RF	Não Linear	Árvore de Decisão	Pvalue	Diferença	Ensemble Bagging	0,51	0,48	0,50	0,50
XGBOOST	Não Linear	Árvore de Decisão	Não	Diferença	Ensemble Boosting	0,50	0,49	0,53	0,51
XGBOOST	Não Linear	Árvore de Decisão	Pvalue	Diferença	Ensemble Boosting	0,50	0,49	0,53	0,51
SARIMA	Linear	Regressão	Não	Não	Não	0,46	0,45	0,42	0,42
SARIMAX	Linear	Regressão	Não	Não	Não	0,46	0,43	0,42	0,49
MLP	Não Linear	Rede Neural	Rfa	Não	Ensemble Stacking	0,41	0,34	0,37	0,39
MLP	Não Linear	Rede Neural	Rfa	Não	Não	0,40	0,37	0,35	0,38
MLP	Não Linear	Rede Neural	Não	Não	Ensemble Stacking	0,38	0,32	0,34	0,35
MLP	Não Linear	Rede Neural	Rta	Diferença	Não	0,38	0,36	0,45	0,35
MLP	Não Linear	Rede Neural	Não	Não	Não	0,35	0,26	0,34	0,35
CNN	Não Linear	Rede Neural	Não	Não	Ensemble Stacking	0,35	0,29	0,30	0,32
MLP	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0,34	0,30	0,36	0,33
MLP	Não Linear	Rede Neural	Pvalue	Não	Não	0,34	0,35	0,34	0,36
MLP	Não Linear	Rede Neural	Não	Polyfit	Não	0,34	0,30	0,34	0,27
MLP	Não Linear	Rede Neural	Rfa	Polyfit	Não	0,34	0,30	0,39	0,27
CNN	Não Linear	Rede Neural	Pvalue	Diferença	Não	0,32	0,29	0,35	0,32
CNN	Não Linear	Rede Neural	Não	Diferença	Não	0,32	0,29	0,35	0,31
MLP	Não Linear	Rede Neural	Não	Diferença	Não	0,31	0,29	0,25	0,25
MLP	Não Linear	Rede Neural	Pvalue	Diferença	Não	0,30	0,31	0,25	0,26
CNN	Não Linear	Rede Neural	Não	Polyfit	Não	0,30	0,24	0,23	0,24
CNN	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0,30	0,27	0,23	0,24
CNN	Não Linear	Rede Neural	Não	Não	Não	0,29	0,32	0,30	0,32
CNN	Não Linear	Rede Neural	Pvalue	Não	Não	0,29	0,32	0,30	0,32
LSTM	Não Linear	Rede Neural	Rfa	Polyfit	Não	0,25	0,21	0,24	0,15
LSTM	Não Linear	Rede Neural	Não	Polyfit	Não	0,22	0,19	0,15	0,15
LSTM	Não Linear	Rede Neural	Pvalue	Polyfit	Não	0,22	0,19	0,15	0,16
LSTM	Não Linear	Rede Neural	Pvalue	Não	Não	0,21	0,24	0,22	0,23
LSTM	Não Linear	Rede Neural	Não	Não	Não	0,21	0,24	0,22	0,24
LSTM	Não Linear	Rede Neural	Rfa	Não	Não	0,21	0,24	0,22	0,28

Fonte: elaboração própria

5. Conclusões

A previsão de demanda tem se tornado uma atividade fundamental para o varejo de moda *vis-à-vis* o aumento de competitividade crescente no setor, o desenvolvimento

de cadeias globais e o crescimento dos níveis de incerteza. Adicionalmente, as peculiaridades pertencentes ao setor são aspectos que tornam o processo de previsão mais complexo do que em outros setores varejistas.

Diante de tais pontos, há consenso de que a geração de previsões de vendas mais assertivas, o que demanda a seleção de métodos de previsão, contribui para uma gestão mais eficiente da cadeia de abastecimento, garantindo tanto uma performance financeira superior como um nível de atendimento mais elevado ao cliente.

Perante esse contexto, este capítulo se propôs responder a duas questões: (1) qual o modelo de previsão de vendas mais preciso para o setor de moda, com base na avaliação de 79 modelos de previsão com foco nos modelos *ensembles*, e (2) qual o modelo de previsão de vendas mais adequado para o setor, o que foi feito através de uma análise AHP, considerando cinco critérios relevantes. Ambas as respostas foram obtidas por meio das previsões realizadas para quatro categorias de produtos da rede de lojas de varejo Lojão do Brás.

