

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO**

Programa de Pós-Graduação em Administração

**Modelo híbrido SOM-ANN/BP para previsão de índices
da NYSE através de redes neurais artificiais**

por
Adriano Beluco

Porto Alegre, junho de 2013

CIP - Catalogação na Publicação

Beluco, Adriano
Modelo híbrido SOM-ANN/BP para previsão de índices
da NYSE através de redes neurais artificiais /
Adriano Beluco. -- 2013.
113 f.

Orientadora: Denise Lindstrom Bandeira.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Escola de Administração, Programa
de Pós-Graduação em Administração, Porto Alegre, BR-RS,
2013.

1. Pesquisa Operacional. 2. Redes Neurais
Artificiais. 3. Modelos de previsão financeira. 4.
Bolsa de valores de New York. I. Lindstrom Bandeira,
Denise, orient. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO

Programa de Pós-Graduação em Administração

**Modelo híbrido SOM-ANN/BP para previsão de índices
da NYSE através de redes neurais artificiais**

Adriano Beluco

Orientadora: Prof^a Dr^a Denise L. Bandeira

Dissertação de mestrado, apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul como requisito para a obtenção do título de Mestre em Administração.

Porto Alegre, junho de 2013

AGRADECIMENTOS

Meus agradecimentos à Escola de Administração e ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul pela oportunidade de cursar o mestrado acadêmico e ter propiciado a construção do conhecimento obtido no período. Igualmente à Universidade por ter oferecido tal curso de grande qualidade de forma gratuita.

Gostaria de agradecer à minha orientadora do trabalho, a Prof.^a Dr.^a Denise Lindstrom Bandeira por se mostrar sempre disponível e honesta quanto à elaboração do trabalho, bem como à sua finalização. Ressalto também sua parceria e perseverança nos momentos de maior dificuldade.

Ao Prof. Dr. João Luiz Becker, pela inestimável capacidade de incentivar a construção do conhecimento em suas aulas, de forma agradável e eficiente. Também ao Prof. Dr. Eugênio Pedrozo pelo compartilhamento de sua larga experiência nos debates desenvolvidos em sala de aula. Ao Prof. Dr. Denis Borenstein pelos grandes ensinamentos na área de pesquisa operacional.

Também sou fortemente agradecido a duas pessoas responsáveis pela minha inserção no caminho da Administração de Empresas, minha esposa Bianca e minha filha Amanda. Da mesma forma, agradeço aos colegas de curso Diego e Camilo que muito auxiliaram no transcorrer dos trabalhos.

Agradeço de forma especial ao meu irmão Alexandre e à minha mãe Sirlei pelo apoio afetivo e intelectual durante todo o período.

“A suprema arte da guerra é derrotar o inimigo sem lutar.”

Sun Tzu

*“No meio da confusão, encontre a simplicidade.
No meio da dificuldade reside a oportunidade.”*

Albert Einstein

RESUMO

Este estudo propõe um modelo híbrido que reúne uma rede neural do tipo SOM (*Self-Organizing Map*) com uma rede neural do tipo Multicamadas com Retropropagação (BPN: *Backpropagation Network*). A utilização da rede SOM tem o intuito de segmentar a base de dados em diversos clusters, onde são ressaltadas suas diferenças. A rede BPN é usada para construir um modelo matemático de previsão que descreve a relação entre os indicadores e o valor de fechamento de cada cluster formado na rede SOM. A viabilidade e o percentual de efetividade do modelo proposto são demonstrados através de experimentos de predição de índices utilizados pelo NYSE (New York Stock Exchange). O modelo foi elaborado a partir de uma base de dados composta pelo índice NYSE Composite U.S. 100 no período entre 02 de abril de 2004 a 08 de novembro de 2012. Como variáveis de entrada para as redes neurais, foram utilizados 10 índices: MA_10, BIAS_20, WMS%R_9, K_9, D_9, MTM_10, ROC_10, CCI_24, AR_26, BR_26. Os resultados obtidos com o modelo híbrido proposto se mostraram superiores aos obtidos com modelos convencionais estatísticos.

Palavras-chave: Previsão de índices financeiros, Redes neurais artificiais, Mapas auto-organizáveis; SOM.

ABSTRACT

This study proposes a hybrid model that combines a neural network SOM (Self-Organizing Map) with a neural network with Multilayer Backpropagation (BPN: Backpropagation Network). The SOM aims to segment the database into different clusters, where they highlight their differences. The BPN network is used to build a predictive mathematical model that describes the relationship between the indicators and the closing value of each cluster formed in the SOM. The percentage of viability and effectiveness of the proposed model are demonstrated through experiments predict index used by the NYSE (New York Stock Exchange). The model was developed from a database composed of 100 U.S. NYSE Composite Index in the period from April, 02, 2004 to November, 08, 2012. As input variables for neural networks, we used 10 indices: MA_10, BIAS_20, WMS%R_9, K_9, D_9, MTM_10, ROC_10, CCI_24, AR_26, BR_26. Results obtained with the proposed hybrid model were higher than those obtained with conventional statistical techniques.

Keywords: Forecast financial indices, Artificial neural networks, Self-organizing maps (SOM).

LISTA DE FIGURAS

Figura 1 - Modelo de neurônio	5
Figura 2 - Mapa auto-organizável: Modelo de Kohonen.....	6
Figura 3 – Conexões entre os neurônios da grade pós-sináptica	7
Figura 4 – Grade bidimensional de neurônios.....	8
Figura 5 – Diagrama de fluxo para o algoritmo de treinamento SOM	13
Figura 6 – Modelo de neurônio artificial	17
Figura 7 – Funções de transferência.....	18
Figura 8 – Estrutura da RNA com uma camada intermediária	19
Figura 9 – Estrutura da RNA com duas camadas intermediárias.....	20
Figura 10 - Perceptron de múltiplas camadas (com duas camadas intermediárias) .	21
Figura 11 - Esquema geral do modelo híbrido SOM-ANN/BP.....	30
Figura 12 - U-Matrix e planos componentes da rede SOM	39
Figura 13 - U-Matrix da rede SOM-ANN/BP.....	39
Figura 14 - Plano componente da rede SOM referente ao indicador MA_10.....	40
Figura 15 - Planos componentes da rede SOM referente aos índices(a) BIAS_20,(b) MTM_10,.(c) ROC_10, (d) K_9, (e) D_9 e (f) CCI_24.....	40
Figura 16 - Plano componente da rede SOM-ANN/BP referente ao indicador WMS%R_9.....	41
Figura 17 - Planos componentes da rede SOM referente aos indicadores AR_26 e BR_26.	41
Figura 18 - Topologia inicial do Mapa Auto-organizável (SOM)	42
Figura 19 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 5 camadas intermediárias	44
Figura 20 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 10 camadas intermediárias	45
Figura 21 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 15 camadas intermediárias	46
Figura 22 - RMSE para o treinamento do modelo neural referente ao Cluster 1 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	48

Figura 23 - RMSE para o treinamento do modelo neural referente ao Cluster 2 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	49
Figura 24 - RMSE para o treinamento do modelo neural referente ao Cluster 3 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	50
Figura 25 - RMSE para o treinamento do modelo neural referente ao Cluster 4 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	51
Figura 26 - RMSE para o treinamento do modelo neural referente ao Cluster 5 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	52
Figura 27 - RMSE para o treinamento do modelo neural referente ao Cluster 6 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	53
Figura 28 - RMSE para o treinamento do modelo neural referente ao Cluster 7 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária.....	54
Figura 29 - Múltiplos diagramas de dispersão entre as variáveis ROC_10, CCI_24, BIAS_20, K_9 e WMS%R_9.	59
Figura 30 - Gráfico de regressão entre o índice financeiro MA_10 e o valor de Fechamento do índice NYSE Composite U.S. 100	60
Figura 31 - Gráfico de regressão entre o valor de Fechamento do índice NYSE Composite U.S. 100 e o índice: (a) BIAS_20; (b) WMS%R_9; (c) K_9; (d) D_9; (e) MTM_10; (f) ROC_10.....	61
Figura 32 - Gráfico de regressão entre o valor de Fechamento do índice NYSE Composite U.S. 100 e o índice: (a) CCI_24; (b) AR_26; (c) BR_26.	62
Figura 33 - Valor de fechamento real (VF Real) e projetado (VF Projetado) pelo modelo de regressão múltipla no período de 23/06/2004 a 08/11/2012...	65
Figura 34 - Valor de fechamento real (VF Real) e projetadoS (VF SOM-ANN/BP, VF Enter, VF Stepwise, VF Forward e VF Backward) no período de 23/06/2004 a 08/11/2012.....	77

LISTA DE TABELAS

Tabela 1 - Pesquisas na área de aplicações financeiras de modelos de RNA	26
Tabela 2 - Resumo estatístico descritivo das variáveis componentes da base de dados.....	33
Tabela 3 - Resultados do processo de clusterização	42
Tabela 4 - Análise do desempenho dos modelos neurais	47
Tabela 5 - Composição da camada intermediária dos modelos de melhor desempenho por cluster	54
Tabela 6 - Avaliação do desempenho dos modelos neurais para o grupo de testes	55
Tabela 7 - Avaliação do desempenho dos modelos para o grupo de validação	55
Tabela 8 - Avaliação do desempenho do modelo híbrido SOM-ANN/BP	56
Tabela 9 - Matriz de correlação entre as variáveis independentes	58
Tabela 10 - Correlação entre as variáveis independentes e a variável dependente .	60
Tabela 11 - Análise dos coeficientes do modelo de regressão através do método Enter.....	63
Tabela 12 - Análise do coeficiente de determinação.....	64
Tabela 13 - Análise de variância (ANOVA)	64
Tabela 14 - Desempenho das projeções do modelo de regressão através do método Enter.....	65
Tabela 15 - Análise dos coeficientes do modelo de regressão através do método Stepwise.....	66
Tabela 16 - Análise de variância (ANOVA)	68
Tabela 17 - Desempenho das projeções do modelo de regressão através do método Stepwise	69
Tabela 18 - Análise dos coeficientes do modelo de regressão através do método Forward	71
Tabela 19 - Análise de variância (ANOVA)	72
Tabela 20 - Desempenho das projeções do modelo de regressão através do método Forward	73
Tabela 21 - Análise dos coeficientes do modelo de regressão através do método Backward.....	74
Tabela 22 - Análise de variância (ANOVA)	75

Tabela 23 - Desempenho das projeções do modelo de regressão através do método Backward.....	75
Tabela 24 - Comparativo dos resultados dos modelos SOM-ANN/BP e de regressão	76

LISTA DE SIGLAS E ABREVIATURAS

ANN	<i>Artificial Neural Network</i>
AR_26	Indicador de momento de compra e venda
BIAS_20	Bias no período de 20 dias anteriores
BPN	<i>Backpropagation Network</i>
BR_26	Indicador de tendência de compra e venda no período de 26 dias anteriores
CCI_24	<i>Comodity Channel Index</i> no período de 24 dias anteriores
CP-ANN	<i>Counter-Propagation Artificial Neural Network</i>
D_9	Índice estocástico D no período de 9 dias anteriores
EMQ	Erro Médio Quadrado
FA	<i>Factorial Analysis</i> (Análise Fatorial)
FFT	<i>Fast Fourier Transform</i>
GA	<i>Genetic Algorithms</i> (Algoritmos Genéticos)
GRNN	<i>Generalized Regression Neural Network</i> (Rede Neural de Regressão Generalizada)
IDFT	<i>Inverse Discrete Fourier Transform</i>
K_9	Índice estocástico K no período de 9 dias anteriores
MA_10	Média móvel no período de 10 dias anteriores
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MATLAB	<i>Matrix Laboratory</i> (Software Matemático Matricial)
MTM_10	Momento no período de 10 dias anteriores
NYSE	<i>New York Stock Exchange</i>
RMSE	<i>Root Mean Square Error</i>
RNA	Redes Neurais Artificiais

RNN	<i>Recurrent Neural Network</i> (Rede Neural Recorrente)
ROC_10	Taxa de variação no período de 10 dias anteriores
SOM	Self-Organizing Map
SPSS	<i>Statistical Package for Social Sciences</i> (Software Estatístico)
SVM	<i>Support Vector Machines</i> (Máquinas de Vetores de Suporte)
TAIEX	<i>Taiwan Stock Exchange Capitalization Weighted Stock Index</i>
WMS%R_9	Índice de Williams de supervalorização e subvalorização no período de 9 dias anteriores

SUMÁRIO

1. INTRODUÇÃO	1
1.1 OBJETIVO GERAL	3
1.2 OBJETIVOS ESPECÍFICOS	3
1.3 QUESTÃO DA PESQUISA	4
1.4 MOTIVAÇÃO E RELEVÂNCIA.....	4
1.5 ORGANIZAÇÃO DO ESTUDO	4
2. REVISÃO DA LITERATURA	5
2.1 REDES NEURAIS.....	5
2.1.1 Mapas auto-organizáveis	6
2.1.2 Perceptron multicamadas	16
2.2. MODELOS DE PREVISÃO DE ÍNDICES FINANCEIROS	26
3. MÉTODOS E PROCEDIMENTOS	29
3.1 BASE DE DADOS.....	30
3.2 ESQUEMA GERAL DO MODELO	31
4. MODELO HÍBRIDO SOM-ANN/BP	34
4.1 PRÉ-PROCESSAMENTO DOS DADOS	34
4.1.1 Índices financeiros técnicos	35
4.1.2 Normalização da base de dados	39
4.2 CONSTRUÇÃO DO MODELO HÍBRIDO SOM-ANN/BP	40
4.2.1 Mapas auto-organizáveis	40
4.2.2 Algoritmo <i>backpropagation</i>	46
4.3 AVALIAÇÃO DO MODELO HÍBRIDO SOM-ANN/BP	59
5. MODELOS DE REGRESSÃO MÚLTIPLA	60
5.1 COMPORTAMENTO DAS VARIÁVEIS	60
5.2 MODELOS DE REGRESSÃO.....	65
6. COMPARATIVO ENTRE OS MODELOS	80
7. CONSIDERAÇÕES FINAIS	82
REFERÊNCIAS	85
ANEXO A – EMPRESAS LISTADAS NO ÍNDICE NYSE COMPOSITE U.S.100	89

APÊNDICE A – GRÁFICOS DE EVOLUÇÃO DAS VARIÁVEIS DA BASE DE DADOS	92
APÊNDICE B – GRÁFICOS DE EVOLUÇÃO DOS ÍNDICES FINANCEIROS	95

1. INTRODUÇÃO

Informação é fundamental para a tomada de decisão de investidores. Cada vez mais novas tecnologias como técnicas estatísticas multivariadas e redes neurais artificiais são utilizadas para a projeção e estimação de dados com o intuito de obter melhores resultados financeiros e estratégicos, baseadas em informações financeiras. As aplicações envolvem tarefas como modelagem, projeção de séries de dados e classificação, algumas típicas da pesquisa operacional.

A previsão de índices é um tópico de pesquisa de extrema importância para o processo de tomada de decisão no mercado financeiro. No entanto, as séries temporais financeiras apresentam propriedades de grande volatilidade, complexidade e ruído. O uso de técnicas simples para obter uma previsão adequada é desaconselhável. Vários pesquisadores têm adotado procedimentos híbridos na busca de melhores resultados (HSU, 2011; LU; WU, 2011; LIU; WANG, 2012; DAÍ et al., 2012).

Desde a década de 1990, vários autores têm utilizado conhecimentos de redes neurais artificiais em aplicações de avaliação e classificação de investimentos. Wang et al. (2011) implementaram uma rede neural do tipo *backpropagation* para a previsão de índices de ações da bolsa, aplicando algoritmos de transformada *wavelet* na correção de ruídos na camada de entrada. Kim et al. (2010) elaboraram um algoritmo baseado em redes neurais para a previsão de dividendos corporativos, elemento fundamental para o cálculo do valor de mercado de uma empresa. Liang et al. (2011) apresentaram um modelo de algoritmo de redes neurais com ênfase processual para a previsão de índices de ações da bolsa de valores, baseado em dados espaciais e temporais de forma síncrona, resultando em um modelo multidimensional para problemas de séries temporais. É possível observar uma preponderância de trabalhos envolvendo redes neurais artificiais focados na previsão de séries temporais dedicados a investimentos em ações.

As técnicas de RNA (Redes Neurais Artificiais) apresentam atributos de aprendizagem, generalização e processamento paralelo, o que as tornam hábeis na solução de problemas de grande complexidade. Em termos teóricos, respeitado o período necessário para o treinamento, a rede neural tem a capacidade de simular qualquer tipologia de séries de dados. Dessa forma, é justificada a utilização em

grande escala em aplicações financeiras que dispõem de grandes quantidades de séries de dados.

A predição de índices financeiros ligados às bolsas de valores tem recebido a atenção de diversos pesquisadores nos últimos anos (LEE, 2009; TSAI; HSIAO, 2010; KARA et al., 2011; DAI et al., 2012). Conforme Hsu (2011), por apresentar características de alta complexidade, além do ruído, o problema é alvo de inúmeras metodologias, podendo ser classificado em 3 categorias: análise fundamentalista, análise técnica e previsão de índices temporais.

A análise fundamentalista enfoca o uso das informações financeiras básicas de uma instituição com o objetivo de prever demandas, fornecimentos e lucro, bem como quaisquer fatores que possam afetar o seu valor de mercado ou índice de crescimento. Do ponto de vista dos investidores, a demonstração de resultados das empresas, relatórios contábeis trimestrais e séries temporais financeiras são os fundamentos técnicos para a previsão de tendências a respeito de índices de crescimento e de lucros (HSU, 2011).

Por outro lado, a análise técnica leva em consideração o valor das ações e a análise de suas séries temporais e fatores que as afetam, como volume de negociações, volatilidade dos índices e tendências do setor. No entanto, a análise técnica também considera indicadores relacionados que possam afetar a tendência histórica das ações (TSAI; HSIAO, 2010; HSU, 2011).

A previsão de índices através de séries temporais por meios tradicionais fazem uso de modelos estocásticos tradicionais como modelos de regressão linear multivariada, ARIMA (*Auto Regressive Integrated Moving Average*) e GARCH (*Generalized Auto Regressive Conditional Heteroskedasticity*). No entanto, técnicas isoladas não são apropriadas para a análise de séries temporais financeiras, visto que são caracterizadas por alta complexidade de fatores relacionados, grande volatilidade e ruído exacerbado (TSAI; HSIAO, 2010).

A idéia de utilizar modelos híbridos é ampliar a acurácia de técnicas isoladas através da complementação dos erros produzidos por modelos de forma individualizada. Ince e Trafalis (2004) elaboraram um modelo híbrido baseado em redes neurais do tipo *backpropagation* e máquinas de vetor suporte, analisando índices técnicos a partir da bolsa NASDAQ. Enke e Thavornwong (2005) e Li et al. (2009) fizeram uso de modelos híbridos baseados em redes *backpropagation* a partir de índices técnicos, variáveis financeiras e econômicas. Li e Kuo (2008) e Hsu (2011)

propuseram modelos híbridos com o uso de mapas auto-organizáveis no papel de clusterização de índices técnicos, aliados a modelos de previsão baseados em algoritmos genéticos a partir da bolsa de valores de Taiwan. Lai et al. (2009) também utilizaram dados da bolsa de Taiwan, projetando um modelo que usa abordagem estatística multivariada e algoritmos genéticos com 7 índices técnicos. Huang e Tsai (2009) estruturaram um modelo híbrido baseado em mapas auto-organizáveis e máquinas de vetor suporte a partir de dados da bolsa de índices futuros de Taiwan demonstrados em 13 indicadores técnicos. De forma geral, os estudos citados obtiveram desempenho superior à aplicação de técnicas isoladas.

Com base nos estudos citados, é fato que a utilização de modelos híbridos para a previsão de séries temporais econômicas e financeiras a partir de indicadores técnicos vem obtendo resultados de maior confiabilidade, oportunizando um processo de tomada de decisão com mais segurança. O propósito deste estudo é estruturar um modelo híbrido de previsão de séries temporais financeiras baseado em 10 indicadores técnicos e comparar seu desempenho com modelos paramétricos obtidos por regressão multivariada.

1.1 OBJETIVO GERAL

Desenvolver um modelo híbrido de previsão de índices de ações da bolsa baseado em redes neurais artificiais, comparando-o com um modelo de regressão multivariada.

1.2 OBJETIVOS ESPECÍFICOS

Este estudo tem como objetivos específicos:

- Pesquisar as diversas variáveis de entrada adequadas à construção do modelo híbrido de previsão de índices da bolsa de Nova York;
- Treinar e validar as estruturas topológicas das redes neurais artificiais selecionadas através das séries temporais das ações;
- Comparar os resultados obtidos no modelo híbrido com modelos estocásticos tradicionais;

1.3 QUESTÃO DE PESQUISA

A utilização de um modelo híbrido composto por duas redes neurais com características díspares é mais eficiente para a predição de índices financeiros do que os métodos estocásticos tradicionais?

1.4 MOTIVAÇÃO E RELEVÂNCIA

Séries temporais relativas a ações de empresas sempre foram informações de maior importância para investidores e administradores. No entanto, seu caráter dinâmico, não linear e não paramétrico produz uma série de ruídos quando existe a necessidade de previsão a curto e médio prazos de tais índices. É fato que as cotações de ações são afetadas por fatores macroeconômicos tais como condições econômicas, escolhas de investidores institucionais, eventos políticos, entre outros. Vários autores (CHEN; LEUNG, 2003, THAVORNWONG; ENKE, 2004; ARMANO et al., 2005; WANG et al., 2011) têm obtido resultados importantes no uso de algoritmos de redes neurais artificiais e de pesquisa operacional na tentativa de modelar séries econômicas.

1.5 ORGANIZAÇÃO DO ESTUDO

Este estudo está organizado em 5 capítulos. Este capítulo apresenta uma introdução ao assunto, construindo uma visão geral do trabalho desenvolvido. O Capítulo 2 apresenta uma revisão bibliográfica de conteúdos e trabalhos desenvolvidos por pesquisadores da área, destacando tópicos de maior relevância para a compreensão do estudo. No Capítulo 3 é possível verificar a metodologia proposta a ser desenvolvida. A apresentação e análise dos resultados é realizada no Capítulo 4 e as considerações finais e sugestões para a continuidade do estudo no Capítulo 5. As referências bibliográficas finalizam a apresentação deste trabalho, juntamente com os anexos e apêndices.

2. REVISÃO DA LITERATURA

Neste tópico são apresentados os referenciais teóricos referentes às técnicas de redes neurais artificiais atualmente usadas em séries temporais de dados financeiro-econômicas, aos métodos de previsão de índices financeiros e alguns estudos realizados em ambas as áreas.

2.1 REDES NEURAIS

As redes neurais artificiais representam um conceito de sistema de processamento onde os modelos são baseados em princípios de processamento neurofisiológico. O cérebro é composto por células diferenciadas denominadas neurônios (Figura 1), os quais possuem um corpo celular (soma) onde se encontram a maioria dos seus orgânulos. Do soma de cada neurônio prolongam-se dendritos e axônios (entradas e saídas). Cada neurônio recebe impulsos elétricos pelos seus dendritos, que são processados no soma e transmitidos através dos axônios até os dendritos de outros neurônios. As conexões entre axônios e dendritos dos neurônios são definidas como sinapses (HAYKIN, 2001), que são unidades funcionais básicas para a constituição de circuitos neurais biológicos.

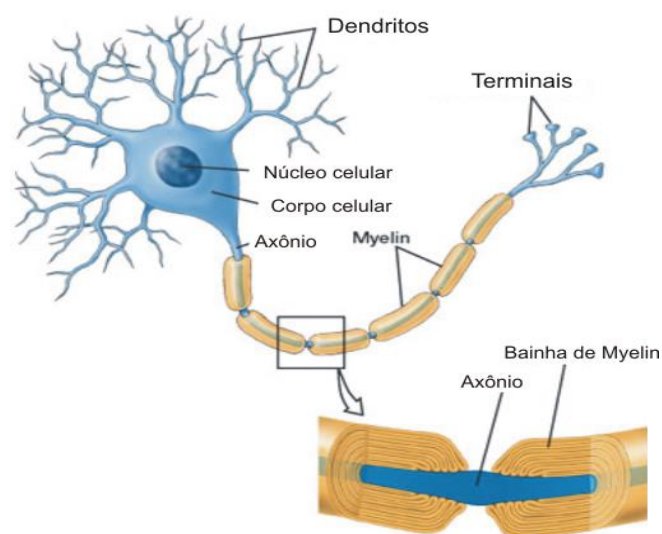


Figura 1 – Modelo de neurônio
Fonte: Adaptada de Haykin (2001)

Nas últimas décadas, as redes neurais artificiais têm sido aplicadas na solução de problemas das mais diversas áreas, porém a maioria concentra-se na utilização de uma única concepção de rede, como a *backpropagation* (retropropagação) para a modelagem de séries temporais. Cada modelo de rede neural artificial possui características específicas que respondem mais adequadamente a determinada classe de problema.

2.1.1 Mapas autoorganizáveis

Os mapas autoorganizáveis são redes neurais com o propósito específico do agrupamento de dados semelhantes entre si ao ponto de formarem *clusters*. Uma rede SOM (*Self-Organizing Maps*) é constituída, basicamente, por duas camadas: a de entrada e a de saída, representada por uma grade pós-sináptica ou mapa de características uni ou bidimensional.

A camada de saída é formada por uma malha de neurônios interligados apenas aos neurônios mais próximos (Figura 2), onde cada neurônio da malha representa um cluster. No entanto, os neurônios que formam a camada de entrada são conectados a todos os neurônios da grade pós-sináptica (Figura 3).

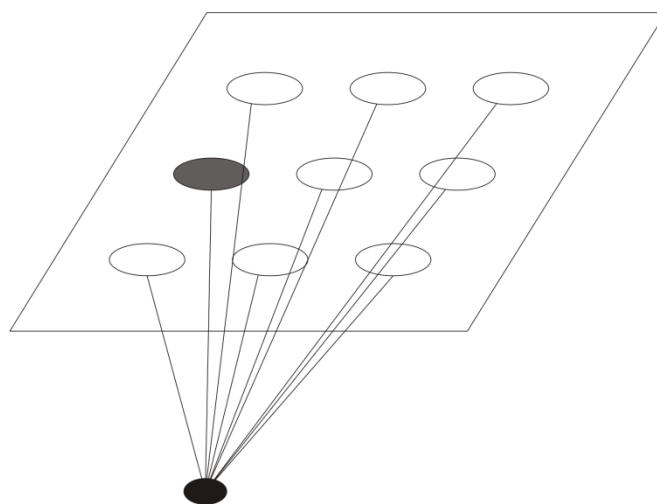


Figura 2 – Mapa auto-organizável: Modelo de Kohonen
Fonte: Adaptada de Haykin (2001)

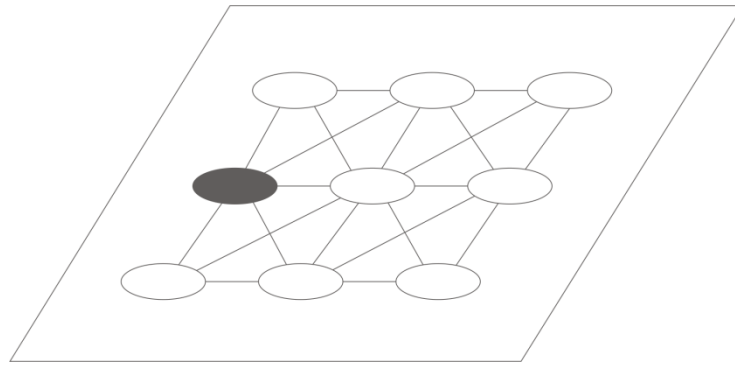


Figura 3 – Conexões entre os neurônios da grade pós-sináptica
Fonte: Adaptada de Haykin (2001)

Uma rede do tipo SOM constitui uma classe de redes neurais artificiais cuja aprendizagem ocorre de forma não-supervisionada. Tal classe também é denominada como Rede de Kohonen (1982).

