

Previsão de vendas no varejo de moda com modelos de redes neurais

**Tema: "Engaging and interactive shopper experience throughout the retail journey"
(tema CLAV 2018)**

Resumo: A melhoria na previsão de vendas é um aspecto crítico para maior parte das organizações, pois permite tornar o processo de planejamento mais eficiente, impactando positivamente os resultados futuros das empresas. Entre as diversas técnicas de previsão, pode-se destacar o grupo de métodos econométricos e o de métodos de inteligência computacional. É neste contexto que surge a questão levantada por este artigo: Quais são as técnicas que apresentam melhor desempenho, quando aplicadas para a previsão de vendas no varejo de moda? Para responder a esta questão, esse trabalho comparou o método atual utilizado pela empresa analisada neste estudo com quatro outros métodos de previsão: dois econométricos, **Naive** e **SARIMA**, e dois de inteligência computacional, rede neural **FDTNN** e rede neural recorrente **LSTM**. A série de tempo escolhida para efetuar as comparações foi a de quantidade semanal de unidades vendidas do produto da categoria blusa de mulher de uma empresa varejista do setor de moda. Essa categoria foi selecionada por ser a mais relevante no faturamento da companhia, correspondendo a cerca de 5,6% do total das vendas. A análise dos resultados gerados pelos métodos aplicados neste estudo, para essa categoria de produto, mostrou que os métodos de inteligência computacional superaram os métodos econométricos e, mais especificamente, a rede neural recorrente LSTM foi a que apresentou o melhor resultado, sendo 50,1% superior ao método atualmente empregado pela empresa. A melhora na previsão de demanda leva à redução dos níveis de estoque, à redução de remarcações, à minimização de rupturas, e ao aumento do giro de estoque, gerando impactos positivos nos resultados financeiros da empresa. No caso deste estudo, a utilização do método LSTM gerou uma redução dos investimentos em estoques de quase 2% e, como consequência, levou a um incremento da margem de contribuição próximo à 4%.

Palavras-chave: Previsão de Vendas; Redes Neurais; Redes Neurais Recorrentes LSTM.

1 INTRODUÇÃO

Diariamente, empresas e organizações enfrentam questões críticas no que se refere a crescimento e a competitividade. Sendo assim, decisões gerenciais exercem um papel fundamental na busca de definições de estratégias e diretrizes com o intuito de atingir maior competitividade e resultados superiores, dentro de um ambiente cada vez mais dinâmico. Para a definição das decisões gerenciais, o processo dos planejamentos em diferentes horizontes de tempo – curto, médio e longo prazo – torna-se essencial.

O planejamento aqui está colocado de uma forma ampla, faz parte dos diversos departamentos de uma empresa, e está presente nos diversos setores de mercado. Para que seja bem-sucedida, a formulação de planejamentos e o direcionamento estratégico das empresas, a identificação e a previsão correta das mudanças no ambiente de negócios são fundamentais, o que torna a previsão de demanda um elemento-chave na tomada de decisão gerencial (Montgomery et al., 1990 e Makridakis et al., 1998). Além disso, as companhias podem melhorar sua eficiência se puderem antecipar problemas e desenvolver planos para evitar ou minimizar seus efeitos. Neste sentido, os métodos de previsão de demanda têm sido desenvolvidos de forma a garantir a antecipação de estados futuros de alguns fatores e/ou variáveis, que afetam o planejamento estratégico das organizações (Armstrong, 1983).

A importância da previsão de demanda está diretamente ligada à racionalização dos recursos envolvidos, e à meta do nível de atendimento aos clientes. Qualquer atividade de planejamento pressupõe previsões sobre o futuro, as quais serão mais assertivas 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).

É fundamental fazer um posicionamento do setor onde este trabalho está inserido. Assim, o 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. O varejo tem forte influência no PIB nacional, uma vez que o seu impacto é de 42,54%, considerando os dados de 2016, conforme a Sociedade Brasileira de Varejo e Consumo (SBVC), e é responsável por 22,3% dos empregos formais do país. Especificamente, em relação ao varejo de moda, setor alvo deste trabalho, ele representa 7,1% do varejo total, é altamente competitivo, uma vez que as cinco maiores empresas detêm 16% do mercado, além disso apresenta alto grau de informalidade, de cerca de 40%, e é afetado por uma série de variáveis exógenas, sendo que as usualmente consideradas como mais relevantes são renda, emprego, crédito e confiança, conforme SBVC (2017).

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 tarefa de previsão de demanda tem ganhado cada vez mais relevância dentro das organizações. A justificativa para isto reside no fato de uma previsão mais assertiva possibilitar o desenvolvimento de uma cadeia de *supply chain* mais integrada e eficiente, gerando menores níveis de estoques, menores remarcações, menores rupturas, 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. Constata-se, então, que o efeito chicote é uma distorção na percepção da demanda pelos participantes de uma cadeia de distribuição, sendo que a intensidade desta distorção é maior para os fornecedores que estão mais distantes dos consumidores finais, representando os elos iniciais de uma rede de fornecimento (Lee et al., 1997).

Cabe ressaltar desde já que a complexidade e a especificidade do varejo têxtil tornam o processo de previsão mais difícil do que em outros setores varejistas. Neste sentido, as características que aumentam o grau de complexidade de previsão neste setor 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 estes fatores tornam o processo de previsão neste setor extremamente complexo e específico de ser realizado (Thomassey, 2010).

Nos últimos anos, temos observado o crescimento de discussões sobre o aumento da importância econômica do consumo baseado na experiência (Backstorme e Johansson, 2006). Trata-se da chamada “economia das experiências”, na qual se busca proporcionar uma experiência memorável para os consumidores. O produto, então, é a própria experiência vivida. Sendo assim, experiências memoráveis, permitem às empresas cobrar um valor superior pela “transformação” vivida pelo consumidor (Pin e Gilmore, 1998).

