FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO

ÉLIA YATHIE MATSUMOTO

**APLICAÇÃO DE REDES NEURAIS NA CLASSIFICAÇÃO DE
RENTABILIDADE FUTURA DE EMPRESAS**

SÃO PAULO
2008

ÉLIA YATHIE MATSUMOTO

**APLICAÇÃO DE REDES NEURAIS NA CLASSIFICAÇÃO DE
RENTABILIDADE FUTURA DE EMPRESAS**

Dissertação apresentada à Escola de Economia da Fundação Getúlio Vargas (FGV/EESP) como requisito para obtenção do título de Mestre em Finanças e Economia Empresarial.

Orientador: Prof. Afonso de Campos Pinto

SÃO PAULO

2008

Matsumoto, Élia Yathie.

Aplicação de Redes Neurais na classificação de rentabilidade futura de empresas / Élia Yathie Matsumoto. - 2008.

55 f.

Orientador: Afonso de Campos Pinto.

Dissertação (Mestrado profissional) - Escola de Economia de São Paulo.

1. Redes neurais (Computação). 2. Investimentos - Processo decisório. 3. Empresas - Lucratividade. I. Pinto, Afonso de Campos. II. Dissertação (Mestrado profissional) - Escola de Economia de São Paulo. III. Título.

CDU 658.155

ÉLIA YATHIE MATSUMOTO

**APLICAÇÃO DE REDES NEURAIS NA CLASSIFICAÇÃO DE
RENTABILIDADE FUTURA DE EMPRESAS**

Dissertação apresentada à Escola de Economia da
Fundação Getúlio Vargas (FGV/EESP) como
requisito para obtenção do título de Mestre em
Finanças e Economia Empresarial.

Data de aprovação:

____/____/____

Banca Examinadora:

Prof. Afonso de Campos Pinto
(Orientador)
FGV-EAESP

Prof. Ricardo Rochman
FGV-EAESP

Prof. Carlos Eduardo Pedreira
UFRJ – Faculdade de Medicina e COPPE (PEE)

Agradecimentos

Aos professores do MPFE-FGV, em especial, ao Prof. Afonso de Campos Pinto, orientador deste trabalho.

Ao Prof. Carlos Eduardo Pedreira.

À minha família.

Resumo

Este trabalho tem por motivação evidenciar a eficiência de redes neurais na classificação de rentabilidade futura de empresas, e desta forma, prover suporte para o desenvolvimento de sistemas de apoio a tomada de decisão de investimentos.

Para serem comparados com o modelo de redes neurais, foram escolhidos o modelo clássico de regressão linear múltipla, como referência mínima, e o de regressão logística ordenada, como marca comparativa de desempenho (*benchmark*).

Neste texto, extraímos dados financeiros e contábeis das 1000 melhores empresas listadas, anualmente, entre 1996 e 2006, na publicação Melhores e Maiores – Exame (Editora Abril).

Os três modelos foram construídos tendo como base as informações das empresas entre 1996 e 2005. Dadas as informações de 2005 para estimar a classificação das empresas em 2006, os resultados dos três modelos foram comparados com as classificações observadas em 2006, e o modelo de redes neurais gerou o melhor resultado.

Abstract

The motivation of this work is to verify the efficiency of neural networks as a tool for classifying forecasts of companies' return on equity, so that they can be used in order to provide support for the development of investment decision support systems.

The results obtained by the proposed neural networks mode are compared to those obtained by the use of the classic multiple linear regression as a minimum reference and, as a benchmark, to those obtained via ordinal logistic multiple regression.

In this work, we gathered top 1000 companies' financial and accounting data, annually listed by Melhores e Maiores – Exame publication (Editora Abril), from 1996 to 2006.

The three models were built using data relative to the period between 1996 and 2004. Using the 2005 data as input in order to forecast companies' classification in 2006, the three models' outputs were compared to the observed 2006 classification, and the neural network model yielded the best results.

Índice

1. Introdução	9
2. Revisão Bibliográfica	10
3. Metodologia	13
3.1. Modelo de Regressão Linear Múltipla (RLM).....	15
3.2. Modelo de Regressão Logística Ordenada (RLO).....	16
3.3. Modelo de Redes Neurais Artificiais (RN).....	18
4. Experimentos	23
4.1. Descrição e Preparação dos Dados	23
4.2. Definição do Modelo de Regressão Linear Múltipla.....	26
4.2.1. Definição do Conjunto de Variáveis Independentes.....	27
4.2.2. Tratamento do Resultado do Modelo	27
4.2.3. Interpretação dos Resultados	28
4.3. Definição do Modelo de Regressão Logística Ordenada.....	32
4.3.1. Definição do Conjunto de Variáveis Independentes.....	32
4.3.2. Tratamento do Resultado do Modelo	33
4.3.3. Interpretação dos Resultados	34
4.4. Definição do Modelo de Redes Neurais.....	35
4.4.1. Tratamento da Variável Dependente e Definição da Saída da RN .	37
4.4.2. Definição da Arquitetura da RN.....	37
4.4.3. Definição do Conjunto de Variáveis Independentes e da RN.....	39
4.4.4. Tratamento do Resultado do Modelo	40
4.4.5. Interpretação dos Resultados	41
4.5. Comparação dos Modelos	42
5. Conclusão	46
REFERÊNCIAS BIBLIOGRÁFICAS	48
ANEXO I.....	53
ANEXO II.....	54
ANEXO III.....	55

1. Introdução

Este trabalho tem por motivação evidenciar a eficiência de redes neurais na classificação de rentabilidade futura de empresas, e desta forma, prover suporte para o desenvolvimento de sistemas de apoio a tomada de decisão de investimentos.

Neste texto, extraímos dados financeiros e contábeis das 1000 melhores empresas listadas anualmente pela publicação Melhores e Maiores – Exame (Editora Abril), entre 1996 e 2006, para classificar a rentabilidade futura das empresas em três categorias: retorno ruim (investimento não recomendado), retorno médio (investimento recomendado) e retorno ótimo (investimento altamente recomendado).

Para serem comparados com o modelo de redes neurais, foram escolhidos o modelo clássico de regressão linear múltipla, como referência mínima, e o de regressão logística ordenada, como marca comparativa de desempenho (*benchmark*).

Os três modelos foram construídos tendo como base as informações das empresas entre 1996 e 2005. A avaliação da qualidade dos modelos foi feita por meio da comparação das estimativas geradas pelos três modelos para classificação das empresas em 2006, dadas as informações de 2005. Os resultados dos três modelos foram comparados com as classificações observadas em 2006, e o modelo de redes neurais gerou o melhor resultado.

Além da introdução, o trabalho é composto por mais quatro capítulos: Capítulo 2 de revisão bibliográfica; Capítulo 3, que apresenta a metodologia e descreve os modelos utilizados no estudo; Capítulo 4, no qual os experimentos desenvolvidos e os resultados obtidos são descritos; Capítulo 5 de considerações finais e proposição de temas para estudos futuros.

2. Revisão Bibliográfica

De acordo com as notas históricas compiladas por Haykin (1994), os primeiros trabalhos sobre Redes Neurais surgiram nos anos 1940 (McCulloch e Pitts, 1943). Pesquisas realizadas durante as duas décadas seguintes como, por exemplo, os trabalhos de Rosenblatt (1958), a quem se atribui a definição do método de aprendizagem supervisionada (*perception*), e Widrow & Hoff (1960), sobre a formulação do *ADALINE* (*Adaptive Linear Element* – elemento linear adaptativo), formaram os fundamentos básicos da teoria de Redes Neurais.

Porém, em 1969, Minsky e Papert publicaram um livro no qual os autores expuseram sérias supostas limitações teóricas dos métodos anteriormente propostos. Esse trabalho e a falta de recursos computacionais para validar os modelos por meio de experimentos práticos, quase interromperam as atividades de pesquisas na área, que só voltaram a tomar fôlego nos anos 1980, com o desenvolvimento de ferramental computacional e, principalmente, com a publicação do trabalho desenvolvido por Rumelhart, Hilton e Williams (1986), com respostas para os problemas apresentados por Minsky quinze anos antes. Posteriormente, foi reconhecido que essas respostas já tinham sido apresentadas dez anos antes por Werbos (1974), em sua tese de doutorado.

A partir de então, além de estudos com contribuições para a evolução da teoria de redes neurais (Abu-Mostafa 1986, Kohonen, 1987), pudemos observar o expressivo desenvolvimento de aplicações de redes neurais na solução de diversos problemas práticos. De acordo com Schürmann (1996), com relação às aplicações, a evolução do uso de redes neurais pouco difere da evolução do uso de outras técnicas estatísticas. Inicialmente, foram utilizadas para resolver problemas na área de automação e controle (Kawato, Furukawa e Suzuki, 1987), diagnóstico na área médica (Saito e Nakano, 1988), porém, a partir dos anos 1990, publicações como a de McNelis (1996), começam a apresentar

usos de técnicas de redes neurais aplicadas à área de finanças como ferramentas de classificação de padrões, *Data Mining* e predição.

Estudos sobre uso de redes neurais em finanças podem ser encontrados em um amplo espectro de aplicações: negociações no mercado financeiro (Baestaens, Bergh & Wood, 1994), previsão no mercado financeiro (Abu-Mostafa & Atiya, 1996), previsão de desempenho financeiro (Gately, 1995 – Jason, 1998), modelamento de taxa de câmbio (Medeiros, Vegas & Pedreira, 2001), previsão de correlação entre moedas (Chen & Leung, 2006), previsão de fundo mútuo (Sahoo & Hathy, 2007), predição de insolvência (Chung, Tan, & Holdsworth, 2008).