A base biológica de um mapa autoorganizável parte do princípio do mapeamento ordenado do córtex cerebral a partir de diferentes entradas sensoriais de proveniência motora, visual e auditiva (HAYKIN, 2001). Kohonen (1990) formula o princípio da formação de mapas topográficos através da correspondência entre a localização espacial de um neurônio de saída em um mapa topográfico e uma característica específica de um dado retirado do espaço de entrada. Este princípio é a base da motivação neurobiológica para modelos de mapeamento de características, entre os quais o modelo proposto por Kohonen (1982), caracterizado pela capacidade de compressão de dados de entrada da rede. A rede SOM transforma um padrão de entrada de dimensão arbitrária de maneira topologicamente ordenada em um mapa discreto uni ou bidimensional (HAYKIN, 2001).

Na Figura 4 é apresentado um diagrama que representa uma grade bidimensional de neurônios na forma de um mapa discreto. Neste diagrama, a grade é representada como uma única camada computacional organizada em linhas e colunas. No caso de uma grade unidimensional, a camada é formada apenas por uma linha ou coluna. Segundo Kohonen (1982), um número suficiente de padrões de entrada deve ser apresentado a todos os neurônios da grade, a fim de que o processo de auto-organização tenha a possibilidade de convergir para a melhor solução.

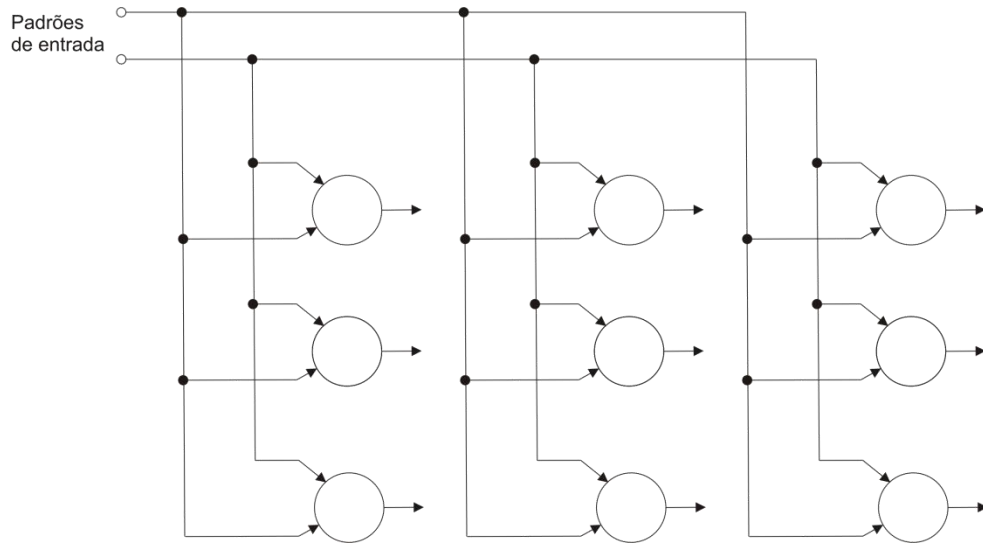


Figura 4 – Grade bidimensional de neurônios
 Fonte: Adaptada de Haykin (2001)

2.1.1.1 Algoritmo de treinamento

A primeira etapa do algoritmo é a inicialização aleatória dos pesos sinápticos da grade. A característica randômica desta etapa garante a ausência de uma organização prévia do mapa. A formação do mapa é constituída por três processos: competição, cooperação e adaptação.

No processo de competição, os neurônios da grade determinam os valores de uma função discriminante correspondentes a cada padrão de entrada. O neurônio que apresenta o resultado mais alto para a função discriminante utilizada é denominado o neurônio vencedor. Cada vetor de entrada do espaço de dados de entrada é submetido a uma função discriminante que será responsável por constituir a base para a competitividade entre os neurônios, onde o neurônio que obtiver o maior valor da função discriminante será designado como o vencedor. O vetor w_j dos pesos de cada neurônio da grade sináptica possui a mesma dimensionalidade do espaço de entrada (HAYKIN, 2001). A representação do vetor de entrada (x) e do vetor dos pesos sinápticos (w_j) pode ser realizada através de:

$$x = [x_1, \dots, x_m]^T \quad (1)$$

$$w_j = [w_{j1}, \dots, w_{jm}]^T \quad j = 1, \dots, l \quad (2)$$

Onde:

m : dimensionalidade do espaço de entrada

j : número de neurônios na grade sináptica

O processo de cooperação específica, a partir do neurônio considerado vencedor, a localização espacial da vizinhança topológica de neurônios em estado de excitação. Para obter a melhor relação do vetor de entrada com os vetores de pesos sinápticos obtém-se o melhor deles através do produto interno $w_j^T x$. A resposta da grade sináptica pode indicar a posição do neurônio vencedor ou o vetor do peso sináptico mais próximo do vetor de entrada, considerando a distância euclidiana. Tal etapa baseia-se na comprovação neurobiológica de interação lateral entre neurônios excitados. Dessa forma, é possível afirmar que a vizinhança topológica ao redor de um neurônio vencedor é reduzida conforme sua distância lateral (KOHONEN, 2013). Sendo $d_{i,j}$ a distância entre um neurônio vencedor i e um neurônio excitado j , e $h_{i,j}$ a vizinhança topológica em torno de um ponto de ativação máxima do neurônio vencedor cuja amplitude decresce com o aumento da distância lateral, uma escolha típica de função para $h_{i,j}$ é a função gaussiana (HAYKIN, 2001):

$$h_{i,j} = \exp\left(-\frac{d_{i,j}^2}{2\sigma^2}\right) \quad (3)$$

Onde σ representa a largura da vizinhança topológica, indicando o grau de participação do neurônio no processo de aprendizagem. O uso da função gaussiana faz com que o algoritmo SOM convirja com maior rapidez que uma função com característica topológica retangular (KOHONEN, 2013). Uma característica única do algoritmo SOM é a variação da vizinhança topológica de acordo com o tempo. Uma forma de tratar tal característica é através da dependência da largura da vizinhança topológica σ com o tempo discreto τ através da equação de decaimento exponencial proposta por Ritter et al. (1992):

$$\sigma = \sigma_o \exp\left(-\frac{n}{\tau}\right) \quad n = 0, 1, \dots \quad (4)$$

Onde σ_0 indica o valor de σ na inicialização do algoritmo SOM e τ é uma constante de tempo.

A etapa de adaptação possibilita que a base de neurônios estimulados em torno do neurônio vencedor aumente os resultados da função discriminante a partir dos padrões de entrada por meio de ajustes nos seus pesos sinápticos (KOHONEN, 2001). Ou seja, para que a grade sináptica seja autoorganizável o vetor dos pesos da grade deve se modificar em relação a cada vetor de entrada. A atualização dos pesos sinápticos ocorre através da equação 5.

$$\Delta w_j = \eta y_j x - g(y_j) w_j \quad (5)$$

Onde

η : parâmetro da taxa de aprendizagem

y_j : sinal de resposta

x : vetor de entrada

$g(y_j)$: função escalar positiva

w_j : vetor de pesos sinápticos

O primeiro termo da equação 5 é conhecido como termo hebbiano e o segundo termo é o de esquecimento (HAYKIN, 2001). O termo hebbiano é assim definido por se basear no postulado de aprendizagem de Hebb, onde o peso de um neurônio é aumentado a partir da simultânea de atividades pré-sináptica e pós-sináptica. O termo de esquecimento é um ajuste realizado em relação ao postulado de Hebb para a aprendizagem não-supervisionada utilizada nos mapas auto-organizáveis.

No entanto, existe uma exigência em relação à função $g(y_j)$:

$$g(y_j) = 0 \quad \text{para } y_j = 0 \quad (6)$$

Para cumprir tal exigência, seleciona-se uma função linear para $g(y_j)$, como por exemplo:

$$g(y_j) = \eta y_j \quad (7)$$

Fazendo:

$$y_j = h_{j,i(x)} \quad (8)$$

Podemos usar as equações 7 e 8 para simplificar a equação 5, obtendo a equação 9.

$$\Delta w_j = \eta h_{j,i(x)}(x - w_j) \quad (9)$$

Fazendo uso do formalismo de tempo discreto e considerando o vetor de pesos sinápticos $w_j(n)$ do neurônio j no tempo n , a equação 10 define o vetor de pesos atualizado $w_j(n + 1)$ no tempo $n + 1$ (KOHONEN, 1982).

$$w_j(n + 1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x - w_j(n)) \quad (10)$$

A equação 10 é aplicada a todos os neurônios da grade sináptica dentro da vizinhança topológica do neurônio vencedor i . Ou seja, a equação 10 provoca o movimento do vetor de pesos w_i do neurônio vencedor i em direção ao vetor de entrada x (HAYKIN, 2001). Assim, tal equação conduz à ordenação topológica do mapa de características no espaço de entrada.

Como a taxa de aprendizagem deve ser variável em função do tempo e, em particular, deve decrescer gradualmente de acordo com o tempo, Haykin (2001) propõe o uso de uma função de decaimento exponencial, obtendo a equação 11.

$$\eta(n) = \eta_0 e^{-\frac{n}{\tau_2}} \quad \text{para } n = 0, 1, 2 \dots \quad (11)$$

Onde

τ_0 : constante de tempo do algoritmo SOM

Kaski (1997) apresenta um modelo padrão de algoritmo de treinamento:

1. Inicialização e definições de parâmetros da rede SOM

- a. Inicialização dos pesos W_{ij}
- b. Definição dos parâmetros de vizinhança
- c. Definição dos parâmetros de aprendizagem

2. Enquanto a condição de parada for falsa, fazer:

- a. Para cada j , calcular:
 - i. $D(j) = \arg \min_j \{\|W_j - X_j\|\}$
 - ii. Encontrar o índice j tal que $D(j)$ seja um mínimo
 - iii. $\forall j \in N_c$ de j e $\forall i$:

$$W_{ij}(\text{nov}) = W_{ij}(\text{antigo}) + \eta [X_j - W_{ij}(\text{antigo})]$$
- b. Atualizar a taxa de aprendizagem
- c. Reduzir o raio da vizinhança

A taxa de aprendizagem apresenta um decaimento lento até a convergência para a solução final.

A Figura 5 apresenta um diagrama de fluxo para o algoritmo de treinamento SOM. O passo inicial do diagrama é a leitura dos vetores dos padrões de entrada X_1, \dots, X_n , considerando a utilização de n padrões durante o processo de treinamento (bloco 1). No bloco 2, ocorre a seleção dos valores iniciais dos pesos sinápticos das conexões W_{ij} , onde $i = 1, \dots, n$ e $j = 1, \dots, m$, e do raio de vizinhança N_c .

Após a inicialização dos pesos W_{ij} e do raio N_c , é processada a função discriminante (ou de ativação) de cada nó de saída. O neurônio vencedor será aquele com maior valor da função discriminante (bloco 6). No bloco 7 é processada a atualização dos pesos das conexões do neurônio vencedor e dos neurônios vizinhos. Tais procedimentos referentes aos blocos 5 a 7 são repetidos para cada padrão de entrada até que todas as n amostras sejam apresentadas, completando uma iteração.

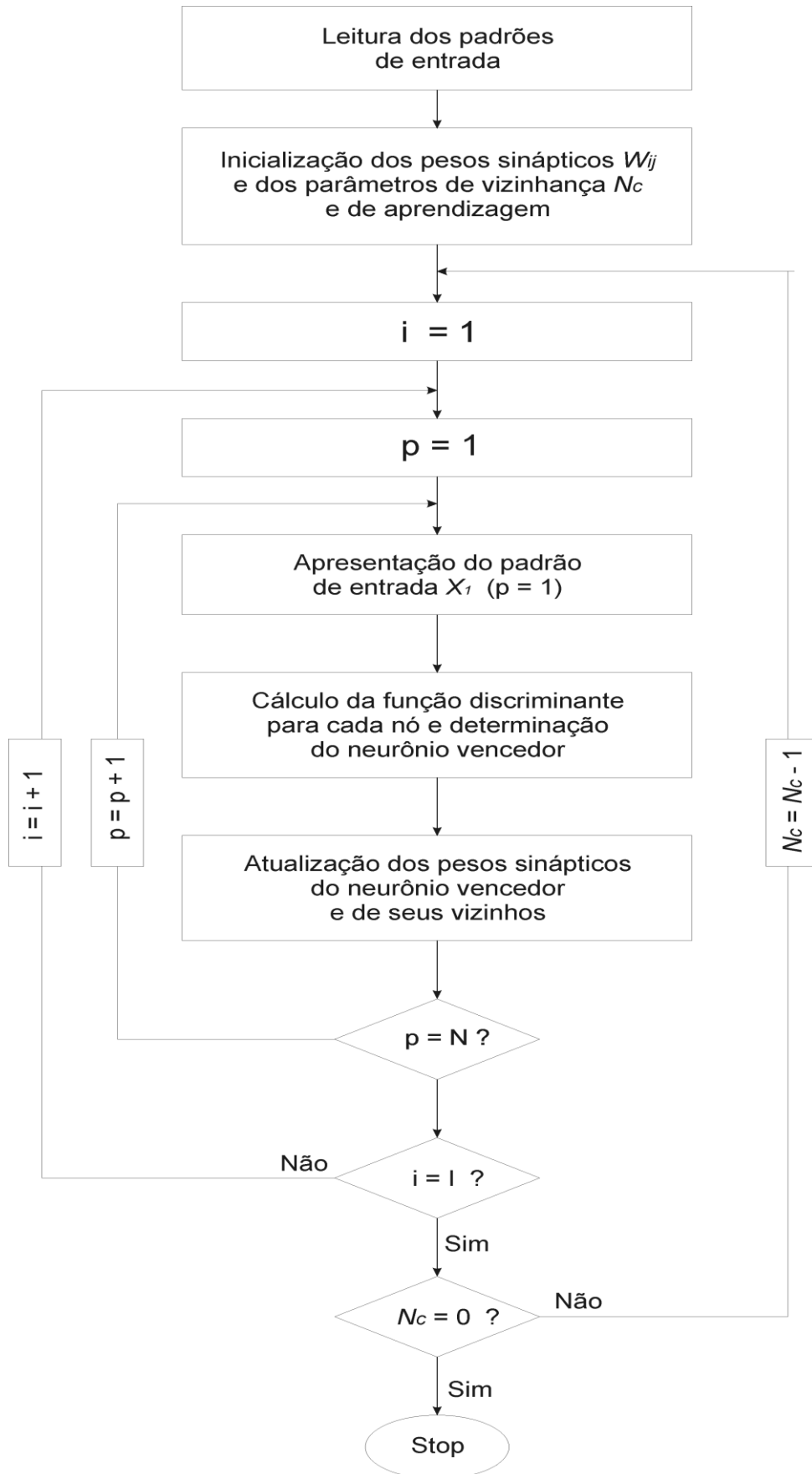


Figura 5 – Diagrama de fluxo para o algoritmo de treinamento SOM
 Fonte: Adaptada de Kaski (1997)

2.1.1.2 Processos de aprendizagem da rede

Um mapa auto-organizável é composto de três etapas no seu processo de formação: processo competitivo, processo colaborativo e processo adaptativo.

a) Aprendizagem sequencial

A forma de atualização dos pesos sinápticos é o que diferencia a aprendizagem sequencial da realizada em lote. O algoritmo SOM que usa o processo de aprendizagem sequencial considera a inserção do vetor de entrada de forma aleatória, com o intuito de garantir a uniformidade na apresentação de todo o espaço de entrada. A taxa de aprendizagem $\eta(n)$ assume um valor pré-fixado, tal que $0 < \alpha(t) < 1$, decaindo com o tempo t até um valor próximo de zero (CHEN, 2012). Da mesma forma, a função de vizinhança também assume um valor pré-fixado, adequando-se de maneira a maximizar a qualidade da formação do mapa SOM (KOHONEN, 2013). A dimensionalidade do SOM depende do problema tratado, onde a versão bidimensional $n \times m$ é a mais utilizada pela capacidade de projeção dos dados. Sua dimensionalidade varia conforme o volume de dados, com mapas grandes para grandes volumes de dados e mapas pequenos para grupos reduzidos de dados (CHEN, 2012).

b) Aprendizagem em lote (*batch*)

Nesta forma de atualização dos pesos da grade sináptica não existe a taxa de aprendizagem $\eta(n)$ e os vetores de entrada x_m não necessitam ser apresentados de forma aleatória. Dessa forma, os vetores dos pesos sinápticos da grade de neurônios são atualizados somente ao final de cada iteração e da apresentação de todos os vetores de entrada.

No presente estudo, primeiramente serão realizados experimentos com o processo de aprendizagem sequencial. Para efeitos de comparação, na sequência será utilizado o processo de aprendizagem em lote.

2.1.1.3 Avaliação da aprendizagem

Existem duas métricas de avaliação do processo de aprendizagem de uma rede SOM que são típicas: o Erro da Quantização Vetorial (E_{qv}) e o Erro Topológico (E_t) (KOHONEN; 2001).

O erro de quantização vetorial indica a média aritmética do erro entre o vetor do espaço de entrada (x_m) e o vetor do neurônio vencedor no processo competitivo do mapa auto-organizável (equação 12).

$$E_{qv} = \frac{\sum_{k=1}^m \|x_k - w_{bmu}\|}{m} \quad (12)$$

Onde

w_{bmu} : vetor do neurônio vencedor no processo competitivo da SOM

Por outro lado, o erro topológico (E_t) indica a proporção na qual a estrutura da grade sináptica aproxima os padrões próximos no espaço de entrada (equação 13). Para cada vetor de entrada x_k existe um BMU_1 que representa o primeiro neurônio na ordem de competição na grade e um BMU_2 que corresponderá ao segundo neurônio. Assim, o erro topológico corresponde ao percentual de padrões BMU_1 e BMU_2 que não são vizinhos na grade.

$$E_t = \frac{1}{m} \sum_{k=1}^m u(x_k) \quad (13)$$

Onde $u(x_k)$ corresponde ao valor 1 se os vetores BMU_1 e BMU_2 não são vizinhos e ao valor 0 no caso contrário.

2.1.1.4 Planos componentes e matriz de distância unificada

A visualização do processo de aprendizagem do mapa auto-organizável é necessária para que seja possível verificar o resultado de sua ordenação topológica. A forma mais usual é a representação dos vetores de pesos sinápticos no espaço n -

dimensional. No entanto, tal processo somente pode ser utilizado quando a dimensionalidade dos dados é inferior a 3. Porém, existem duas outras alternativas: a matriz de distâncias unificada (U-Matrix) e os planos componentes (ULTSCH, 1993).

A U-Matrix facilita o processo de visualização, pois pode ser representada através de uma imagem. A matriz de distâncias unificada permite a visualização das relações topológicas entre os neurônios da grade sináptica (ULTSCH; 1993). Calcula-se a distância euclidiana entre os vetores dos pesos sinápticos dos neurônios adjacentes e os resultados gerados a partir da U-Matrix sobre o mapa autoorganizável será uma imagem $f(x,y)$ cujo nível de intensidade e cada pixel representa uma distância calculada. No caso de um mapa bidimensional hexagonal, obtém-se a U-Matrix através do cálculo das distâncias dx , dy e dz para cada neurônio. A U-Matrix pode ser entendida como uma imagem na qual a coloração dos pixels ocorre de acordo com a intensidade de cada componente da matriz. Dessa forma, valores mais altos correspondem a neurônios vizinhos dissimilares e valores mais baixos correspondem a neurônios vizinhos similares.

Apesar da U-Matrix gerar uma imagem complexa, ela permite visualizar a separação topológica dos agrupamentos. Tal representação é de extrema utilidade quando a dimensionalidade é superior a 3.

Já os planos componentes utilizam os valores de cada componente dos vetores de pesos sinápticos através de uma representação por cores, permitindo a avaliação da distribuição do componente no mapa após o processo de aprendizagem. Considerando um mapa bidimensional $M \times N$, para um determinado componente j gera-se uma imagem $f(x,y)$ com as mesmas dimensões do mapa bidimensional, na qual cada pixel corresponderá ao valor do componente j na posição (x,y) .

2.1.2 Perceptron multicamadas

As redes neurais artificiais são modelos que fazem uso do paradigma conexionista, que procura entender e emular as propriedades decorrentes do alto grau de paralelismo e conectividade, na solução de determinada espécie de problemas. Dessa forma, uma RNA é constituída por um alto número de elementos processadores (soma), amplamente conectados entre si. Cada conexão interliga dois elementos

processadores, num único sentido, através de um valor (peso) que determina o grau de conectividade entre os elementos, determinando o peso da conexão em questão.

O processamento é distribuído entre todos os elementos processadores da rede, onde cada um o realiza de forma isolada e paralela, enviando seu resultado para as outras unidades da camada seguinte através de suas conexões. Por este motivo, as redes neurais artificiais são denominadas sistemas de processamento paralelo e distribuído.

$$Saída = \varphi\left(\sum x_i \cdot p_i\right) \quad (14)$$

Onde:

x_i : é o sinal de entrada do neurônio i da camada de entrada

p_i : é o peso associado ao sinal de entrada x_i

φ : é a função de transferência aplicada ao sinal de saída do neurônio

Cada elemento processador normalmente possui diversas entradas e apenas uma saída, como pode ser visto na Figura 6. O seu processamento consiste em transferir para a saída um valor calculado a partir dos valores apresentados nas entradas, por meio de uma função de transferência.

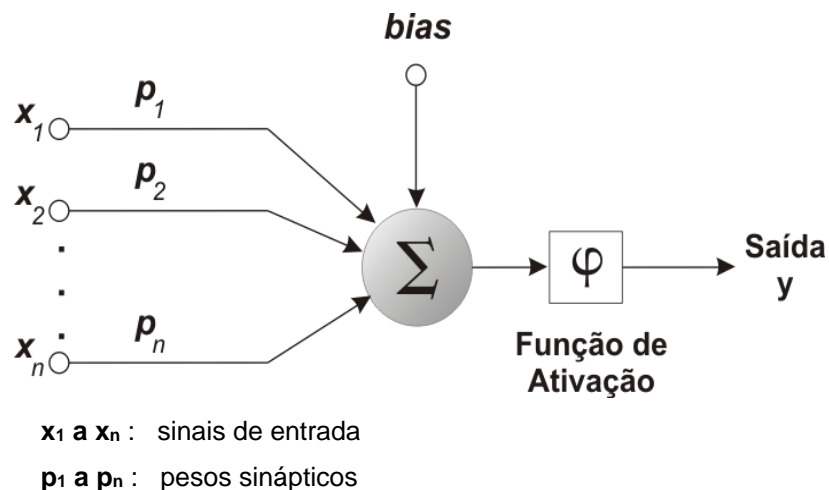
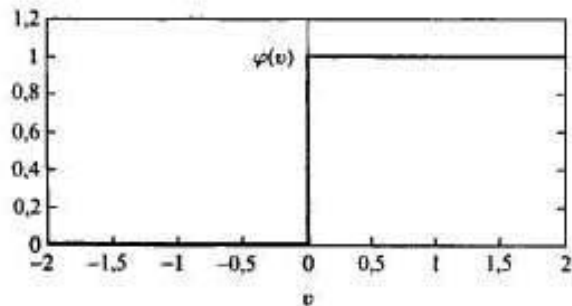


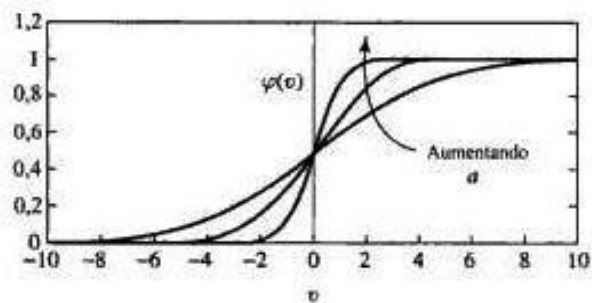
Figura 6 – Modelo de neurônio artificial
 Fonte: Elaborada pelo autor.

Normalmente, as entradas são combinadas através de uma soma ponderada, sendo transferidas para a saída por uma função de limiar (RUMELHART, 1986). Além

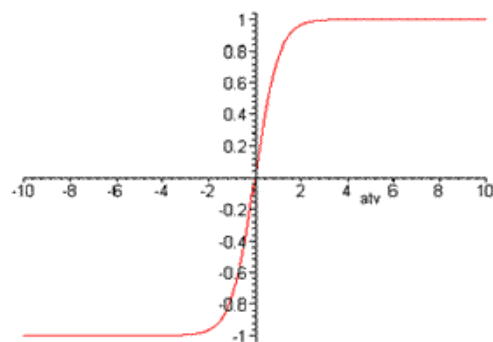
da função de limiar, é possível fazer uso da função sigmóide ou da função tangente hiperbólica (Figura 7).



(a)



(b)



(c)

Figura 7 – Funções de transferência: (a) Função de limiar; (b) Função sigmóide; (c) Função tangente hiperbólica.

Fonte: Elaborada pelo autor.

O padrão de interconexão da rede, também conhecido como topologia da rede, é a forma como os elementos processadores são interligados. A topologia pode ser mantida rígida ou flexível, conforme a pré-disposição. As topologias rígidas são mais simples, geralmente utilizando uma metodologia de aprendizagem de tentativa e erro. Por outro lado, as topologias flexíveis são úteis quando as variáveis do problema não são totalmente conhecidas ou quando ocorrem variações demasiadas conforme a

passagem do tempo. A capacidade de uma rede neural resolver um problema determinado é diretamente proporcional ao seu padrão de interconexão dos neurônios.

O potencial de ativação do elemento processador é o valor da saída em um determinado instante de tempo. O conjunto dos estados de ativação de cada um dos elementos processadores é definido como a função de ativação da rede neural ou estado de ativação da rede neural.

A estrutura da rede neural artificial é determinada pela forma de agrupamento dos elementos processadores em camadas. Em geral, a estrutura é composta de uma camada de entrada, onde ocorre a simples apresentação dos dados à rede, uma (Figura 8) ou mais camadas intermediárias (Figura 9), amplamente conectadas à camada de saída, responsável pela obtenção dos resultados. As camadas são processadas no sentido da camada de entrada para a camada de saída, não existindo conexões entre elementos de uma mesma camada. Usualmente, as conexões sinápticas ocorrem com os elementos da camada anterior e da camada posterior, formando as redes *feedforward*.

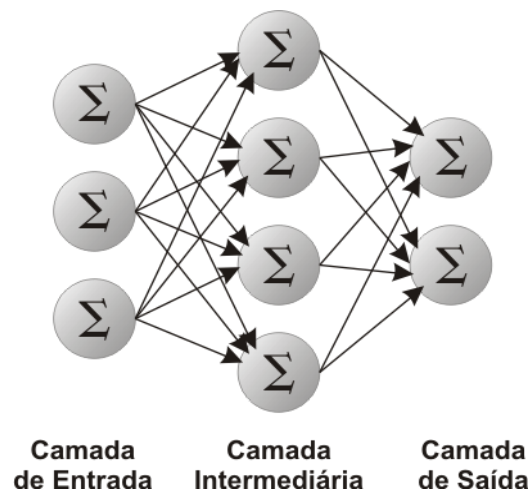


Figura 8 – Estrutura da RNA com uma camada intermediária
Fonte: Elaborada pelo autor

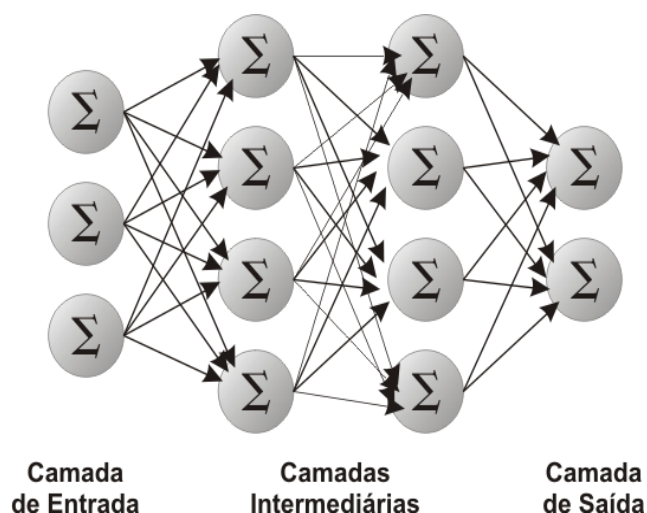


Figura 9 – Estrutura da RNA com duas camadas intermediárias
Fonte: Elaborada pelo autor

O processo de síntese de uma RNA consiste na correta inicialização, tornando-a capaz de solucionar o problema a que se destina, gerando um padrão de interconexão suficiente para modelar a resolução. Do contrário, a rede deve passar por um processo de treinamento que modifique gradativamente seu padrão de interconexão inicial, adaptando-se à resolução do problema. A inicialização dos pesos das conexões (síntese) é um fator determinante do sucesso ou do fracasso do treinamento dos neurônios para a solução dos problemas.