Com base nas previsões feitas para os 79 modelos estudados, em relação à pergunta sobre o modelo mais preciso para o setor, é fundamental destacar que foram obtidos desempenhos distintos entre as métricas de acurácia escolhidas (RMSE, RMSLE, MAE, MAPE, MAEP e RMSEP), assim como por categoria de produto. Ademais, conforme a média realizada para as seis métricas de precisão estudadas, o modelo de previsão mais preciso médio evidenciou uma acurácia maior em 82,9% que o pior modelo e 48,7% superior à média de todos os modelos de previsão.

Tendo em vista o maior poder de interpretação das métricas RMSLE e MAPE e buscando evidenciar o nível de acurácia alcançado, obteve-se um resultado médio com base nos níveis de precisão dos modelos mais assertivos para cada categoria de produto de 12,76% de RMSLE e de 10,08% de MAPE.

De forma mais detalhada, levando em conta os resultados de todas as categorias de produtos, para os indicadores RMSE e RMSEP, a maior precisão foi observada pela aplicação do modelo de média simples de combinação das diversas previsões geradas. Já para RMSLE, o modelo de previsão mais preciso foi o LGBM sem tratamento sazonal e com tratamento de seleção de variáveis por RFA. No caso do MAE, a maior acurácia se deu com o algoritmo XGBOOST com tratamento sazonal com ajuste da curva e com seleção de variáveis por RFA. Para o MAPE, a maior precisão ocorreu com a aplicação da técnica RF sem tratamento sazonal e com tratamento de seleção de variáveis por

RFA e, por último, para a métrica MAEP, a melhor performance foi alcançada com o algoritmo XGBOOST sem tratamento sazonal e com seleção de variáveis por RFA.

Na análise por categoria de produto, verificou-se que os modelos com maior precisão, tendo em vista uma análise de frequência feita para as categorias de produto camiseta de homem, blusa de mulher, camiseta de menino e blusa de menina, respectivamente, foram: LGBM sem tratamento sazonal e com tratamento de seleção de variáveis RFA; XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA; MLP com tratamento sazonal por diferença e com tratamento de seleção de variáveis RFA; XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA.

Mediante os resultados mencionados e considerando que fosse obrigatório a seleção de um único modelo de previsão para a previsão de vendas para o varejo de moda, a sugestão é a escolha de um modelo XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA. A escolha é justificada por essa especificação de modelo de previsão aparecer em 37,5% dos casos como modelo de previsão com maior acurácia, tendo em vista as seis métricas de precisão calculadas para as quatro categorias de produtos estudadas.

Uma ressalva fundamental é que, com base no teste MCS, que tem por objetivo verificar qual modelo de previsão apresenta maior capacidade preditiva, foram obtidos desempenhos distintos para cada métrica de precisão e para cada tipo de produto estudado. Através, mais uma vez, de uma análise de frequência, pode-se afirmar que o modelo com maior capacidade preditiva para o setor de varejo de moda foi o algoritmo RF sem tratamento sazonal e com tratamento de seleção de variáveis por RFA.

A comparação do resultado do teste MCS com o resultado obtido em termos do *ranking* dos níveis de precisão para cada indicador contemplado permite constatar que não ocorreu uma sobreposição dos resultados. Sendo assim, houve divergência tanto no tipo de método de combinação de previsão como no tratamento sazonal recomendado, tendo somente como ponto comum o método de seleção de variáveis. O ideal buscado em um estudo de previsão é que ocorra uma sobreposição total entre o teste MCS e a análise numérica simples dos indicadores de acurácia calculados, pois isso garante a robustez da conclusão apurada.

Após verificado qual o modelo de previsão de venda mais preciso para o setor de varejo de moda, a próxima resposta a ser dada é qual o melhor modelo para esse setor. Como já visto, a definição do melhor modelo foi feita com base em uma análise

AHP, que considera cinco critérios e não somente a acurácia. Não seria adequado recomendar um modelo somente por tal critério, já que existem outros tão importantes quanto, que são velocidade, requerimentos de dados, facilidade de uso e estabilidade.