No Brasil, muitos trabalhos têm sido desenvolvidos abordando, especificamente, as condições do mercado brasileiro: análise de crédito (Lemos, Steiner & Nievola 2005), precificação de opções no mercado brasileiro (Freitas & Souza 2002, Chagas 2006), recuperação de receita com prevenção de insolvência (Pinheiro, Evsukoff & Ebecken 2006), avaliação de crédito bancário em instituição financeira brasileira (Arnauld, Cunha, Ludermir & Adeodato 2005), variação de cotação de ações da Telebrás (Zaneti & Almeida 1998).

Boa parte desses estudos elabora comparações entre redes neurais com outros tipos de modelos: lineares (Medeiros, Veiga & Pedreira 2001), K-means (Arnauld, Cunha, Ludermir & Adeodato 2005), Black & Scholes (Freitas & Souza 2002, Chagas 2006).

Particularmente, com relação à aplicação de redes neurais em estudos sobre previsão de rentabilidade e desempenho de empresas, encontramos trabalhos com abordagem de classificação binária, como o artigo de Almeida e Dumontier (1996), no qual os autores, tendo como dados de entrada, informações contábeis das empresas do setor de carga rodoviário francês, classificam as mesmas como possivelmente inadimplentes ou não. A conclusão desse estudo foi que o modelo de redes neurais não conseguiu produzir resultados superiores aos modelos desenvolvidos com métodos

estatísticos tradicionais. Uma das explicações dada pelos autores para esse insucesso foi a limitação para explorar soluções mais sofisticadas imposta pelo excessivo esforço de desenvolvimento e tempo de processamento exigido pelos modelos de redes neurais. Se levarmos em conta que o experimento foi realizado há mais de doze anos, talvez, atualmente, essas barreiras já tenham sido superadas.

Sobre previsão de desempenho, o trabalho escrito por LAM (2004) descreve um modelo de redes neurais que recebe como dados de entrada um conjunto de informações financeiras e contábeis de empresas e um conjunto de informações macroeconômicas, e produz como resultado as taxas de retorno de suas ações para tomada de decisão de investimento. Para verificar o desempenho do modelo de redes neurais, os resultados dos investimentos indicados pela rede neural foram comparados com um *portfolio* de rendimento mínimo (*minimum benchmark*), composto por uma cesta de empresas bem diversificada (*portfolio* média do mercado) e com um *portfolio* de rendimento máximo (*maximum benchmark*), composto pelas empresas presentes na região de 1/3 do topo da lista de taxas de retorno mais altas (*portfolio* informação perfeita). A conclusão do estudo foi que o modelo de redes neurais conseguiu superar o *portfolio minimum benchmark*, mas ficou abaixo do *maximum benchmark*. O estudo não encontrou evidências de que a introdução de variáveis macroeconômicas melhora o modelo de redes neurais.

Seguindo a linha de previsão de desempenho futuro de empresas, neste trabalho, também utilizamos informações financeiras e contábeis como dados de entrada para os modelos de redes neurais, de forma semelhante aos dois estudos acima mencionados, porém ao invés de tentar prever um valor numérico, como no caso do trabalho de LAM (2004), propomos classificar a rentabilidade futura das empresas em três classes (rentabilidade ruim, média ou ótima), ao invés de duas como no caso do estudo de Almeida e Dumontier (1996).

3. Metodologia

Para solução do problema de classificação de padrões, utilizamos a metodologia descrita por Duda, Stork e Hart (2005), na qual esse processo é dividido em seis tarefas:

- I. Definir um conjunto de indivíduos suficientemente grande e representativo.

Trata-se do clássico problema de amostragem presente nos experimentos estatísticos: como fazer a coleta de dados, qual o tamanho da amostra, etc.

- II. Identificar os atributos dos indivíduos.

Idealmente, o conjunto de atributos deve conter informações suficientes para generalizar o processo de classificação. Atributos irrelevantes têm o mesmo efeito que ruídos e tendem a diminuir o poder de estimação do modelo (Wooldridge, 2003).

- III. Definir um subconjunto de dados de treinamento e outro de verificação.

A amostra deve ser dividida em dois subconjuntos: um subconjunto para ser utilizado na construção do modelo (dados de treinamento), e outro, para comparação da classificação estimada pelo modelo com a classificação real dos dados (dados de verificação).

- IV. Definir o modelo de estimação.

No caso de estudos comparativos, a escolha é de mais de um modelo de estimação, por exemplo: modelos lineares, não lineares, baseados em regras. Além disso, é necessário observar o tipo de organização dos dados (*cross-section*, séries de tempo ou painel).

V. Construir o modelo utilizando os dados de treinamento.

A maioria dos métodos de construção de modelos trabalha com algum processo de otimização no qual, dado o conjunto de dados de treinamento, X_{trein} , com atributos e classificações conhecidos, Y_{trein} , busca-se encontrar o modelo F , tal que o erro entre o resultado estimado por F , $Y_{trein}' = F(X_{trein})$, e a classificação real, Y_{trein} , seja mínimo.

VI. Avaliar os modelos calibrados por meio da comparação dos resultados da classificação dos dados de verificação.

O conjunto de dados de verificação também é formado por atributos, X_{verif} , e classificações, Y_{verif} , conhecidos. A avaliação do modelo é feita confrontando-se a classificação conhecida, Y_{verif} , com os resultados gerados pelo modelo para o conjunto de dados de verificação, $Y_{verif}' = F(X_{verif})$. O melhor modelo será o que apresentar o menor erro.

Com relação ao Passo IV, para serem comparados com o modelo de RNs, escolhemos o modelo clássico de regressão linear múltipla como referência base e o de regressão logística ordenada (uma generalização do modelo de regressão linear) como marca comparativa de desempenho (*benchmark*), cujas características seguem descritas nos próximos itens deste capítulo.

Os demais passos, acima apresentados, nortearam a preparação dos dados e os experimentos realizados neste estudo, que estão descritos no Capítulo 4.

Todos os modelos foram criados e executados em um mesmo equipamento, com a seguinte configuração:

- Microcomputador PC Intel Dual Core, 1.83GHz, 667MHZ, 2GB de RAM.
- Sistema operacional: Windows Vista Home Premium.
- Software: MATLAB 7.6 com os módulos Statistics Toolbox e Neural Network Toolbox.

3.1. Modelo de Regressão Linear Múltipla (RLM)

O modelo de regressão linear múltipla (RLM) considera apenas as relações lineares entre as variáveis independentes.

Apesar da aparente limitação, esta formulação permite interpretar o fenômeno analisando-o sob a condição '*Ceteris Paribus*', isto é, avaliar o efeito individual de cada variável independente, supondo todas as demais constantes. Na prática, o modelo RLM é muito utilizado, sendo capaz de produzir bons resultados. Outra razão para a popularidade do modelo RLM é o fato de o mesmo possuir uma metodologia robusta de construção e de fácil implantação computacional.

Características do modelo:

$$Y^t = \beta^t * X^{t-1} + \varepsilon$$

- Y^t : vetor da variável dependente (classificação em t).
- X^{t-1} : matriz das variáveis independentes (dados em t-1).
- β : vetor de coeficientes da regressão.
- ε : vetor de erro ou resíduo da regressão.
- Premissa: Y^t , variável dependente, tem distribuição normal $N(\mu, \sigma^2)$.
- Método de otimização: calcular o conjunto β tal que a soma dos erros ao quadrado, $\varepsilon^2 = (Y^t - \beta^t * X^{t-1})^2$, seja mínima.

De acordo com Wooldridge (2003), atendidas as condições dadas pelas hipóteses de Gauss-Markov, o RLM provê testes estatísticos que permitem escolher, para um dado grau de confiabilidade, o modelo estatisticamente melhor:

- Modelo cujos coeficientes são estatisticamente significativos. A hipótese nula deste teste é 'o coeficiente é nulo', e Estatística-T e P-valor são as duas estatísticas de verificação.
- Modelo relativamente melhor que outro. Espera-se que um modelo trabalhe com variáveis independentes o suficiente para explicar a variável dependente, mas não demais a ponto de perder sua capacidade de generalização. A Estatística-F fornece uma medida dessa qualidade, pois pondera esses dois aspectos. Ao comparar dois modelos, o que tiver a Estatística-F com valor mais alto é considerado o melhor.

3.2. Modelo de Regressão Logística Ordenada (RLO)

O modelo de Regressão Logística Ordenada (RLO) é uma generalização do modelo de regressão linear para tratar variáveis dependentes discretas (*limited dependent variable*) e ordenadas.

Este modelo estima a probabilidade de a variável assumir um valor menor ou igual a uma classe. Por este motivo, apresenta-se como uma alternativa teoricamente melhor que o RLM, que trata apenas variáveis dependentes contínuas.