As metodologias de aprendizado permitem a modificação do padrão de interconexão de uma rede neural, habilitando-a a resolver determinado problema. Três mecanismos geralmente são usados no aprendizado: supervisionado, quando são fornecidos os resultados desejados; por reforço, quando um parâmetro externo de comparação é apresentado; e não supervisionado, quando a própria rede é hábil em ajustar o seu funcionamento.

De uma forma geral, são utilizadas três fases para a modelagem de uma rede neural a um problema. Primeiramente, uma fase de treinamento, onde se ensina a rede a modelar um conjunto de padrões de saída associados a padrões de entrada. Numa segunda fase, são apresentados padrões de entrada à rede e as saídas são comparadas às saídas desejadas. Numa última fase, a rede é utilizada para implementar a solução do problema.

As redes neurais artificiais são sistemas computacionais com o propósito de representar circunstâncias do meio ambiente com base em conceitos ou inspiração

neurobiológica. Analogamente à neurobiologia, existem diversas concepções computacionais com base em tais conceitos, conforme a aplicação e sua usabilidade. Entre os sistemas de melhor performance, está o *perceptron* multicamadas com retropropagação.

Uma rede neural do tipo *perceptron* multicamadas é composta por um conjunto de unidades sensoriais (neurônios artificiais) que compõem uma camada de entrada, uma ou mais camadas intermediárias ou ocultas e uma camada de saída de nós computacionais (HAYKIN, 2001). Os sinais de entrada são propagados através das camadas de forma distribuída e paralela, camada após camada. O *perceptron* multicamadas é uma rede neural artificial do tipo totalmente conectada. A Figura 10 ilustra uma rede *perceptron* multicamadas com duas camadas intermediárias.

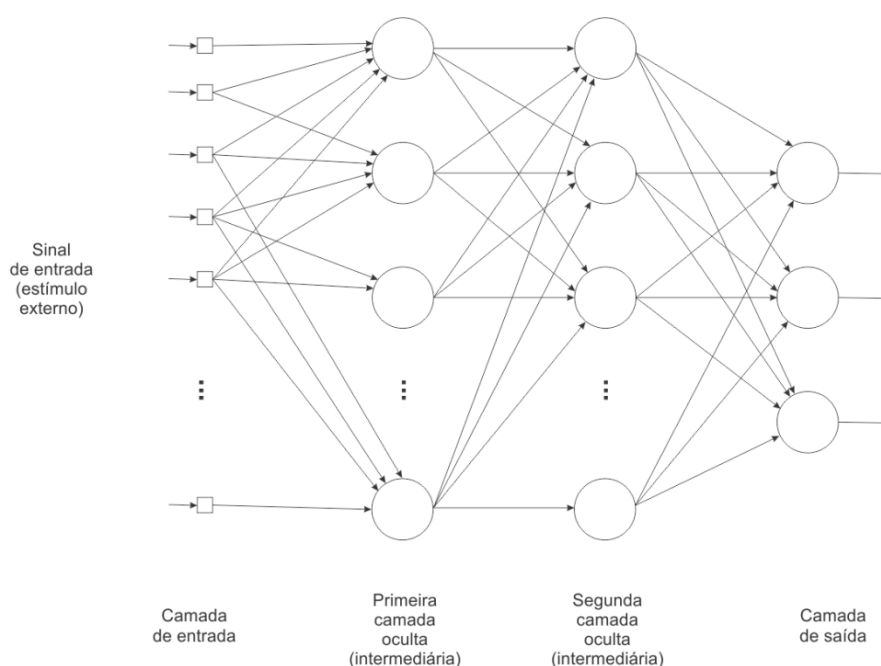


Figura 10 - Perceptron de múltiplas camadas (com duas camadas intermediárias)
Fonte: HAYKIN (2001, p. 186)

Um conceito de aprendizagem muito utilizado pelo *perceptron* multicamadas é o de retropropagação de erro. Na etapa de propagação, os pesos sinápticos que assinalam a importância do neurônio no resultado final são fixos e determinados estocasticamente. Na etapa de retropropagação, ocorre a correção dos pesos sinápticos com base na correção dos erros nos resultados. A aprendizagem por retropropagação tem se tornado o algoritmo padrão para utilização em redes neurais

do tipo *perceptron* multicamadas pois soluciona de forma concisa o problema de atribuição de crédito através da correção por retropropagação dos erros nos resultados.

A topologia do *perceptron* multicamadas possui características extremamente plásticas para o desenho de sua estrutura no que tange ao número de neurônios artificiais a serem utilizados em cada camada ou mesmo ao número de camadas intermediárias a serem consideradas. Na verdade, não existem estudos que especifiquem o número de neurônios por camada ou o número de camadas a serem utilizadas. A flexibilidade biológica adotada pelo cérebro humano para a solução de problemas, neste aspecto, é de grande similaridade.

2.1.2.1 Algoritmo de treinamento

Seja X_p um vetor de entrada que representa o padrão p ; W_{ji} o peso de conexão entre o neurônio j da camada escondida e o neurônio i da camada de entrada; Y_{pk} a saída desejada e O_{pk} a saída obtida pela rede para o padrão p do neurônio k de saída; é possível descrever o algoritmo de *backpropagation* através do seguinte processo (HAYKIN, 2001; KAASTRA; BOYD, 1996):

- a) Inicialização dos pesos e os parâmetros da rede (coeficiente de aprendizado e parâmetro do momento);
- b) Cálculo dos potenciais de ativação dos neurônios j da camada oculta h :

$$net_{pj}^h = \sum_{i=1}^N W_{ji}^h \cdot X_{pi} + \theta_j^h \quad (15)$$

Onde:

N : número de neurônios da camada de entrada

W_{ji}^h : peso do sinal de entrada i que se conecta ao neurônio j da camada escondida

X_{pi} : sinal de entrada do neurônio i da camada de entrada

θ_j : valor de limiar do neurônio j

O termo θ_j , conhecido por *bias*, tem a função de contribuir na convergência da rede. Na verdade, ele funciona como um peso a mais a ser adicionado ao somatório. Cada neurônio da camada intermediária e da camada de saída possui o seu próprio *bias*.

c) Cálculo das saídas dos neurônios da camada intermediária:

$$i_{pj} = f(\text{net}_{pj}^h) \quad (16)$$

Onde:

f : função de ativação

Em geral, utiliza-se como função de ativação a função sigmóide ou a função tangente hiperbólica. Para a função tangente hiperbólica, os neurônios geram valores de saída dentro do intervalo $[-1;+1]$. Dessa forma:

$$i_{pj} = \tanh(\text{net}_{pj}^h) \quad (17)$$

d) Cálculo dos potenciais de ativação para os neurônios k da camada de saída o :

$$\text{net}_{pk}^o = \sum_{j=1}^L W_{kj}^o \cdot i_{pj} + \theta_k^o \quad (18)$$

Onde:

L : número de neurônios da camada intermediária

W_{kj}^o : peso do sinal do neurônio j da camada intermediária que se conecta ao neurônio k da camada de saída

i_{pj} : sinal de entrada do neurônio j da camada intermediária

θ_j : valor de limiar do neurônio k

e) Cálculo dos valores de ativação dos neurônios de saída de forma semelhante à camada intermediária:

$$o_{pk} = \tanh(\text{net}_{pk}^o) \quad (19)$$

f) Cálculo dos termos de erro (gradiente local do erro) para as unidades de saída:

$$\varphi_{pk}^o = (Y_{pk} - O_{pk}) \cdot f' \quad (20)$$

onde f' é a derivada da função de ativação escolhida. No caso da função tangente hiperbólica, tem-se:

$$f' = \tanh'(x) = 1 - (\tanh(x))^2 \quad (21)$$

E, como para a camada de saída, $\tanh(x) = O_{pk}$, obtém-se

$$f' = \tanh'(x) = 1 - (O_{pk})^2 \quad (22)$$

g) Cálculo dos termos de erro para as unidades intermediárias:

$$\varphi_{pj}^h = [1 - (i_{pj})^2] \cdot \sum_k \varphi_{pk}^o \cdot W_{kj}^\theta \quad (23)$$

h) Atualização dos pesos na camada de saída:

$$W_{kj}^o(t+1) = W_{kj}^o(t) + \eta \cdot \varphi_{pk}^o \cdot i_{pj} \quad (24)$$

Onde η é o parâmetro da taxa de aprendizado, que determina a velocidade do aprendizado.

i) Atualização dos pesos na camada intermediária:

$$W_{ji}^h(t+1) = W_{ji}^h(t) + \eta \cdot \varphi_{pj}^h \cdot X_{pi} \quad (25)$$

j) Repetição dos passos b a i para todos os P padrões.

O processo se encerra quando o erro médio quadrado (EMQ), dos P padrões, for menor ou igual a um erro mínimo previamente estipulado, ou ao atingir-se um número determinado de iterações.

$$EMQ = \frac{1}{P} \cdot \sum_{p=1}^P E_p \quad (26)$$

Onde E_p é o erro instantâneo obtido ao apresentar o padrão X_p na entrada da rede. Este erro é dado por

$$EMQ = \frac{1}{n} \cdot \sum_{k=1}^M (Y_{pk} - O_{pk})^2 \quad (27)$$

2.1.2.2 Avaliação do desempenho de modelos neurais

Existem diversas possibilidades para a análise da performance de um modelo neural no processo de modelagem de uma série temporal. A raiz do erro médio quadrado (*Root Mean Square Error*) é frequentemente utilizada para avaliar o desempenho obtido por um modelo neural (Equação 28).

$$RMSE = \sqrt{EMQ} = \sqrt{\frac{1}{n} \cdot \sum_{k=1}^M (Y_{pk} - O_{pk})^2} \quad (28)$$

O MAE (*Mean Absolute Error*) é outra possibilidade para a avaliação de desempenho de modelos neurais (Equação 29). O MAE ou o Erro Médio Absoluto é uma medida relativamente comum do erro de previsão para a análise de séries temporais.

$$MAE = \frac{1}{n} \cdot \sum_{k=1}^n |Y_{pk} - O_{pk}| \quad (29)$$

O Erro Médio Absoluto é considerado uma alternativa interessante ao RMSE e ao EMQ, visto que seu cálculo é simplificado quando em comparação.

Aliado ao MAE, tem-se o Erro Médio Absoluto Percentual, ou MAPE (*Mean Absolute Percentage Error*), que também é conhecido como o desvio médio percentual absoluto. O MAPE é uma medida relativa referente ao erro entre o valor previsto e o valor real (Equação 30).

$$MAPE = \frac{1}{n} \cdot \sum_{k=1}^n \left| \frac{Y_{pk} - O_{pk}}{Y_{pk}} \right| \quad (30)$$

Neste estudo, com o objetivo de compor uma análise mais completa, serão usados os índices RMSE, MAE e MAPE na análise dos modelos neurais.

2.2. MODELOS DE PREVISÃO DE ÍNDICES FINANCEIROS

As redes neurais artificiais são funções de ampla aplicação e de alta flexibilidade para a modelagem e aproximação de funções em áreas cognitivas ou de engenharias. Tais características têm aproximado sua aplicação a áreas financeiras que dispõem de grandes bases de dados na forma de séries temporais, como classificação de risco de investimentos, aplicações em renda fixa e variável, simulação do comportamento de mercados, seleção e diversificação de portfólio e previsões econômicas. Desde os anos 1990, as redes neurais têm sido utilizadas para solucionar problemas em diversas áreas, como a empresarial, a comercial e a industrial. Na área de marketing, por exemplo, as redes neurais vêm sendo usadas em aplicações para processar a análise de mercados, através da segmentação dos clientes de acordo com características que incluem dados demográficos, perfil socioeconômico, localização geográfica, padrões de aquisição e comportamento referente a um determinado produto (ARMANO et al., 2005; BILDIRICI et al., 2009; GURESEN et al., 2011; RAMEZANIAN et al., 2011).

A Tabela 1 demonstra algumas características e resultados de pesquisas desenvolvidas na área de RNA aplicadas ao setor de finanças nos últimos seis anos.

Tabela 1 - Pesquisas na área de aplicações financeiras de modelos de RNA

Ano	Autores	Método	Período dos dados	Tipo dos dados	Objetivo do estudo	Resultado
2007	Preminger & Franck	RAR (<i>Robust linear Autoregressive</i>) RNN (<i>Robust Neural Network</i>)	1971-2004	JPY / GBP / USD	Obtenção de melhores previsões	Modelos robustos obtêm melhores resultados do que modelos clássicos, mas inferiores a modelos do tipo RW (<i>RandomWalk</i>)
2007	Celik & Karatepe	RNA (Rede Neural Artificial)	1989-2004	Séries temporais de dados bancários	Previsão de crises bancárias através de índices financeiros	Previsão de índices financeiros de até 4 meses
2007	Hassan, Nath & Kirley	HMM (<i>Hidden Markov Model</i>) RNA e Algoritmos Genéticos	2003-2004	Ações da bolsa: IBM, Apple e Dell	Previsão do valor das ações	Modelos híbridos são superiores a modelos clássicos
2008	Zhu, Wang, Xu & Li	RNA básica RNA ampliada	1989-2005	Índices Nasdaq, Dow Jones e Bolsa de Cingapura	Investigar o efeito do volume de negociações na previsão de índices com RNAs	É possível aprimorar a performance da RNA de forma significativa e modesta pela adição de informações a respeito do volume de negociações
2009	Egrioglu, Aladag, Youlcu, Uslu & Basaran	RNA com retropropagação	1974-2004	Acidentes por ano em rodovias na Bélgica	Novo método que dispense o uso de lógica fuzzy para analisar séries temporais fuzzy de altas ordens	O modelo forneceu previsões mais acuradas que métodos tradicionais
2009	Leu, Lee & Jou	RBFNN (<i>Radial Basis-function neural network</i>) Modelo <i>Random Walk</i> RNA básica	2006-2007	TAIEX, NTD/USD, KRW/USD, CNY/USD, JPY/USD	Previsão de índices e comparação de métodos	O modelo RBFNN superou o desempenho dos modelos <i>RandomWalk</i> e RNA com base no erro médio quadrado
2009	Majhi, Panda & Sahoo	FLANN (<i>Functional Link ANN</i>) CFLANN (<i>Cascaded Functional Link ANN</i>) Modelo LMS (-	Dólar americano, Libras esterlinas, Rúpias indianas, Ienes	Comparativo sobre a performance dos métodos	O modelo CFLANN obteve um melhor desempenho que os modelos FLANN e LMS
2009	Bildirici & Ersin	ARCH, GARCH, EGARCH, TGARCH, PGARCH, APGARCH RNA	1987-2008 Dados diários	Índice da bolsa de Valores de Istambul	Aprimorar previsões através de RNAs	Modelos baseados em RNAs fornecem aprimoramentos significativos em previsões de índices de mercado
2011	Ramezani	RNA do tipo <i>Perceptron</i> <i>Multicamadas</i>	2000-2009	Bolsa de valores de Teerã (Irã)	Desenvolver um modelo de previsão baseado em RNA	Resultados significativos quanto ao uso de variáveis baseadas em indicadores financeiros

Fonte: Elaborada pelo autor

Tabela 1 (Continuação) - Pesquisas na área de aplicação de modelos de RNA a séries temporais financeiras

2012	Höglund	SOM (<i>Self-Organizing Map</i>) MLP (<i>Multilayer Perceptron</i>) GRNN (<i>General Regression Neural Network</i>)	2006-2007	Dados financeiros de indústrias norte-americanas	Avaliar se os modelos baseados em RNA são superiores aos modelos lineares na detecção de gerenciamento de resultados	Os melhores resultados foram obtidos pela GRNN.
2012	Tsai, Lu & Yen	RNA do tipo <i>Perceptron</i> Multicamadas	1996-2007	Dados financeiros de 1380 empresas de Taiwan	Identificar fatores de importância que afetam os ativos intangíveis	Resultados mais significativos foram obtidos com a extração de variáveis através de um modelo híbrido GA – <i>Stepwise</i>

Fonte: Elaborada pelo autor

Recentemente, frente aos resultados obtidos a partir de modelos neurais e estocásticos para a previsão de séries temporais, pesquisadores (CHEN, 2012; DAI et al., 2012; HSU, 2011; ESPÍNOLA et al., 2011; BILDIRICI, 2009) vêm se dedicando à composição de modelos híbridos com diferentes naturezas, como SOM (*Self-Organizing Maps*) e algoritmos genéticos, SOM e GRNN (*General Regression Neural Network*), MLP (*Multilayer Perceptron*) e modelos GARCH (*Generalized Autoregressive Conditional Heteroskedastic*). Com essa premissa, optou-se neste estudo pela concepção de um modelo de natureza híbrida conjugando duas formas de redes neurais artificiais, a SOM (Mapas Autoorganizáveis) e a BPN (Rede Perceptron Multicamadas com Retropropagação).

3. MÉTODOS E PROCEDIMENTOS

Os métodos utilizados no presente estudo estão baseados em técnicas de pesquisa operacional (PO) e em técnicas de implementação de modelos de predição através de redes neurais artificiais.

As pesquisas baseadas em técnicas de PO devem se balizar em cinco etapas principais, segundo Arenales et al.(2007), as quais são:

- a) Definição do problema
- b) Construção do modelo
- c) Solução do modelo
- d) Validação do modelo
- e) Análise dos resultados

Os autores também afirmam que tais etapas não são rígidas, podendo ser reorganizadas conforme o problema a ser analisado.

As redes neurais artificiais possuem comportamento de caixa-preta, visto que não se possui acesso a detalhamento específico a respeito do seu funcionamento. Dessa forma, a compreensão sobre o processo de utilização ocorre através de estímulos de entrada e análise das saídas. Hsu (2011) e Mostafa (2010) elencam alguns dos aspectos essenciais na modelagem de um sistema baseado em redes neurais artificiais:

- a) Seleção das variáveis de entrada
- b) Determinação da quantidade de variáveis de entrada
- c) Definição da topologia das redes
- d) Especificação do algoritmo de treinamento
- e) Previsão da saída da rede

No entanto, Setyawati et al. (2003) atentam para a inexistência de regras para a definição da topologia das redes neurais. Não existem evidências que comprovem a prevalência da arquitetura hexagonal sobre a retangular nas redes SOM, ou mesmo a relação inversa. Da mesma forma, ainda prevalece o questionamento a respeito do número de neurônios na camada oculta de uma rede *backpropagation*, ou mesmo do número de camadas ocultas.

3.1 BASE DE DADOS

Os dados desta pesquisa são constituídos pelas cotações diárias do índice NYSE Composite U.S. 100 no período de 5 de abril de 2004 até 8 de novembro de 2012. Os dados foram obtidos através do software Economatica® e de cotações disponíveis pela New York Stock Exchange (NYSE Composite, 2012).

A Bolsa de Nova York estabeleceu a NYSE Composite Index (New York Stock Exchange) em 1966 para fornecer uma medida abrangente do desempenho de todas as ações ordinárias listadas. O NYSE Composite representa 77% da capitalização total de mercado de todas as empresas de capital aberto nos Estados Unidos e 66% da capitalização total de mercado de todas as empresas de capital aberto no mundo. O NYSE Composite U.S. 100 retrata a capitalização das 100 empresas com maior participação no índice. A lista com as 100 empresas com maior participação no índice pode ser visualizada no Anexo A.

Um dos objetivos da NYSE para o desenvolvimento de um índice é o de mostrar a força das empresas listadas e negociadas na bolsa de valores, além de prover aos investidores uma maneira de acompanhar as tendências de mercado de forma mais líquida. Por esses motivos, a NYSE Composite lançou em junho de 2002 quatro novos índices para auxiliar os investidores: NYSE Composite U.S. 100, NYSE International100, NYSE World Leaders e NYSE TMT. Em 2004 foi criado o índice NYSE Financial, Energy and Healthcare. Com ampla participação das maiores empresas do mundo, a NYSE é referência de mercado para o posicionamento de investidores globais.

O NYSE Composite Index foi desenvolvido para medir o desempenho de todas as ações ordinárias listadas na NYSE, consistindo em mais de 2.000 ações norte-americanas e estrangeiras. É uma medida das alterações no valor total de mercado de todas as ações listadas na NYSE, ajustadas para eliminar os efeitos das mudanças de capitalização. A opção pela utilização do índice NYSE para a concepção do modelo híbrido SOM-ANN/BP deve-se à fidedignidade da fonte dos dados. Historicamente, modelos de previsão de dados baseados em informações oriundas de bolsas de valores necessitam de dados confiáveis a fim de que a sua eficiência seja maximizada.

O uso do índice específico NYSE Composite U.S. 100 é justificado pelo número reduzido de dados em relação a outros índices voltados a bolsas. Para um

modelo baseado em algoritmos de redes neurais artificiais, um conjunto excessivo de informações pode prejudicar o treinamento, levando à não convergência dos resultados.

As variáveis utilizadas para o banco de dados envolvem os valores diários de negociação, considerando o intervalo de oscilação diária. Dessa forma, as variáveis que compõem a base de dados são:

- Valor de abertura e de fechamento
- Valor mínimo e máximo diários
- Volume de negociações

Tais variáveis foram selecionadas por demonstrarem de forma abrangente as variações cotidianas do índice NYSE Composite U.S. 100, sendo usadas de forma ampla no cálculo dos índices financeiros técnicos que serão inseridos na camada de entrada do modelo híbrido SOM-ANN/BP.

3.2 ESQUEMA GERAL DO MODELO

Conforme as premissas apresentadas a respeito das metodologias utilizadas nas técnicas de pesquisa operacional e de redes neurais artificiais, o esquema proposto no presente estudo para o modelo híbrido de previsão de índices financeiros SOM-ANN/BP é composto por quatro fases ilustradas na Figura 11. Cada fase do esquema geral do modelo híbrido SOM-ANN/BP possui diversas sub-etapas de execução.

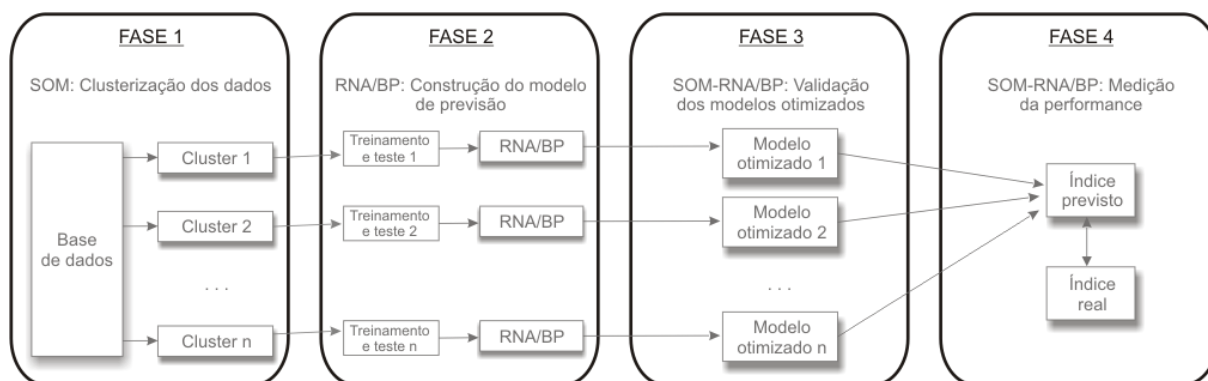


Figura 11 - Esquema geral do modelo híbrido SOM-ANN/BP

A fase 1 do modelo híbrido SOM-ANN/BP é caracterizada pela coleta dos dados e seu pré-processamento. Primeiramente é realizada a coleta da base de dados por meio das séries históricas diárias dos valores mínimo e máximo, de abertura e de fechamento, bem como de volume de negócios no período de 5 de abril de 2004 até 8 de novembro de 2012. Numa segunda etapa, ocorre o cálculo dos 10 índices financeiros técnicos com base nos dados coletados e, numa terceira etapa, os índices são normalizados para o intervalo $[-1;1]$. O processo de cálculo dos índices e subsequente normalização são processados através do software Excel[®] devido à dimensionalidade da amostra.

Após o pré-processamento dos dados, a segunda fase caracteriza-se pela clusterização por meio dos mapas autoorganizáveis (SOM). Inicialmente, é efetuada a segmentação da base de dados de forma aleatória em grupo de treinamento (80% dos dados), grupo de teste (10% dos dados) e grupo de validação (10% dos dados). A segmentação aleatória para o treinamento das redes é baseada em trabalhos publicados por diversos pesquisadores (HSU; 2011; WANG et al.; 2011; LU et al.; 2011; TSAI et al.; 2010). Após a clusterização utiliza-se a variável normalizada do valor de fechamento para o treinamento da rede neural artificial *perceptron* multicamadas com *backpropagation* em cada cluster definido pelo mapa autoorganizável. Finalizando a segunda etapa temos a inserção do grupo de testes para uma análise prévia dos resultados obtidos pelos modelos neurais. O processamento da rede SOM foi realizado pelo software Matlab[®], a partir da toolbox SOM, desenvolvida pelo *Laboratory of Computer and Information Science* (CIS). Da mesma forma, todas as etapas referentes ao desenvolvimento da rede *perceptron* multicamadas foi desenvolvida através da toolbox de redes neurais, nativa do software Matlab[®].

Na terceira fase processa-se a validação dos modelos otimizados na fase dois. São selecionados os modelos otimizados em cada cluster para a construção do modelo híbrido SOM-ANN/BP, onde serão inseridos os grupos de validação. O processo de seleção é baseado nos resultados obtidos para a Raiz do Erro Quadrado Médio (RMSE).

A quarta e última etapa é destinada para a avaliação do desempenho do modelo construído. Efetua-se o processo inverso da normalização dos resultados para obter o índice previsto com o objetivo de avaliar o desempenho do modelo com base na Raiz do Erro Quadrado Médio (RMSE), no Erro Médio Absoluto (MAE) e no Erro

Médio Percentual (MAPE). Tais medidas de avaliação são típicas na análise de eficiência de redes neurais (HAYKIN, 2001).

A efetividade do modelo híbrido de previsão SOM-ANN/BP é avaliada através dos modelos de regressão múltipla com metodologia *Enter*, *Stepwise*, *Forward* e *Backward*, desenvolvidos através do software SPSS®. Um comparativo é estabelecido para aprimorar a análise dos resultados obtidos através do modelo baseado em redes neurais e dos modelos de regressão.

4. MODELO HÍBRIDO SOM-ANN/BP

As análises processadas neste capítulo visam investigar os índices que influenciam no comportamento das séries diárias do índice NYSE Composite U.S. 100, com o intuito de desenvolver um modelo eficiente de projeção do índice a curto e a médio prazo. Para tanto, é apresentada a fase inicial de pré-processamento para a construção do modelo híbrido SOM-ANN/BP, que se compõe da coleta das variáveis referente ao índice no período elencado, o cálculo dos índices financeiros baseados nas variáveis coletadas e a normalização dos índices calculados. Na sequência, a fase 2 é abordada através da clusterização da base de dados normalizada, sua segmentação em grupos de treinamento, teste e validação, bem como o treinamento das redes *backpropagation* em cada cluster formado. Em seguida, serão realizadas as fases de validação das redes e avaliação do modelo híbrido. Não obstante, serão analisadas variações topológicas nas redes SOM e BP, além de supressões de índices financeiros técnicos. Finalmente, é realizada a análise de regressão múltipla e a comparação com o desempenho do modelo SOM-ANN/BP.

4.1 PRÉ-PROCESSAMENTO DOS DADOS

Os dados coletados são compostos de 2128 observações diárias no período entre 5 de abril de 2004 até 8 de novembro de 2012 das variáveis: valor de abertura e de fechamento do índice NYSE Composite U.S. 100, além do valor mínimo e máximo e do volume de negócios do mesmo índice. Um resumo estatístico descritivo pode ser visualizado na Tabela 2.

A base de dados coletada está organizada na forma matricial de ordem 2128 x 5, referente às cinco variáveis das 2128 observações diárias (BELUCO, 2013a).