A partir da aplicação da metodologia AHP, foi possível evidenciar que o melhor modelo de previsão, de forma unânime, para os quatro tipos de categoria de produtos analisados, foi o modelo *Naive* ajustado, o qual pode ser visto como um modelo *Random Walk*, sendo que para as semanas do mês de dezembro são considerados os valores dessas mesmas semanas no ano imediatamente anterior.

A mesma congruência observada para a constatação do modelo *Naive* ajustado como melhor modelo de previsão de vendas entre as quatro categorias de produtos estudadas foi verificada para os dez melhores modelos para todas as séries de produto, com exceção da categoria de blusa de menina, que teve uma mudança no último colocado. Assim, os dez melhores modelos, mas não necessariamente nessa ordem, foram: *Naive*, *Naive* ajustado, modelo *ensemble* de média simples de todas as previsões geradas, modelo *ensemble stacking* com regressão Ridge com tratamento por seleção de variáveis e sem tratamento sazonal, modelo *ensemble stacking* com regressão Ridge sem tratamento por seleção de variáveis por RFA e sem tratamento sazonal, HWM, HWA, regressão Ridge, regressão Lasso e regressão Elastic Net.

Os resultados apresentados sobre os melhores modelos para previsão de vendas no varejo de moda, por meio da análise AHP, comprova que, apesar da acurácia inferior dos modelos *Naives* frente a apurada em outros modelos de previsão, a sua facilidade uso, em companhia da sua baixa requisição de dados, da sua estabilidade e da sua alta velocidade, fazem dele uma opção de modelo de previsão praticamente imbatível diante dos demais modelos de previsão contemplados neste trabalho. Portanto, a consistência de uma avaliação de mediana a superior no maior número de critérios, já que há uma baixa diferença entre os pesos dos critérios participantes da análise da AHP, justificam a liderança dos modelos *Naives* nessa competição entre modelos de previsão de vendas para o setor em foco.

Se os modelos de previsão *Naives* fossem excluídos como opção, a sugestão para a realização da previsão de vendas no setor de varejo de moda seria o modelo *ensemble* de média simples, o qual contempla todos os modelos de previsão presentes neste trabalho e significa efetivamente a inclusão do aspecto da diversidade de modelos de previsão em termos de algoritmos, de tratamentos de sazonalidade e de tratamentos de seleção de variáveis.

REFERÊNCIAS

- ABLOW, C. M.; KAYLOR, D. J. A committee solution of the pattern recognition problem. **IEEE Transaction on Information Theory**, v. 11, n. 3, p. 453-455, 1965.
- BATES, J. M., GRANGER, C. W. J. The combination of forecasts. **Journal of the Operational Research Society**, v. 20, p. 309-325, 1969.
- BREIMAN, Leo. Bagging Predictors. **Machine Learning**, v. 24, p. 123-140, 1996a.
- BREIMAN, Leo. Stacked Regressions. **Machine Learning**, v. 24, p. 49-64, 1996b.
- BUTTER, Frank; JANSEN, Pieter. Beating the random walk: A performance assessment of long-term interest rate forecasts. **Applied Financial Economics**, v. 23, p. 749-765, 2013.
- CLAESKENS, Gerda *et al.* **The forecast combination puzzle: a simple theoretical explanation**. Research Gate, 2014.
- DEMIGUEL, Victor; GARLAPPI, Lorenzo; UPPAL, Raman. How inefficient are simple asset-allocation strategies? **SSRN Electronic Journal**, 2005.
- DEMIGUEL, Victor; GARLAPPI, Lorenzo; UPPAL, Raman. Optimal versus Naive diversification: how inefficient is the 1/N portfolio strategy? **Review of Financial Studies**, v. 22, n. 5, p. 1915-1953, 2009.
- DIETTERICH, T. G. **Ensemble learning**. In: ARBIB, M. A. (ed.). *The Handbook*. 2002.
- FREUND, Yoav, SCHAPIRE, Robert E. Decision-theoretic generalization of on-line learning and an application to boosting. **Journal of Computer and System Sciences**, v. 55, n. 1, p. 119-139, 1997.
- GRANGER, Clive W.J.; RAMANATHAN, Ramu. Improved methods of combining forecasts. **Journal of Forecasting**, v. 3, n. 2, p. 197-204, 1984.
- HAYKIN, Simon S. **Redes neurais: princípios e práticas**. Porto Alegre: Bookman, 2001.
- HANSEN, Lars Kai; SALAMON, Peter. Neural networks ensembles. **IEEE Transactions on Patterns Analysis and Machine Intelligence**, v. 12, n. 10, p. 993-1001, 1990.
- ISSLER, João; LIMA, Luiz. A panel data approach to economic forecasting: the bias-corrected average forecast. **Journal of Econometrics**, v. 152, p. 153-164, 2008.
- LIU, Na *et al.* Sales forecasting for fashion retailing service industry: a review. **Mathematical Problems in Engineering**, v. 4, p. 1-9, 2013.