Características do modelo:

$$\log\left(\frac{P_1}{1-P_1}\right) = \alpha_1 - \beta^t * X^{t-1} \Rightarrow P_1 = P(Y^t \leq 1 | X^{t-1})$$

$$\log\left(\frac{\sum_{j=1}^k P_j}{1 - \sum_{j=1}^k P_j}\right) = \alpha_k - \beta^t * X^{t-1} \Rightarrow P_k = P(Y^t \leq k | X^{t-1}), k = 1, \dots, n-1$$

$$P_n = 1 - \sum_{j=1}^k P_j \Rightarrow P_n = P(Y^t \leq n | X^{t-1})$$

- $P_k = P(Y^t \leq k \mid X^{t-1})$: dado o conjunto de atributos X^{t-1} (dados em t-1), P_k é o vetor de probabilidade de o vetor da variável dependente Y^t (classificação em t, elemento a elemento) ser menor ou igual a k , $k = 1, \dots, n-1$.
- X^{t-1} : matriz de variáveis independentes (dados em t-1).
- α_k : coeficientes auxiliares da regressão, $k = 1, \dots, n-1$.
- β : vetor de coeficientes da regressão.
- Premissa: P_k tem distribuição normal $N(\mu, \sigma^2)$, $k = 1, \dots, n-1$.

De acordo com McCullagh e Nelder (1994), para a estimação das probabilidades P_k , não há restrições sobre a distribuição de probabilidade de Y . A solução do método RLO implica encontrar a função de máxima log-verossimilhança da variável Y , ou seja, encontrar a distribuição de probabilidade com maior chance de ser a distribuição de Y .

Para o modelo RLO, existem testes estatísticos, semelhantes aos do modelo RLM, com os quais também é possível escolher, para um dado grau de confiabilidade, o modelo estatisticamente melhor:

- Modelo cujos coeficientes são estatisticamente significantes. As mesmas Estatística-T e P-valor utilizadas nos testes para modelos RLM também se aplicam para os modelos RLO.
- Modelo relativamente melhor que outro. Para modelos RLO, o Critério de Informação de Akaike (AIC) ou o Critério de Informação de Schwartz (SIC) podem ser utilizado como medidas dessa qualidade. Ao comparar dois modelos, o que tiver o critério de informação mais baixo é considerado o melhor (McCullagh e Nelder, 1994).

Os resultados dos modelos RLO também podem ser analisados sob a condição '*Ceteris Paribus*'.

3.3. Modelo de Redes Neurais Artificiais (RN)

Tanto o modelo RLM quanto o RLO, supõem que as variáveis independentes são não-correlacionadas. A presença de correlação faz com que os estimadores calculados pelos modelos não estimem os valores corretamente (Wooldridge, 2003).

No caso do modelo RLM, o mesmo ocorrerá, se a variável independente não tiver distribuição de probabilidade normal.

Nesses casos, é interessante experimentar modelos com abordagens diferentes, com restrições mais fracas, ou mesmo, sem restrições, e verificar se os mesmos conseguem fornecer resultados superiores.

Redes Neurais Artificiais, ou simplesmente Redes Neurais (RNs), são um exemplo dessa classe de modelo: não impõem restrições de não-correlação entre as variáveis independentes (exigem, apenas, ausência de colinearidade perfeita), nem estabelecem premissas sobre a distribuição de probabilidade da variável dependente.

Além disso, são potencialmente capazes de tratar relações não-lineares entre as variáveis independentes.

Suponha o problema: estimar Y , considerando um conjunto X de variáveis independentes. Uma abordagem, chamada de força bruta por Duda, Stork e Hart (2005), para encontrar uma solução, sem estabelecer premissas, seria testar todas as formas funcionais possíveis e escolher a melhor.

RNs podem ajudar a achar essa solução, isto é, uma forma funcional que relacione as variáveis de entrada X com a variável de saída Y .

Sejam as seguintes definições e nomenclaturas relativas à RNs, conforme Hagan, Demuth e Beale (1996):

Sobre o fluxo de dados:

- RN Acíclica (*feedforward*): nesses tipo de RNs, o fluxo de dados é unidirecional e não há ciclos internos de processamento, isto é, os dados não retornam para estágios de processamento anteriores.
- RN Recorrente (*recurrent*): ao contrário da RN acíclica, RNs recorrentes possuem ciclos internos de processamento, nos quais dados retornam para estágios de processamento anteriores.

Sobre a arquitetura da RN (Topologia da RN):

- Camada (*layer*): nome dado aos estágios de processamento aos quais os dados de entrada da RN são submetidos. RNs têm pelo menos duas camadas, uma de entrada e outra de saída. Camadas internas são chamadas 'camadas escondidas' (*hidden layers*). RNs com mais de duas camadas são chamadas de *Multilayer Networks*.

Hagan, Demuth e Beale (1996) destacam que, em teoria, não há restrições para o número de camadas que uma RN pode ter, mas que, em geral, RNs são criadas com três camadas: uma camada de entrada, um de saída e apenas uma camada escondida.

- Neurônio (*neuron*): em cada camada, o processamento é executado em unidades de processamento que recebem o nome de neurônio.
- Peso (*weight*) e Inclinação (*bias*): dado uma entrada p , o neurônio gera os parâmetros Peso (w – *weight*) e Inclinação (b – *bias*) e aplica $f(w \cdot p + b)$ para gerar uma saída, que será o parâmetro de entrada para a próxima camada.

Sobre as funções executadas pelos neurônios:

- Função de Transferência (*transfer function*): são funções executadas pelos neurônios. De acordo com Haykin (1994), as funções de transferência mais comuns são: *tansig* e *logsig*, funções sigmóides que

possibilitam composições não-lineares entre os valores dos neurônios; *purelin*, função linear tipicamente utilizada na camada de saída para produzir valores com módulo maior que um. Os coeficientes (pesos) que multiplicam os parâmetros de entrada alteram a derivada das funções no ponto (0,0) e as constantes (inclinação) deslocam o ponto de inflexão das funções ao longo do eixo Y.

Sobre os algoritmos (funções) utilizados pela RN:

- Função de Treinamento (*training function*): funções de treinamento buscam relacionar os valores de entrada com o valor de saída e desta forma ‘ensinam’ os neurônios da RN a calibrar os coeficientes w e b .
- Função de Aprendizado (*learning function*): as funções de aprendizado controlam a intensidade das alterações dos coeficientes w e b .
- Função de Desempenho (*performance function*): as funções de desempenho são a função objetivo do processo de otimização utilizado para treinar a rede. A função de desempenho mais comum é a média dos quadrados dos erros. Neste caso, a RN calculará os coeficientes w e b de forma que a média dos quadrados dos erros seja mínima.

Sobre o processo de treinamento da RN:

- Objetivo (*Goal*): valor objetivo do treinamento da RN, isto é, o processo de calibração dos coeficientes continua até que a função de desempenho atinja um valor menor que o valor objetivo.
- Época (*Epochs*): indica o número máximo de vezes que o processo de treinamento será executado.

De acordo com Nascimento e Yoneyama (2002), essa complexa arquitetura tem sua capacidade de aproximação validada pelo Teorema de Kolmogorov (1957) que afirma que qualquer função pode ser implementada por uma RN de três camadas, dado um número suficiente de neurônios na camada interna e funções de transferência e pesos apropriados.

No entanto, o Teorema de Kolmogorov não fornece qualquer indicação de qual seria um número de neurônios suficiente, nem tão pouco, quais as funções de transferência e os pesos apropriados.

Além disso, Duda, Stork e Hart (2005) afirmam que, para um conjunto de dados de entrada e saída, $\{X, Y\}$, é possível, construir uma RN capaz de implementar uma função F , tal que, $F(X) = Y$. Essa solução pode ser trivialmente encontrada, porém sua utilidade é equivalente a conseguir interpolar n pontos com um polinômio de grau $n-1$, isto é, quase nula, pois não tem capacidade de generalização.

Esse problema é conhecido como *overfitting*, isto é, a RN, por utilizar um número excessivo de coeficientes, acaba modelando dados e ruídos e não a relação entre dados de entrada e de saída.

Duda, Stork e Hart (2005) também descrevem algumas estratégias desenvolvidas para tratar o problema de *overfitting* no treinamento de RNs. Uma das mais utilizadas é a *Early Stopping*, que significa interromper o treinamento antes que a RN atinja o valor objetivo, caso sua capacidade de generalização diminua.

A implementação desta estratégia consiste dividir o conjunto de dados de entrada e de saída em três subconjuntos:

- Subconjunto de Treinamento: subconjunto fornecido para o treinamento da RN para a calibração dos coeficientes w e b (pesos e inclinações).
- Subconjunto de Validação: subconjunto usado para monitorar o treinamento, isto é, o treinamento continua enquanto a RN, treinada com o subconjunto de treinamento, permaneça gerando bons resultados, isto é, que o valor da função de desempenho continue caindo para o subconjunto de validação.
- Subconjunto de Teste: subconjunto usado para comparar modelos. Não é usado no treinamento.

Este é um dos motivos pelo qual a massa de dados para modelos de RNs não pode ser pequena, pois terá que ser dividida em três partes.

Além disso, a representatividade dos dados é crítica, pois a eficiência da RN é muito sensível à massa de dados fornecida para a calibração do modelo. A RN conseguirá classificar melhor indivíduos com características semelhantes aos indivíduos da amostra com a qual RN foi treinada.

Porém, indivíduos com características diferentes podem ser adicionados ao conjunto de dados de treinamento, e a RN re-calibrada com o novo conjunto de dados, em teoria, será capaz de classificá-los.

A seguir, descrevemos os experimentos realizados utilizando-se a metodologia apresentada neste capítulo.

4. Experimentos

O problema proposto para a realização dos experimentos consistiu em estimar a classificação da rentabilidade futura de empresas, utilizando modelos construídos com base em informações passadas.