Tabela 2 - Resumo estatístico descritivo das variáveis componentes da base de dados

	Volume de negociação	Fechamento	Abertura	Mínimo	Máximo
Média	781629,89	69,09	69,10	68,72	69,42
Desvio padrão	1521566,23	10,34	10,36	10,39	10,32
Erro padrão	32984,13	0,22	0,22	0,23	0,22
Mediana	351327,26	69,25	69,34	69,02	69,57
Moda	62550,00	56,75	62,35	60,55	63,43
Variância	2315163798516	106,94	107,39	107,91	106,42
Curtose	71,86	-0,21	-0,21	-0,16	-0,26
Range	22327249,00	53,42	54,20	53,23	53,52
Valor mínimo	5765,00	38,18	37,93	37,74	38,78
Valor máximo	22333014,00	91,60	92,13	90,97	92,30

Fonte: Elaborada pelo autor no software SPSS®

4.1.1 Índices financeiros técnicos

As variáveis de entrada selecionadas para o modelo híbrido de previsão de índices financeiros derivam de índices técnicos utilizados por diversos autores (LIU; WANG, 2012; HSU, 2011; MOSTAFA, 2010; LI; KUO, 2008) para avaliar o desempenho de bolsas de valores. Tais índices são calculados com base no valor diário de abertura e fechamento do NYSE Composite U.S. 100, no valor máximo e no valor mínimo atingido pelo índice diariamente e no volume de negociações diárias realizadas (BELUCO, 2013b).

As variáveis selecionadas para a concepção do modelo híbrido SOM-ANN/BP são oriundas de conjuntos de variáveis usadas na construção de modelos por pesquisadores (HSU, 2011; GHIASSI et al, 2005; HULME; XU, 2001; KIM; HAN, 2000; LIAO; WANG, 2010; SMITH; GUPTA, 2000). O número de índices técnicos utilizados neste estudo é baseado no sucesso obtido nos modelos adotados ou propostos pelos pesquisadores supracitados. Dessa forma, os dez índices financeiros técnicos abordados como camada de entrada para o modelo SOM-ANN/BP (MA_10: 10-day moving average, BIAS_20: 20-day bias, WMS%R_9: 9-day William overbought/overday, K_9: 9-day stochastic indicator K, D_9: : 9-day stochastic indicator D, MTM_10: 10-day momentum, ROC_10: 10-day rate of change, CCI_24: 24-day commodity channel index, AR_26: 26-day buying/selling momentum indicator

e BR_26: 26-day buying/selling willingness indicator) foram selecionados a partir dos resultados que se mostraram significativos nos estudos citados.

Para o cálculo dos índices, são utilizadas as seguintes notações:

i índice do dia i

HP_i Valor máximo do índice NYSE 100 Composite no dia i

LP_i Valor mínimo do índice NYSE 100 Composite no dia i

OP_i Valor de abertura do índice NYSE 100 Composite no dia i

CP_i Valor de fechamento do índice NYSE 100 Composite no dia i

TV_i Volume de negociações do índice NYSE 100 Composite no dia i

a) MA_10

Índice que analisa a média móvel do valor de fechamento do índice NYSE Composite U.S. 100 no período de 10 dias anteriores. É calculado através da equação:

$$MA_{10}_i = \frac{\sum_{t=i-9}^i CP_t}{10} \quad (31)$$

b) BIAS_20

Índice que calcula o desvio entre o valor de fechamento e a média móvel do valor de fechamento dos últimos 20 dias. O BIAS_20 pode ser calculado através da equação:

$$BIAS_{20}_i = \frac{CP_i - MA_{20}_i}{MA_{20}_i} \quad (32)$$

c) WMS%R_9

O índice de sobrecompra e sobrevenida de Williams é um indicador de momento baseado na relação entre a diferença do valor máximo e do valor de fechamento e a diferença entre os valores máximo e mínimo no período dos últimos 9 dias, sendo calculado através da equação:

$$WMS\%R_{9i} = \frac{HP_{9i} - CP_i}{HP_{9i} - LP_{9i}} \quad (33)$$

Na qual HP_{9i} representa o valor máximo no período dos últimos 9 dias e LP_{9i} representa o valor mínimo no período dos últimos 9 dias.

d) K_9

O índice estocástico K referente aos últimos 9 dias é definido por:

$$K_{9i} = \frac{2}{3} \times K_{9i-1} + \frac{1}{3} \times \frac{CP_i - LP_{9i}}{HP_{9i} - LP_{9i}} \times 100 \quad (34)$$

Onde HP_{9i} representa o valor máximo no período dos últimos 9 dias e LP_{9i} representa o valor mínimo no período dos últimos 9 dias.

e) D_9

O índice estocástico D referente aos últimos 9 dias pode ser calculado através da equação:

$$D_{9i} = \frac{2}{3} \times D_{9i-1} + \frac{1}{3} \times K_{9i} \quad (35)$$

Onde K_{9i} representa o índice estocástico K no período dos últimos 9 dias.

f) MTM_{10}

Índice de momento que mede as alterações no valor de fechamento no período dos últimos 10 dias. É calculado por:

$$MTM_{10i} = CP_i - CP_{i-10} \quad (36)$$

g) ROC_10

O Índice ROC_10 representa uma taxa de variação que mede as alterações percentuais entre o valor de fechamento atual e o valor de fechamento de 10 dias atrás, sendo calculado da forma:

$$ROC_{10}_i = \frac{CP_i - CP_{i-10}}{CP_{i-10}} \times 100 \quad (37)$$

h) CCI_24

O índice conhecido como Commodity Channel é utilizado para identificar ciclos no valor de fechamento das commodities.

Inicialmente, obtém-se o valor típico (TP) através da equação:

$$TP_i = \frac{HP_i + LP_i + CP_i}{3} \quad (38)$$

Em seguida, calcula-se o valor típico médio dos últimos 24 dias:

$$SMATP_{24}_i = \frac{\sum_{j=i-23}^i TP_j}{24} \quad (39)$$

O desvio médio dos últimos 24 dias pode ser calculado por:

$$MD_{24}_i = \frac{\sum_{j=i-23}^i |TP_j - SMATP_{24}_i|}{24} \quad (40)$$

E, finalmente, com as variáveis obtidas, é possível calcular o índice CCI_24:

$$CCI_{24}_i = \frac{TP_i - SMATP_{24}_i}{0,015 \times MD_{24}_i} \quad (41)$$

i) AR₂₆

O índice que indica o momento de compra e de venda no período dos últimos 26 dias.

O AR₂₆ é calculado através de:

$$AR_{26_i} = \frac{\sum_{j=i-25}^i (HP_j - OP_j)}{\sum_{j=i-25}^i (OP_j - LP_j)} \quad (42)$$

j) BR₂₆

O BR₂₆ é um índice que indica o momento de propensão à compra e à venda no período dos últimos 26 dias, sendo calculado por:

$$BR_{26_i} = \frac{\sum_{j=i-25}^i (HP_j - CP_{j-1})}{\sum_{j=i-25}^i (CP_{j-1} - LP_j)} \quad (43)$$

4.1.2 Normalização da base de dados

A base de dados definitiva está organizada na forma matricial de ordem 2102 x 10, referente aos 10 índices financeiros técnicos calculados com base nas cinco variáveis das 2128 observações diárias (BELUCO, 2013c). Devido a alguns índices considerarem períodos anteriores de até 26 dias, foram descartados valores de todos os índices no período entre 05/04/2004 e 10/05/2004 com o intuito de minimizar ruídos no processo de treinamento das redes neurais.

O processo de normalização é aplicado a cada um dos 10 vetores que compõem a base de dados. A normalização para o intervalo [-1;1] ocorre conforme a equação:

$$V_n = 2 \times \frac{V - V_{\text{mínimo}}}{V_{\text{máximo}} - V_{\text{mínimo}}} - 1 \quad (44)$$

Onde:

V_n Vetor normalizado

V Vetor a ser normalizado

$V_{mínimo}$	Valor mínimo do vetor V
$V_{máximo}$	Valor máximo do vetor V

O processo de normalização aplicado aos vetores de entrada e de saída que são aplicados às redes neurais que compõem o modelo híbrido de previsão de índices financeiros SOM-ANN/BP auxilia na convergência dos resultados (HAYKIN, 2001).

4.2 CONSTRUÇÃO DO MODELO HÍBRIDO SOM-ANN/BP

Este estudo propõe um modelo híbrido de previsão de índices financeiros que reúne uma rede neural do tipo SOM (*Self-Organizing Map*) com uma do tipo *Perceptron* Multicamadas com Retropropagação. A utilização da rede SOM tem o intuito de segmentar a base de dados em diversos clusters, onde são ressaltadas suas características. A rede BPN é usada para construir um modelo de previsão para cada cluster, a fim de compor um modelo baseado em 10 indicadores financeiros calculados com base em informações do NYSE Composite U.S. 100

4.2.1 Mapas auto-organizáveis

Os mapas autoorganizáveis possuem a característica de dividir bases de dados em agrupamentos que apresentem características afins. Aplicados aos índices financeiros técnicos, os mapas autoorganizáveis processarão a clusterização da base de dados em função de características em comum. O propósito de tal processo é maximizar a eficiência do modelo híbrido SOM-ANN/BP a partir da otimização dos modelos parciais referentes a cada cluster. Da mesma forma, o processo de clusterização é hábil em apontar características similares na base de dados que reflitam eventos econômicos que afetam os resultados nas bolsas de valores.

A rede neural artificial do tipo SOM desenvolvida consiste de 10 neurônios em uma camada de entrada para os índices financeiros normalizados e diversos neurônios arranjados em uma camada bidimensional. O peso inicial de cada neurônio é determinado de forma aleatória e o número total de iterações na etapa de aprendizagem da rede neural é de 62.400, aproximadamente 30 vezes o número de

registros da base de dados, ou seja, $30 \times 2.080 = 62.400$ (KOHONEN, 2013). O parâmetro de aprendizagem inicial é de 0,06, reduzindo gradualmente para 0,03 após 31.156 iterações de aprendizagem, alcançando um valor de 0,01 após 52.640 iterações de aprendizagem. Foram criados 7 clusters conforme pode ser visualizado na U-Matrix, onde o mapa de pesos sinápticos foi gerado de forma aleatória nas dimensões 25×9 . O processo de aprendizagem utilizado neste experimento é sequencial.

A U-Matrix e os planos componentes gerados referentes a cada índice financeiro estão dispostos na Figura 12. O processo de implementação da rede neural SOM e a geração da U-Matrix e dos planos componentes foi realizado através do software Matlab®, versão 2010 e gerou 7 clusters ou agrupamentos de dados.

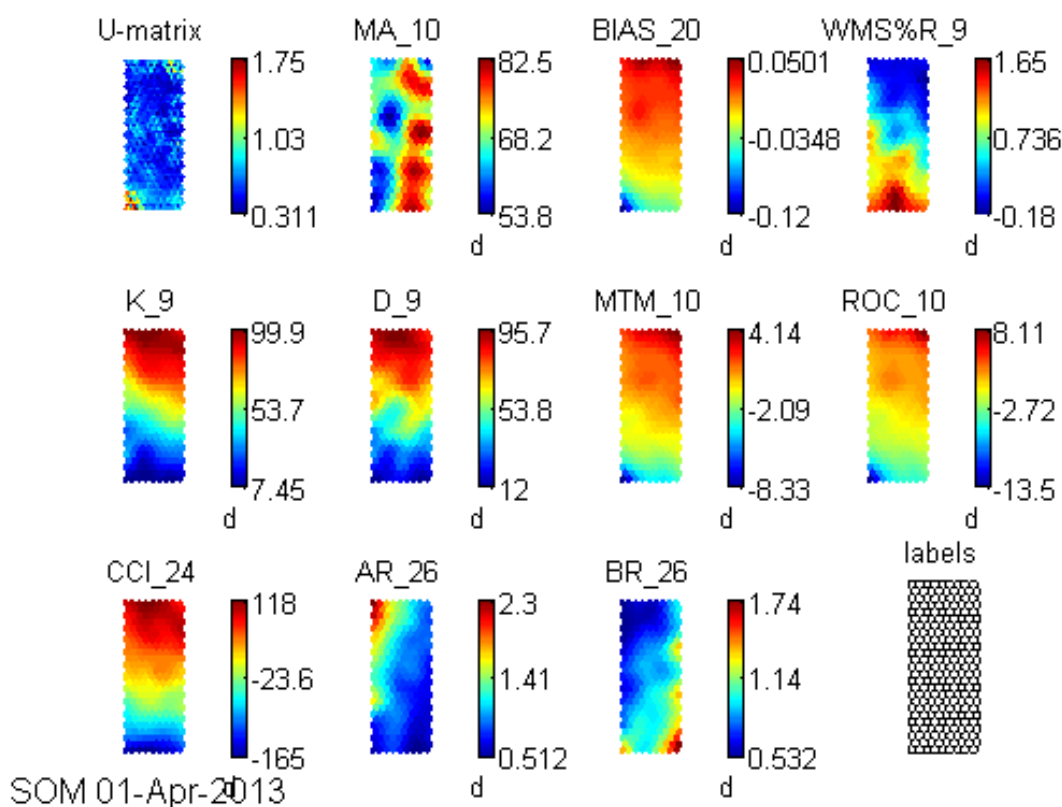


Figura 12 - U-Matrix e planos componentes da rede SOM
Fonte: Elaborada pelo autor no software Matlab®

A matriz de distância unificada, ou U-Matrix, pode ser interpretada como uma imagem na qual a coloração dos pixels ocorre de acordo com a intensidade de cada componente da matriz. Neurônios vizinhos díspares são representados por valores altos (avermelhados) e neurônios vizinhos similares por valores baixos (azulados).

Dessa forma, regiões com baixos valores de gradientes correspondem a vales que agrupam neurônios especializados em padrões semelhantes. Assim, regiões formadas por altos valores indicam fronteiras entre agrupamentos. A Figura 13 demonstra a U-Matrix do modelo SOM-ANN/BP.

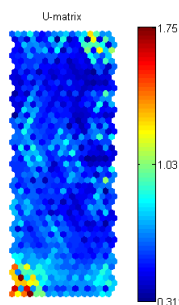


Figura 13 - U-Matrix da rede SOM-ANN/BP
Fonte: Elaborada pelo autor no software Matlab®

O plano componente apresentado para o índice financeiro MA_10 é diferenciado quando em comparação com outros planos componentes. Este apresenta uma divisão de clusters bem definida e pode ser visualizado na figura 14.

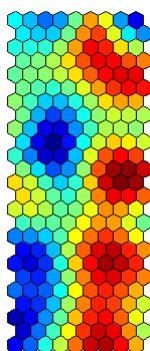
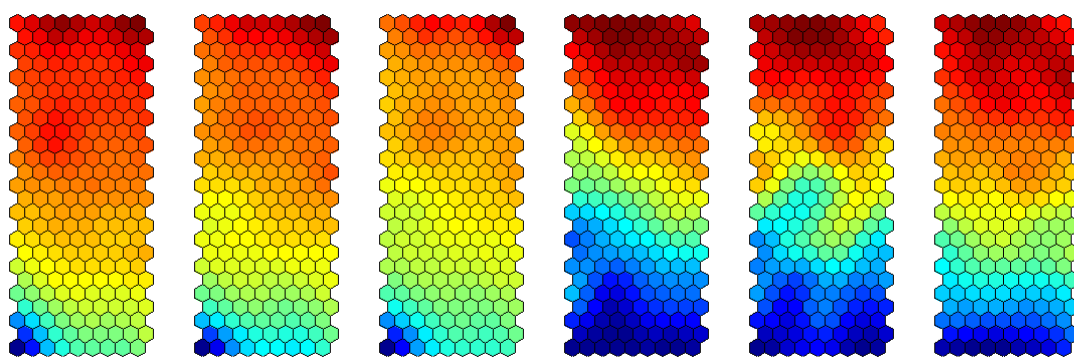


Figura 14 - Plano componente da rede SOM referente ao indicador MA_10
Fonte: Elaborada pelo autor no software Matlab®

Alguns planos componentes apresentaram clusterização equivalente, com baixo grau de diferenciabilidade de suas características. Entre os quais estão os planos componentes dos índices BIAS_20, K_9, D_9, MTM_10, ROC_10 e CCI_24 (Figura 15). Mais especificamente, é possível perceber áreas em comum nos processos de clusterização dos índices BIAS_20, MTM_10 e ROC_10 e entre os índices K_9, D_9 e CCI_24.



(a) (b) (c) (d) (e) (f)
 Figura 15 - Planos componentes da rede SOM referente aos índices (a) BIAS_20, (b) MTM_10, (c) ROC_10, (d) K_9, (e) D_9 e (f) CCI_24
 Fonte: Elaborada pelo autor no software Matlab®

Quanto ao índice WMS%R_9, seu plano componente apresentou agrupamentos específicos, de forma diferenciada em relação aos demais (Figura 16).

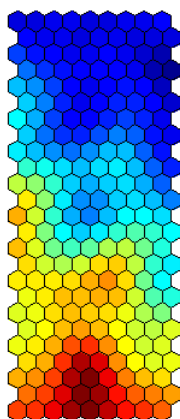
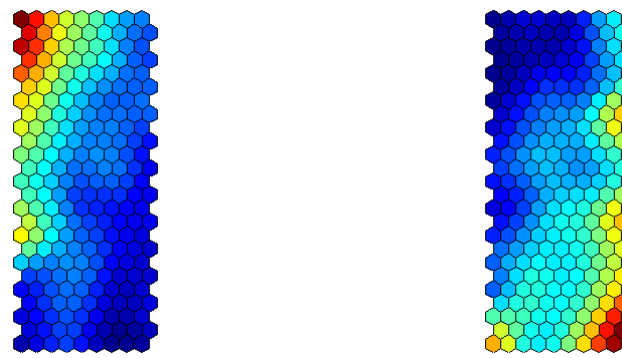


Figura 16 - Plano componente da rede SOM-ANN/BP referente ao indicador WMS%R_9
 Fonte: Elaborada pelo autor no software Matlab®

Os planos componentes dos índices AR_26 e BR_26 denotaram uma correlação inversa, esperada devido à natureza dos seus cálculos (Figura 17).



(a) (b)
 Figura 17 - Planos componentes da rede SOM referente aos indicadores AR_26 e BR_26
 Fonte: Elaborada pelo autor no software Matlab®

Todo o processo de desenvolvimento do algoritmo SOM foi realizado através do software Matlab®. Na Figura 18 é possível visualizar a identificação dos clusters após o processo de treinamento da rede SOM, onde: AR indica o índice AR_26, BIAS indica o índice BIAS_20, BR indica o índice BR_26, CCI representa o índice CCI_24, D representa o índice D_9, K indica o índice K_9, MA representa o índice MA_10, MTM indica o índice MTM_10, ROC representa o índice ROC_10 e WMS indica o índice WMS%R_9.

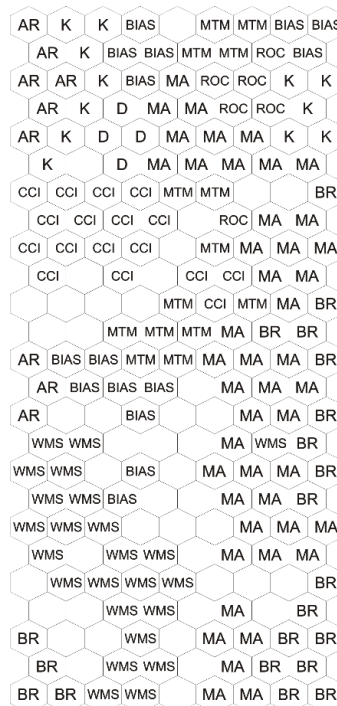


Figura 18 - Identificação visual dos clusters após o treinamento da rede
 Fonte: Elaborada pelo autor no software Matlab®

A Figura 19 demonstra a topologia inicial e final do mapa autoorganizável antes e após o processo de clusterização,. A topologia ao final da etapa de treinamento é equivalente a uma malha de 25 x 9 pontos.

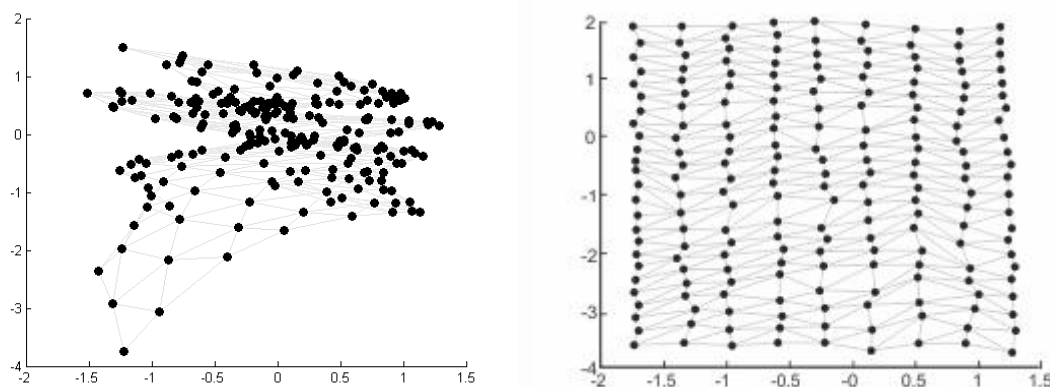


Figura 19 - Topologia inicial e final do Mapa Autoorganizável (SOM)
Fonte: Elaborada pelo autor no software Matlab®

O processo de clusterização resultou em 7 agrupamentos, cuja quantificação de dados de cada cluster é descrita na Tabela 3. Cabe ressaltar que o processo de clusterização através do algoritmo SOM não objetiva a divisão equânime de dados por cluster.

Tabela 3 – Resultados do processo de clusterização

Cluster	Número de variáveis
1	228
2	279
3	345
4	308
5	248
6	390
7	282

Fonte: Elaborada pelo autor

A próxima etapa do modelo híbrido SOM-ANN/BP buscará os melhores modelos neurais de retropropagação para a previsão do valor de fechamento do índice NYSE Composite U.S. 100 para o dia seguinte. Para tanto, serão realizados testes com diferentes topologias dos modelos neurais baseados em algoritmos do tipo perceptron multicamadas com retropropagação.

4.2.2 Algoritmo backpropagation

A partir da criação dos clusters na etapa de aplicação dos mapas auto-organizáveis, são criados modelos de redes neurais do tipo perceptron com retropropagação para cada cluster, com o propósito de obter a melhor solução entre eles. O método de aprendizagem dos modelos neurais selecionado é o supervisionado, onde são fornecidos os valores de entrada e o valor de saída desejada no que se refere ao conjunto de dados de treinamento. A ferramenta de software utilizada para a implementação dos algoritmos é o Mathworks Matlab®, em sua versão 10.0. A escolha do software é justificada pela robustez do software em termos de programação.

A base de dados normalizada (BELUCO, 2013c) foi segmentada da seguinte forma:

- Grupo de treinamento, contendo 80% da totalidade dos dados;
- Grupo de testes, contendo 10% da totalidade dos dados
- Grupo de validação, contendo 10% da totalidade dos dados

A topologia das redes utilizadas baseia-se em testes realizados com diferentes composições da camada intermediária, através de uma camada simples com 5, 10 e 15 neurônios em sua estrutura. A estrutura dos testes é baseada em trabalhos similares publicados por pesquisadores (HULME; XU, 2001; ARMANO et al, 2005; ENKE; THAVORNWONG, 2005; DAI et al, 2005; LIANG et al, 2011).

A camada de entrada foi definida com 10 neurônios, a partir dos 10 índices financeiros técnicos. A camada de saída é responsável por projetar o Valor de Fechamento do índice NYSE Composite U.S. 100 para o dia seguinte em relação aos índices financeiros técnicos do dia anterior.

A determinação do número de neurônios a ser utilizado na camada intermediária não possui regra específica (HAYKIN, 2001). Desta forma, a especificação utilizada nos experimentos é a escolha de três modelos que apresentassem quantidades de neurônios inferior, superior e equânime à dimensão da camada de entrada.

A Figura 20 apresenta a topologia com 5 neurônios na camada intermediária usada nas baterias de testes.

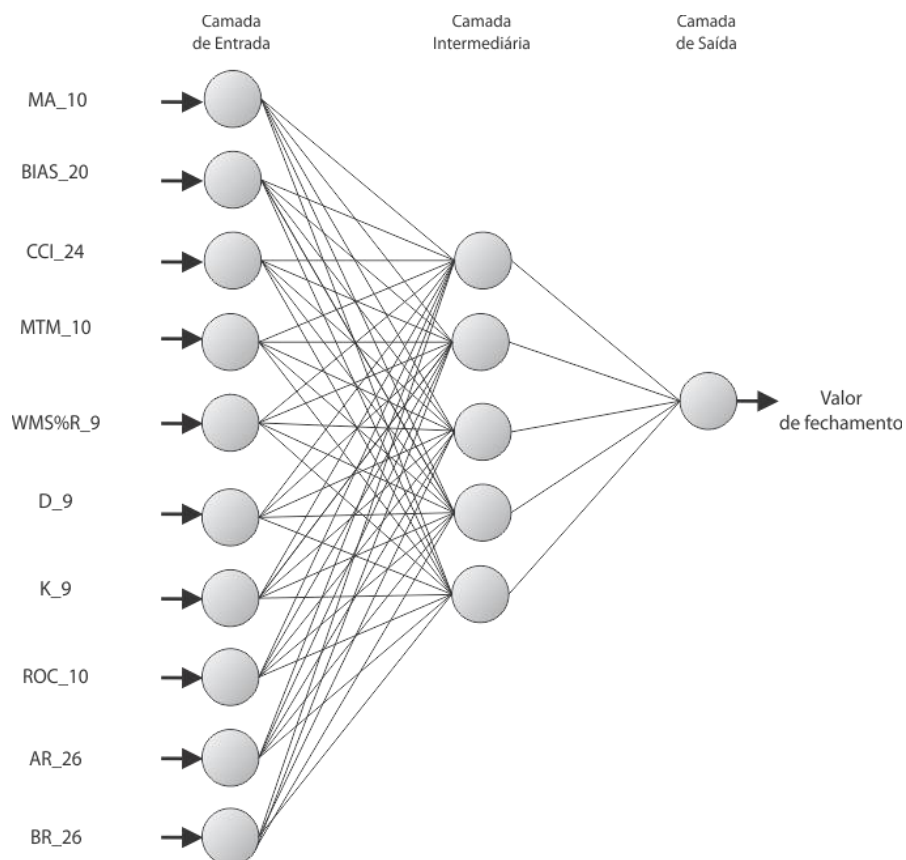


Figura 20 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 5 camadas intermediárias
Fonte: Elaborada pelo autor no software Matlab®

A topologia para o modelo que utiliza 10 neurônios na camada intermediária é mostrada na Figura 21 e a topologia para o modelo contendo 15 neurônios na camada intermediária é ilustrada na Figura 22.

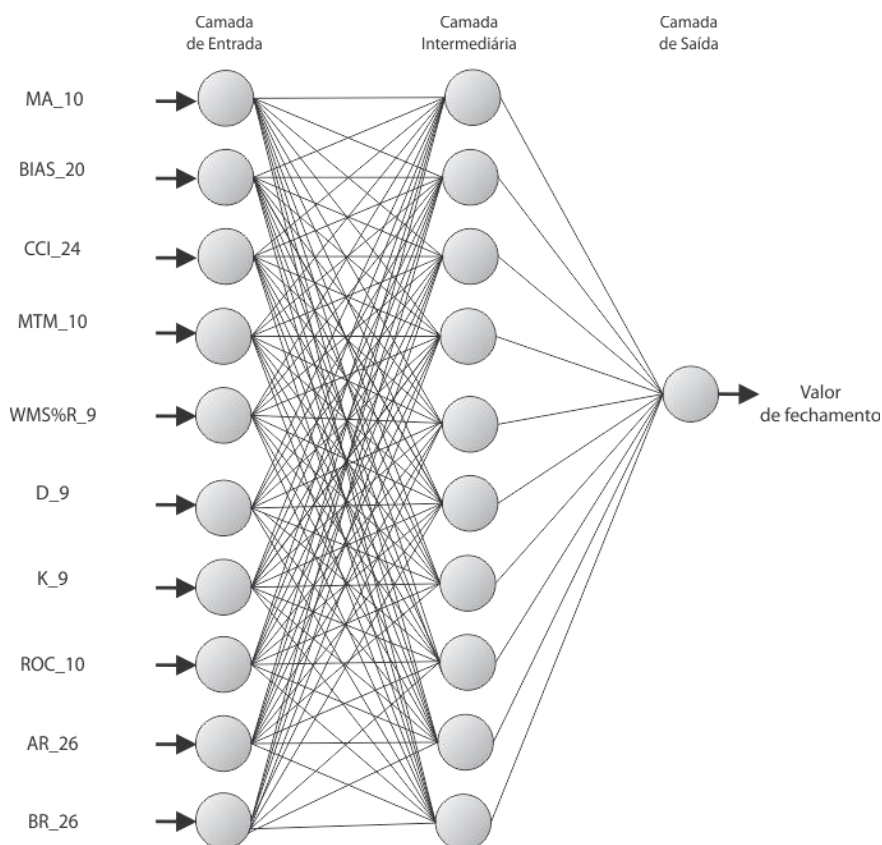


Figura 21 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 10 camadas intermediárias
Fonte: Elaborada pelo autor no software Matlab®

Inicialmente foram realizados testes considerando todos os índices financeiros técnicos como parâmetros de entrada. Nesta condição, ocorreram três baterias de testes considerando diferentes composições da camada intermediária com 5, 10 e 15 neurônios para cada cluster.