Criar uma experiência de compra superior é um dos objetivos centrais do varejo atual. Varejistas ao redor do mundo todo perceberam a relevância do conceito da gestão da experiência de compra, incorporando-a às suas missões. A experiência de compra compreende uma visão holística por

natureza e envolve respostas emocionais, afetivas, cognitivas, sociais e físicas aos varejistas. A experiência de compra é criada por elementos controlados ou não pelo varejo, e engloba a experiência por completo, pesquisa, compra, consumo, pós-venda e os diferentes canais de compra no varejo (Verhoef et al., 2009). A seguir, é apresentada a Figura 1, que evidencia um modelo conceitual da criação da experiência de compra, com os seus antecedentes e variáveis moderadoras.

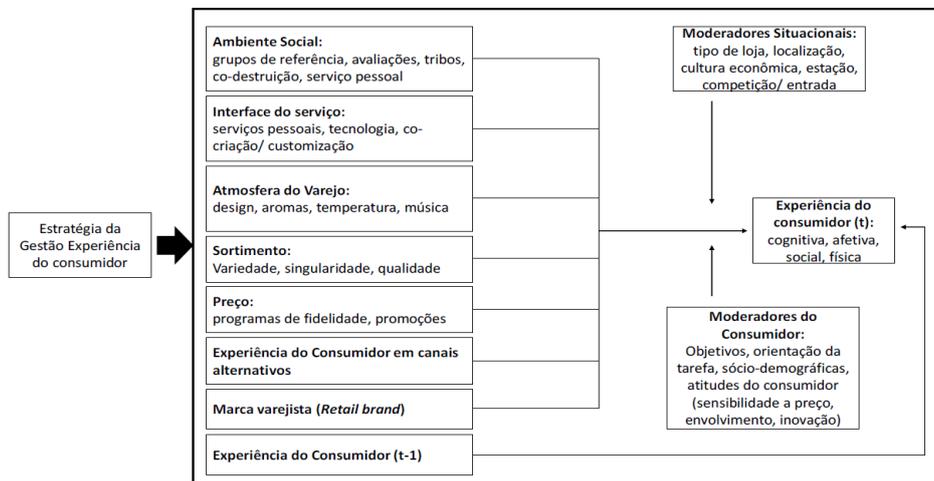


Figura 1 – Modelo conceitual da criação da experiência de compra

Fonte: Verhoef et al., 2009.

O tema de previsão de vendas deste artigo está diretamente conectado com a questão do sortimento adequado de produtos dentro da loja, visto que a variedade de produtos insuficiente contribui negativamente para a experiência, assim como uma variedade em conformidade ao esperado pelo consumidor possui retornos positivos. A falta de produtos básicos (*commodities*) ou a falta de estoque de ofertas especiais afetam de forma negativa (Backstrom e Johansson, 2006). O *display* de produtos, facilitando a compra e o atingimento dos objetivos do consumidor, bem como uma disponibilidade adequada de tamanhos, cores, estilos, etc., proporcionam experiências positivas.

Além disso, um amplo sortimento de categorias pode criar valor para o cliente ao oferecer comodidade e facilidade ao consumidor. Desta forma, quanto maior o sortimento de categorias de produtos, maior a possibilidade de o varejista ser lembrado nas diferentes situações de compra. E, realizar as compras num único lugar (*one-stop shopping*) tem se tornado mais relevante para o consumidor com o tempo mais limitado (Messinger e Narasimhan, 1997; Ailawadi e Keller, 2004).

Em termos de análise de profundidade de produto, pode-se constatar que ela influencia na imagem percebida da loja, na escolha da loja, e na satisfação com a loja em si. Contudo, um grande número de SKUs (*Stock Keeping Unit*) não necessariamente melhora essas percepções. A redução do número de SKUs não reduz a percepção de variedade do consumidor, caso seja mantido seu produto preferido ou não reduzindo o espaço total da categoria. Nesse caso, uma redução moderada no número de SKUs pode, inclusive, aumentar as percepções dos consumidores sobre o sortimento da loja (Broniarczyk et al., 1998).

A partir do observado acima, verifica-se que previsões mais assertivas contribuem para manter o nível adequado de estoques, mais do que isto, tem como objetivo principal garantir que o produto certo e na quantidade adequada esteja no ponto de venda correto, em conformidade com as demandas de consumo do cliente, em termos de sortimento e profundidade, contribuindo para retornos positivos da experiência de compra do consumidor.

Diante deste quadro, pode-se dizer que a principal motivação do trabalho é encontrar modelos de previsão mais assertivos, com o intuito de garantir uma margem de erro menor do que os modelos tradicionais, os quais, conseqüentemente, assegurarão melhores resultados para as companhias deste setor. Neste sentido, espera-se que técnicas mais contemporâneas como as de redes neurais, que são técnicas mais flexíveis, genéricas e abrangentes, sejam também mais precisas.

Os demais objetivos deste trabalho são: (i) verificar as restrições e limitações existentes para a aplicação dos métodos de previsão de vendas para o mesmo setor; (ii) identificar os métodos de previsão de vendas empregados no varejo de moda; (iii) aplicar os diversos métodos de previsão para a série temporal escolhida para este trabalho; (iv) comparar a acuracidade obtida com o uso dos métodos de inteligência computacional em relação aos métodos econométricos; e (v) comparar os modelos econométricos e de inteligência computacional com o método atual vigente na empresa Lojão do Brás, que forneceu os dados utilizados nos experimentos desenvolvidos neste trabalho.

2 REFERENCIAL TEÓRICO

A indústria da moda adota técnicas de previsão de vendas semelhantes às utilizadas por outros ramos de atividades, e esse repertório de técnicas pode ser agrupado em estratégias que fazem uso de métodos heurísticos, de métodos econométricos, ou de métodos de inteligência computacional (Choi et al., 2014).