Para tanto, usamos dados disponíveis no Portal Exame Melhores & Maiores de Editora Abril (<http://app.exame.abril.com.br/servicos/melhoressemaiores/>).

4.1. Descrição e Preparação dos Dados

Utilizamos os dados financeiros e contábeis de empresas classificadas na pesquisa ‘1000 Melhores & Maiores’, no período entre 1996 e 2006. Foram selecionadas empresas estatais e do setor privado e, excluídas instituições financeiras, bancos e seguradoras.

Dos 24 parâmetros disponíveis (listados no Anexo I), adotados os critérios de exclusão listados a seguir, 11 foram selecionados para a construção dos modelos.

Critérios de exclusão:

- Parâmetro não numérico. Exemplo: Controle Acionário.
- Parâmetro não disponível em todos os anos. Exemplo: EBITDA só existe a partir de 2001.
- Parâmetro diretamente derivado de outro. Exemplo: Patrimônio Líquido Legal é o Patrimônio Líquido Ajustado sem o efeito da inflação.

Os parâmetros selecionados, com as respectivas descrições fornecidas pela publicação, estão listados a seguir:

1. Capital Circulante Líquido (CCL): Representa o total de recursos (em US\$ Milhões) de curto prazo disponíveis para financiamento das atividades da empresa.
2. Lucro Líquido Ajustado (LLA): Lucro líquido apurado (em US\$ Milhões) depois de reconhecidos os efeitos da inflação nas demonstrações contábeis.
3. Patrimônio Líquido Ajustado (PLA): Patrimônio líquido legal (em US\$ Milhões) atualizado pelos efeitos da inflação.
4. Riqueza Criada por Empregado (RCE): Total da riqueza criada pela empresa dividido pela média aritmética do número de empregados, sem levar em conta eventuais serviços terceirizados (expressa em US\$ Mil).
5. Crescimento das Vendas (CV): Mostra a evolução da receita bruta de vendas em reais (% em relação ao ano anterior), descontada a inflação média apontada pela variação do IGP-M.
6. Endividamento Geral (EG): Soma do passivo circulante (isto é, dívidas e obrigações de curto prazo, incluindo-se as duplicatas descontadas) com o exigível a longo prazo (o resultado é mostrado em % em relação ao ativo total ajustado).
7. Margem das vendas (MV): Divisão do lucro líquido ajustado pelas vendas, expressa em porcentagem.
8. Índice de Giro de Ativo (IGA): Receita bruta de vendas dividida pelo ativo total ajustado. Mede a eficiência operacional da empresa e deve ser comparado com a margem de lucro sobre vendas.
9. Índice de Liquidez Geral (ILG): Mostra uma relação entre os recursos da empresa que não estão imobilizados (tecnicamente, aplicados no ativo permanente) e o total de sua dívida.
10. Vendas em Dólares (VD): Foram apuradas com base nas vendas em reais e convertidas adequadamente (expressa em US\$ Milhões).

11. Valor da Rentabilidade do patrimônio ajustado (VR): Principal indicador de excelência empresarial, porque mede o retorno do investimento para os acionistas. Resulta da divisão dos lucros líquidos, legal e ajustado, pelos respectivos patrimônios líquidos, legal e ajustado. O produto é multiplicado por 100 para ser expresso em %.

A partir desse conjunto de parâmetros, criamos as observações da seguinte forma:

- Variável Dependente: a partir do parâmetro Valor da Rentabilidade do patrimônio ajustado dos anos de 1997 até 2006 (VR^t), construímos a variável dependente do modelo, Classificação da Rentabilidade da Empresa (CR^t), com a seguinte divisão:

CR^t	Condição	Avaliação
1	$VR^t < 2,85\%$	Retorno de rentabilidade ruim
2	$2,85\% \leq VR^t < 12,5\%$	Retorno de rentabilidade médio
3	$VR^t \geq 12,5\%$	Retorno de rentabilidade ótimo

Quadro 1 – Definição da classificação do valor da rentabilidade do patrimônio ajustado.

As classes foram definidas de modo que todas tivessem, aproximadamente, o mesmo número de observações (~740).

- Variáveis Independentes: 11 parâmetros de uma mesma empresa, no mesmo ano, a partir de 1996 até 2005. O Anexo II exibe uma tabela de estatística descritiva das variáveis independentes.
- Observações: cada observação foi composta com a variável dependente Classificação da Rentabilidade da Empresa no ano (CR^t) e 11 parâmetros da mesma empresa (variáveis independentes) no ano anterior ($t-1$), conforme o ilustrado na Figura 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Informação Extra			Variável dependente	Variáveis Independentes: Atributos										
2	Sector	Empresa	Ano		CCL	LLA	PLA	RCE	CV	EG	MV	IGA	ILG	VD	VR
3	ALIMENTOS, BEBIDAS E F	Açúcar União	1997	1	61,27	1,42	1,61	198,26	-19,31	61,29	0,24	2,25	0,75	726,45	1,4
4	ALIMENTOS, BEBIDAS E F	Antarctica / PB	1997	2	-3,21	5,74	6,52	239,99	-1,08	22,21	3,33	0,86	1,36	174,08	3,62
5	ALIMENTOS, BEBIDAS E F	Antarctica do Piauí	1997	3	-2,94	23,96	27,24	217,75	-20,01	24,52	12,14	0,93	2,05	246,76	13,57
6	ALIMENTOS, BEBIDAS E F	ANTARCTICA DO SUDE	1997	1	22,07	-2,23	-2,54	164,89	-10,08	56,43	-0,89	0,84	0,38	278,54	-1,71
7	ALIMENTOS, BEBIDAS E F	Antarctica Norte/Nordest	1997	2	-19,9	56,08	63,78	193,71	-5,77	29,01	14	0,87	0,61	425,11	15,42
8	ALIMENTOS, BEBIDAS E F	Antarctica Polar	1997	2	2,67	16,24	18,47	149,94	-27,92	26,28	5,49	0,84	1,9	410,3	5,82
9	ALIMENTOS, BEBIDAS E F	Antarctica RJ	1997	2	18,19	19,11	21,74	328,69	-9,05	41,43	3,08	0,9	0,46	681,56	4,65
10	ALIMENTOS, BEBIDAS E F	AURORA	1997	1	1,26	2,51	2,86	80,46	8,25	49,24	0,65	1,98	0,74	355,09	2,55
11	ALIMENTOS, BEBIDAS E F	AVIPAL	1997	2	33,63	14,76	16,79	67,1	29,9	43,32	4,62	0,53	0,81	245,81	4,22
12	ALIMENTOS, BEBIDAS E F	BIANCHINI	1997	3	-1,87	1,1	1,25	488,71	15,22	50,82	0,6	2,72	0,96	159,91	3,3
13	ALIMENTOS, BEBIDAS E F	Brahma	1997	3	-182,58	342,37	389,34	640,31	10,57	54,5	11,18	1,09	0,41	2770,51	25,29
14	ALIMENTOS, BEBIDAS E F	BRASWEY	1997	1	7,74	-3,63	-4,13	303,81	22,63	38,21	-0,99	1,4	1,67	298,29	-2,24

Figura 1 - Observações do experimento

Pela própria construção das observações, só foram utilizadas, empresas que apareceram na lista em pelos menos dois anos consecutivos.

O conjunto total de observações foi separado em dois subconjuntos, conforme o descrito no Passo III, no capítulo anterior:

- Dados de treinamento: 2215 observações com dados de empresas de 1996 a 2004 com classificação de rentabilidade de 1997 a 2005.
- Dados de verificação: 410 observações com dados de empresas em 2005 com classificação de rentabilidade em 2006.

E como o indicado no Passo V, os dados de treinamentos foram utilizados para encontrar modelos, f , tais que:

$$CR^t = f(CCL^{t-1}, LLA^{t-1}, PLA^{t-1}, RCE^{t-1}, CV^{t-1}, EG^{t-1}, MV^{t-1}, IGA^{t-1}, ILG^{t-1}, VD^{t-1})$$

4.2. Definição do Modelo de Regressão Linear Múltipla

Neste item, descrevemos os procedimentos executados para definir o modelo de Regressão Linear Múltipla (RLM):

- Definição do conjunto de variáveis independentes.
- Tratamento do resultado do modelo.
- Interpretação dos resultados obtidos com o modelo para os dados de treinamento.

4.2.1. Definição do Conjunto de Variáveis Independentes

Para selecionar as variáveis do modelo RLM, foi utilizado o método de eliminação (Wooldridge, 2003), com coeficiente de confiabilidade $\alpha = 5\%$.

Para comparar os modelos, foram calculadas as estatísticas R^2 -Ajustado e F, conforme o exibido na Tabela 1 no Anexo III. O Quadro 2 exhibe o conjunto escolhido no final do processo.

CONJUNTO ESCOLHIDO
CCL^{t-1} – Capital Circulante Líquido
LLA^{t-1} – Lucro Líquido Ajustado
PLA^{t-1} – Patrimônio Líquido Ajustado
EG^{t-1} – Endividamento Geral
MV^{t-1} – Margem de Vendas
IGA^{t-1} – Índice de Giro de Ativo
VR^{t-1} – Valor da Rentabilidade

Quadro 2 – Conjunto de variáveis de entrada do modelo RLM para os dados de treinamento.

4.2.2. Tratamento do Resultado do Modelo

Conforme ilustrado no Quadro 3, os valores de saída do modelo foram arredondados (aproximados para o inteiro mais próximo), pois o RLM é um modelo contínuo e a classificação é dada por um número inteiro.