O processo de segmentação da base de dados foi aleatória e a composição percentual de cada grupo é a utilizada por diversos pesquisadores da área de redes neurais artificiais (HSU; 2011; WANG et al.; 2011; LU et al.; 2011; TSAI et al.; 2010). Os gráficos das séries históricas dos índices financeiros técnicos utilizados como parâmetros de entrada dos modelos neurais podem ser visualizadas no Apêndice B.

A função de ativação usada nos neurônios do modelo foi a tangente hiperbólica, comprovadamente mais eficiente na convergência dos resultados quando a base de dados é normalizada para o intervalo entre -1 e +1.

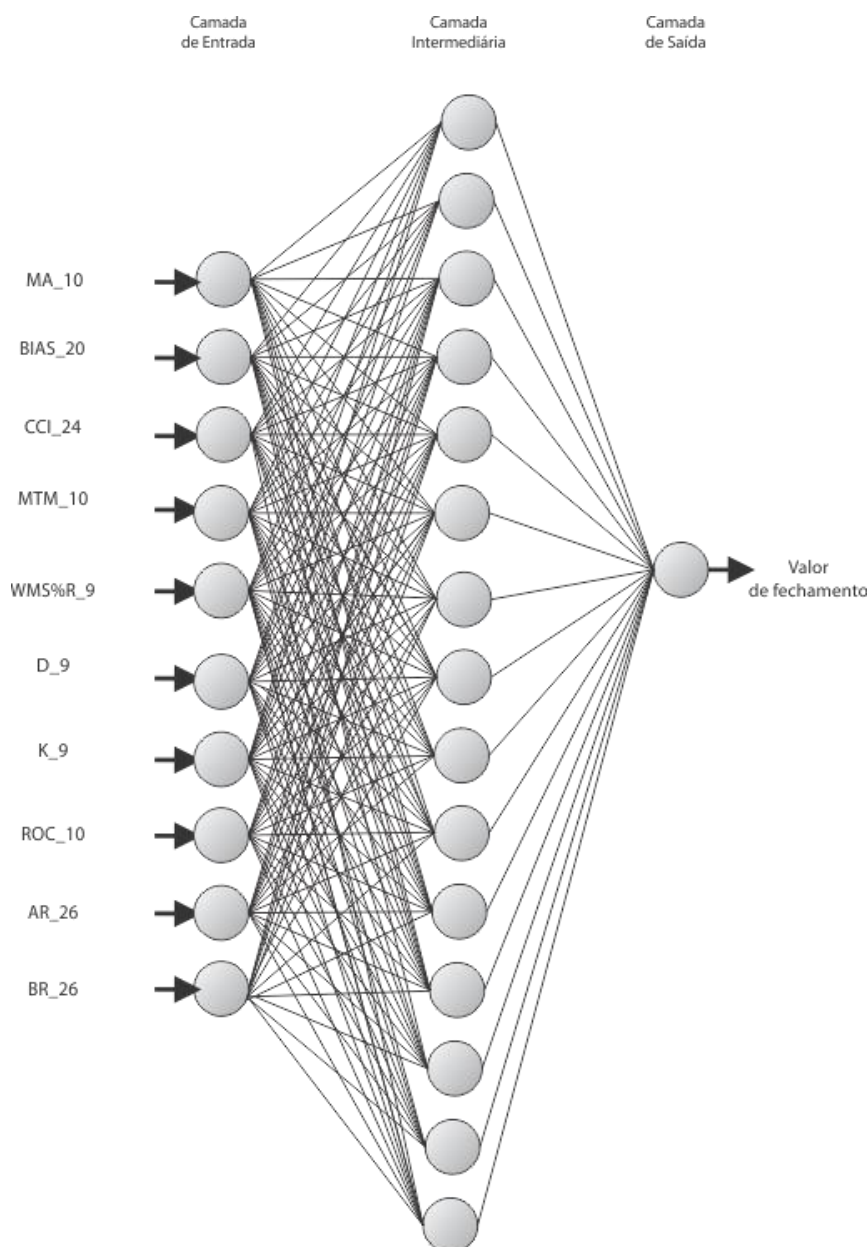


Figura 22 - Modelo proposto de rede neural do tipo perceptron multicamadas com retropropagação com 15 camadas intermediárias
Fonte: Elaborada pelo autor no software Matlab®

A Tabela 4 demonstra os resultados obtidos na etapa de treinamento dos modelos neurais em cada cluster analisado. O número de iterações da fase de treinamento foi estipulado em 2500, baseado em estudos similares (HSU, 2011; MOSTAFA, 2010).

Tabela 4 - Análise do desempenho dos modelos neurais

Cluster	Número de neurônios na camada intermediária	Número de iterações	RMSE
1	5	2500	0,5417
	10	2500	0,4220
	15	2500	0,5826
2	5	2500	0,3075
	10	2500	0,2619
	15	2500	0,1432
3	5	2500	0,2734
	10	2500	0,5092
	15	2500	0,1249
4	5	2500	0,3582
	10	2500	0,3103
	15	2500	0,3491
5	5	2500	0,2211
	10	2500	0,0995
	15	2500	0,4117
6	5	2500	0,3320
	10	2500	0,5273
	15	2500	0,2439
7	5	2500	0,3714
	10	2500	0,3178
	15	2500	0,4005

Fonte: Elaborada pelo autor

Os resultados para o treinamento do modelo neural gerado para o Cluster 1 são demonstrados através dos gráficos da raiz do erro médio quadrático (RMSE) na Figura 23. O modelo contendo 10 neurônios na camada intermediária demonstrou ser o que apresentou o menor valor da raiz do erro médio quadrático (RMSE = 0,4220).

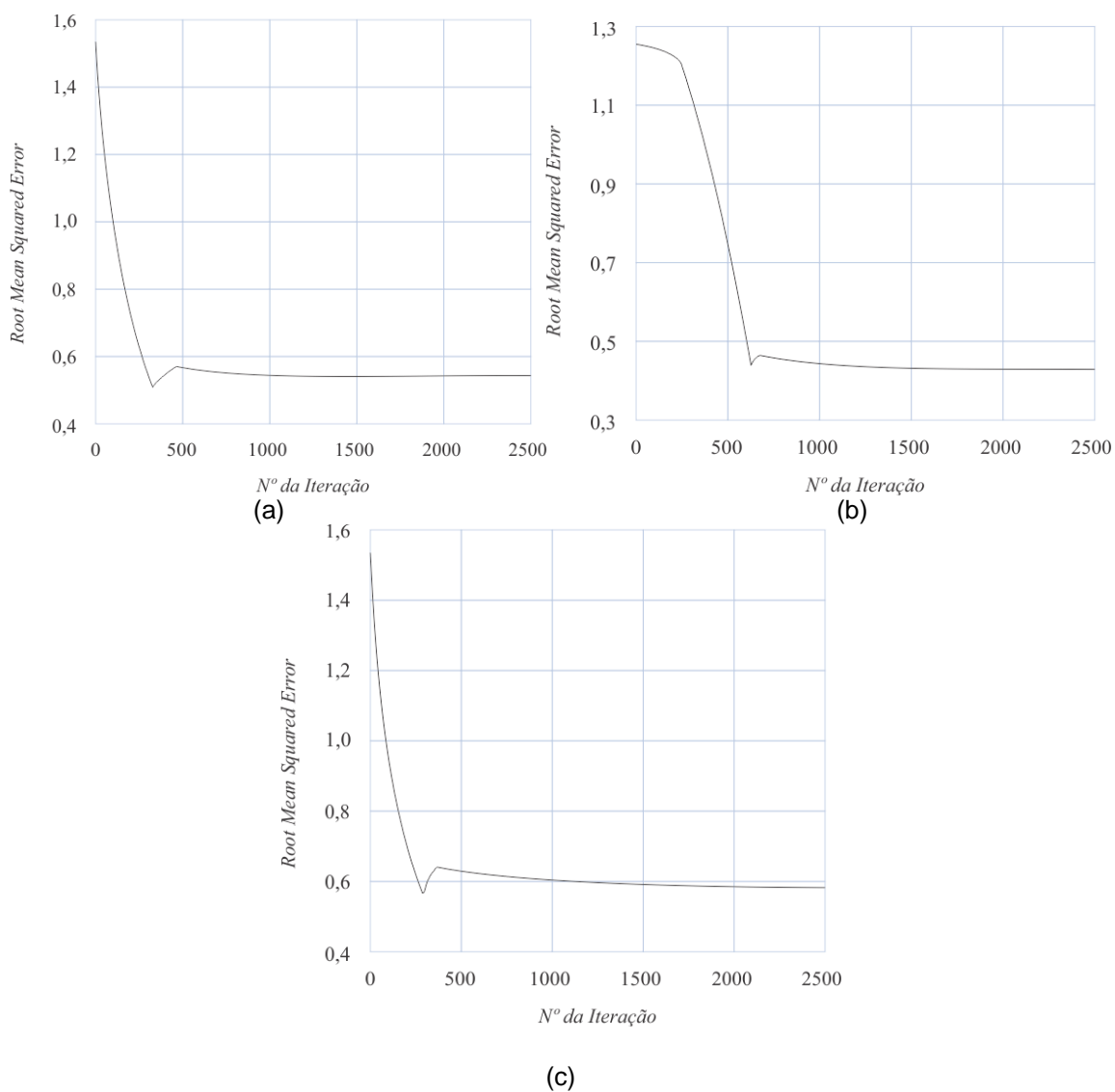


Figura 23 - RMSE para o treinamento do modelo neural referente ao Cluster 1 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
Fonte: Elaborada pelo autor no software Matlab®

Os resultados da etapa de treinamento para o modelo neural do Cluster 2 estão ilustrados na Figura 24. O modelo neural que obteve o menor valor para a raiz do erro médio quadrático é o composto por 15 neurônios na camada intermediária (RMSE = 0,1432).

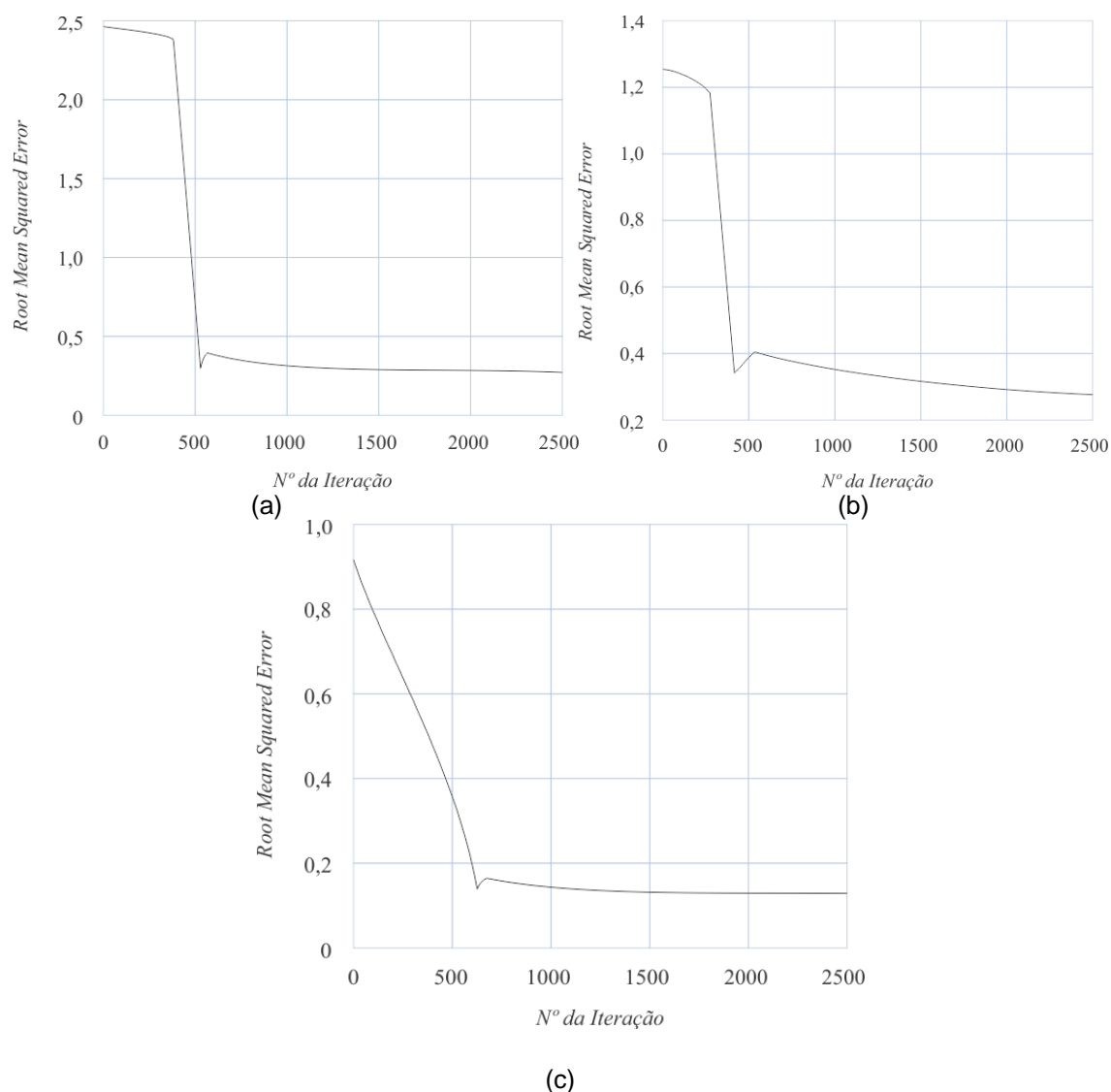


Figura 24 - RMSE para o treinamento do modelo neural referente ao Cluster 2 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
 Fonte: Elaborada pelo autor no software Matlab®

A Figura 25 demonstra os gráficos dos resultados obtidos no treinamento dos modelos neurais para o Cluster 3. Da mesma forma como o modelo neural obtido para o Cluster 2, a opção por 15 neurônios na camada intermediária apresentou um RMSE menor em relação às opções com 5 e 10 neurônios na mesma camada (RMSE = 0,1249).

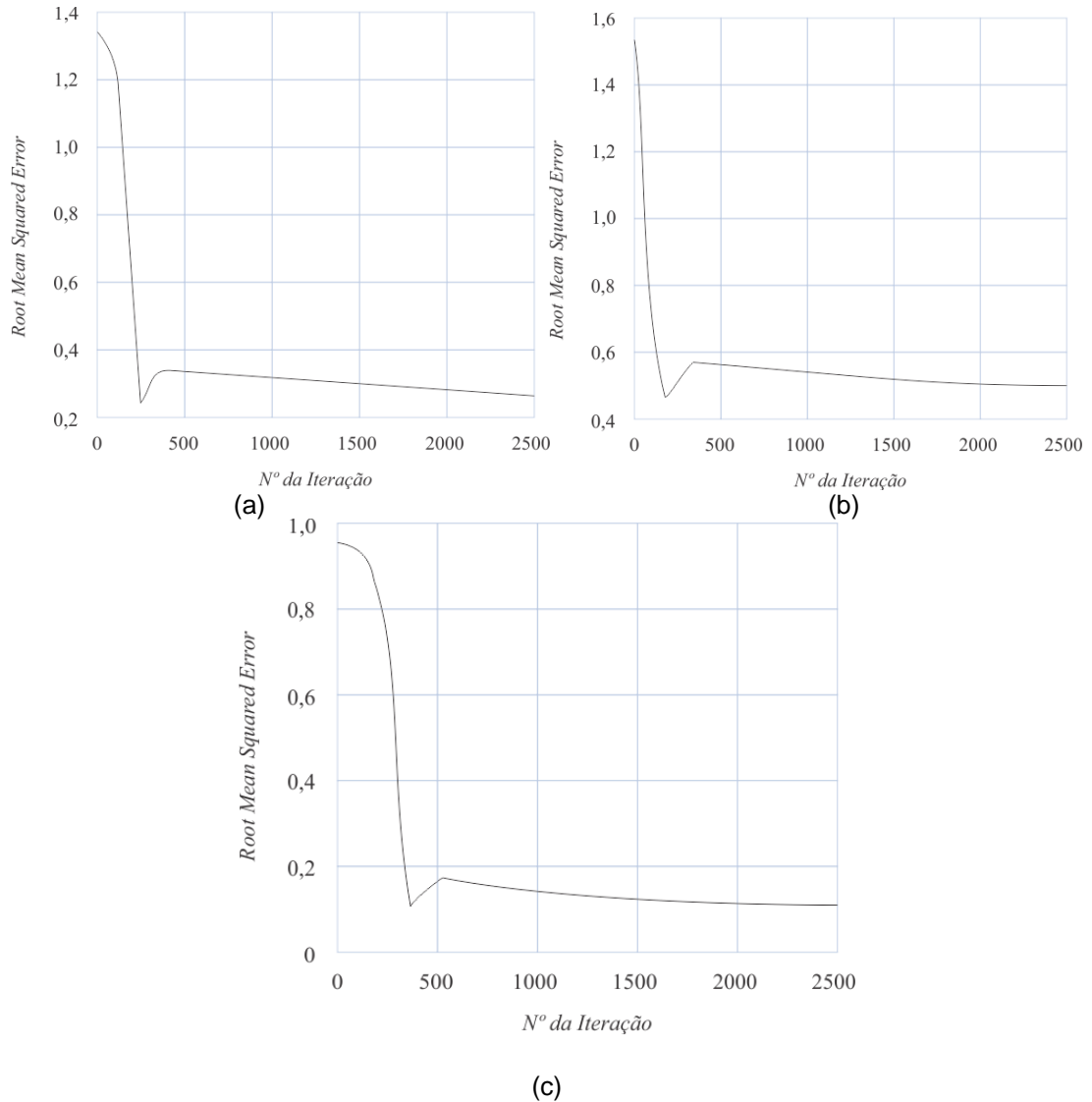


Figura 25 - RMSE para o treinamento do modelo neural referente ao Cluster 3 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
 Fonte: Elaborada pelo autor no software Matlab®

No modelo neural elaborado após a etapa de treinamento para o Cluster 4 o valor da raiz do erro médio quadrático para o que obteve o melhor desempenho foi igual a 0,3103, referente ao modelo com 10 neurônios na camada intermediária. A Figura 26 ilustra os gráficos com os resultados.

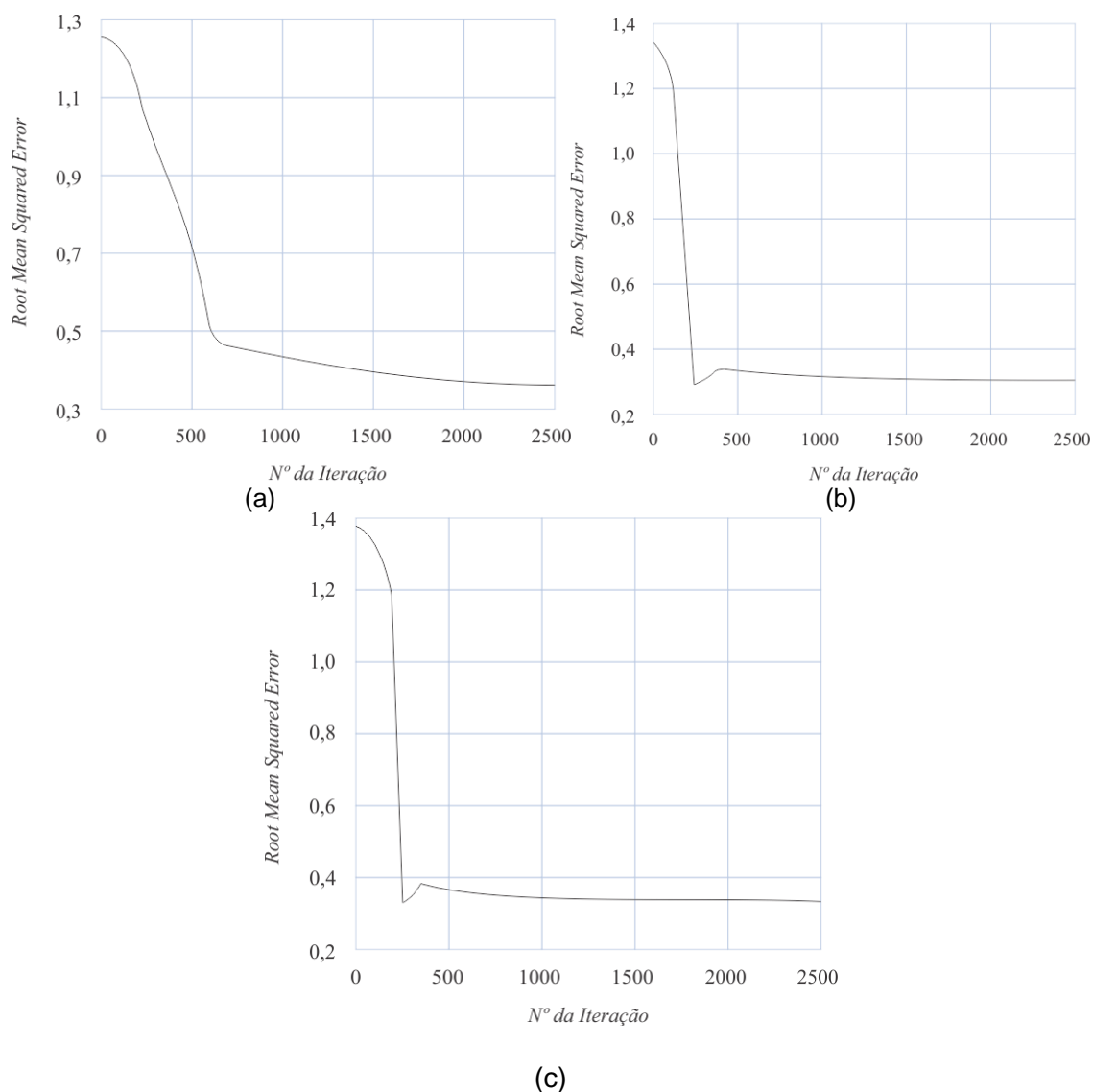


Figura 26 - RMSE para o treinamento do modelo neural referente ao Cluster 4 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
 Fonte: Elaborada pelo autor no software Matlab®

A figura 27 apresenta os gráficos dos resultados da raiz do erro médio quadrático relativos à etapa de treinamento par a obtenção dos modelos neurais referentes ao Cluster 5. O modelo neural com melhor desempenho foi o composto por 10 neurônios na camada intermediária, obtendo $RMSE = 0,0995$.

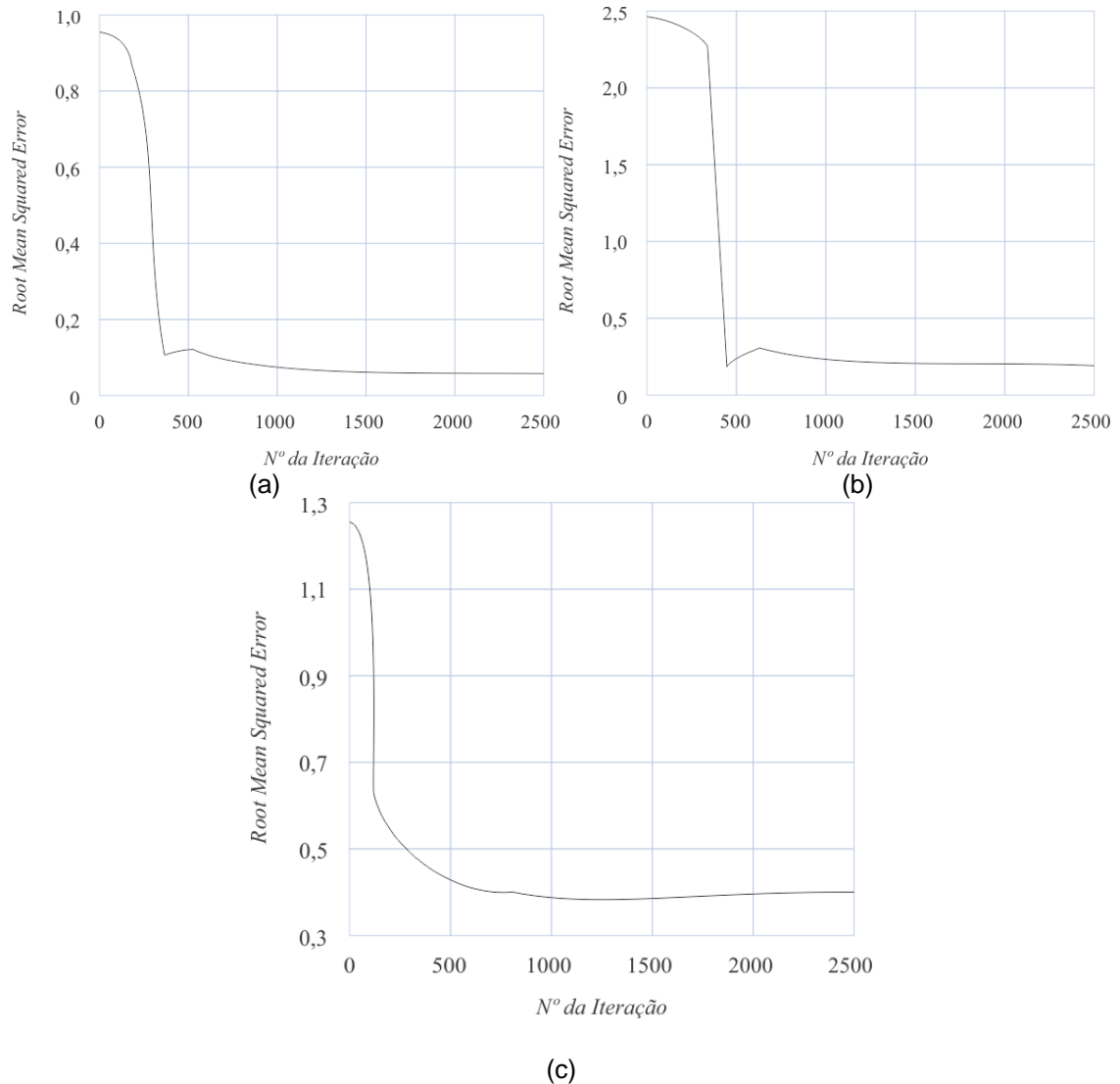


Figura 27 - RMSE para o treinamento do modelo neural referente ao Cluster 5 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
Fonte: Elaborada pelo autor no software Matlab®

O modelo neural estruturado para o Cluster 6 é composto de 15 neurônios na camada intermediária, obtendo uma raiz do erro médio quadrático no valor de 0,2439. A Figura 28 demonstra a convergência mais eficiente no processo de treinamento para a obtenção do modelo neural.