MEESE, Richard A.; ROGOFF, Kenneth. Empirical exchange rate models of the seventies: do they fit out of sample? **Journal of international economics**, v. 14, n. 1, p. 3-24, 1983.

NILSSON, Nils J. **Learning Machines**: foundations of trainable pattern-classifying systems. Nova York: McGraw-Hill, 1965.

REN, Shuyun; CHAN, Hau-Ling; RAM, Pratibha. A comparative study on fashion demand forecasting models with multiple sources of uncertainty. **Annals of Operations Research**, v. 257, n. 1, p. 335-355, out. 2017.

SAATY, Thomas L. **Analytic hierarchy process**. Nova York: McGraw-Hill, 1980

SCHAPIRE, Robert E. The strength of weak learnability. **Machine Learning**, v. 5, n. 2, p. 197-227, 1990.

SMITH, Jeremy; WALLIS, Kenneth F. A simple explanation of the forecast combination puzzle. **Oxford Bulletin of Economics and Statistics**, v. 71, p. 331-355, 2009.

THOMASSEY, Sébastien. Sales forecasts in clothing industry: the key success factor of the supply chain management. **International Journal of Production Economics**, v. 128, n. 2, p. 470-483, 2010.

WOLPERT, David H. Stacked generalization. **Neural Networks**, v. 5, n. 2, p. 241-259, 1992.

YOON, Gawon. Forecasting with structural change: why is the random walk model so damned difficult to beat? **Applied Economics Letters, Taylor & Francis Journals**, v. 5, n. 1, p. 41-42, 1998.

ZHANG, Cha; Ma, Yunqian. **Ensemble machine learning**: methods and applications. Nova York: Springer, 2012.

V. Considerações Finais

"Forecasts create the mirage that the future is knowable."

Peter L. Bernstein

O varejo de moda, último elo na cadeia da moda e responsável diretamente pelo atendimento ao consumidor final, nos últimos anos, passou a ser pressionado pelo aumento da competitividade, por clientes mais exigentes, pelo crescimento do nível de incerteza e pela maior dependência de cadeias globais de abastecimento. Nesse cenário, fica evidente a relevância da gestão eficiente da cadeia de abastecimento como um todo, o que passa, necessariamente, pela construção de sistemas de previsão de vendas robustos. Tal aspecto se torna ainda mais relevante, tendo em vista que o setor possui uma cadeia de abastecimento longa, composta por vários fornecedores.

Além das adversidades mencionadas, certas características intrínsecas à indústria de moda fazem com que a previsão de vendas seja mais complexa de ser realizada quando comparada a outros setores. As limitações mais relevantes para o processo de previsão são: a influência do aspecto sazonal; a alta volatilidade da demanda pela crescente mudança nos hábitos dos consumidores, fato alimentado pela própria indústria por meio do lançamento frequente de novos produtos; diversos fatores exógenos que afetam a previsão em si; a granularidade em que a previsão muitas vezes precisa ser gerada em termos de cor e tamanho, aumentando o nível de dificuldade pela limitação do volume de dados existente; o baixo histórico de informação em geral.

No Brasil, o comércio varejista de vestuário é uma atividade econômica muito importante, com faturamento de R\$ 124,5 bilhões, em 2018, o que corresponde a 6,8%, aproximadamente, do total apurado pelo comércio varejista, segundo Pesquisa Anual do Comércio (PAC), enquanto a atividade varejista total equivale a cerca de 11,5% do PIB brasileiro, segundo apuração da Sociedade Brasileira de Varejo e Consumo (2020).