O presente trabalho analisou e implementou modelos baseados em cada um desses três grupos de técnicas. Esta seção começa com a descrição do modelo heurístico atualmente utilizado pelo Lojão do Brás, empresa foco deste estudo. Na seqüência, estão destacadas as referências teóricas dos métodos econométricos e de inteligência computacional.

2.1. Método heurístico

Dada a relativa complexidade de implantação de métodos quantitativos (econométricos e de inteligência computacional), muitas empresas decidem pela implantação de métodos heurísticos, que se baseiam na experiência acumulada pela empresa na compreensão do setor de atuação, e empregam este conhecimento acumulado para construir modelos próprios com bases empíricas.

Tal fato pôde ser observado no Lojão do Brás. O processo de previsão de vendas na empresa estudada, atualmente é feito com periodicidade semanal, no horizonte de um ano, e usa, como referência, as vendas do ano anterior. Em termos de agregação, a previsão é realizada no nível de categoria de produto – por exemplo, blusa de mulher – e, depois, desdobrada em termos da quantidade de modelos que cada loja tem capacidade de receber. Não ocorre a realização de previsões no nível de item (SKU).

Após a consideração da referência das vendas do ano anterior, é verificada a previsão de crescimento da economia para o ano da previsão, por meio do relatório Focus do Banco Central, o qual é ajustado de acordo com a expectativa da diretoria da empresa, considerando o histórico da relação entre o PIB e o crescimento obtido pela empresa ao longo dos anos. Então, esta variação é aplicada ao valor de referência para se determinar a previsão final.

2.2. Métodos econométricos

Os métodos econométricos têm como estratégia básica modelar comportamentos de séries de tempo estacionárias, utilizando 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 utilizados neste artigo foram **Naive** e **SARIMA**.

O método **Naive** é o que utiliza o pressuposto mais simples entre todos os métodos econométricos. Ele considera que a previsão nada mais é do que o último valor observado, sem acréscimo de choques ou incertezas. Nesse caso, a previsão do valor y_{t+1} para o período $t + 1$, é dada por:

$$y_{t+1} = y_t,$$

em que y_t , é o valor observado no período anterior, t .

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 apresentada a fórmula deste método. A previsão do valor y_t para o período t , é dada por:

$$y_t = \phi_1 y_{t-s} + \phi_2 y_{t-2s} + \dots + \phi_P y_{t-Ps} + \varepsilon_t - \theta_1 \varepsilon_{t-s} - \theta_2 \varepsilon_{t-2s} - \theta_Q \varepsilon_{t-Qs},$$

em que:

y_τ : valor previsto para o período τ , se $\tau = t$; ou valor observado no período τ , se $\tau < t$;

s : termo do período sazonal;

P : ordem (número de defasagens) do modelo autorregressivo sazonal;

Q : ordem do modelo de média móvel sazonal;

ϕ_i : i -ésimo parâmetro da parte autorregressiva sazonal do modelo, $i = 1, \dots, P$;

ε_{t-js} : j -ésimo termo de erro, $j = 0, \dots, Q$ para dados t e s ;

θ_k : k -ésimo parâmetro da parte de média móvel sazonal, $k = 1, \dots, Q$.

2.3. Métodos de inteligência computacional

Dentre o vasto repertório de métodos de inteligência computacional para a previsão de séries de tempo, para este trabalho foram escolhidos aqueles baseados em redes neurais *Multi-Layer Perceptron* (MLP), de dois tipos diferentes: a rede neural *Focused Time Delay Neural Network* (FTDNN) e a rede neural recorrente *Long-Short Term Memory* (LSTM).

A escolha da classe de redes neurais MLP foi baseada no bom desempenho, robustez e capacidade de generalização desses tipos de modelos atestados pela literatura especializada, que destaca como característica positiva das MLP, a capacidade de captar relações não-lineares entre as variáveis de entrada, a ponto de serem consideradas como potenciais aproximadores universais de funções (Haykin, 2001).

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.

As redes FTDNN são ditas do tipo *feedforward*, isto é, durante o treinamento da rede, as variáveis de entrada são submetidas ao modelo sequencialmente, apenas uma vez. Redes FTDNN podem ser interpretadas como modelos autorregressivos com formulação não-linear (Kim e Shin, 2007).

As redes LSTM, por sua vez, são do tipo recorrente, ou seja, neste caso, o modelo pode ser configurado para ser realimentado com as variáveis de entrada mais de uma vez. Além disso, as redes LSTM possuem uma estrutura interna a mais do que as demais redes MLP, denominada *forget gate* (portão de esquecimento), o que permite decidir se um valor deverá ser ignorado pelo processo de treinamento ou não. Dentro do contexto de previsão de séries de tempo, a capacidade de descartar valores permite operar diferentemente informações mais recentes e mais antigas ao longo da série de tempo, o que pode de ser interpretado como uma capacidade de dar tratamentos distintos para as memórias de curto e de longo prazo (Gers et al., 2000).

3 MÉTODO

A metodologia utilizada é a CRISP-DM (*Cross industry standard process for data mining*). Ela foi concebida em 1996 por um consórcio formado por empresas como Daimler-Chrysler, SPSS e NCR para ser uma metodologia estruturada e robusta para planejamento e realização de projetos de mineração de dados. Esta metodologia é composta de seis etapas: (1) entendimento do negócio, (2) compreensão dos dados, (3) preparação dos dados, (4) modelagem, (5) avaliação e (6) desenvolvimento (Chapman et al., 1996). A seguir, é apresentada a Figura 2, que resume esta metodologia, e detalha cada um dos seus passos.