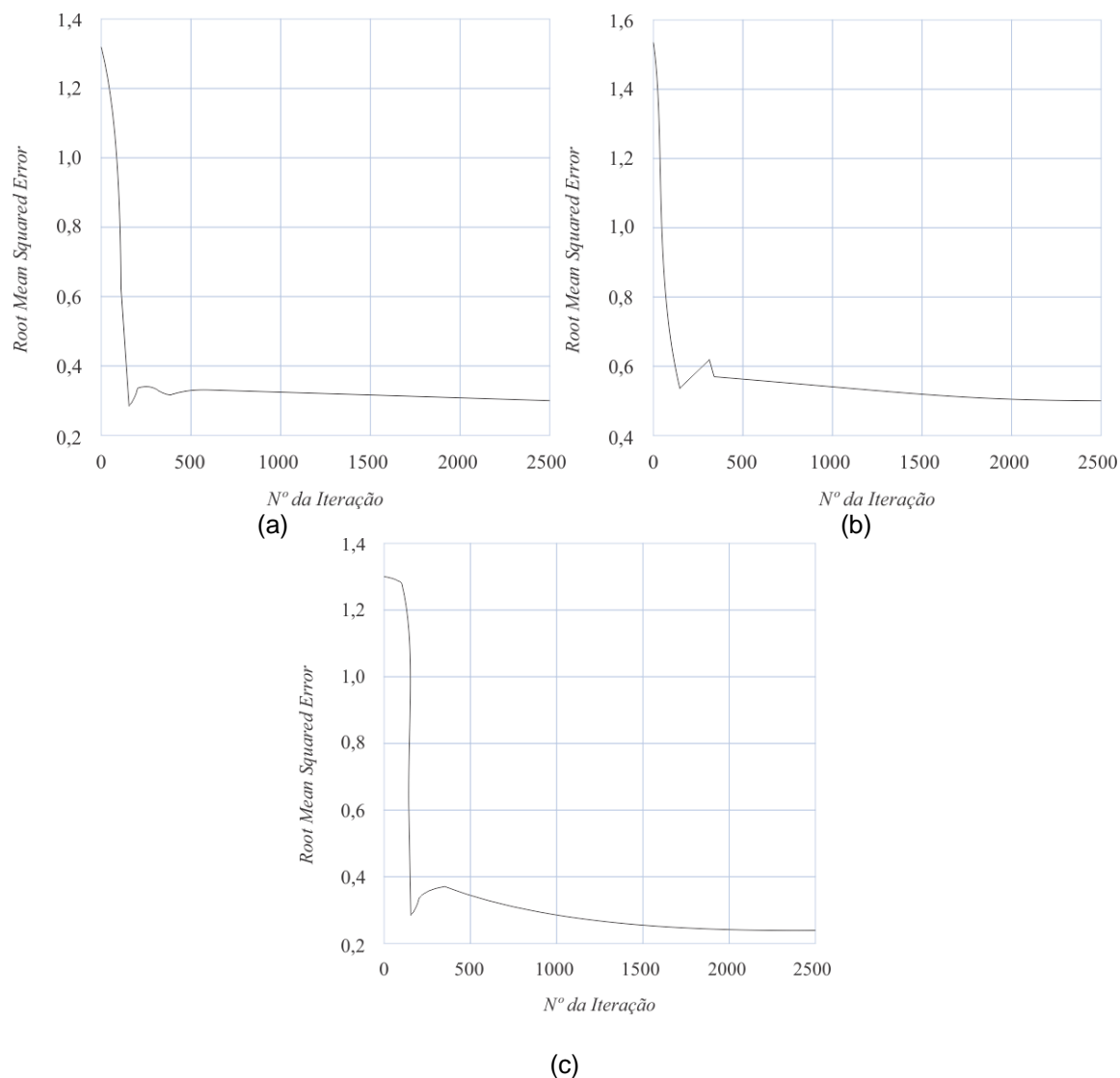


Figura 28 - RMSE para o treinamento do modelo neural referente ao Cluster 6 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
 Fonte: Elaborada pelo autor no software Matlab®

Por fim, a Figura 29 apresenta os resultados para a etapa de treinamento do modelo neural do Cluster 7. O melhor desempenho foi obtido pelo modelo contendo 10 neurônios na camada intermediária, que alcançou um valor igual a 0,3178 para a raiz do erro médio quadrático.

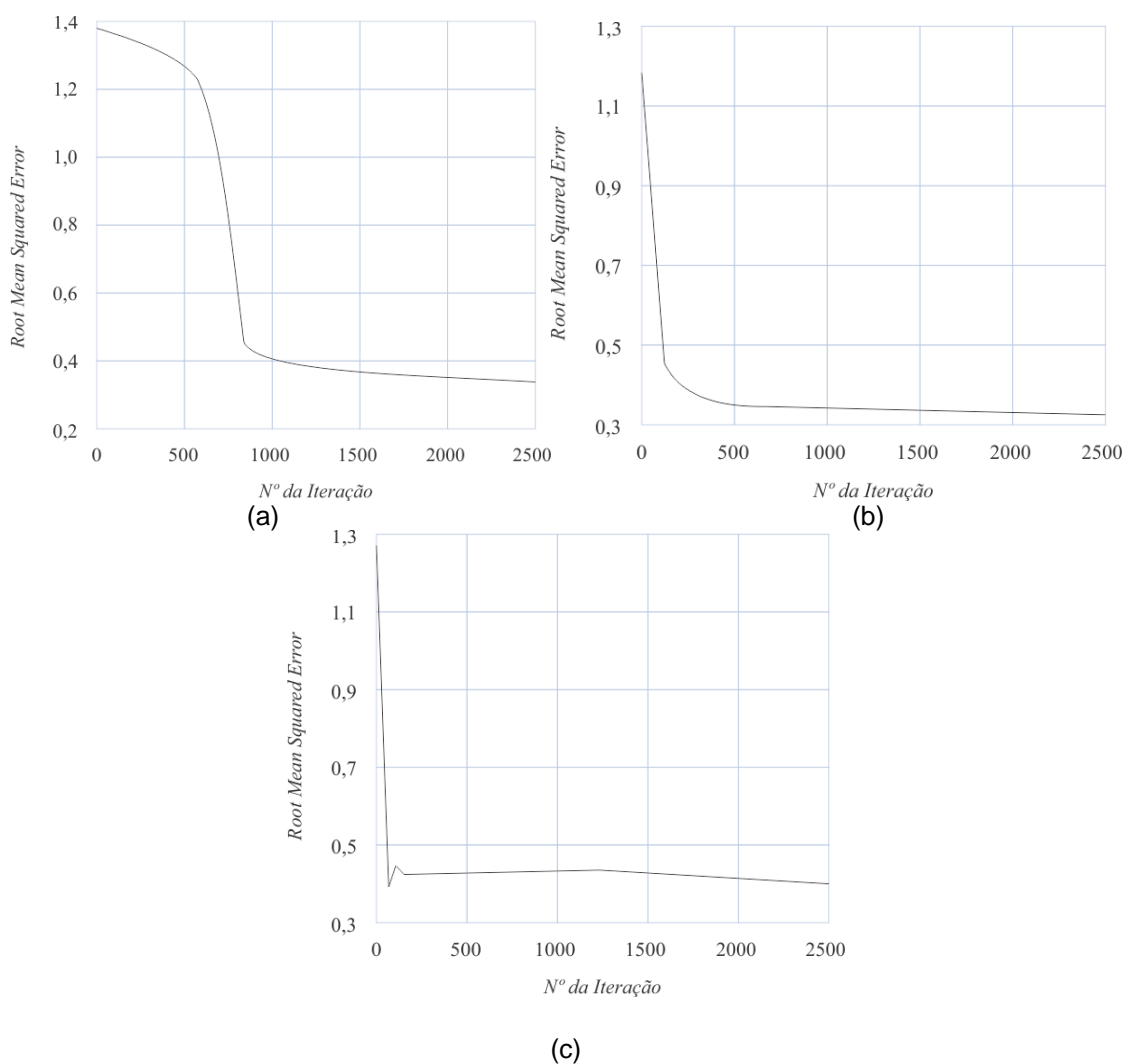


Figura 29 - RMSE para o treinamento do modelo neural referente ao Cluster 7 com: (a) 5 neurônios, (b) 10 neurônios e (c) 15 neurônios na camada intermediária
Fonte: Elaborada pelo autor no software Matlab®

Dessa forma, após a execução da etapa de treinamento dos modelos neurais com diversas topologias é possível verificar na Tabela 5 a composição dos que apresentaram o melhor desempenho.

Tabela 5 - Composição da camada intermediária dos modelos de melhor desempenho por cluster

Cluster	1	2	3	4	5	6	7
Número de neurônios na camada intermediária	10	15	15	10	10	15	10

Fonte: Elaborada pelo autor

Após a definição dos modelos neurais que obtiveram os melhores desempenhos em cada cluster é possível realizar uma fase de teste com base nos grupos de teste previamente selecionados de forma aleatória. Os resultados para a raiz do erro médio quadrático de cada cluster são demonstrados na Tabela 6. É possível perceber que os resultados obtidos se enquadram num intervalo entre 0,1 e 0,5. Tais resultados podem ser considerados de bom desempenho em virtude da proximidade ao valor zero.

Tabela 6 - Avaliação do desempenho dos modelos neurais para o grupo de testes

Cluster	1	2	3	4	5	6	7
RMSE	0,4175	0,1489	0,1523	0,2850	0,1281	0,2511	0,3209

Fonte: Elaborada pelo autor

4.3 AVALIAÇÃO DO MODELO HÍBRIDO SOM-ANN/BP

O processo de validação dos modelos neurais ocorre com base na análise dos resultados após a inserção dos dados do grupo de validação selecionados previamente a partir da base de dados. A avaliação dos resultados é baseada na raiz do erro médio quadrático. Os resultados podem ser visualizados na Tabela 7.

Tabela 7 - Avaliação do desempenho dos modelos para o grupo de validação

Desempenho	RMSE	MAE	MAPE
Modelo SOM-ANN/BP (cluster 1)	0,5074	0,2942	0,3837 %
Modelo SOM-ANN/BP (cluster 2)	0,3222	0,1571	0,2140 %
Modelo SOM-ANN/BP (cluster 3)	0,2939	0,1410	0,2075 %
Modelo SOM-ANN/BP (cluster 4)	0,3568	0,2594	0,1543%
Modelo SOM-ANN/BP (cluster 5)	0,2519	0,1480	0,1937 %
Modelo SOM-ANN/BP (cluster 6)	0,3845	0,2745	0,2562 %
Modelo SOM-ANN/BP (cluster 7)	0,4718	0,2531	0,2355 %

Fonte: Elaborada pelo autor

Os resultados para a Raiz do Erro Médio Quadrado (RMSE) são um tanto superiores aos obtidos com o grupo de testes. Os modelos neurais apresentaram resultados distintos para o RMSE, variando num intervalo entre 0,25 e 0,50, aproximadamente. Quanto ao Erro Médio Absoluto (MAE), os resultados estão num intervalo entre 0,14 e 0,30, aproximadamente. O Erro Médio Absoluto Percentual (MAPE) resultante foi representativo, no instante em que se localizou no intervalo entre 0,15% e 0,25%, com apenas um resultado fora do intervalo (0,3837%).

5. MODELOS DE REGRESSÃO MÚLTIPLA

Para a obtenção do modelo de regressão foi utilizada a equação 45 cuja variável Y_{REG} corresponde à variável dependente Valor do Fechamento presente na base de dados do índice NYSE Composite U.S. 100 (BELUCO, 2013a).

$$Y_{REG} = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_4 \cdot X_4 + \beta_5 \cdot X_5 + \beta_6 \cdot X_6 + \beta_7 \cdot X_7 + \beta_8 \cdot X_8 + \beta_9 \cdot X_9 + \beta_{10} \cdot X_{10} \quad (45)$$

Onde

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9, \beta_{10}$ são as constantes

$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}$ representam as variáveis independentes MA_10, BIAS_20, WMS%R_9, K_9, D_9, MTM_10, ROC_10, CCI_24, AR_26 e BR_26, respectivamente. Todas as variáveis são apresentadas na forma normalizada.

A construção do modelo de regressão foi processada com o auxílio dos softwares SPSS®, versão 18.0, e Microsoft Excel®, versão 2013.

5.1 COMPORTAMENTO DAS VARIÁVEIS

Através da matriz de correlação entre as variáveis independentes, é possível perceber uma correlação alta (superior a 0,9) entre as variáveis BIAS_20 e ROC_10. A matriz de correlação entre as variáveis independentes pode ser verificada na Tabela 8. Algumas correlações altas já eram esperadas devido à natureza do cálculo dos índices financeiros técnicos. Por exemplo, entre os índices MTM_10 e ROC_10, que consideram os resultados dos 10 dias anteriores. O índice K_9 apresentou uma sequência de correlações superiores a 0,75 com vários índices: BIAS_20, D_9, MTM_10, ROC_10 e CCI_24. O mesmo índice K_9 também apresentou correlação superior a 0,75, porém negativa, com o índice WMS%R_9.

Tabela 8- Matriz de correlação entre as variáveis independentes

Índices financeiros		MA_10	BIAS_20	WMS%R_9	K_9	D_9	MTM_10	ROC_10	CCI_24	AR_26	BR_26
MA_10	Pearson Correlation	1	0,006	0,048*	0,007	0,043*	-0,009	-0,024	0,007	-0,181**	0,004
	Sig. (2-tailed)		0,795	0,029	0,746	0,049	0,677	0,281	0,759	0,000	0,873
BIAS_20	Pearson Correlation	0,006	1	-0,640**	0,759**	0,735**	0,890**	0,900**	0,828**	0,409**	-0,375**
	Sig. (2-tailed)	0,795		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
WMS%R_9	Pearson Correlation	0,048*	-0,640**	1	-0,801**	-0,581**	-0,621**	-0,586**	-0,720**	-0,262**	0,305**
	Sig. (2-tailed)	0,029	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000
K_9	Pearson Correlation	0,007	0,759**	-0,801**	1	0,921**	0,789**	0,755**	0,864**	0,401**	-0,452**
	Sig. (2-tailed)	0,746	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000
D_9	Pearson Correlation	0,043*	0,735**	-0,581**	0,921**	1	0,761**	0,726**	0,836**	0,459**	-0,515**
	Sig. (2-tailed)	0,049	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
MTM_10	Pearson Correlation	-0,009	0,890**	-0,621**	0,789**	0,761**	1	0,979**	0,791**	0,358**	-0,340**
	Sig. (2-tailed)	0,677	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000
ROC_10	Pearson Correlation	-0,024	0,900**	-0,586**	0,755**	0,726**	0,979**	1	0,746**	0,344**	-0,301**
	Sig. (2-tailed)	0,281	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
CCI_24	Pearson Correlation	0,007	0,828**	-0,720**	0,864**	0,836**	0,791**	0,746**	1	0,457**	-0,454**
	Sig. (2-tailed)	0,759	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
AR_26	Pearson Correlation	-0,181**	0,409**	-0,262**	0,401**	0,459**	0,358**	0,344**	0,457**	1	-0,573**
	Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000
BR_26	Pearson Correlation	0,004	-0,375**	0,305**	-0,452**	-0,515**	-0,340**	-0,301**	-0,454**	-0,573**	1
	Sig. (2-tailed)	0,873	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

* . Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

A Figura 30 demonstra os diagramas de dispersão entre os índices financeiros com as correlações mais altas entre si. Os índices ROC_10, CCI_24, BIAS_20, K_9 e WMS%R_9 obtiveram valores de correlação superiores a 0,8.

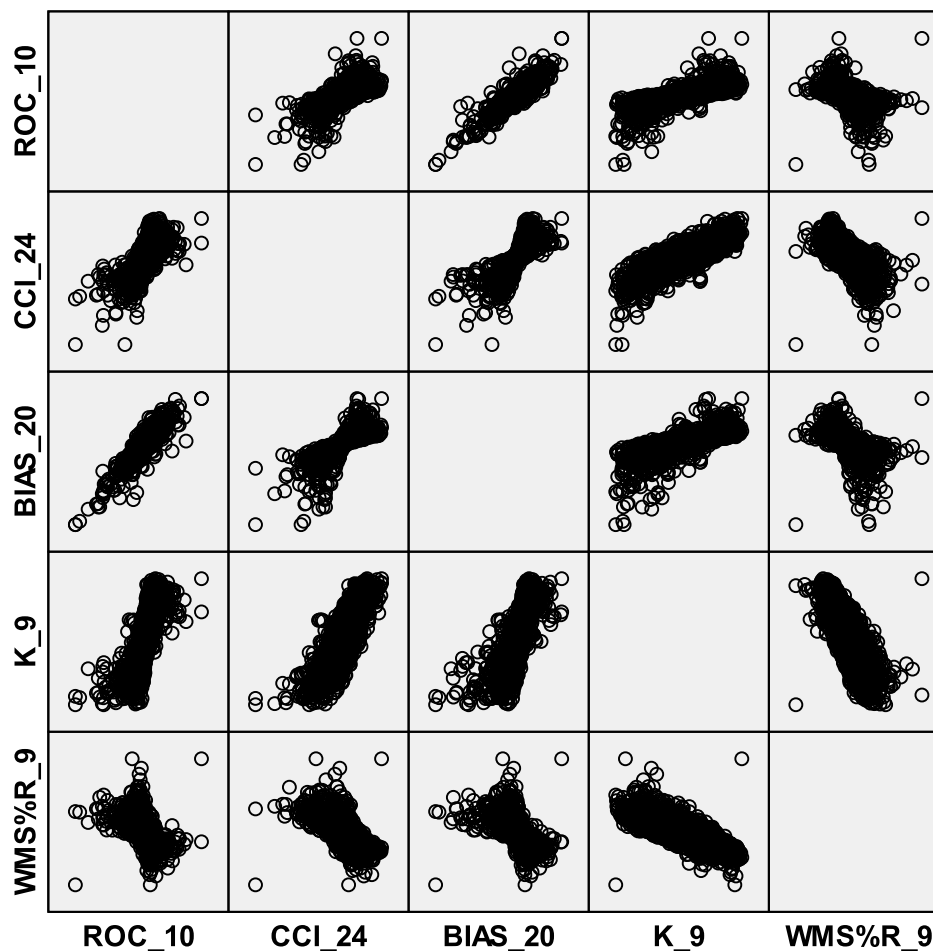


Figura 30 - Múltiplos diagramas de dispersão entre as variáveis ROC_10, CCI_24, BIAS_20, K_9 e WMS%R_9

Fonte: Elaborada pelo autor no software SPSS®

Por meio dos múltiplos diagramas de dispersão ilustrados na Figura 30 é possível perceber as correlações positivas e negativas existentes entre alguns dos índices financeiros técnicos utilizados no modelo híbrido que demonstraram maior correlação.

A Tabela 9 demonstra a correlação entre as variáveis independentes (índices financeiros) e a variável dependente (Valor de fechamento). É perceptível a alta correlação entre o índice financeiro MA_10 e o valor de fechamento visto que o resultado foi muito próximo ao valor 1.

Tabela 9 - Correlação entre as variáveis independentes e a variável dependente

Correlação	Valor de fechamento	Significância
MA_10	0,989	0,000
BIAS_20	0,124	0,000
WMS%R_9	-0,085	0,000
K_9	0,120	0,000
D_9	0,131	0,000
MTM_10	0,107	0,000
ROC_10	0,088	0,000
CCI_24	0,115	0,000
AR_26	-0,149	0,000
BR_26	-0,051	0,020

Fonte: Elaborada pelo autor

Considerando a variável dependente Y_{REG} referente ao valor de fechamento do índice NYSE Composite U.S. 100, as Figuras 31, 32 e 33 apresentam os gráficos de regressão linear entre cada variável independente e dependente.

O gráfico de regressão entre a variável independente MA_10 e a variável dependente Valor de Fechamento já demonstra a forte correlação positiva existente entre elas (Figura 31).

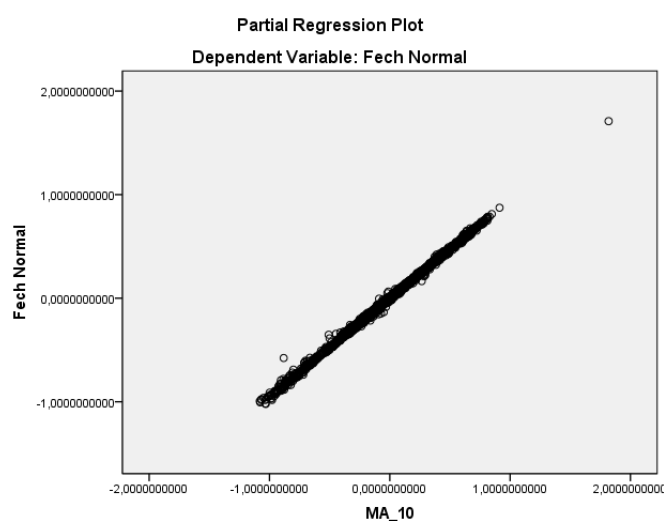


Figura 31 - Gráfico de regressão entre o índice financeiro MA_10 e o valor de Fechamento do índice NYSE Composite U.S. 100

Fonte: Elaborada pelo autor no software SPSS®

A partir dos diagramas de dispersão, percebe-se uma correlação positiva mais alta entre as variáveis BIAS_20 e MTM_10 e o valor de fechamento (variável dependente).

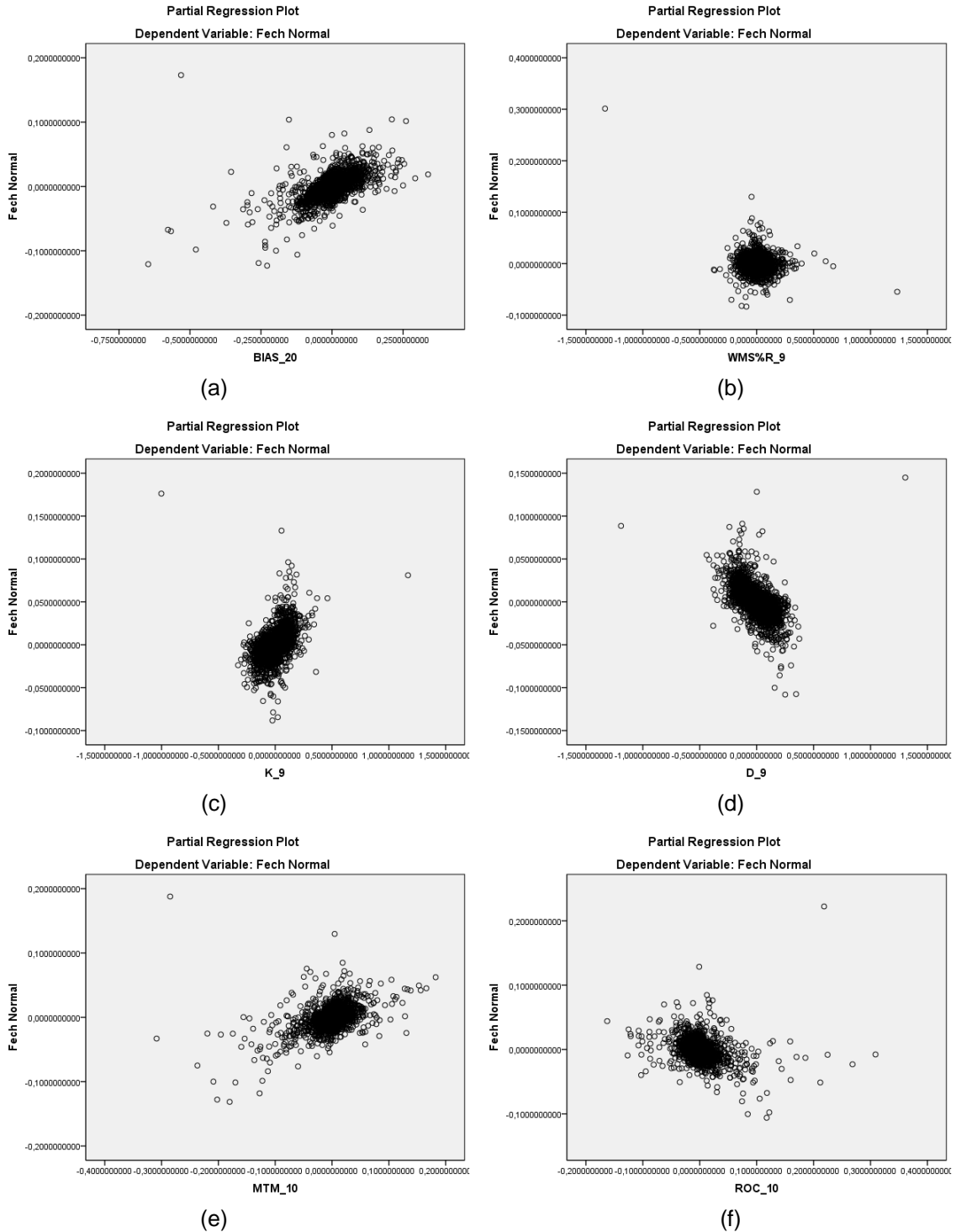


Figura 32 - Gráfico de regressão entre o valor de Fechamento do índice NYSE Composite U.S. 100 e o índice: (a) BIAS_20; (b) WMS%R_9; (c) K_9; (d) D_9; (e) MTM_10; (f) ROC_10

Fonte: Elaborada pelo autor no software SPSS®

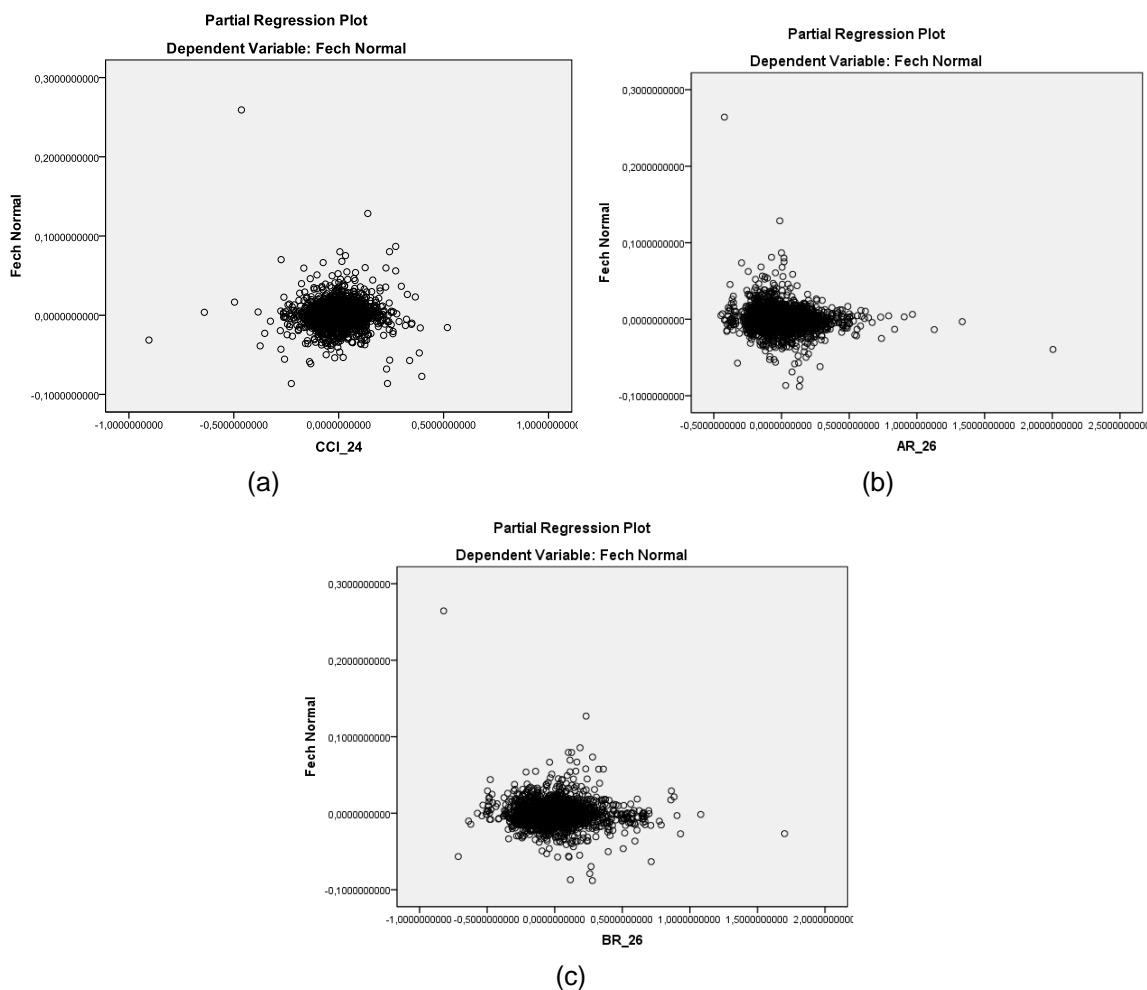


Figura 33 - Gráfico de regressão entre o valor de Fechamento do índice NYSE Composite U.S. 100 e o índice: (a) CCI_24; (b) AR_26; (c) BR_26

Fonte: Elaborada pelo autor no software SPSS®

5.2 MODELOS DE REGRESSÃO

Através da matriz de correlação (Tabela 10) entre as variáveis independentes, é possível perceber uma correlação alta (superior a 0,9) entre as variáveis BIAS_20 e ROC_10. Algumas correlações altas já eram esperadas devido à natureza do cálculo dos índices financeiros técnicos. Por exemplo, entre os índices MTM_10 e ROC_10, que consideram os resultados dos 10 dias anteriores. O índice K_9 apresentou uma sequência de correlações superiores a 0,75 com vários índices: BIAS_20, D_9, MTM_10, ROC_10 e CCI_24. O mesmo índice K_9 também apresentou correlação superior a 0,75, porém negativa, com o índice WMS%R_9.