Modelos de previsão de vendas assertivos são essenciais para a gestão das empresas do varejo de moda. De fato, a previsão de vendas permite a organização das compras, dos estoques e da distribuição, de maneira a ampliar a eficiência e a eficácia dos recursos empregados nas atividades comerciais e garantir que o nível de serviço de atendimento, junto ao consumidor, seja mantido em patamares satisfatórios.

O contexto apresentado acima foi o pano de fundo desta Tese, um ensaio sobre previsão de vendas para o setor de varejo de moda, tendo como estudo de caso a rede Lojão do Brás, que forneceu os dados, sendo possível gerar os estudos necessários.

O ensaio é constituído por quatro capítulos, sendo o primeiro introdutório e visa ambientar o leitor sobre o assunto e as especificidades da indústria estudada. Os demais capítulos tiveram propósitos específicos com o intuito de averiguar tratamentos e métodos que ajudassem no processo de previsão para a indústria abordada.

Assim, o segundo capítulo teve como objetivo comprovar o uso de tratamentos sazonais para a geração das previsões em si. Considerando os resultados alcançados, o tratamento sazonal não é recomendado para modelos de árvores de decisão, pois não trouxe ganhos de desempenho em termos de precisão, devendo ser aplicado somente nos modelos baseados em redes neurais, especificamente para os algoritmos CNN e LSTM.

Já o terceiro capítulo buscou verificar o impacto da aplicação dos tratamentos de seleção de variáveis para geração de previsões de vendas mais assertivas. Foi comprovado que o uso de métodos de seleção de variáveis contribuiu para o alcance de previsões mais precisas, sendo o tratamento RFA o mais indicado, apesar do seu alto custo computacional.

Por fim, o objetivo do quarto capítulo foi recomendar o modelo de previsão de vendas com maior acurácia, assim como o melhor modelo de previsão para o setor, baseado em uma análise AHP e considerando como foco o uso de modelos de previsão *ensembles*. Desse modo, a sugestão, em termos de modelo com maior precisão, seria a escolha pelo XGBOOST com tratamento sazonal por ajuste de curva e com tratamento de seleção de variáveis RFA. De acordo com a análise AHP, o modelo de previsão indicado como o melhor para o setor do varejo de moda, englobando cinco critérios e não somente a acurácia, seria o *Naïve* ajustado.

Apesar desta pesquisa ter sido concebida sobre um tema e um setor muito específico, o varejo de moda, e ser fundamentada somente nos dados de uma empresa representante do mesmo, além dos resultados apresentados em cada capítulo, foi possível chegar a algumas conclusões relevantes que podem colaborar para estudos de previsão de uma forma mais ampla em diversas áreas de pesquisa.

Desse modo, o primeiro aspecto diz respeito à questão da incerteza. Conceitualmente, a incerteza, diferentemente do risco, ocorre quando não são conhecidas previamente todas as alternativas e suas respectivas probabilidades de ocorrência. Adicionalmente, para a econometria, a incerteza é vista como um ruído presente nos dados. Segundo Machado (2003), o ruído pode ser definido como o conjunto das influências não-sistemáticas sobre o comportamento de um sistema, não

estando compreendido no modelo determinístico desse sistema. Logo, a presença de ruído afeta negativamente os métodos de previsão.

Como já observado, a demanda da indústria da moda é extremamente incerta e volátil, sendo influenciada por vários critérios. Perante esse contexto e a necessidade de realizar previsões, a aplicação de métodos *ensembles* se mostrou uma solução bastante interessante, pois minimiza exatamente esse aspecto da incerteza, por levar em conta um conjunto de modelos heterogêneos, os quais, por sua diversidade e por estarem juntos, são capazes de gerar previsões mais robustas com menor variabilidade.

Diversos filósofos estudaram o impacto sobre o comportamento humano tanto da presença da incerteza como do conhecimento limitado sobre o mundo. O que mudou atualmente foi que a presença de um ambiente marcado pela incerteza tem se tornado cada vez mais uma constante para uma ampla gama de setores da economia. Yuval Harari (2018), reconhecido historiador e filósofo do século 21, afirma que a mudança é a única constante, o que torna o futuro difícil de prever e ainda mais difícil de planejar. Essa reflexão fica ainda exacerbada com a presença da Covid-19, que gerou uma crise econômica e sanitária para toda humanidade e, mais do que isso, mostrou que qualquer perspectiva de previsibilidade caiu por terra.