Resultado do modelo	Classificação atribuída
1,3839	1
1,8907	2
2,7391	3

Quadro 3 – Tratamento do resultado do modelo RLM.

Para este conjunto, o percentual de acerto do modelo RLM foi 37,65% (834 acertos em 2215 observações). A Figura 2 representa graficamente o resultado do modelo de RLM escolhido para os dados de treinamento.

Figura 2 - Comparação do resultado do modelo RLM com a classificação real, para os dados de treinamento.

Exemplo de interpretação do gráfico: o eixo horizontal representa a classificação gerada pelo modelo RLM e o vertical, a classificação real. No retângulo superior esquerdo de coordenada [1,3], os números 8 e 0.36% indicam, respectivamente, a quantidade absoluta e a porcentagem de ocorrências na qual a classificação atribuída pelo modelo RLM é 1 (rendimento ruim) e a classificação real da empresa é 3 (rendimento ótimo). As diagonais (posições [1,1], [2,2] e [3,3]) indicam acertos.

4.2.3. Interpretação dos Resultados

O percentual de acerto do modelo RLM para os dados de treinamento (37,65%) ficou muito próxima ao percentual de acerto teórico de um sorteio aleatório (33%), para a composição do conjunto de dados de treinamento com, aproximadamente, um terço das observações para cada classe. Este baixo resultado também foi indicado pelo baixo valor da estatística R^2 -Ajustado (0,1683).

O Quadro 4 seguir exibe os coeficientes calculados pelo modelo.

CONJUNTO FINAL	Coeficientes	Estatística-T	P-Valor
Constante	1,55215	31,25269	0,00000
CCL^{t-1} – Capital Circulante Líquido	-0,00014	-2,30015	0,02153
LLA^{t-1} – Lucro Líquido Ajustado	0,00026	3,91740	0,00009
PLA^{t-1} – Patrimônio Líquido Ajustado	-0,00004	-3,40715	0,00067
EG^{t-1} – Endividamento Geral	0,00536	6,51043	0,00000
MV^{t-1} – Margem de Vendas	0,02229	15,18037	0,00000
IGA^{t-1} – Índice de Giro de Ativo	0,07252	5,29801	0,00000
VR^{t-1} – Valor da Rentabilidade	0,00186	4,77646	0,00000

Quadro 4 – Coeficientes e estatísticas do modelo RLM.

O valor 1,55215 estimado para constante indica que, se nenhuma variável fosse considerada, qualquer empresa seria classificada como de retorno de rentabilidade médio (1,55215 arredondado para 2).

Apesar do baixo percentual de acerto, conforme o mencionado no item 3.1, podemos utilizar os coeficientes calculados pelo modelo para interpretar o fenômeno sob a condição '*Ceteris Paribus*', que permite interpretar o efeito individual de cada variável independente, supondo todas as demais constantes.

O modelo estimou coeficientes positivos para cinco variáveis: LLA^{t-1} (Lucro Líquido Ajustado), EG^{t-1} (Endividamento Geral), MV^{t-1} (Margem de Vendas), IGA^{t-1} (Índice de Giro de Ativo) e VR^{t-1} (Valor da Rentabilidade). Isto indica que aumentos nos valores dessas variáveis contribuiriam para melhorar a classificação das empresas.

Nesse conjunto, a variável que mais contribuiu (coeficiente com valor mais alto) para a melhora da classificação das empresas foi a IGA^{t-1} , índice que mede eficiência operacional, o que estaria de acordo com a intuição de que, uma empresa mais eficiente é mais rentável.

As contribuições positivas das demais variáveis são suportadas pela teoria de finanças corporativas e contabilidade empresarial, conforme apresentado em Brealy, Myers & Allen (2006) e Marion (2002):

- LLA^{t-1} : lucro líquido aparece no numerador da definição da fórmula de rentabilidade, *ROE - Return on Equity* (Brealy, Myers & Allen, 2006):

$$ROE = \frac{\text{Lucro Líquido}}{\text{Patrimônio}}$$

- EG^{t-1} : endividamento geral aparece no numerador da fórmula de rentabilidade (*ROE - Return on Equity*) dada pela Identidade de Du Pont (Marion, 2002):

$$ROE = 1 + \frac{\text{Endividamento Geral}}{\text{Patrimônio}}$$

Além disso, de acordo com a publicação Melhores e Maiores, EG^{t-1} é um bom indicador de risco de negócio, desta forma, o coeficiente positivo poderia indicar maior risco implicando maior expectativa de retorno.

- MV^{t-1} : margem de venda afeta positivamente o lucro líquido da empresa
- VR^{t-1} : um valor de rentabilidade alto deve afetar positivamente o valor no próximo período.

Além disso, as duas fórmulas de rentabilidade apresentadas também justificam o coeficiente negativo estimado pelo modelo para a variável PLA^{t-1} (Patrimônio Líquido Ajustado). Nos dois casos, o patrimônio aparece no denominador, ou seja, o aumento dessa variável diminui o valor do ROE.

No caso da variável CCL^{t-1} (Capital Circulante Líquido), que representa o total de recursos de curto prazo disponíveis para financiamento das atividades da empresa, um aumento dessa variável poderia indicar melhora na capacidade de liquidação de compromissos de curto prazo, menor risco de insolvência, e portanto, favorecer a rentabilidade da empresa.

Porém, de acordo com Brealy, Myers & Allen (2006), a relação entre capital circulante líquido e rentabilidade também depende da composição do ativo circulante da empresa, que é formado por: estoque e contas a receber.

$$\text{Capital Circulante Líquido} = \text{Ativo Circulante} - \text{Passivo Circulante} = \\ \text{Estoque} + \text{Contas a Receber} - \text{Passivo Circulante}$$

No caso de empresas com altos níveis de estoque de realização demorada, o aumento no valor de capital circulante líquido não significa necessariamente capacidade para liquidar compromissos de curto prazo, além disso, um alto valor de capital circulante conseqüente de um alto valor de ativo circulante poderia indicar ineficiência na gestão de recursos financeiros, o que prejudicaria a rentabilidade da empresa.

Sendo assim, a teoria indica que a influência do Capital Circulante Líquido sobre a rentabilidade depende de características da gestão de capital de giro das empresas.

Para este experimento, o modelo de RLM estimou um coeficiente negativo para a variável CCL^{t-1} . Em valor absoluto, é o coeficiente de menor valor, portanto o menos influente, o que reforça a dualidade descrita pela teoria.

Caso fosse possível identificar as características da gestão de capital de giro das empresas, uma variável instrumental *dummy*, indicando o tipo de composição do ativo circulante, poderia ser adicionada ao modelo para tentar captar, com mais eficiência, as diferentes influências do capital circulante líquido sobre a rentabilidade futura.

4.3. Definição do Modelo de Regressão Logística Ordenada

Para este modelo foram executados os mesmos procedimentos listados no item anterior.

4.3.1. Definição do Conjunto de Variáveis Independentes

Para selecionar as variáveis do modelo RLO, foi utilizado o método de eliminação, com coeficiente de confiabilidade $\alpha = 5\%$.

Para comparar os modelos, foi utilizado o critério de informação de Schwarz (SIC), pois este critério impõe uma penalidade maior pela inclusão de coeficientes (McCullagh e Nelder, 1994).

O Quadro 5 exibe o conjunto escolhido, isto é, o conjunto, composto apenas por variáveis com coeficientes estatisticamente significativos, e com o SIC mais baixo. Os valores estão exibidos na Tabela 2 – Anexo III.

CONJUNTO ESCOLHIDO
LLA^{t-1} – Lucro Líquido Ajustado
PLA^{t-1} – Patrimônio Líquido Ajustado
EG^{t-1} – Endividamento Geral
MV^{t-1} – Margem de Vendas
IGA^{t-1} – Índice de Giro de Ativo
VR^{t-1} – Valor da Rentabilidade

Quadro 5 – Conjunto de variáveis de entrada do modelo RLO para os dados de treinamento.

4.3.2. Tratamento do Resultado do Modelo

Para cada observação, o modelo produz três valores correspondentes às probabilidades de a empresa pertencer a uma classe menor ou igual à posição examinada.

Para obter a classificação da rentabilidade das empresas, escolhemos a classe com valor de probabilidade mais alta, conforme o Quadro 6, a seguir.

Saída do Modelo			Classificação Atribuída
Prob(Classe $\leq$ 1)	Prob(Classe $\leq$ 2)	Prob(Classe $\leq$ 3)	
0,734	0,203	0,063	1
0,302	0,397	0,302	2
0,066	0,208	0,726	3

Quadro 6 – Tratamento do resultado do modelo RLO.

Para este conjunto, o percentual de acerto do modelo RLO foi de 54,54% (1208 acertos em 2215 observações).

Figura 03 – Comparação do resultado do modelo RLO com a classificação real, para os dados de treinamento.

4.3.3. Interpretação dos Resultados

O fato de este melhor resultado ter sido obtido com um conjunto menor de variáveis independentes (todas as variáveis selecionadas pelo RLM menos a CCL – Capital Circulante Líquido) indica que a capacidade de generalização do modelo RLO é superior a do RLM.

Com relação à análise sob a condição '*Ceteris Paribus*', para o modelo RLO, os valores dos coeficientes são interpretados de acordo com a explicação descrita a seguir.