Para obter um parâmetro de comparação com o modelo híbrido SOM-ANN/BP foram construídos os modelos de regressão com base em quatro métodos: Enter, Stepwise, Forward e Backward.

a) Método Enter

A regressão múltipla com método de inserção de variáveis do tipo Enter realizada pelo software SPSS® é processada através da apresentação simultânea de todas as variáveis independentes. Dessa forma, a inserção das variáveis ocorre em bloco.

Os coeficientes obtidos no processo de construção do modelo de regressão pelo método Enter estão demonstrados na Tabela 10. Todos os coeficientes apresentaram um erro padrão próximo de zero, assim como a constante β_0 . O teste t possui como hipótese nula a suposição de que os coeficientes obtidos para o modelo de regressão são iguais a zero e a hipótese não nula indica a possibilidade de que ao menos um dos coeficientes seja diferente de zero. Neste caso, o teste t revela que o índice CCI_24 possui o coeficiente igual a zero, visto que seu valor para o teste t é inferior à estatística teste ($t = 1,96$).

O modelo 2 apresentado na Tabela 10 desconsidera o índice CCI_24. Após sua retirada, o teste t indica também a baixa significância dos índices WMS%R_9 e BR_26. Dessa forma, o modelo 3 presente na Tabela 10 registra os coeficientes do modelo de regressão linear múltipla com a ausência dos índices CCI_24, WMS%R_9 e BR_26.

Tabela 10 - Análise dos coeficientes do modelo de regressão através do método Enter

Modelo		Coeficientes Unstandardized		t	Significância
		B	Erro padrão		
1	Constante	-0,122	0,003	-35,862	0,000
	MA_10	0,948	0,001	1.043,190	0,000
	BIAS_20	0,160	0,005	29,362	0,000
	WMS%R_9	-0,031	0,004	-8,031	0,000
	K_9	0,082	0,004	23,391	0,000
	D_9	-0,088	0,003	-31,197	0,000
	MTM_10	0,248	0,011	22,100	0,000
	ROC_10	-0,165	0,013	-12,352	0,000
	CCI_24	0,003	0,004	0,872	0,384
	AR_26	-0,012	0,002	-6,177	0,000
	BR_26	-0,007	0,002	-3,812	0,000
2	Constante	0,738	0,104	7,102	0,000
	MA_10	0,997	0,001	1.128,292	0,000
	BIAS_20	27,055	0,686	39,450	0,000
	WMS%R_9	-0,020	0,039	-0,509	0,610
	K_9	0,042	0,001	29,370	0,000
	D_9	-0,047	0,001	-37,869	0,000
	MTM_10	0,548	0,019	28,163	0,000
	ROC_10	-0,218	0,013	-17,179	0,000
	AR_26	-0,108	0,026	-4,190	0,000
	BR_26	-0,046	0,037	-1,253	0,210
	3	Constante	0,738	0,073	8,827
MA_10		0,997	0,001	1.133,277	0,000
BIAS_20		27,229	0,643	42,357	0,000
K_9		0,042	0,001	56,151	0,000
D_9		-0,047	0,001	-56,951	0,000
MTM_10		0,550	0,019	28,667	0,000
ROC_10		-0,221	0,012	-17,842	0,000
AR_26		-0,092	0,023	-4,056	0,000

Fonte: Elaborada pelo autor no software SPSS®

É possível expressar o modelo de regressão resultante após a retirada do índice CCI_24, ilustrado na equação 46.

$$Y_{REG} = 0,738 + 0,997MA_{10} + 27,229BIAS_{20} + 0,042K_9 - 0,047D_9 + 0,550MTM_{10} - 0,221ROC_{10} - 0,092AR_{26} \quad (46)$$

O teste para o coeficiente de determinação R^2 indica que o modelo é responsável por explicar 99,8% da variação dos dados com um erro padrão de 0,016 (Tabela 11).

Tabela 11 - Análise do coeficiente de determinação

Modelo	R	R^2	R^2 Ajustado	Erro padrão para a estimativa
1	0,999a	0,998	0,998	0,016

a. Preditores: Constante, BR_26, MA_10, ROC_10, WMS%R_9, AR_26, D_9, CCI_24, BIAS_20, K_9, MTM_10

b. Variável dependente: Valor do Fechamento

Fonte: Elaborada pelo autor no software SPSS®

A análise de variância ou Anova do modelo criado está apresentada na Tabela 12. O valor calculado para a estatística F foi igual a 120.417,057, com um nível de significância igual a 0,000. O resultado confirma que pelo menos um dos coeficientes do modelo é significativamente diferente de zero.

Tabela 12 - Análise de variância (ANOVA)

Modelo	Soma dos Quadrados	df	Mean Square	F	Significância
Regressão	309,328	10	30,933	120.417,057	0,000a
Residual	0,532	2.071	0,000		
Total	309,860	2.081			

a. Preditores: Constante, BR_26, MA_10, ROC_10, WMS%R_9, AR_26, D_9, CCI_24, BIAS_20, K_9, MTM_10

b. Variável dependente: Valor do Fechamento

Fonte: Elaborada pelo autor no software SPSS®

A análise dos resultados projetados pelo modelo de regressão é realizada através da Raiz do Erro Médio Quadrado (RMSE: *Root Mean Square Error*), Valor Médio dos Erros Absolutos (MAE: *Mean Absolute Error*) e pelo Percentual do Valor Médio dos Erros Absolutos (MAPE: *Mean Absolute Percentage Error*). Tais resultados estão descritos na Tabela 13.

Tabela 13 -Desempenho das projeções do modelo de regressão através do método *Enter*

Modelo	RMSE	MAE	MAPE
Resultados	0,3996	0,2695	0,4183 %

Fonte: Elaborada pelo autor

A Figura 34 demonstra a relação entre os valores observados e os valores projetados da variável Valor de Fechamento por intermédio de um gráfico de linha.

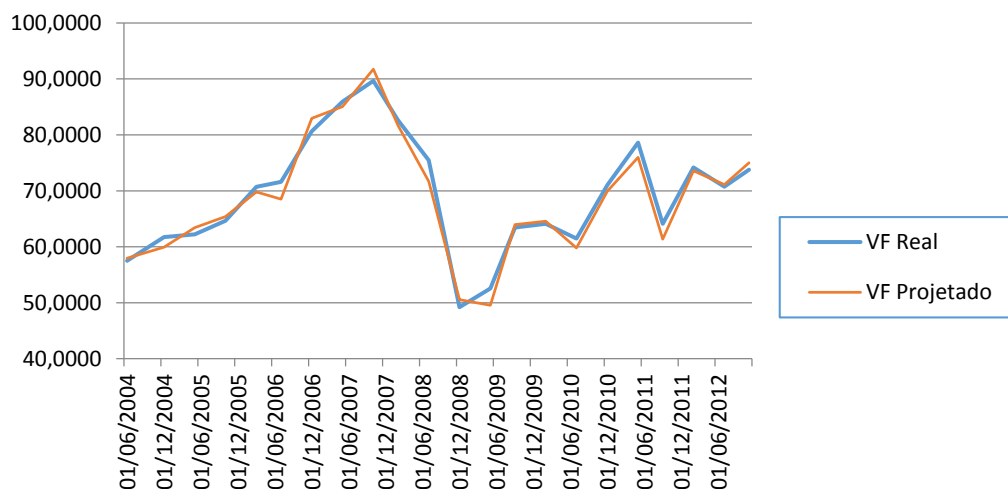


Figura 34 - Valor de fechamento real (VF Real) e projetado (VF Projetado) pelo modelo de regressão múltipla no período de 23/06/2004 a 08/11/2012
Fonte: Elaborada pelo autor no software Microsoft Excel®

b) Método Stepwise

O método de regressão por etapas, ou stepwise, consiste na adição ou remoção de variáveis em cada etapa. O critério utilizado é expresso com base no teste parcial F. O princípio do procedimento é a formação de um modelo baseado em apenas uma variável que possua a mais alta correlação com a variável dependente (MONTGOMERY; RUNGER, 2008). Da mesma forma, a variável selecionada inicialmente é a que apresenta o maior resultado para a estatística F. No modelo 1 construído, a variável dependente que apresentou a maior estatística F foi o índice MA_10 (Tabela 14).

A análise residual dos modelos construídos através do modelo de regressão pelo método Stepwise indica o modelo 10 como o de melhor desempenho.

Tabela 14 - Análise dos coeficientes do modelo de regressão através do método Stepwise

Modelo		Coeficientes		t	Significância
		Unstandardized			
		B	Erro padrão		
1	Constante	0,584	0,232	2,514	0,012
	MA_10	0,992	0,003	299,438	0,000
2	Constante	0,440	0,113	3,908	0,000
	MA_10	0,993	0,002	617,993	0,000
	BIAS_20	41,741	0,508	82,197	0,000
3	Constante	0,909	0,090		0,000
	MA_10	0,995	0,001	0,992	0,000
	BIAS_20	28,564	0,546	0,090	0,000
	WMS%R_9	-1,065	0,030	-0,061	0,000
4	Constante	0,814	0,083		0,000
	MA_10	0,996	0,001	0,992	0,000
	BIAS_20	15,642	0,826	0,049	0,000
	WMS%R_9	-0,969	0,028	-0,056	0,000
	MTM_10	0,211	0,011	0,050	0,000
5	Constante	1,530	0,080		0,000
	MA_10	0,998	0,001	0,994	0,000
	BIAS_20	18,222	0,748	0,057	0,000
	WMS%R_9	-1,054	0,025	-0,061	0,000
	MTM_10	0,288	0,010	0,069	0,000
	D_9	-0,014	0,001	-0,038	0,000
6	Constante	0,435	0,081		0,000
	MA_10	0,999	0,001	0,995	0,000
	BIAS_20	22,117	0,664	0,070	0,000
	WMS%R_9	-0,139	0,041	-0,008	0,001
	MTM_10	0,247	0,009	0,059	0,000
	D_9	-0,044	0,001	-0,120	0,000
	K_9	0,039	0,001	0,122	0,000
7	Constante	0,518	0,076		0,000
	MA_10	0,998	0,001	0,994	0,000
	BIAS_20	26,669	0,677	0,084	0,000
	WMS%R_9	-0,003	0,039	0,000	0,930
	MTM_10	0,545	0,019	0,130	0,000
	D_9	-0,048	0,001	-0,131	0,000
	K_9	0,042	0,001	0,132	0,000
	ROC_10	-0,214	0,013	-0,079	0,000

Fonte: Elaborada pelo autor no software SPSS®

Tabela 14 - Análise dos coeficientes do modelo de regressão através do método Stepwise (continuação)

Modelo	Coeficientes Unstandardized		t	Significância	
	B	Erro padrão			
8	Constante	0,515	0,066		0,000
	MA_10	0,998	0,001	0,994	0,000
	BIAS_20	26,690	0,631	0,084	0,000
	MTM_10	0,546	0,019	0,130	0,000
	D_9	-0,048	0,001	-0,131	0,000
	K_9	0,043	0,001	0,132	0,000
	ROC_10	-0,214	0,012	-0,080	0,000
9	Constante	0,642	0,073		0,000
	MA_10	0,997	0,001	0,993	0,000
	BIAS_20	27,229	0,643	0,086	0,000
	MTM_10	0,550	0,019	0,131	0,000
	D_9	-0,047	0,001	-0,129	0,000
	K_9	0,042	0,001	0,131	0,000
	ROC_10	-0,221	0,012	-0,082	0,000
AR_26	-0,092	0,023	-0,004	0,000	
10	Constante	0,720	0,075		0,000
	MA_10	0,997	0,001	0,993	0,000
	BIAS_20	25,750	0,742	0,081	0,000
	MTM_10	0,524	0,020	0,125	0,000
	D_9	-0,048	0,001	-0,130	0,000
	K_9	0,041	0,001	0,127	0,000
	ROC_10	-0,200	0,013	-0,074	0,000
	AR_26	-0,102	0,023	-0,004	0,000
CCI_24	0,001	0,000	0,009	0,000	

Fonte: Elaborada pelo autor no software SPSS®

Após 10 etapas e a inclusão de oito variáveis (MA_10, BIAS_20, MTM_10, D_9, K_9, ROC_10, AR_26 e BR_26), o modelo de regressão através do método stepwise resultante é descrito na equação 47. Cabe ressaltar a falta de representatividade do índice CCI_24, de apenas 0,1%.

$$Y_{REG} = 0,72 + 0,997MA_{10} + 25,75BIAS_{20} + 0,524MTM_{10} - 0,048D_9 + 0,041K_9 - 0,2ROC_{10} - 0,102AR_{26} + 0,001CCI_{24} \quad (47)$$

A análise de variância dos modelos indica através dos resultados para a significância dos modelos obtidos que em cada modelo resultante existe ao menos um dos coeficientes diferente de zero (Tabela 15).

Tabela 15 - Análise de variância (ANOVA)

	Modelo	Soma dos Quadrados	df	Mean Square	F	Significância
1	Regressão	214.632,151	1	214.632,151	89.663,069	0,000a
	Residual	4.974,240	2.078	2,394		
	Total	219.606,391	2.079			
2	Regressão	218.436,787	2	109.218,394	193.951,696	0,000b
	Residual	1.169,604	2.077	0,563		
	Total	219.606,391	2.079			
3	Regressão	218.879,607	3	72.959,869	208.403,913	0,000c
	Residual	726,784	2.076	0,350		
	Total	219.606,391	2.079			
4	Regressão	218.993,790	4	54.748,448	185.443,776	0,000d
	Residual	612,601	2.075	0,295		
	Total	219.606,391	2.079			
5	Regressão	219.115,646	5	43.823,129	185.206,489	0,000e
	Residual	490,745	2.074	0,237		
	Total	219.606,391	2.079			
6	Regressão	219.239,501	6	36.539,917	206.457,826	0,000f
	Residual	366,890	2.073	0,177		
	Total	219.606,391	2.079			
7	Regressão	219.284,520	7	31.326,360	201.659,405	0,000g
	Residual	321,871	2.072	0,155		
	Total	219.606,391	2.079			
8	Regressão	219.284,519	6	36.547,420	235381,980	0,000h
	Residual	321,872	2.073	0,155		
	Total	219.606,391	2.079			
9	Regressão	219.287,055	7	31326,722	203.262,111	0,000i
	Residual	319,336	2.072	0,154		
	Total	219.606,391	2.079			
10	Regressão	219.289,442	8	27411,180	179.109,681	0,000j
	Residual	316,949	2.071	0,153		
	Total	219.606,391	2.079			

a. Predictors: (Constant), MA_10

b. Predictors: (Constant), MA_10, BIAS_20

c. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9

d. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10

e. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9

f. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9, K_9

g. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9, K_9, ROC_10

h. Predictors: (Constant), MA_10, BIAS_20, MTM_10, D_9, K_9, ROC_10

i. Predictors: (Constant), MA_10, BIAS_20, MTM_10, D_9, K_9, ROC_10, AR_26

j. Predictors: (Constant), MA_10, BIAS_20, MTM_10, D_9, K_9, ROC_10, AR_26, CCI_24

k. DependentVariable: Fechamento

Fonte: Elaborada pelo autor no software SPSS®

Através da análise do RMSE, do MAE e do MAPE de cada modelo obtido pela regressão múltipla stepwise, é possível perceber a melhora gradual na performance do modelo resultante em relação aos modelos parciais através da redução do MAPE (Tabela 16).

Tabela 16 -Desempenho das projeções do modelo de regressão através do método Stepwise

Modelo	RMSE	MAE	MAPE
1	1,6594	1,3072	1,9597 %
2	1,2964	0,6923	1,0451 %
3	1,2619	0,9818	1,4567 %
4	1,0312	0,8536	1,2663 %
5	1,6581	1,5542	2,2941 %
6	0,7673	0,5058	0,7579 %
7	0,8547	0,5470	0,8179 %
8	0,8224	0,4926	0,7399 %
9	0,3938	0,2716	0,4207 %
10	0,3905	0,2679	0,4146 %

Fonte: Elaborada pelo autor

O modelo 10 apresenta o menor valor para os índices de erro analisados, obtendo um MAPE de apenas 0,4146%, demonstrando uma redução interessante quando comparado ao modelo 1, que obteve um MAPE de 1,9597%. É possível perceber variações e picos locais até a convergência do modelo.

c) Método Forward

O método de processamento de regressão múltipla através do algoritmo forward também é denominado como método de seleção progressiva de variáveis. O algoritmo forward é considerado como uma variação da regressão stepwise, na qual se baseia na ideia de que os regressores devem ser inseridos no modelo um de cada vez, até que não existam mais candidatos que produzam aumento significativo na soma quadrática da regressão. Dessa forma, as variáveis independentes são adicionadas gradualmente à medida que seu valor parcial de F seja superior ao F_{IN} .

Após 7 etapas e a exclusão de três variáveis (CCI_24, WMS%R_9 e BR_26), o modelo de regressão através do método forward resultante é descrito na equação 48.

$$Y_{REG} = 0,72 + 0,998MA_{10} + 26,669BIAS_{20} + 0,545MTM_{10} - 0,048D_9 + 0,042K_9 - 0,214ROC_{10} - 0,105AR_{26} \quad (48)$$

A Tabela 17 ilustra os resultados parciais do processo de regressão múltipla através do modelo forward, ou de seleção progressiva de variáveis independentes.

Tabela 17 - Análise dos coeficientes do modelo de regressão através do método Forward

Modelo		Coeficientes Unstandardized		t	Significância
		B	Erro padrão		
1	Constante	0,584	0,232		0,012
	MA_10	0,992	0,003	0,989	0,000
2	Constante	0,440	0,113		0,000
	MA_10	0,993	0,002	0,990	0,000
	BIAS_20	41,741	0,508	0,132	0,000
3	Constante	0,909	0,090		0,000
	MA_10	0,995	0,001	0,992	0,000
	BIAS_20	28,564	0,546	0,090	0,000
	WMS%R_9	-1,065	0,030	-0,061	0,000
4	Constante	0,814	0,083		0,000
	MA_10	0,996	0,001	0,992	0,000
	BIAS_20	15,642	0,826	0,049	0,000
	WMS%R_9	-0,969	0,028	-0,056	0,000
	MTM_10	0,211	0,011	0,050	0,000
5	Constante	1,530	0,080		0,000
	MA_10	0,998	0,001	0,994	0,000
	BIAS_20	18,222	0,748	0,057	0,000
	WMS%R_9	-1,054	0,025	-0,061	0,001
	MTM_10	0,288	0,010	0,069	0,000
	D_9	-0,014	0,001	-0,038	0,000
6	Constante	0,435	0,081		0,000
	MA_10	0,999	0,001	0,995	0,000
	BIAS_20	22,117	0,664	0,070	0,000
	WMS%R_9	-0,139	0,041	-0,008	0,656
	MTM_10	0,247	0,009	0,059	0,000
	D_9	-0,044	0,001	-0,120	0,000
	K_9	0,039	0,001	0,122	0,000
7	Constante	0,720	0,076		0,000
	MA_10	0,998	0,001	0,994	0,000
	BIAS_20	26,669	0,677	0,084	0,000
	MTM_10	0,545	0,039	0,000	0,000
	D_9	-0,048	0,019	0,130	0,000
	K_9	0,042	0,001	-0,131	0,000
	ROC_10	-0,214	0,001	0,132	0,000
	AR_26	-0,105	0,013	-0,079	0,000

Fonte: Elaborada pelo autor no software SPSS®

A Tabela 18 demonstra os resultados da análise de variância para os modelos parciais do processamento do modelo de regressão pelo método forward.

Tabela 18 - Análise de variância (ANOVA)

	Modelo	Soma dos Quadrados	df	Mean Square	F	Significância
1	Regressão	214.632,151	1	214.632,151	89.663,069	0,000a
	Residual	4.974,240	2.078	2,394		
	Total	219.606,391	2.079			
2	Regressão	218.436,787	2	109.218,394	193.951,696	0,000b
	Residual	1.169,604	2.077	0,563		
	Total	219.606,391	2.079			
3	Regressão	218.879,607	3	72.959,869	208.403,913	0,000c
	Residual	726,784	2.076	0,350		
	Total	219.606,391	2.079			
4	Regressão	218.993,790	4	54.748,448	185443,776	0,000d
	Residual	612,601	2.075	0,295		
	Total	219.606,391	2.079			
5	Regressão	219.115,646	5	43.823,129	185.206,489	0,000e
	Residual	490,745	2.074	0,237		
	Total	219.606,391	2.079			
6	Regressão	219.239,501	6	36.539,917	206.457,826	0,000f
	Residual	366,890	2.073	0,177		
	Total	219.606,391	2.079			
7	Regressão	219.284,520	7	31.326,360	201.659,405	,000g
	Residual	321,871	2.072	0,155		
	Total	219.606,391	2.079			

a. Predictors: (Constant), MA_10

b. Predictors: (Constant), MA_10, BIAS_20

c. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9

d. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10

e. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9

f. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9, K_9

g. Predictors: (Constant), MA_10, BIAS_20, WMS%R_9, MTM_10, D_9, K_9, ROC_10

h. Dependent Variable: Fechamento

Fonte: Elaborada pelo autor no software SPSS®

Foram analisados o RMSE, o MAE e o MAPE de cada modelo obtido pela regressão múltipla forward (Tabela 19). É possível perceber uma melhora significativa

na performance do modelo resultante em relação aos modelos parciais através da redução expressiva do MAPE.

Tabela 19 -Desempenho das projeções do modelo de regressão através do método Forward

Modelo	RMSE	MAE	MAPE
1	1,5465	1,1275	1,7113 %
2	0,7502	0,5451	0,8363 %
3	0,5915	0,4126	0,6379 %
4	0,5427	0,3875	0,5938 %
5	0,4859	0,3384	0,5194 %
6	0,4200	0,2861	0,4426 %
7	0,3939	0,2688	0,4164 %

Fonte: Elaborada pelo autor

d) Método Backward

Também conhecido como método de regressão múltipla por eliminação regressiva, este algoritmo inicia o processamento considerando todas as variáveis independentes. Assim, a variável independente que possuir a menor estatística parcial F é removida, desde que que essa estatística F seja insignificante, ou seja, atenda a condição de eliminação $F < F_{out}$.

A partir da retirada de uma das variáveis, o modelo com n-1 variáveis é ajustado e a próxima variável independente, também denominada como regressor, é encontrada e eliminada. O algoritmo se encerra quando a condição de eliminação não for verificada para nenhum dos regressores resultantes.

A Tabela 20 indica os resultados parciais obtidos na construção do modelo de regressão múltipla através do método de eliminação regressiva ou backward. Destaca-se que foram necessários apenas 3 passos para a obtenção do modelo de regressão, onde o modelo ideal fez a eliminação dos índices financeiros técnicos WMS%R_9 e BR_26.

Tabela 20 - Análise dos coeficientes do modelo de regressão através do método Backward

Modelo	Coeficientes Unstandardized		t	Significância	
	B	Erro padrão			
1	(Constant)	0,781	0,104	7,498	0,000
	MA_10	0,997	0,001	1.131,693	0,000
	BIAS_20	25,748	0,759	33,905	0,000
	WMS%R_9	0,015	0,040	,378	0,706
	K_9	0,042	0,001	29,430	0,000
	D_9	-0,048	0,001	-37,832	0,000
	MTM_10	0,523	0,020	25,693	0,000
	ROC_10	-0,199	0,013	-14,763	0,000
	CCI_24	0,001	0,000	3,946	0,000
	AR_26	-0,116	0,026	-4,491	0,000
	BR_26	-0,046	0,037	-1,248	0,212
2	(Constant)	0,796	0,096	8,286	0,000
	MA_10	0,997	0,001	1.132,104	0,000
	BIAS_20	25,690	0,743	34,558	0,000
	K_9	0,041	0,001	51,410	0,000
	D_9	-0,048	0,001	-56,515	0,000
	MTM_10	0,522	0,020	25,728	0,000
	ROC_10	-0,199	0,013	-14,804	0,000
	CCI_24	0,001	0,000	3,962	0,000
	AR_26	-0,117	0,026	-4,538	0,000
	BR_26	-0,047	0,037	-1,269	0,205
3	(Constant)	0,720	0,075	9,584	0,000
	MA_10	0,997	0,001	1.136,775	0,000
	BIAS_20	25,750	0,742	34,704	0,000
	K_9	0,041	0,001	51,390	0,000
	D_9	-0,048	0,001	-57,222	0,000
	MTM_10	0,524	0,020	25,854	0,000
	ROC_10	-0,200	0,013	-14,998	0,000
	CCI_24	0,001	0,000	3,950	0,000
	AR_26	-0,102	0,023	-4,460	0,000

Fonte: Elaborada pelo autor no software SPSS®

Após 3 etapas e a eliminação de duas variáveis, o modelo de regressão através do método backward resultante é descrito na equação 49.

$$Y_{REG} = 0,72 + 0,967MA_{10} + 25,75BIAS_{20} + 0,041K_9 - 0,048D_9 + 0,524MTM_{10} - 0,2ROC_{10} + 0,001CCI_{24} - 0,102AR_{26} \quad (49)$$

A Tabela 21 demonstra os resultados da análise de variância para os modelos parciais do processamento do modelo de regressão pelo método *backward*.

Tabela 21 - Análise de variância (ANOVA)

	Modelo	Soma dos Quadrados	df	Mean Square	F	Significância
1	Regressão	219.289,711	10	21.928,971	143.270,749	0,000a
	Residual	316,680	2.069	0,153		
	Total	219.606,391	2.079			
2	Regressão	219.289,689	9	24.365,521	159.255,653	0,000b
	Residual	316,702	2.070	0,153		
	Total	219.606,391	2079			
3	Regressão	219.289,442	8	27.411,180	179.109,681	0,000c
	Residual	316,949	2.071	0,153		
	Total	219.606,391	2.079			

a. Predictors: (Constant), BR_26, MA_10, WMS%R_9, AR_26, ROC_10, D_9, CCI_24, BIAS_20, K_9, MTM_10

b. Predictors: (Constant), BR_26, MA_10, AR_26, ROC_10, D_9, CCI_24, BIAS_20, K_9, MTM_10

c. Predictors: (Constant), MA_10, AR_26, ROC_10, D_9, CCI_24, BIAS_20, K_9, MTM_10

d. DependentVariable: Fechamento

Fonte: Elaborada pelo autor no software SPSS®

Para efeito de comparação, foram analisados o RMSE, o MAE e o MAPE de cada modelo obtido pela regressão múltipla *backward*. É possível perceber uma sensível melhora na performance do modelo resultante em relação aos modelos parciais através da redução do MAE e do MAPE (Tabela 22).

Tabela 22 - Desempenho das projeções do modelo de regressão através do método Backward

Modelo	RMSE	MAE	MAPE
1	0,3905	0,2796	0,4322 %
2	0,3907	0,2686	0,4159 %
3	0,3975	0,2680	0,4146 %

Fonte: Elaborada pelo autor

6. COMPARATIVO ENTRE OS MODELOS

A predição de um índice resultante da bolsa de valores é uma tarefa desafiadora, visto que se caracteriza por apresentar alta volatilidade, complexidade e turbulência através do tempo. A análise fundamentalista, a análise técnica e o estudo de séries temporais formam o conjunto de técnicas frequentemente utilizadas na análise de metodologias de previsão de resultados de índices de bolsas de valores. Neste estudo, um mapa autoorganizável e uma rede *backpropagation* com múltiplas camadas foram usadas para desenvolver uma abordagem integrada denominada SOM-ANN/BP para realizar a previsão de séries temporais financeiras, mais especificamente o índice NYSE Composite U.S. 100.

A fim de distinguir o modelo de previsão de índice financeiro NYSE Composite U.S. 100 mais eficiente, é interessante comparar os coeficientes RMSE, MAE e MAPE. Tais coeficientes têm sido de grande utilidade na avaliação de desempenho de modelos de previsão baseados em redes neurais artificiais. Para obter um comparativo uniforme, os mesmos coeficientes foram usados para a avaliação de performance dos modelos multivariados.