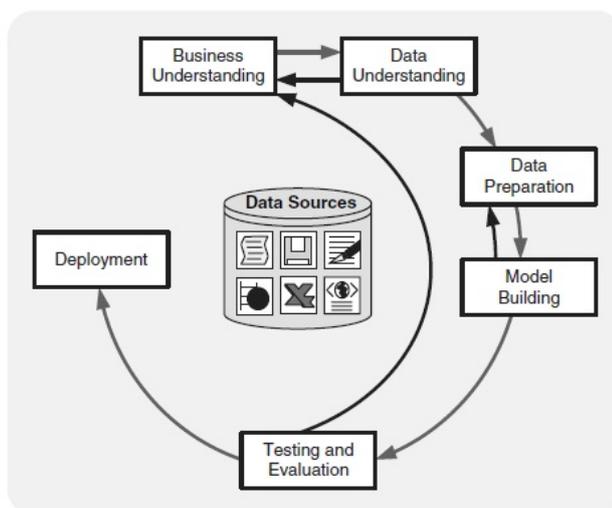


Figura 2 – Metodologia CRISP-DM
Fonte: Chapman et al. (1996), p.10.

Etapa 1 - Entendimento do negócio

Nesta etapa, o profissional deve buscar um entendimento profundo do problema que necessita ser resolvido. É preciso buscar detalhes sobre como a questão afeta a organização, e quais são os principais objetivos e expectativas em relação ao trabalho como um todo.

No caso deste trabalho, o problema da empresa varejista Lojão do Brás é realizar a previsão de vendas na periodicidade semanal, no nível agrupado, para a categoria de produto blusa de mulher, no curto prazo. De forma mais detalhada, o objetivo é conseguir gerar previsões para a próxima semana, implementando métodos econométricos e métodos de inteligência computacional, com o intuito de encontrar o método com a maior acuracidade, a qual será mensurada pela média dos

erros percentuais absolutos (MAPE) e pela raiz da média dos erros quadrados (RMSE).

Esta série foi a escolhida para ser estudada, pois é a série mais importante para a empresa Lojão do Brás, já que possui a maior participação em termos de vendas dentre as 965 séries de produtos existentes, representando especificamente 5,6% do faturamento total da companhia em 2017 e 5,2% em 2016.

Etapa 2 – Compreensão dos dados: todas as variáveis

Após a primeira etapa, o próximo passo passa a ser analisar, organizar e descrever todos os dados disponíveis para que sejam utilizados na solução do problema em questão. A escolha dos dados relevantes para modelar a solução do problema selecionado é uma das funções relevantes do gerente do projeto.

Para o problema de previsão de vendas escolhido, foram selecionados os dados de quantidade vendida em unidades para a categoria de produto blusa de mulher. Para esta variável, foi adotada a periodicidade semanal, contemplando as semanas de 2007 a 2016, sendo que as semanas de 2007 a 2014 foram utilizadas para treinar os modelos propostos e as semanas de 2015 e 2016 para testar os respectivos modelos. Desde já, é fundamental destacar a presença de uma forte sazonalidade na série, sendo que os principais picos de vendas são observados nas semanas de dezembro, novembro e maio respectivamente.

A Figura 3 exibe o histograma da série de quantidade vendida de blusa de adulto feminina, no qual é possível constatar a presença de assimetria à esquerda, já que a média supera a mediana, o que é justificado, sobretudo, pelas vendas das semanas de dezembro, as quais apresentam valores muito superiores, elevando o valor da média e a presença de uma alta dispersão, o que também pode ser explicada pelos picos de vendas.

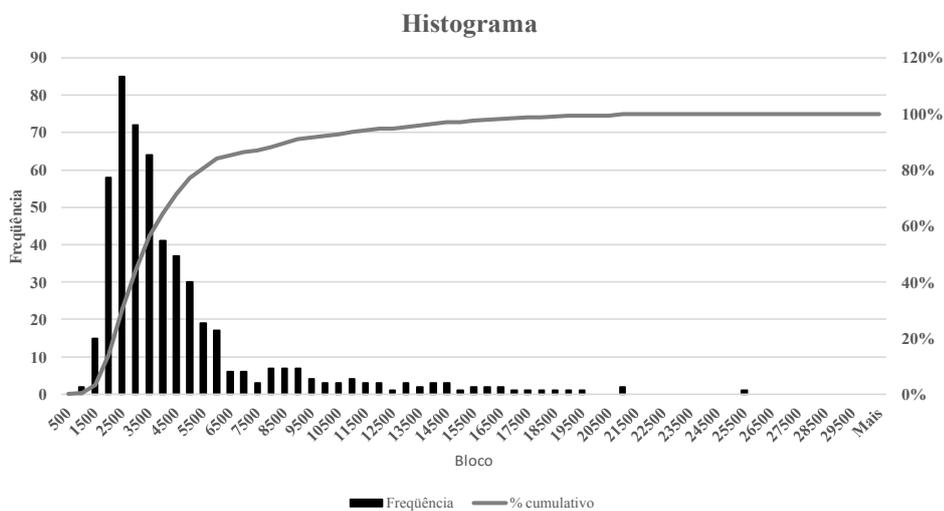


Figura 3 – Histograma série blusa de mulher
Fonte: elaboração própria.

Etapa 3 – Preparação dos dados

Depois de ter sido feita uma análise descritiva da série, o próximo passo consiste em realizar a preparação efetiva dos dados a serem utilizados, o que significa realizar os tratamentos e ajustes prévios para que os dados possam ser utilizados.

Neste sentido, para o problema de previsão de vendas tratado, no caso dos métodos econométricos, foi necessário verificar se a série de blusa de mulher era estacionária por meio dos testes de raiz unitária: ADF, PP e KPSS. De forma surpreendente, verificou-se que a série

estudada era estacionária, não sendo necessário aplicar o tratamento de diferenciação para implementação dos métodos econométricos. Vale ressaltar que a estacionariedade é um pré-requisito para a utilização da metodologia Box-Jenkins.