O Quadro 7 exhibe os coeficientes calculados pelo modelo:

CONJUNTO FINAL	Coeficientes	Estatística-T	P-Valor
α_1 valor de corte 1	0,60057	4,25168	0,00002
α_2 valor de corte 2	2,27927	15,29821	0,00000
LLA ^{t-1} – Lucro Líquido Ajustado	-0,00228	-3,14480	0,00166
PLA ^{t-1} – Patrimônio Líquido Ajustado	0,00010	3,50202	0,00046
EG ^{t-1} – Endividamento Geral	-0,01523	-6,69247	0,00000
MV ^{t-1} – Margem de Vendas	-0,06592	-8,98779	0,00000
IGA ^{t-1} – Índice de Giro de Ativo	-0,17925	-4,72320	0,00000
VR ^{t-1} – Valor da Rentabilidade	-0,02088	-5,71044	0,00000

Quadro 7 – Coeficientes e estatísticas do modelo RLO.

Os coeficientes α_1 e α_2 fornecem informações sobre os valores de probabilidades gerados pelo modelo RLO caso todas as variáveis independentes fossem zero:

$$\log\left(\frac{P_1}{1-P_1}\right) = \alpha_1 - \beta^* X^{t-1} = 0,60057 \Rightarrow P_1 = P(y^t \leq 1 | X^{t-1}) = 0.64579$$

$$\log\left(\frac{P_1 + P_2}{1 - (P_1 + P_2)}\right) = \alpha_2 - \beta^* X^{t-1} = 2,27927 \Rightarrow P_2 = P(y^t \leq 2 | X^{t-1}) = 0.26136$$

$$P_3 = 1 - (P_1 + P_2) \Rightarrow P_3 = P(y^t \leq 3 | X^{t-1}) = 0.09285$$

Desta forma, se nenhuma variável fosse fornecida, o modelo classificaria a empresa com Classe = 1 (classe com retorno de rentabilidade ruim).

O modelo RLO estimou coeficiente positivo apenas para a variável PLA (Patrimônio Líquido Ajustado). O sinal negativo no $\beta^* X^{t-1}$ indica que o aumento no valor dessa variável contribui para piorar a classificação das empresas. Por outro lado, aumentos nos valores das demais variáveis, para as quais foram estimados coeficientes negativos, contribuem para melhorar a classificação das empresas.

Ou seja, o modelo RLO apontou as mesmas relações encontradas pelo modelo RLM, com exceção da variável CCL^{t-1}. O resultado produzido indica que o modelo RLO não encontrou evidências para considerar essa variável influente, pois, conforme o exposto no item anterior, a rentabilidade pode ser influenciada pelo capital circulante líquido tanto positiva quanto negativamente.

4.4. Definição do Modelo de Redes Neurais

A teoria de RNs não provê critérios para seleção de variáveis de entrada, semelhantes aos utilizados nos modelos anteriores. Desta forma, a arquitetura de uma RN acaba sendo escolhida com base na experiência e conhecimento dos modeladores.

Essa experiência acumulada aparece na literatura especializada, como Bishop Bailer-Jones(2008), na forma de um conjunto de recomendações que podem ser utilizadas para orientar o processo de definição da arquitetura de uma RN, quais sejam:

- Normalizar as variáveis de entrada e saída de forma que todas fiquem em um mesmo intervalo específico[*IMin* *IMax*]:

$$ValorNormalizado = (Valor - ValorMin) * (IMax - IMin) / (ValorMax - ValorMin) + IMin$$

Usualmente, utiliza-se[-1,1], e neste caso temos:

$$ValorNormalizado = (Valor - ValorMin) * 2 / (ValorMax - ValorMin) - 1$$

Este procedimento tem como objetivo limitar o domínio da solução da função de otimização e desta forma tornar o treinamento da rede mais eficiente.

- Aplicar análise de componentes principais nas variáveis de entrada pode melhorar o treinamento da RN pois possibilita reduzir as dimensões do problema, com a eliminação de componentes menos significantes.
- Testar arquitetura com diferentes números de neurônios no *hidden layer*.
- Executar vários treinamentos com pontos iniciais diferentes. As soluções encontradas pelas RNs dependem do ponto inicial utilizado no algoritmo de otimização. Na prática, o mínimo local raramente é encontrado.

Para definir o modelo de RNs, foram executados os seguintes procedimentos:

- Tratamento da variável dependente e definição do *output layer* da RN.
- Definição da arquitetura da RN.
- Definição do conjunto de variáveis independentes e da arquitetura da RN.
- Tratamento do resultado do modelo.
- Interpretação dos resultados obtidos com o modelo para os dados de treinamento.

4.4.1. Tratamento da Variável Dependente e Definição da Saída da RN

Para que a RN pudesse devolver a classificação da rentabilidade da empresa e não um número real, definimos a RN com *output layer* com 3 neurônios: um para cada classe, representando seu estado de ativação, de acordo com o quadro abaixo onde, 1 indica o mais alto estado de ativação e -1, o mais abaixo. Para a construção das RNs, a variável dependente foi recodificada para essa forma.

Codificação da Classificação para o <i>Output Layer</i> da RN	Classificação
(1 -1 -1)	1
(-1 1 -1)	2
(-1 -1 1)	3

Quadro 8 – Codificação para o Output Layer do modelo de RNs.

4.4.2. Definição da Arquitetura da RN

No caso do experimento desenvolvido neste estudo, para as características gerais da arquitetura da RN, adotamos configurações consideradas convencionais, por Hagan, Demuth e Beale (1996), e adotadas em trabalhos como os de Freitas e Souza (2002), Lam (2004) e McNellis (2005):

- Tipo da Rede: acíclica.
- Número de camadas: 3 (de entrada, de saída e escondida)
 - Camada de entrada: definida com número de neurônios igual ao número de variáveis de entrada, com pré-processamento de normalização de dados e redução de dimensão com aplicação de análise de componentes principais.

- Camada de saída: definida com 3 neurônios (número de saídas).
 - Camada escondida: o número de neurônios na camada escondida foi determinado pela função de treinamento escolhida, conforme descrito a seguir.
- Função de transferência:
 - Camadas de entrada e escondida: *tansig*.
 - Camada de saída: *purelin*.

Para os parâmetros de treinamento, os valores foram arbitrados:

- Valor objetivo: 1000
- Número máximo de execuções: 100.
- Tamanho dos subconjuntos:
- Treinamento (usado para calibrar a RN): 50%.
- Validação (usado para evitar o *overfitting*): 20%.
- Teste (usado para comparar modelos): 30%.

Com relação à função de treinamento, para evitar ter que arbitrar o número de neurônios na camada escondida, ou seja, ter que testar arquiteturas com diferentes números de neurônios, escolhemos a função de treinamento baseada no método de regularização *bayesiana*, que, de acordo com o trabalho desenvolvido por MacKay (1992), e detalhado por Foresee e Hagan (1997), estima os parâmetros ótimos de regularização de uma RN.

Regularização é o nome dado ao processo de otimização que tem como objetivo encontrar valores de pesos e *bias* que minimizam a função de desempenho. A regularização *bayesiana*, com o objetivo evitar o *overfitting* e melhorar a capacidade de generalização da RN, acrescenta, ao processo de otimização, um componente que penaliza a função com o aumento do número de coeficientes, o que permite que o número de neurônios da camada escondida seja redefinido automaticamente durante o processo de treinamento.

4.4.3. Definição do Conjunto de Variáveis Independentes e da RN

Trabalhar todos os conjuntos possíveis gerados pelas 11 variáveis de entrada significaria testar um grande número de combinações, mais exatamente 2047, todas as permutações possíveis com 1, 2 até 11 elementos.

No entanto, aproveitamos o fato de os modelos RLM e RLO já terem apontado as variáveis com relações lineares significativas para tomá-las como base de composição do conjunto de variáveis de entrada, e desta forma, ao invés de 2047, trabalhamos com 17 conjuntos de variáveis (Tabela 3 – Anexo III):

- Conjunto selecionado pelo modelo RLO.
- Conjunto selecionado pelo modelo RLM.
- 15 conjuntos gerados pela combinação das outras 4 variáveis acrescidas ao conjunto selecionado pelo modelo RLM.

Para determinar o conjunto de variáveis de entrada e a arquitetura da RN, foram realizadas 100 execuções para cada uma das 17 combinações, e, para comparação, usamos o número de acerto no subconjunto de teste. O quadro a seguir exibe o conjunto escolhido no final do processo.

CONJUNTO ESCOLHIDO
CCL^{t-1} – Capital Circulante Líquido
LLA^{t-1} – Lucro Líquido Ajustado
PLA^{t-1} – Patrimônio Líquido Ajustado
RCE^{t-1} – Riqueza Criada por Empregado
CV^{t-1} – Crescimento em Vendas
EG^{t-1} – Endividamento Geral
MV^{t-1} – Margem de Vendas
IGA^{t-1} – Índice de Giro de Ativo
VR^{t-1} – Valor da Rentabilidade

Quadro 9 – Conjunto de variáveis de entrada do modelo de RNs para os dados de treinamento.

O resultado do experimento indica que o modelo de RNs conseguiu captar mais relações entre as variáveis que os modelos RLM e RLO, com a inclusão das variáveis RCE^{t-1} e CV^{t-1} , e desta forma melhorar o percentual de acerto.

4.4.4. Tratamento do Resultado do Modelo

Para cada observação, o modelo de RNs produziu 3 valores correspondentes ao estado de ativação de cada classe. Para obter a classificação da rentabilidade das empresas, analogamente a construção exibida no Quadro 8, o escolhemos a classe com valor de ativação mais alto, conforme o exemplo no Quadro 10.