Portanto, a incerteza será algo latente e lançar mão de modelos de previsão que consigam lidar melhor com esse aspecto passa ser essencial. Vale salientar que, especificamente, a categoria de *ensemble stacking*, a princípio, seria a mais apropriada por utilizar modelos de previsão heterogêneos, e não homogêneos, como ocorre nas categorias de *ensemble bagging* e *boosting*.

Outro aspecto relevante diz respeito à quantidade de informação que temos acesso nos dias de hoje. Harari (2018) afirma que, em contraste com outras épocas, no século 21, somos inundados por enormes quantidades de informações, tornando-se um ponto crítico filtrar aquelas que são efetivamente relevantes.

Esse aspecto também está presente no que tange à previsão de uma infinidade de séries temporais. A solução proposta para lidar com esse desafio, com base nos resultados alcançado por este trabalho, é composta por três recomendações: (1) alocar a maior parte do tempo de pesquisa em estudos de aprendizagem de máquina, considerando a metodologia KDD, na fase de definição do problema e de preparação dos dados para garantir um entendimento profundo daquilo que se busca modelar, não sendo a aplicação por si só de uma série de algoritmos, mesmo que sejam os mais avançados, que serão capazes de produzir as melhores estimativas, (2) empregar

feature engineering para criar variáveis relevantes para o modelo de previsão proposto, embasadas nos dados brutos e (3) aplicar tratamentos de seleção de variáveis que, apesar de ser uma das possibilidades dentro da fase de preparação de dados, aqui é posto como um elemento essencial para o processo de previsão. Certamente, o uso dessas três recomendações contribui para a construção de modelos de previsão mais parcimoniosos e precisos.

Neste trabalho, a sazonalidade foi considerada de forma determinística e se buscou tratá-la de antemão com o intuito de minimizar seu impacto sobre os modelos de previsão. Com base nos resultados alcançados e, especificamente, para os métodos de aprendizagem de máquina, não foi possível constatar que a eliminação desse aspecto contribuiu efetivamente para alcançar modelos de previsão mais precisos. Isso pode ser justificado pela modelagem da sazonalidade não ter sido realizada da maneira mais adequada ou pelo fato de a sazonalidade ter se tornado um processo menos determinístico, configurando-se parcialmente como um ruído, o que aumenta significativamente a dificuldade da sua especificação.

Outro ponto identificado a ser destacado é que os métodos de aprendizagem de máquina, sobretudo aqueles baseados em redes neurais, têm a sua performance comprometida na presença de séries muito ruidosas, como é o caso das séries estudadas neste ensaio. A questão já havia sido explorada por Haykin (2001), que enfatizou a dificuldade das redes neurais de diferenciar a informação pura do ruído.

Último ponto a ser realçado está relacionado à recomendação final de modelo de previsão mais apropriado para o setor de varejo de moda, que foram os métodos de previsão *Naïves*. Nesse sentido, assim como visto em vários estudos de previsão de séries temporais econômicas e financeiras, muitas vezes não é fácil suplantar o desempenho dos modelos de previsão mais simples. Portanto, tais modelos devem ser vistos como referências a serem usadas como comparação diante de novas propostas.

A maior limitação desta Tese foi a quantidade de dados disponíveis para cada série temporal, sendo que cada uma contou com 680 observações. O volume de dados é aspecto crítico para o treinamento de modelos de aprendizagem de máquina em geral. Outro aspecto que pode ser visto como limitação é a seleção dos métodos de previsão contemplados, tendo em vista a grande quantidade de métodos disponíveis.

Em termos de pesquisas futuras, existe ainda uma série de estudos que poderiam ser realizados no âmbito de previsão de vendas e para o setor de varejo de moda. Esses estudos poderiam ser divididos nas seguintes vertentes: (1) tratamentos

de pré-processamento, (2) melhorias nos métodos de previsão já utilizados e (3) aplicação de outros métodos de previsão.