No caso da utilização dos métodos de inteligência computacional, foi aplicado o processo de normalização para a série em questão, que consiste no redimensionamento dos dados para que eles fiquem entre o intervalo de zero e um, tendo como base os valores máximo e mínimo. A normalização tem como objetivo reduzir o tempo de treinamento das redes neurais.

Etapa 4 – Modelagem:

Nesta etapa, são selecionadas e aplicadas as técnicas de previsão mais apropriadas, dependendo dos objetivos levantados na primeira fase. Conforme destacado na Seção 2, para ser comparado com o método heurístico utilizado atualmente pela empresa, foram selecionados os quatro métodos abaixo:

- Métodos econométricos: Naive e SARIMA;
- Métodos de inteligência computacional: rede FDTNN e rede LSTM.

Mais especificamente, foram processados 261 modelos no total para a série em estudo.

No caso do método econométrico SARIMA, diante da possibilidade de serem produzidas várias alternativas de modelos, é importante que seja estabelecida uma métrica, que possa determinar o número adequado de parâmetros para o modelo e, desta forma, tornar possível a seleção do modelo mais apropriado. Segundo Bueno (2008), a ideia por trás do critério de informação é a de minimização de uma função baseada nos resíduos e penalizada pelo número de repressores. O melhor modelo será o mais parcimonioso, satisfazendo a condição de os resíduos apresentarem os menores valores possíveis. Conforme o autor, há três principais critérios de informação: a estatística de Schwarz (BIC), a estatística de Akaike (AIC) e a estatística Hanna-Quinn (HC). Neste trabalho foi adotado o critério de Akaike.

No caso dos métodos de inteligência computacional, foi realizado um procedimento de validação, que consiste na replicação de cem arquiteturas do modelo que apresentou o menor MAPE de teste, com o intuito de efetivamente conferir os respectivos resultados encontrados. Mais especificamente, a replicação foi feita através da alteração do valor da semente. Posteriormente, foi 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 e Chakraborti, 2003). Este 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 foi utilizado para comparar se as médias obtidas no processo de validação com as cem arquiteturas para os dois tipos de redes neurais eram distintas.

Em termos de software, foram utilizados o R para os métodos econométricos e o Python, com as bibliotecas do Tensorflow e do Keras, para os métodos de inteligência computacional. Os dois softwares são amplamente utilizados para *machine learning* e *data Science*, e são linguagens amigáveis, simples, produtivas, abertas e extensíveis. A seguir, na Tabela 1, é apresentado um quadro resumo com o detalhamento das configurações avaliadas em cada método.

Tabela 1 – Alternativas de configurações para cada método de previsão

Método	Modelos Testados	Software
1. Naive	Avaliada 1 alternativa, que considera a última defasagem como próxima previsão	R
2. SARIMA	Avaliadas 7 alternativas, por meio da implantação da metodologia Box-Jenkins	R
3. Atual	Avaliada 1 alternativa, em que o modelo considera o valor do ano anterior com baseline e ajusta por um fator	R
4. FTDNN	Avaliadas 126 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Número de defasagens de entrada: 52; • Número de camadas escondidas: 1; • Número de camadas de saída: 1; • Neurônios na camada escondida: 6,13,26,52,104,156,208; • Taxa de aprendizado: 0.001, 0.01, 0.1; • Funções de ativação da camada escondida: Sigmoid e RELU; • Otimizadores: SGD, ADM e RMSProp. 	Python
5. LSTM	Avaliadas 126 alternativas com base na variação dos seguintes parâmetros: <ul style="list-style-type: none"> • Número de defasagens de entrada: 52; • Número de camadas escondidas: 1; • Número de camadas de saída: 1; • Blocos de memória: 6,13,26,52,104,156,208; • Taxa de aprendizado: 0.001, 0.01, 0.1; • Funções de ativação da camada escondida: Sigmoid e RELU; • Otimizadores: SGD, ADM e RMSProp. 	Python

Fonte: elaboração própria.

Etapa 5 – Avaliação:

Nesta fase é realizada a apresentação dos resultados alcançados, a avaliação efetiva do modelo construído, e são evidenciadas as conclusões obtidas.

Para avaliar a acuracidade dos modelos construídos, adotamos as métricas de média dos erros percentuais absolutos (MAPE) e de raiz da média dos erros quadrados (RMSE) por serem medidas usualmente utilizadas em trabalhos relacionados com redes neurais (Zhang et al., 1998). Tendo em vista a presença de picos de vendas na série estudada, optou-se pela utilização do MAPE como principal medida, pois é uma medida que relativiza o erro. A seguir, são apresentadas as equações do MAPE e do RMSE:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|, \quad RMSE = \left[\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right]^{\frac{1}{2}}$$

em que, y_t é o valor observado da variável y no período t , \hat{y}_t é o valor previsto pelo modelo da variável y no período t , e n é o número de observações.

Etapa 6 - Desenvolvimento:

O conhecimento adquirido por meio do processo de mineração dos dados e do desenvolvimento dos modelos de predição deve ser aplicado de forma prática. Assim, nesta última etapa da metodologia, são feitos os ajustes necessários para a realização de uma entrega mais palpável e aplicável ao cliente, a partir das análises dos dados feitas pela equipe do projeto.

Para este trabalho, foi realizada uma integração entre as previsões feitas e o sistema de *business intelligence* utilizado pela varejista estudada, de forma que a previsão de demanda realizada para a série selecionada pudesse ser analisada e consumida pelo setor comercial da companhia, tornando possível ajustar o nível de estoque de forma apropriada.

4. RESULTADOS

Antes de serem discutidos os resultados quantitativos (em valores de MAPE e RMSE), na Tabela 2, apresentamos as especificações das melhores configurações encontradas para cada método.

Tabela 2– Configurações dos modelos

Método	Especificação do Melhor Modelo (determinado na fase de treino)
2. SARIMA	SARIMA(2,0,2)(1,1,0)52
4. FTDNN	MLP - n=104, $\eta = 0.001$, relu, adam
5. LSTM	LSTM - n=156, $\eta = 0.01$, relu, adam

Fonte: elaboração própria.