Saída do Modelo (Estado de Ativação)			Classificação Atribuída
Classe 1	Classe 2	Classe 3	
0,880	-0,830	-1,050	1
-0,990	0,704	-0,715	2
-0,774	-0,731	0,505	3

Quadro 10 – Tratamento do resultado do modelo de RNs.

Para este conjunto, o percentual acerto do modelo de RNs foi de 60,00% (1329 acertos em 2215 observações).

Figura 04 – Comparação do resultado do modelo de RNs com a classificação real, para os dados de treinamento.

4.4.5. Interpretação dos Resultados

O Quadro 11 mostra que o modelo de RNs produziu o melhor resultado para os dados de treinamento.

Dados de Treinamento	Modelo		
	RLM	RLO	RN
Acertos	834	1208	1329
% Acertos	37,65%	54,54%	60,00%

Quadro 11 – Comparação entre os resultados dos modelos para dados de treinamento.

No entanto, de acordo com a teoria de RNs conhecida até o hoje, não existem procedimentos para a extração das regras utilizadas pela RN, e nem uma metodologia, como a de análise '*Ceteris Paribus*', para relacionar efeitos de variações nas variáveis de entrada sobre as variáveis de saída.

Isto poderia levantar a suspeita de que o número de acertos mais alto resultante do modelo de RNs tenha sido apenas uma coincidência.

Para verificar esta questão, examinamos a distribuição dos percentuais de acerto para os dados de treinamento nas 1700 execuções realizadas (Figura 5).

Figura 5 - Distribuição dos percentuais de acerto para os dados de treinamento.

E contamos quantos resultados da RN ficaram abaixo do percentual de acerto do modelo RLO (54,54%), e chegamos aos seguintes números:

- 7 resultados do modelo de RNs ficaram abaixo do modelo RLO (~ 0,5%).
- 1693, acima (~ 99,5%).

Verificamos desta forma que, mesmo considerando qualquer conjunto de variáveis de entrada, os modelos de RNs foram, em sua maioria, capazes de gerar resultados superiores ao modelo RLO.

4.5. Comparação dos Modelos

Para comparar a capacidade de generalização dos modelos, utilizamos os dados de verificação para estimar a classe de rentabilidade de 410 empresas em 2006, a partir de informações de 2005, e obtivemos resultados exibidos no Quadro 12.

Dados de Verificação	Modelo		
	RLM	RLO	RN
Acertos	144	245	290
% Acertos	35,12%	59,76%	70,73%

Quadro 12 – Comparação entre os resultados dos modelos para dados de verificação.

Para os dados de verificação, o percentual de acerto do modelo RLM de 35,12% (144 acertos em 410 observações) ficou abaixo do percentual obtido com os dados de treinamento (37,65%). Esta baixa capacidade de generalização indica que esse tipo de modelo não se mostra adequado para tratar o tipo de problema apresentado neste estudo.

No gráfico de distribuição dos resultados do modelo RLM, para os dados de verificação, podemos observar a concentração de respostas na Classe 2.

Figura 6 – Comparação do resultado do modelo RLM com a classificação real, para os dados de verificação.

No caso do modelo RLO, o percentual de acerto para os dados de verificação subiu 5,22% (de 54,54% para 59,76%), o que reforça a teoria de que modelos RLO são mais adequados para tratar problemas de classificação.

Figura 07 – Comparação do resultado do modelo RLO com a classificação real, para os dados de verificação.

Da mesma forma que no modelo RLO, o resultado do modelo de RNs para os dados de verificação foi superior. A diferença é que, além de partir de um percentual de acerto mais alto, o aumento relativo foi maior que o observado no RLO, isto é o percentual de acerto da RN passou de 60,00% para 70,73%, subindo 10,73%, contra os 5,22% do modelo RLO.

Este resultado indica que o modelo de RNs, comparado aos RLM e RLO, conseguiu captar melhor o relacionamento entre as variáveis de entrada e saída, sem perder a capacidade de generalização, conforme o resumo dos resultados comparativos apresentado no Quadro 11.

Modelo	Aumento % Absoluto	Aumento % Relativo
RLM	35,61%	101,39%
RLO	10,98%	18,37%

Quadro 11 – Comparação percentual entre os resultados dos modelos RLM e RLO com o RN para dados de verificação.

O melhor desempenho do modelo de RNs também pode ser observado no gráfico de distribuição dos resultados do modelo de RNs para os dados de verificação.

Figura 08 – Comparação do resultado do modelo de RNs com a classificação real, para os dados de verificação.

Novamente, com a finalidade de verificar a consistência do resultado obtido, examinamos a distribuição dos percentuais de acerto para os dados de verificação nas 1700 execuções realizadas (Figura 9).

Figura 9 - Distribuição dos percentuais de acerto para os dados de verificação.

E contamos quantos resultados da RN ficaram abaixo do resultado dado pelo modelo RLO (59,76%) e chegamos aos seguintes números:

- 20 resultados do modelo de RNs inferiores ao do modelo RLO (~ 1,2%)
- 1680 superiores (~ 98,8%).

Ou seja, também para os dados de verificação, mesmo considerando qualquer conjunto de variáveis de entrada, os modelos de RNs foram, em sua maioria, capazes de gerar resultados superiores ao modelo RLO.

5. Conclusão

Os experimentos descritos no Capítulo 4 fornecem evidências de que, para previsão de classificação de rentabilidade empresas, modelos de redes neurais podem gerar resultados satisfatórios e melhores do que os modelos de regressão linear e de logística ordenada.

Este estudo ilustra a aplicação prática de técnicas de classificação de padrões que podem ser utilizadas para desenvolver e aprimorar qualquer tipo de ferramenta de apoio à escolha de investimentos, bem como, fazer parte de sistemas de gerenciamento de risco, avaliação de crédito, identificação de comportamento, ou seja, qualquer processo que envolva tomada de decisão.

Para melhorar os percentuais de acerto, com relação ao conjunto de informações, estudos semelhantes podem ser realizados levando-se em conta as empresas por setor. Outra possibilidade seria acrescentar variáveis que, de alguma forma, pudessem expressar influências de fatores macroeconômicos ou mesmo que conseguissem quantificar subjetividade.

Sobre a escolha das variáveis de entrada do modelo de redes neurais, seria interessante experimentar o uso de procedimentos estatísticos, como por exemplo, a medida de informação mútua que estima a quantidade de informação que duas variáveis aleatórias têm em comum.

Quanto às redes neurais, outros tipos, como por exemplo, *Localist Attractor Network*, que trabalha com variável dependente discreta, ou *Self-Organization Map*, que utiliza método de aprendizado não supervisionado poderiam ser testados.

Com relação à combinação de técnicas, *Clustering* ou *Support Vector Machine* poderiam ser avaliados para refinar a separação de empresas, e a teoria de lógica difusa (*Fuzzy Logic*) poderia ser explorada para quantificar subjetividade.

Como conclusão, destacamos que os resultados dos modelos RLM e RLO não serviram apenas como marcas comparativas, mas também foram fundamentais para a definição do conjunto de variáveis de entrada dos modelos de RNS.

Desta forma, os experimentos descritos também demonstram que o uso combinado de técnicas pode ajudar a produzir resultados melhores.

REFERÊNCIAS BIBLIOGRÁFICAS

Abu-Mostafa, Y. – *Neural Networks for Computing* – American Institute of Physics, pp. 1-6 (1986).

Abu-Mostafa, Y. & **Atiya**, A. - *Introduction to financial forecasting* - Applied Intelligence, 6(3):205-213, Kluwer Academic Publishers (1996).

Almeida, F. & **Dumontier**, P. - *O uso de redes neurais em avaliação de riscos de inadimplência* - Revista de Administração FEA/USP, Vol.31-1 (1996).

Arnauld, A., **Cunha**, R., **Ludermir**, T. & **Adeodato**, P. - *Modelo Híbrido com Redes Neurais Artificiais e Técnicas não-supervisionadas para o problema de credit scoring* - XXV Congresso da Sociedade Brasileira de Computação (2005).

Baestaens, D., **Bergh**, W. & **Wood**, D. - *Neural network solutions for trading in financial markets* - London Pitman (1994).

Bailer-Jones, C. - *Introduction to Machine Learning and Pattern Recognition* - Lecture Hörsaal, Max-Planck, Institut für Astronomie Fev/Mar (2008).

Bigus, J. - *Data mining with neural networks : solving business problems: from application development to decision support* - New York McGraw-Hill (1996).

Bishop, C. *Pattern Recognition and Machine Learning* (Information Science and Statistics). Springer Science+Business (2006).

Brealy, R., **Myers**, S. & **Allen**, F. - *Principles of Corporate Finance, 8th Edition* - Irwin/McGraw-Hill (2006).

Chagas, G. - Precificação de opções de dólar no mercado brasileiro utilizando redes neurais e algoritmos genéticos – Escola de Economia de São Paulo, Fundação Getúlio Vargas (2006).

Chen, A & Leung, M. - Performance Evaluation of Neural Network Architectures: The Case of Predicting Foreign Exchange Correlations - Journal of Forecasting, Vol. 24 (2006).

Chung, K, Tan, S. & Holdsworth, D. - Insolvency Prediction Model Using Multivariate Discriminant Analysis and Artificial Neural Network for the Finance Industry in New Zealand - International Journal of Business and Management, Vol. 39-1 (2008).