Na trilha de tratamentos de pré-processamento, poderiam ser aplicados outros tratamentos sazonais, diferentes técnicas de seleção de variáveis, o filtro de Kalman para reduzir a quantidade de ruído presente nas séries estudadas e a transformada *wavelet*, uma ferramenta matemática para decompor funções hierarquicamente. A transformação *wavelet* é utilizada para extrair os componentes de tendência e sazonalidade das séries temporais, permitindo, então, a previsão de dados em diferentes escalas.

Já na trilha de melhorias nos métodos de previsão de vendas utilizados, poderiam ser testados outros modelos de validação dos dados. Ou seja, existe a possibilidade de realizar o treinamento dos modelos nos diversos anos disponíveis de forma separada, tendo em vista a presença de acontecimentos específicos em cada ano. Adicionalmente, isso permitiria realizar também uma análise de consistência.

Em relação a outros métodos de previsão, poderiam ser considerados diferentes modelos híbridos, tais como: a combinação do método ensemble de maior acurácia com o Naive ajustado, a aplicação de lógica *fuzzy* com RNs e o uso de modelos ELM com *grey model*. Ademais, poderia ser considerada a possibilidade de inclusão de métodos qualitativos, a utilização da rede convolucional TCN e a implementação de redes neurais pulsadas (RNPs).

Conforme Maass (1997), as RNPs são aproximadores universais, já que podem aproximar, por meio da codificação temporal, qualquer função contínua de diversas variáveis. A RNP pode ser considerada um modelo apropriado para o problema de previsão de vendas estudado, fundamentado nos seguintes aspectos: forte aspecto temporal, presença de não estacionariedade e alta presença de ruídos.

Cabe ressaltar que a tarefa de fazer previsões não se resume apenas em aplicar modelos em séries de dados e, assim, gerar estimativas que serão utilizadas em decisões relevantes dentro das organizações. Essa atividade é mais complexa do que parece, pois os colaboradores participantes do processo de previsão devem compreender o comportamento da série de dados, o funcionamento dos algoritmos envolvidos, mesmo que de forma superficial, e buscar, assim, através das observações dos resultados produzidos, a melhoria contínua do modelo implantado. Tal fato colabora, efetivamente, tanto para o aumento da confiança sobre o modelo de previsão adotado como para o nível de adesão ao mesmo.

Além disso, no caso dos métodos de previsão baseados em aprendizagem de máquina, o fato desses modelos serem vistos como “caixas pretas” acaba por reforçar o aspecto mencionado sobre a dificuldade de aceitação desse tipo de modelo pelos colaboradores das empresas do varejo de moda. Logo, buscar abordagens diferenciadas e mais convenientes para implementação dessas técnicas de inteligência computacional junto às equipes, que atuam no processo de previsão de vendas no varejo de moda, de forma a garantir um maior nível de aderência, certamente permitirá capturar os ganhos de precisão em sua totalidade de métodos mais avançados.

Por fim, outro enfoque que poderia ser explorado diz respeito à forma de mensurar os ganhos efetivos financeiros da escolha de determinados métodos de previsão no varejo de moda, já que a adoção de métodos com maior acurácia não garante por si só um aumento do lucro obtido pelos varejistas. Segundo Reid, Hussain e Tawfik (2014), para a previsão de séries temporais financeiras, o objetivo também é obter lucros comerciais, além do aspecto da precisão em si.

REFERÊNCIAS

- HAYKIN, Simon S. **Redes neurais: princípios e práticas**. Porto Alegre: Bookman, 2001.
- HARARI, Yuval Noah. **21 lições para o século 21**. São Paulo: Companhia das Letras, 2018.
- MAASS, Wolfgang. Networks of spiking neurons: the third generation of neural network models. **Neural Networks**, Elsevier Publishing, v. 10, n. 9, p. 1659-1671, 1997.
- MACHADO, Karla Fedrizzi. **Módulo de auto-localização para um agente exploratório usando filtro de Kalman**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Rio Grande do Sul, Porto Alegre. 63 p. 2003.
- REID, David; HUSSAIN, Abir; TAWFIK, Hissam. Financial time series prediction using spiking neural networks. **Plos One**, v. 9, n. 8, 2014.
- SBVC. **O papel do varejo na economia**. Disponível em: <http://www.sbvc.com.br/wpcontent/uploads/2015/04/O-Papel-do-Varejonaeconomia_apresenta%C3%A7%C3%A3ooficial1.pdf>. Acesso em: 10 jan. 2020.