Com base na Tabela 2, é possível afirmar que, no caso dos métodos econométricos, foram necessários poucos termos autorregressivos e de médias móveis para se determinar as configurações mais satisfatórias. Em relação às redes neurais, constatou-se a necessidade de pelo menos 104 defasagens para entender o padrão da série estudada, a preponderância da função de ativação *Relu* na camada intermediária e do otimizador *Adam* nas duas arquiteturas de redes neurais utilizadas. Com relação à taxa de aprendizado, no caso da FTDNN, a melhor configuração foi treinada com uma taxa de aprendizado de 0,001, já na rede neural recorrente LSTM, a taxa de aprendizado foi de 0,01.

Após a especificação das melhores configurações para cada método da série de blusa de mulher, pôde-se verificar que os métodos de inteligência computacional conseguiram superar os resultados alcançados pelos métodos econométricos e pelo método atual utilizado pela empresa. É importante ressaltar que tal fato ocorreu tanto na fase de treino como na fase de teste. Um aspecto interessante é que os métodos Naive e o atual conseguiram apresentar resultados melhores que os métodos econométricos na fase de teste; o mesmo não ocorreu no período de treinamento. A Figura 4 exibe um gráfico auxiliar para a análise comparativa dos resultados.

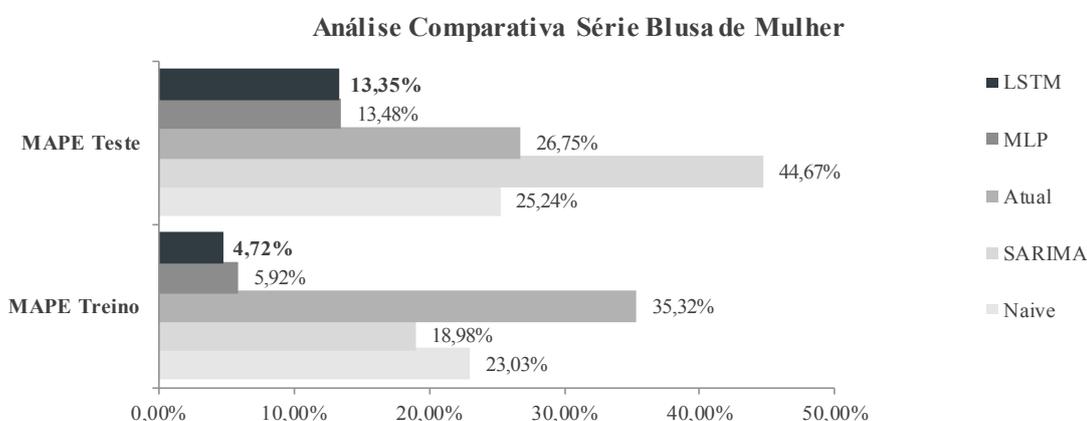


Figura 4 – Análise comparativa dos métodos de previsão

Fonte: elaboração própria.

No caso dos métodos de inteligência computacional, o método que produziu o melhor resultado (menor valor de MAPE) na fase de treino foi o método LSTM, seguido de perto pelo método FTDNN. Tal resultado foi o mesmo observado na fase de teste. Além disso, como previamente descrito na etapa de modelagem, foi realizado o processo de validação por meio da execução de cem treinamentos, com modificação dos pesos iniciais, das melhores configurações dos modelos FTDNN e LSTM, sendo constatado pelo teste de Wilcoxon, com nível de significância de 5%, que a média dos valores MAPE produzidos pelos modelos LSTM é estatisticamente menor do que a gerada pelos modelos FTDNN, confirmando que o método LSTM realmente superou o FTDNN em termos de acuracidade. A seguir, é apresentada a Tabela 3 com os resultados médios obtidos e o resultado do teste de Wilcoxon:

Tabela 3 – Validação das redes neurais

	FTDNN		LSTM	
	Média	Desvio Padrão	Média	Desvio Padrão
MAPE	13,48%	0,93%	13,35%	1,22%
RMSE	787,30	85,83	765,12	67,31

Teste Wilcoxon:

Statistics = 1359.00, p = 0.000

Rejeita-se Ho: as amostras possuem distribuições distintas

Fonte: elaboração própria.

O método atual utilizado pela empresa foi suplantado pelos métodos Naive, FTDNN e LSTM. Em relação ao método com melhor desempenho, o LSTM, houve um incremento de precisão de 50,1% na fase de teste, quando comparado ao método atual vigente na empresa. A Tabela 4 mostra os valores MAPE e RMSE produzidos pelos modelos no período de teste. É possível observar que os métodos de inteligência computacional (LSTM e FTDNN) superam todos os demais, e que o método atual supera os métodos econométricos (SARIMA e Naive).

Tabela 4 – Análise comparativa por meio do MAPE (%) e do RMSE dos métodos empregados na amostra de teste

	MAPE (%) de Teste	RMSE de Teste
Naive	25,24%	1.497,92
SARIMA	44,67%	1.974,12
Atual	26,75%	1.197,23
FTDNN	13,48%	787,30
LSTM	13,35%	765,12

Fonte: elaboração própria.

Por fim, deve-se reiterar que o método que apresentou mais acuracidade para a série de blusa de mulher foi o método LSTM, segundo o critério de menor valor MAPE na fase de teste dos diversos métodos utilizados. Na Figura 5, é apresentado um gráfico evidenciando os dados reais contra os dados previstos do método LSTM, na fase de treino e de teste.