Duda, R., Stork, D & Hart, P. - *Pattern Classification* - John Wiley & Sons (2005)

Exame, Revista – *Melhores e Maiores* – Editora Abril (2007).

Foresee, F.D & Hagan, M.T. – *Gauss-Newton approximation to Bayesian regularization* - Proceedings of the 1997 International Joint Conference on Neural Networks, pp. 1930–1935 (1997).

Freitas, S. & Souza, A. - *Utilização de redes neurais na precificação de opções* – Resenha BM&F, Vol.150 (2002).

Gately, E. – *Neural Networks for Financial Forecasting* - John Wiley & Sons (1995).

Hagan, M., Demuth, H. & Beale, M. - *Neural Network Design* – PWS Publishing (1996).

Haykin, S. – *Neural networks : a comprehensive foundation* - New York: Macmillan (1994).

Jason, E.K. – Forecasting Financial Markets Using Neural Networks: an analysis of methods and accuracy – Naval Postgraduate School, Monterey, Califórnia/EUA (1998).

Lam, M – Neural Network techniques for financial performance prediction: integrating fundamental and technical analysis - Decision Support Systems, Vol 37-4 (2004).

Lemos, E., **Steiner**, M. & **Nievola**, J. – Análise de crédito bancário por meio de redes neurais e árvores de decisão : uma aplicação simples de data mining - Revista de Administração, São Paulo - Vol. 40-3 (2005).

Kawato, M, **Furukawa**, K. & **Suzuki**, R. – A hierarchical neural-network model for control and learning of voluntary movement – Springer Berlin/Heidelberg (1987).

Kohonen, T. – *Self-Organization and Associative Memory* – Spring-Verlag, Berlim (1987).

MacKay, D.J.C. – *Bayesian Interpolation* – Neural Computation, Vol. 4-3 – pp. 415-447 (1992).

Marion, J.C. – *Análise das Demonstrações Contábeis: contabilidade empresarial* – 2a Edição - Editora Atlas, São Paulo (2002).

McCullagh, P. & **Nelder**, J. - *Generalized Linear Models* – Chapman & Hall (1994).

McCulloch, W. & **Pitts**, W. – *A logical calculus of the ideas immanent in nervous activity* - Bulletin of Mathematical Biophysics 7:115 – 133 (1943).

McNelis, P. – *A Neural Network Analysis of Brazilian Stock Prices* – Journal of Emerging Markets, Vol. 1-2 (1996).

McNelis, P. – Neural networks in finance-gaining predictive edge in the market - Elsevier Academic Press (2005) .

Medeiros, M., **Veiga**, A. & **Pedreira**, C. - *Modelling Exchange Rates: Smooth Transitions, Neural Networks, and Linear Models* – IEEE Transactions on Neural Networks, 12, 755-646 (2001).

Miller, W., **Sutton**, R.& **Werbos**, P. – *Neural Networks for Control* - MIT Press (1990).

Minsky, M. & **Papert**, S. – *Perceptrons* - MIT Press, Cambridge (1969).

Nascimento, C. & **Yoneyama**, T. - *Inteligência Artificial em Controle e Automação* – Editora Edgard Blücher (2002).

Pinheiro, C., **Evsukoff**, A. & **Ebecken**, N. - *Revenue Recovering with Insolvency Prevention on a Brazilian* - ACM SIGKDD Explorations Newsletter, Vol. 8-1(2006).

Rosenblatt, F. – The Perception, a Probabilistic Model for Information Organization and Storage in the Brain – Psychological Review V65: 363-408 (1958).

Rumelhart, D.E, **Hinton**, G. E. & **Williams**, R.J. - *Learning Representations by Backpropagation Errors* – Nature (London) N. 323, pp. 533-536 (1986).

Sahoo, P. & **Hathy**, P. – *Prediction of Mutual Funds: Use of Neural Network Technique* - The Icfai Journal of Applied Finance, Vol 13-11 (2007).

Saito, K. & **Nakano**, R. – *Medical diagnostic expert system based on PDP model* – Neural Network IEEE International Conference (1988).

Schürmann, J. – Pattern Classification, a unified view of statistical and neural approaches – John Wiley & Sons (1996).

The Mathworks – Manual do software MATLAB e do aplicativo Neural Network Toolbox – The Mathworks, Inc. (2008).

Widrow, B. & Hoff, M. E. - *Adaptive switching circuits* - IRE WESCON, 96-104, New York. Convention Record (1960).

Wooldridge, J. – Introductory Econometrics: A Modern Approach. 2nd Ed. - Thomson South-Western (2003).

Werbos, Paul - Beyond regression :new tools for prediction and analysis in the behavioral sciences – Universidade de Harvard (1974).

Zaneti, L. & Almeida, F. – Exploração do uso de redes neurais na previsão do comportamento de ativos financeiros - III SEMEAD, FEA/USP (1998)

Zemel, R. & Mozer, M. – *Localist Attractor Network* - Neural Computation. 2001-13 – MIT Press (2001).

ANEXO I

Lista de Parâmetros da Exame – Melhores & Maiores

Parâmetro	Numérico	Selecionado
01. Ativo Total Ajustado	Sim.	Não. Correlacionado com 02.
02. Capital Circulante Líquido	Sim.	Sim.
03. Controle Acionário	Não.	Não.
04. Crescimento das Vendas	Sim.	Sim.
05. EBITDA	Sim.	Não. Disponível a partir de 2002.
06. Endividamento a longo prazo	Sim.	Não. Correlacionado com 07.
07. Endividamento geral	Sim.	Sim.
08. Exigível a longo prazo	Sim.	Não. Correlacionado com 07.
09. Exigível total	Sim.	Não. Correlacionado com 07.
10. Exportação	Sim.	Não. Disponível a partir de 1998.
11. Índice de Giro de Ativo	Sim.	Sim.
12. Liderança de mercado	Não.	Não.
13. Liquidez corrente	Sim.	Não. Correlacionado com 15.
14. Índice de Liquidez Geral	Sim.	Sim.
15. Lucro Líquido Ajustado	Sim.	Sim.
16. Lucro Líquido Geral	Sim.	Não. Correlacionado com 15.
17. Margem das Vendas	Sim.	Sim.
18. Passivo Circulante	Sim.	Não. Correlacionado com 07.
19. Patrimônio Líquido Ajustado	Sim.	Sim.
20. Patrimônio Líquido Legal	Sim.	Não. Correlacionado com 19.
21. Produtividade	Não.	Não.
22. Riqueza Criada por empregado	Sim.	Sim.
23. Vendas em dólares	Sim.	Sim.
24. Rentabilidade de Patrimônio Ajustada	Sim.	Sim.

ANEXO II

Tabela de estatística descritiva das variáveis independentes.

Variáveis	Média	Desvio Padrão	Assimetria	Curtose	Mínimo	Máximo	Mediana
CCL ^{t-1}	5,92	322,41	-9,04	124,00	-5185,07	1199,57	21,19
LLA ^{t-1}	34,64	295,43	18,54	381,02	-576,05	6863,18	7,83
PLA ^{t-1}	741,29	2133,53	6,46	56,06	-580,58	29055,26	182,74
RCE ^{t-1}	186,88	462,61	16,42	372,57	-55,02	12835,29	97,44
CV ^{t-1}	9,34	34,55	13,85	337,57	-79,50	973,83	5,25
EG ^{t-1}	51,07	20,34	0,08	2,33	5,27	116,22	50,83
MV ^{t-1}	3,61	11,83	-1,95	19,82	-110,25	65,81	2,91
IGA ^{t-1}	1,42	1,22	2,16	10,17	0,02	10,02	1,08
ILG ^{t-1}	1,15	1,02	5,86	62,81	0,05	17,21	0,95
VD ^{t-1}	919,46	2384,14	12,29	196,85	64,61	47040,11	414,79
VR ^{t-1}	5,25	43,67	-15,05	339,61	-1220,47	200,92	7,06

ANEXO III

Tabela 1: Estatísticas R^2 Ajustado e F dos modelos de RL

Passo	Conjunto											R ² -Ajust	Estat.-F
1	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}	0,1689	41,9073
2	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}		0,1693	46,1133
3	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}			0,1696	51,2350
4	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}				0,1692	57,3521
5	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}					0,1683	65,0059*

Tabela 2: Critério de Informação de Schwartz dos modelos RLO

Passo	Conjunto											SIC
1	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}	1,9901
2	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}		1,9866
3	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}			1,9837
4	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}				1,9812
5	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}					1,9789
6		LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}					1,9772*

Tabela 3: Conjunto de variáveis utilizados na escolha do modelo de RNs.

Ind	Conjunto de variáveis de entrada									
1		LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}			
2	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}			
3	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VD ^{t-1}	VR ^{t-1}		
4	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}		
5	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}		
6	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}		
7	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}	
8	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VD ^{t-1}	VR ^{t-1}	
9	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}	
10	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VD ^{t-1}	VR ^{t-1}	
11	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}	
12	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VR ^{t-1}	
13	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}
14	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1}	VR ^{t-1}
15	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	VD ^{t-1}	VR ^{t-1}
16	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VR ^{t-1}
17	CCL ^{t-1}	LLA ^{t-1}	PLA ^{t-1}	RCE ^{t-1}	CV ^{t-1}	EG ^{t-1}	MV ^{t-1}	IGA ^{t-1}	ILG ^{t-1}	VD ^{t-1} VR ^{t-1}