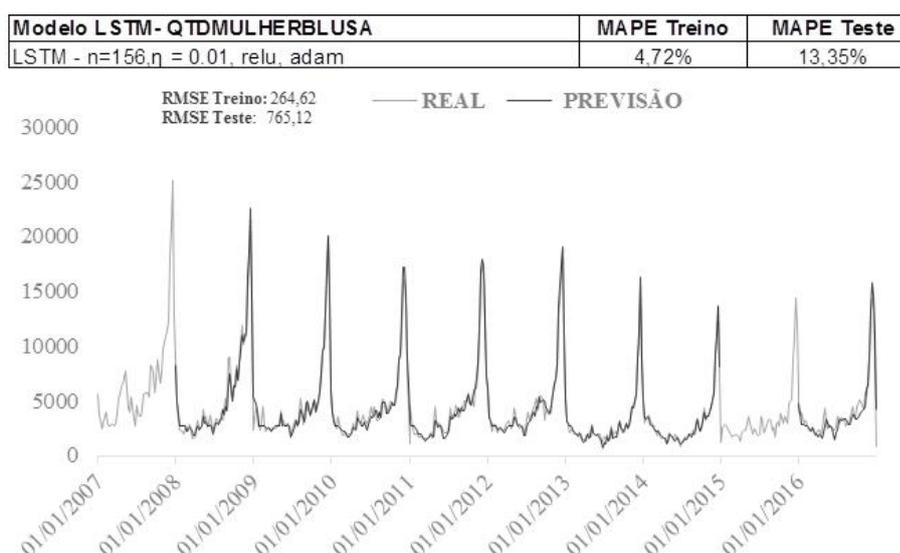


Figura 5 – Gráfico com resultados do método LSTM

Fonte: elaboração própria.

5. CONCLUSÕES

Os objetivos deste trabalho foram: (i) verificar as restrições e limitações existentes para a aplicação dos métodos de previsão de vendas para o mesmo setor; (ii) identificar os métodos de previsão de vendas empregados no varejo de moda; (iii) aplicar os métodos de previsão para a série temporal que compõe este trabalho; (iv) comparar a acuracidade obtida com o uso dos métodos de inteligência computacional em relação aos métodos econométricos; e (v) comparar os modelos econométricos e de inteligência computacional com o método atual vigente na empresa.

Foi constatado que já existe a aplicação de diversos métodos de previsão de vendas para o varejo de moda além dos métodos heurísticos, que vão desde os métodos econométricos até os métodos de inteligência computacional, que têm as redes neurais como um dos seus representantes. Em relação às restrições e dificuldades de previsão de vendas no varejo de moda, verificou-se que este setor é afetado por diversos fatores, tais como clima, sazonalidade, preço e promoções praticados pela empresa e pela concorrência, definição do nível de agregação para a realização da previsão, ciclo de vida de produtos, que muitas vezes limita o volume de histórico de dados existente para a efetivação da previsão, e por mais uma série de variáveis exógenas exploratórias.

Um aspecto extremamente relevante foi a constatação de que a série de quantidade vendida de blusa de mulher analisada neste estudo é estacionária, não sendo necessário passar pelo processo de diferenciação. Tal fato é fundamental para a aplicação adequada e satisfatória dos métodos de previsão de séries de tempo.

Após a construção dos cinco métodos de previsão avaliados neste trabalho – atual, Naive, SARIMA, FTDNN e LSTM – foi realizada a comparação da acuracidade entre os métodos, tendo como critério o erro percentual médio absoluto (MAPE). Com base na avaliação da acuracidade, pode-se afirmar que os métodos de inteligência computacional e, mais especificamente, a LSTM apresentou o melhor resultado.

Em comparação ao modelo atual de previsão de vendas vigente na empresa Lojão do Brás, verificou-se que somente os métodos de inteligência computacional conseguiram alcançar resultados superiores. O acréscimo de precisão do método LSTM foi de 50,1%, já que foi observado a queda no valor do MAPE produzido no período de teste exatamente nesta proporção, quando comparado com o método corrente. Em outras palavras, o erro gerado no processo de previsão de vendas foi reduzido pela metade com a aplicação do método LSTM comparado ao método atual. O ganho apresentado foi bastante relevante, entretanto foi apenas observado para a categoria de produto selecionada, não existindo garantia, a priori, da repetição de tal comportamento para outras categorias de produto. Em termos de impactos financeiros, a maior acuracidade da previsão de vendas trouxe uma redução de investimento em estoques de 1,9%, e um incremento da margem de contribuição de 3,9%, evidenciando os benefícios da adoção de melhores práticas de previsão.

5.1 Limitações

Provavelmente, a maior limitação enfrentada por este trabalho foi a pequena quantidade de dados disponíveis para a série temporal (524 observações). De acordo com a teoria, para a utilização de métodos baseados em redes neurais, quanto maior o volume de dados existente, maior a chance de melhora no processo de aprendizagem. Desta forma, seria importante ter um quantidade maior de observações para a aplicação dos métodos de inteligência computacional.

Outra limitação da pesquisa foi o tempo gasto para o treinamento das redes neurais recorrentes LSTM. Certamente, a utilização de GPUs mais potentes possibilitaria a construção e análise de

uma variedade maior de modelos.

5.2 Pesquisas futuras

Em termos de pesquisas futuras, existem muitos estudos que podem ser realizados como uma continuação deste artigo. Tais trabalhos podem ser divididos em três vertentes: (1) tratamento dos dados de entradas e inclusão de variáveis exógenas; (2) variações e ajustes dos métodos utilizados; e (3) aplicação de outros métodos.

Na vertente de tratamento de dados, uma possibilidade seria dessazonalizar os dados e verificar se as redes neurais produzir previsões mais precisas. Além disso, seria interessante repetir os experimentos utilizando outras categorias de produtos, visando validar os resultados obtidos neste estudo. Outra melhoria importante seria testar a inclusão de variáveis exógenas com o objetivo de encontrar *inputs* que pudessem efetivamente ajudar na construção de modelos de previsão mais precisos.

Já na vertente de variações e ajustes dos métodos utilizados, mais especificamente para os métodos de inteligência computacional, poder-se-ia aplicar alguma técnica de regularização, como a de *dropout*, com o objetivo de verificar se as redes neurais terão uma maior generalização e, possivelmente, apresentar uma maior acuracidade. Outra sugestão seria a construção de RNNs recorrentes LSTM empilhadas, que são as redes LSTM com mais de uma camada intermediária e/ou as redes LSTM *Stateful*, que têm as suas memórias reiniciadas em determinados momentos, como, por exemplo, no início de cada ano.

Em relação à vertente de aplicação de outros métodos, deve-se considerar a existência de uma grande quantidade de técnicas de previsão de vendas e a possibilidade da utilização de modelos híbridos, que levam em conta, na sua implementação, a utilização de mais de um modelo. O campo de estudos de métodos de previsão é bastante produtivo, constantemente propondo novos métodos e soluções para alcançar resultados superiores, e contribuir para aumentar a eficiência das companhias. Assim, existem muitos métodos que podem ser testados, tais como: Lógica *Fuzzy*; *Extreme Learning Machine* (ELM), que é uma rede neural MLP que não utiliza o *backpropagation* como algoritmo de treinamento; Algoritmos Genéticos e alguns métodos híbridos, como o MLP com otimização por meio de Algoritmo Genético; e o método *Neuro-Fuzzy*, que utiliza de forma híbrida as técnicas de Rede Neural e Lógica *Fuzzy*.

Por fim, cabe ressaltar que um grande desafio será a implementação efetiva dos métodos de inteligência computacional na própria organização estudada, tendo em vista sua complexidade e a falta de processo para integração das previsões geradas com os sistemas da empresa. Tal aspecto, provavelmente, é um ponto crítico não apenas para o Lojão do Brás, mas também para outras empresas do setor. Isto poderia motivar um trabalho de cunho exploratório com o intuito de verificar quais abordagens possíveis e mais convenientes para implementação destas técnicas mais avançadas, de forma a garantir um maior nível de aderência por parte das equipes presentes nas empresas, permitindo então, que os ganhos de precisão destes novos métodos fossem aproveitados em sua totalidade.

REFERÊNCIAS

- ARMSTRONG, J. Strategic Planning and Forecasting Fundamentals. In: ALBERT, K. The Strategic Management Handbook. New York: MacGraw Hill, 1983, p.1-32.
- AILAWADI K.L., KELLER, K.L. Understanding retail branding: conceptual insights and research priorities. *Journal of Retailing*, Vol. 80 pp.331-42.2004
- BACKSTORM, Kristina; JOHANSSON, Ulf. Creating and consuming experiences in retail

- store environments: Comparing retailer and consumer perspectives. *Journal of Retailing and Consumer Services*, 13, p. 417-430, 2006.
- BRONIARCZYK, Susan, HOYER, Wayne, McAlister, Leigh. (1998). Consumers' perceptions of the assortment offered in a grocery category: The impact of item reduction. *Journal of Marketing Research*, 35(May), 166–176.
- BUENO, R. L. S. *Econometria de séries temporais*. 2. ed. São Paulo: Cengage Learning, 2008.
- CHOI, Tsan-Ming; HUI, Chi-Leung; YU, Young. *Intelligent fashion forecasting systems: models and applications*. New York: Springer, 2014.
- CHAPMAN, Pete; CLINTON, Julian; KERBER, Randy; Khabaza, THOMAS; REINARTZ, Thomas; SHEARER, Colin; WIRTH, Rüdiger. *CRISP-DM 1.0 – Step-by-step data mining guide*. Disponível em : < <https://the-modeling-agency.com/crisp-dm.pdf>> Acesso em: 27 abr. 2018.
- DIAS, George Paulus Pereira. *Proposta de Processo de Previsão de Vendas para Bens de Consumo*. Disponível em: <www.abepro.org.br/biblioteca/ENECEP1999_A0687.PDF> Acesso em: 10 mar. 2018.
- GERS, Felix A.; SCHMIDHUBER, Jürgen; CUMMINS, Fred. Learning to Forget: Continual Prediction with LSTM. *Neural Computation*, v. 12, n. 10, p. 2451–2471, 2000.
- GIBBONS, J.D; CHAKRABORTI, S. *Nonparametric Statistical Inference - Fourth Edition, Revised and Expanded*. Statistics Textbooks and Monographs, 168. Nueva York: Marcel Decker, Inc.
- HAYKIN, Simon S. *Redes neurais: princípios e práticas*. Porto Alegre: Bookman, 2001.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997.
- 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*, [S.L.], v. 7, n. 2, p.569-576, 2007.
- LEE, H. L.; PADMANABHAN, V.; WHANG, S. Information distortion in a supply chain: the bullwhip effect. *Management Science*, v. 43, n. 4, p. 546-558, 1997.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. *Forecasting – methods and applications*. 3. ed. Nova York: John Wiley, 1998.
- MESSINGER, Paul R., NARASIMHAN, Chakravarthi. (1997). A model of retail formats based on consumers' economizing on shopping time. *Marketing Science*, 16(1), 1–23.
- MONTGOMERY, D.; JOHNSON, L.; GARDINER, J. *Forecasting and Time Series Analysis*. Nova York: McGraw-Hill, 1990.
- PINE II, Joseph B; GILMORE, James H. *The Experience Economy*. Harvard Business School, 1998.
- SOCIEDADE BRASILEIRA DE VAREJO E CONSUMO. *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: 20 out. 2017.
- 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.
- VERHOEF, Peter C., LEMON, Katherine N., PARASURAMAN A., ROGGEVEEN, Anne, TSIROS, Michael, SCHLESINGER, Leonard A. Customer Experience Creation: Determinants, Dynamics and Management Strategies. *Journal of Retailing*, 85 (1, 2009), 31-41.
- ZHANG, G. P.; PATUWO, E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, v. 14, p. 35-62, 1998.