Os resultados obtidos pelo modelo híbrido SOM-ANN/BP e pelos modelos de regressão (Enter, Stepwise, Forward e Backward) estão condensados na Tabela 23. Os resultados referentes à regressão multivariada foram obtidos pelos modelos com melhor desempenho em cada método. Desta forma, os resultados para o modelo de regressão pelo método stepwise fazem referência ao modelo obtido na 10ª etapa de processamento do algoritmo. O melhor modelo obtido a partir do método forward é referente ao obtido após 7 etapas e o modelo resultante após o processamento pelo método backward é o modelo 3.

Tabela 23 - Comparativo dos resultados dos modelos SOM-ANN/BP e de regressão

	Modelo	RMSE	MAE	MAPE
SOM – ANN/BP	Cluster 1	0,7074	0,3942	0,1837 %
	Cluster 2	0,3222	0,1571	0,2140 %
	Cluster 3	0,2939	0,1410	0,2075 %
	Cluster 4	0,3568	0,2594	0,1543%
	Cluster 5	0,2519	0,1480	0,1937 %
	Cluster 6	0,3845	0,2745	0,2562 %
	Cluster 7	0,6718	0,3531	0,2355 %
REGRESSÃO	Método Enter	0,3996	0,2695	0,4183 %
	Método Stepwise (Mod. 10)	0,3905	0,2679	0,4146 %
	Método Forward (Mod. 7)	0,3939	0,2688	0,4164 %
	Método Backward (Mod. 3)	0,3975	0,2680	0,4146 %

Fonte: Elaborado pelo autor

Os resultados obtidos pelos modelos de regressão múltipla registraram a ausência de duas a três das variáveis elencadas para a inserção na camada de entrada do modelo híbrido SOM-ANN/BP. As variáveis desprezadas por alguns dos modelos de regressão oscilaram entre BR_26, WMS%R_9, CCI_24 e ROC_10. O modelo de regressão baseado no método Enter apresentou um MAPE que pode ser considerado alto quando comparado com os outros métodos de regressão, no valor de 6,0622 %. Em relação aos modelos de regressão através dos métodos Stepwise, Forward e Backward, o coeficiente MAPE não apresentou grandes discrepâncias, visto que orbitaram em torno de 0,415%.

Na comparação do desempenho do modelo híbrido SOM-ANN/BP com os modelos de regressão baseados nos métodos Stepwise, Forward e Backward, o modelo neural mostrou uma performance suficientemente superior (MAPE=0,2518%) para afirmar que pode ser considerado como uma ferramenta viável e eficiente para a previsão de índices financeiros. Os baixos valores de MAPE refletem a pequena variação entre as formas gráficas.

7. CONSIDERAÇÕES FINAIS

Este estudo foi desenvolvido com o propósito de construir um modelo híbrido baseado em redes neurais artificiais para a previsão do índice NYSE Composite U.S. 100, que se apresente como uma ferramenta que auxilie o processo de tomada de decisão de investidores financeiros, resultando em vantagens em termos de versatilidade e precisão, quando comparado a métodos estatísticos multivariados. As redes neurais artificiais são apontados por diversos pesquisadores como ferramenta de maior eficiência na previsão de séries temporais financeiras. A comparação resultante entre o modelo híbrido neural e os modelos estatísticos de regressão múltipla evidenciou a alta capacidade de modelagem dos dados com grande ruído. Por outro lado, o modelo de regressão através do método Enter demonstrou uma baixa capacidade de processamento na presença de ruído, ao contrário dos modelos de regressão pelos métodos Stepwise, Forward e Backward. O modelo Enter apresentou tal limitação por incorporar todas as informações referentes às variáveis independentes, inclusive as irrelevantes aos resultados, reduzindo sua precisão. A superioridade do modelo híbrido baseado em redes neurais artificiais deve-se à capacidade de generalização e à imunidade aos ruídos presentes nas séries temporais financeiras (CHEN, 2012).

A principal característica dos modelos neurais é a versatilidade na modelagem de séries temporais com diversos cenários. Desta forma, a mudança de cenário ocasiona apenas a necessidade de um novo processo de treinamento dos pesos sinápticos, ao passo que um modelo estatístico exigiria uma reformulação de todo o processo. Novos cenários que incluam índices financeiros diferenciados requerem somente um novo processamento dos pesos sinápticos.

O presente estudo enfocou as características sugeridas por Haykin (2001) para a concepção e o bom desempenho das redes neurais, a saber: determinação das variáveis mais adequadas ao problema em questão; a topologia da rede; e o processo de treinamento da rede quanto à minimização do erro de aprendizagem. Quanto a tais fatores, a implementação do modelo obteve resultados deveras satisfatórios, ressaltando a influência das variáveis utilizadas na projeção do resultado do valor de fechamento do dia subsequente. O índice MA_10 demonstrou grande influência nos resultados projetados, assim como os índices BIAS_20, MTM_10 e

ROC_10. A partir dos resultados obtidos pelos modelos multivariados, é possível concluir que alguns índices como o WMS%R_9 e CCI_24 não possuem influência significativa.

No que tange à topologia da rede, as arquiteturas que se apresentaram mais estáveis envolvem a composição de 10 e 15 neurônios na única camada intermediária. A topologia da rede demonstra uma sensibilidade alta em relação ao número de neurônios nas camadas, sob pena de não haver a convergência do aprendizado. A dimensionalidade da amostra é outro fator fundamental para a convergência. Amostras pequenas podem não ser suficientes, ao passo que amostras demasiadamente grandes podem incorrer no aprendizado de padrões específicos, com a perda da capacidade de generalização da rede.

Para o presente estudo, procurou-se utilizar uma abordagem híbrida neural para elaborar um modelo de previsão do índice NYSE Composite U.S. 100. Entretanto, inúmeras questões podem ser aperfeiçoadas na tentativa de otimizar a pesquisa. Algumas possibilidades estão descritas a seguir:

- No modelo SOM-ANN/BP foram utilizadas uma rede do tipo autoorganizável para a clusterização dos dados e uma rede do tipo *backpropagation* para a modelagem dos dados. Uma alternativa viável para a otimização dos resultados obtidos é o uso de uma rede neural do tipo GRNN (*General Regression Neural Network*) ao invés da rede *backpropagation*. A principal vantagem do uso da GRNN é a necessidade de uma amostra inferior para o treinamento dos pesos sinápticos com uma performance superior.

- Uma alternativa a ser explorada é a aplicação de meta-heurísticas de busca e otimização de soluções, como os algoritmos genéticos, após o processo de clusterização pelos mapas autoorganizáveis. Os algoritmos genéticos têm sido de grande utilidade na modelagem de séries temporais em diversas áreas, como contabilidade, economia e administração.

- Não foi verificado o efeito de variáveis macroeconômicas de forma conjunta aos índices financeiros técnicos. O uso de tais variáveis pode ser considerado um caminho a ser investigado com mais propriedade.

Para a avaliação do modelo SOM-ANN/BP utilizou-se uma análise comparativa com quatro métodos de regressão multivariada. Cada método foi responsável pela criação de um modelo de regressão específico com a presença e a ausência de índices financeiros técnicos conforme sua lógica de processamento.

Entre as limitações encontradas no estudo está o grande número de índices financeiros técnicos disponíveis no mercado. Para o modelo SOM-ANN/BP foram usados 10 índices, dos quais alguns se mostraram insignificantes para os modelos de previsão multivariados.

A maior contribuição deste trabalho foi demonstrar que os modelos de previsão de séries temporais financeiras baseados em algoritmos de redes neurais são alternativas mais eficientes que os métodos estatísticos multivariados tradicionais. Através de comparações, este estudo evidenciou as vantagens no uso das redes neurais em detrimento aos modelos estatísticos.

REFERÊNCIAS

ARMANO, G.; MARCHESI, M.; MURRU, A.. A hybrid genetic-neural architecture for stock indexes forecasting. **Information Sciences**, v.170, p. 3-33, 2005.

ARENALES, M.N.; ARMENTANO, V.; MORABITO, R.; YANASSE, H.. **Pesquisa operacional para cursos de engenharia**. Rio de Janeiro: Elsevier, 2007.

ESPÍNOLA, R.; MATÉ, C.. Different approaches to forecast interval time series: a comparison in finance. **Computational Economics**, n. 37, p. 169-191, 2011.

BELUCO, A.. Base de dados NYSE Composite U.S. 100. Disponível em: <<http://www.beluco.com.br/eaufrgs/index1.html>>. Acesso em: 10 fev.2013a.

BELUCO, A.. Base de dados NYSE Composite U.S. 100: Índices financeiros técnicos. Disponível em: <<http://www.beluco.com.br/eaufrgs/index2.html>>. Acesso em: 12 fev.2013b.

BELUCO, A.. Base de dados NYSE Composite U.S. 100: Índices financeiros normalizados. Disponível em: <<http://www.beluco.com.br/eaufrgs/index3.html>>. Acesso em: 15 fev.2013c.

BILDIRICI, M.; ERSIN, O. O.. Improving forecasts of GARCH family models with the artificial neural networks: an application to the daily returns in Istanbul stock exchange. **Expert Systems with Applications**, v. 36, n.4, p. 7355-7362, 2009.

CELIK, A. E.; KARATEPE, Y.. Evaluating and forecasting banking crises through neural network models: an application for Turkish banking sector. **Expert Systems with Applications**, v. 33, n.4, p. 809-815, 2007.

CHEN, A.S.; LEUNG, M. T.. Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan stock index. **Computers and Operations Research**, v.30, p. 901-923, 2003.

CHEN, J-H. Developing SFNN models to predict financial distress of construction companies. **Expert Systems with Applications**, v.39, n.1, p. 823-827, 2012.

HSU, C-M.. A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. **Expert Systems with Applications**, v.38, n.11, p. 14026-14036, 2011.

DAI, W.; WU, J-Y.; LU, C-J.. Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. **Expert Systems with Applications**, v.39, n.4, p. 4444-4452, 2012.

ENKE, D.; THAWORNWONG, S.. The use of data mining and neural networks for forecasting stock market returns. **Expert Systems with Applications**, v. 29, n.4, p. 927-940, 2005.

GHIASSI, M.; SAIDANE, H.; ZIMBRA, D. K.. A dynamic artificial neural network model for forecasting time series events. **International Journal of Forecasting**, n.21, p. 341-362, 2005.

GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U.. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, v. 38, n.8, p. 10389-10397, 2011.

HASSAN, M. R.; NATH, B.; KIRLEY, M.. A fusion model of HMM, ANN and GA for stock market forecasting. **Expert Systems with Applications**, v. 33, n.1, p. 171-180, 2009.

HAYKIN, S.. **Redes neurais: Princípios e prática**. 2ª ed. Porto Alegre: Bookman, 2001.

HSU, C-M.. A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. **Expert Systems with Applications**, v.38, n.11, p. 14026-14036, 2011.

HUANG, C.L.; TSAI, C.Y.. A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. **Expert Systems with Applications**, v. 36, n.2, p. 1529-1539, 2009.

HULME, D.; XU, S.. Application of genetic algorithm to the optimization of neural network configuration for stock market forecasting. **Lecture Notes in Artificial Intelligence**, v. 2256, p. 285-296, 2001.

INCE, H.; TRAFALIS, T.B.. Kernel principal component analysis and support vector machines for stock price prediction. **IEEE International Joint Conference on Neural Networks**, p. 2053-2058, 2004.

KAASTRA, I.; BOYD, M.. Designing a neural network for forecasting financial and economic time series. **Neurocomputing**, n. 10, p. 215-236, 1996.

KARA, Y.; BOYACIOGLU, M. A.; BAYKAN, O. K.. Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange. **Expert Systems with Applications**, v.38, n.5, p. 5311-5319, 2011.

KASKI, S.. Data exploration using self-organizing maps. **Mathematics, Computing and management in Engineering Series**, n. 82, 1997.

KIM, J.; WON, C.; BAE, J. K.. A knowledge integration model for the prediction of corporate dividends. **Expert Systems with Applications**, v.37, n.2, p. 1344-1350, 2010.

KOHONEN, T.. Essentials of the self-organizing map. **Neural Network**, n. 37, p. 52-65, 2013.

KOHONEN, T.. Self-organized formation of topologically correct feature maps. **Biological Cybernetics**, n. 43, p. 59-69, 1982.

KOHONEN, T.. **Self-organizing maps**. 3^a ed. Berlin-Heidelberg: Springer, 2001.

LAI, R.K.; FAN, C.Y.; HUANG, W.H.; CHANG, P.C.. Evolving and clustering fuzzy decision tree for financial time series data forecasting. **Expert Systems with Applications**, v. 36, n.2, p. 3761-3773, 2009.

LEE, M-C.. Using support vector machine with a hybrid feature selection method to the stock trend prediction. **Expert Systems with Applications**, v.36, n.8, p. 10896-10904, 2009.

LI, S.T.; KUO, S. C.. Knowledge discovery in financial investment for forecasting and trading strategy through wavelet-based SOM networks. **Expert Systems with Applications**, v. 34, n.2, p. 935-951, 2008.

LIANG, J.; SONG, W.; WANG, M.. Stock price prediction based on procedural neural networks. **Advances in Artificial Neural Systems**, v. 2011, p. 1-11, 2011.

LI, X.; YANG, Z.; SONG, Z.. Short-term stock price prediction based on echo state networks. **Expert Systems with Applications**, v. 36, n.3, p. 7313-7317, 2009.

LIU, F.; WANG, J.. Fluctuation prediction of stock market index by Legendre neural network with random time strength function. **Neurocomputing**, n.83, p. 12-21, 2012.

LU, C-J; WU, J-Y.. An efficient CMAC neural network for stock index forecasting. **Expert Systems with Applications**, v.38, n.12, p. 15194-15201, 2011.

MAJHI, R.; PANDA, G.; SAHOO, G.. Efficient prediction of exchange rates with low complexity artificial neural network models. **Expert Systems with Applications**, v. 36, n.1, p. 181-189, 2009.

MONTGOMERY, D. C.; RUNGER, G. C.. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC, 2008.

MOSTAFA, M. M.. Forecasting stock exchange movements using neural networks: Empirical evidence from Kuwait. **Expert Systems with Applications**, v.37, n.9, p. 6302-6309, 2010.

NYSE Composite. New York Stock Exchange (Base de dados dos índices NYSE). Disponível em: <<http://nyse.nyx.com>>. Acesso em: 10 outubro.2012

RAMEZANIAN, M. R.. Combination neural network and financial indices for stock price prediction. **Journal of Applied Sciences**, v. 11, n. 19, p. 3429-3435, 2011.

RUMELHART, D. E., Mc CLELLAND, J. L.. **Parallel distributed processing**. Cambridge: MIT Press, 1986.

SETYAWATI, B.R.; CREESE, R.C.; SAHIRMAN, S.. Neural network for cost estimation. **AACE International Transactions**, v.14, p.1-10, 2003.

SMITH, K. A.; GUPTA, J. N. D.. Neural networks in business: techniques and applications for the operations researcher. **Computers & Operations Research**, v. 27, p. 1023-1044, 2000.

THAVORNWONG, S.; ENKE, D.. The adaptive selection of financial and economic variables for use with artificial neural networks. **Neurocomputing**, v. 56, p. 205-232, 2004.

TSAI, C-F.; HSIAO, Y-C.. Combining multiple feature selection methods for stock prediction: Union, intersection and multi-intersection approaches. **Decision Support Systems**, n.50, p. 258-269, 2010.

WANG, J-Z; WANG, J-J; ZHANG, Z-G; GUO, S-P.. Forecasting stock indices with back propagation neural network. **Expert Systems with Applications**, v.38, n.11, p.14346-14355, 2011.

ZHU, X.; WANG, H.; XU, L.; LI, H.. Predicting stock index increments by neural networks: the role of trading volume under different horizons. **Expert Systems with Applications**, v. 34, n.4, p. 3043-3054, 2008.

ANEXO A – EMPRESAS LISTADAS NO ÍNDICE NYSE COMPOSITE U.S. 100

O NYSE Composite Index foi concebido para medir o desempenho de todas as ações ordinárias listadas na NYSE, consistindo em mais de 2.000 ações norte-americanas e estrangeiras. É uma medida das alterações no valor total de mercado de todas as ações listadas na NYSE, ajustadas para eliminar os efeitos das mudanças de capitalização.

Empresa	Código	Participação no índice (%)	Ramo de atividade
General Electric	GE	2,13	Industrial
Exxon Mobil	XOM	2,05	Energia
Citigroup	C	1,45	Financeiro
BP	BP	1,33	Energia
Procter & Gamble	PG	1,14	Consumo Não-cíclico
Bank of America	BAC	1,07	Financeiro
HSBC Holdings	HBC	1,05	Financeiro
Johnson & Johnson	JNJ	1,03	Saúde
Pfizer	PFE	1,00	Saúde
Toyota Motor	TM	0,92	Consumo cíclico
American Int Group	AIG	0,89	Financeiro
Altria Group	MO	0,89	Consumo Não-cíclico
Glaxo Smith Kline	GSK	0,84	Saúde
Total S.A.	TOT	0,83	Energia
JP Morgan Chase	JPM	0,80	Financeiro
Mitsubishi UFJ Finantial	MTU	0,79	Financeiro
Vodafone Group	VOD	0,76	Telecomunicações
Int Business Machines	IBM	0,75	Tecnologia
Novartis	NVS	0,75	Saúde
Chevron	CVX	0,73	Energia
Royal Dutch Shell	RDS.A	0,70	Energia
Wal-Mart Stores	WMT	0,69	Consumo cíclico
Wells Fargo	WFC	0,60	Financeiro
PepsiCo	PEP	0,56	Consumo Não-cíclico
AT&T	T	0,55	Telecomunicações
UBS	UBS	0,54	Financeiro
Sanofi-Aventis	SNY	0,54	Saúde
Coca-Cola	KO	0,51	Consumo Não-cíclico
Royal Dutch Shell	RDS.B	0,51	Energia
Home Depot	HD	0,50	Consumo cíclico
United Health	UNH	0,49	Saúde
Verizon Communications	VZ	0,48	Telecomunicações

Empresa	Código	Participação no índice (%)	Ramo de atividade
Banco Santander Hispano	STD	0,47	Financeiro
Wachovia	WB	0,47	Financeiro
Hewlett-Packard	HPQ	0,47	Tecnologia
Nokia	NOK	0,47	Tecnologia
Astra Zeneca	AZN	0,44	Saúde
Time Warner	TWX	0,44	Consumo cíclico
Conoco Phillips	COP	0,42	Energia
E.ON AG	EON	0,41	Utilidades
Siemens	SI	0,41	Industrial
ENI S.p.A.	E	0,41	Energia
Berkshire Hathaway Inc.	BRK.A	0,40	Financeiro
Merck	MRK	0,40	Saúde
Medtronic	MDT	0,40	Saúde
ING Group	ING	0,40	Financeiro
Barclays	BCS	0,39	Financeiro
Telefonica S.A.	TEF	0,38	Telecomunicações
Sprint Nextel	S	0,38	Telecomunicações
Credit Suisse	CSR	0,36	Financeiro
Wyeth	WYE	0,35	Saúde
Abbott Laboratories	ABT	0,35	Saúde
Banco Bilbao Vizcaya	BBV	0,35	Financeiro
Allianz	AZ	0,35	Financeiro
BHP Billiton	BHP	0,34	Materiais Básicos
Tyco International	TYC	0,34	Industrial
Schlumberger	SLB	0,33	Energia
Merrill Lynch	MER	0,33	Financeiro
American Express	AXP	0,33	Financeiro
Morgan Stanley	MWD	0,32	Financeiro
Motorola	MOT	0,32	Tecnologia
Ely Lilly	LLY	0,31	Saúde
Boeing	BA	0,29	Industrial
Royal Bank of Canada	RY	0,29	Financeiro
Deutsche Bank	DB	0,29	Financeiro
Bell South	BLS	0,28	Tecnologia
ABN AMRO Holding	ABN	0,28	Financeiro
Goldman Sachs	GS	0,28	Financeiro
Rio Tinto	RTP	0,28	Materiais Básicos
Canon	CAJ	0,28	Tecnologia
Daimler Chrysler	DCX	0,28	Consumo cíclico
Walt Disney	DIS	0,28	Consumo cíclico
United Parcel Service	UPS	0,27	Industrial
Lloyds TSB Group	LYG	0,27	Financeiro
Fannie Mae	FNM	0,27	Financeiro
Lowe's	LOW	0,27	Consumo cíclico
Manulife Financial	MFC	0,27	Financeiro
Viacom Inc. CI B	VIA.B	0,26	Consumo cíclico
Walgreen	WAG	0,26	Consumo cíclico
Freddie Mac	FRE	0,26	Financeiro
Bristol-Myers Squibb	BMJ	0,26	Saúde

Empresa	Código	Participação no índice (%)	Ramo de atividade
Deutsche Telekom	DT	0,25	Telecomunicações
Target	TGT	0,25	Consumo cíclico
Matsushita Electric Industrial	MC	0,25	Consumo cíclico
Genentech	DNA	0,25	Saúde
McDonald's	MCD	0,25	Consumo cíclico
Washington Mutual	WM	0,24	Financeiro
E.I, DuPont de Nemours	DD	0,24	Materiais básicos
Dow Chemical	DOW	0,24	Materiais básicos
Diageo	DEO	0,24	Consumo Não-cíclico
France Telecom	FTE	0,24	Telecomunicações
BASF	BF	0,23	Materiais básicos
Sony	SNE	0,23	Consumo cíclico

Fonte: Adaptada de NYSE Composite (2012)

APÊNDICE A – GRÁFICOS DE EVOLUÇÃO DAS VARIÁVEIS DA BASE DE DADOS

Volume de negócios do NYSE Composite U.S. 100

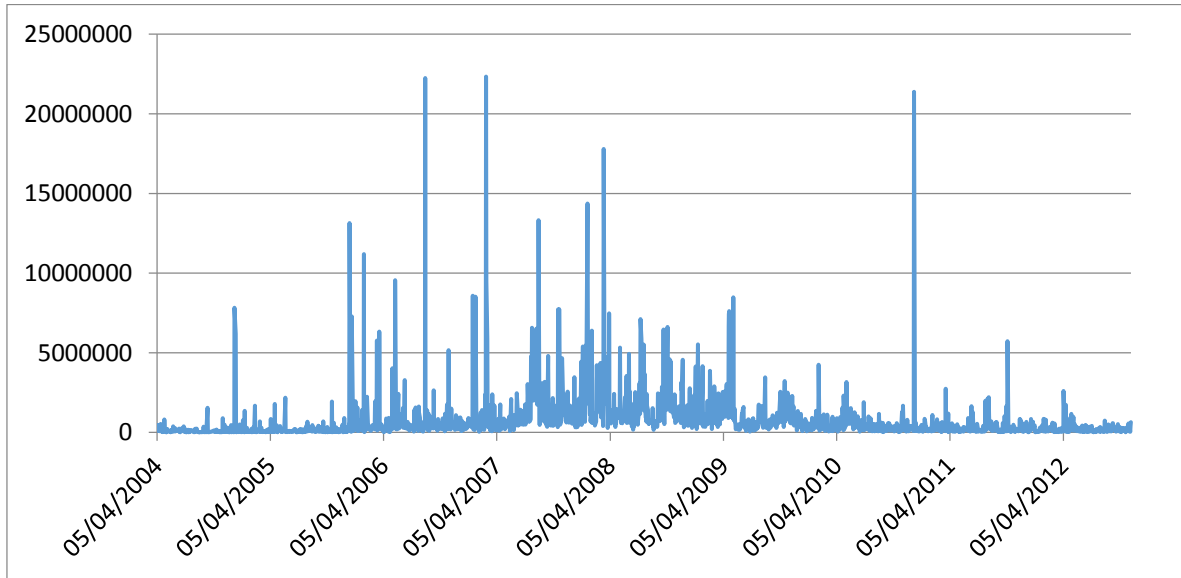


Figura 35 - Gráfico da evolução da série histórica do volume de negócios do NYSE Composite U.S. 100 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

Valor de abertura do NYSE Composite U.S. 100

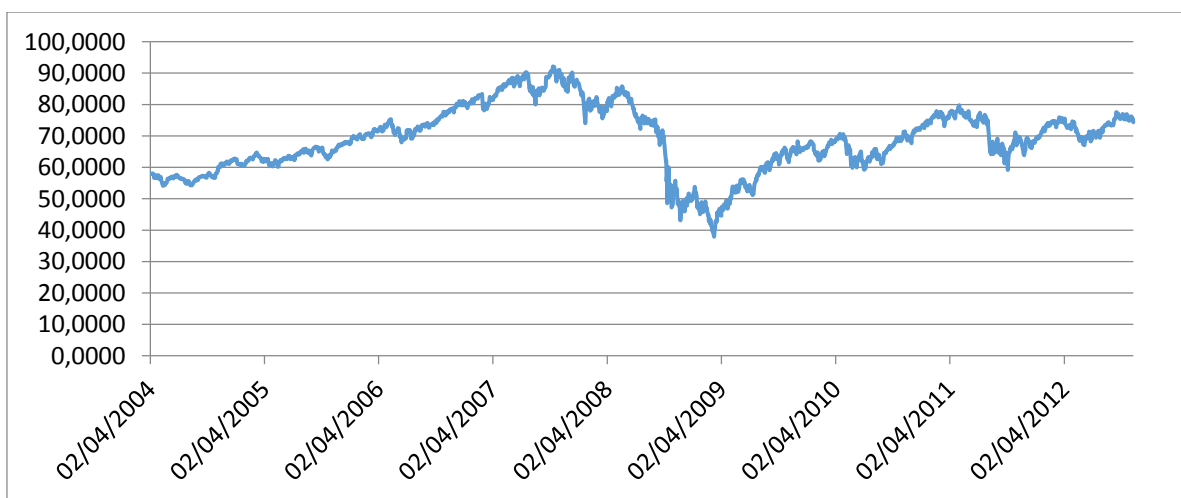


Figura 36 - Gráfico da evolução da série histórica do valor de abertura do índice NYSE Composite U.S. 100 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

Valor de fechamento do NYSE Composite U.S. 100

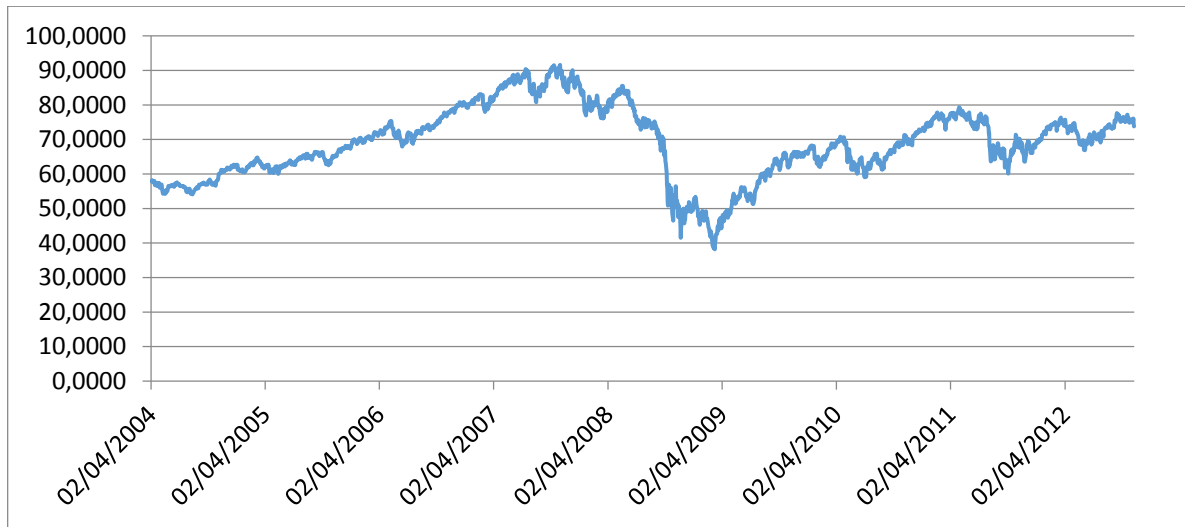


Figura 37 - Gráfico da evolução da série histórica do valor de fechamento do índice NYSE Composite U.S. 100 no período entre 05/04/2004 e 08/11/2012.
Fonte: Elaborada pelo autor

Valor mínimo do NYSE Composite U.S. 100

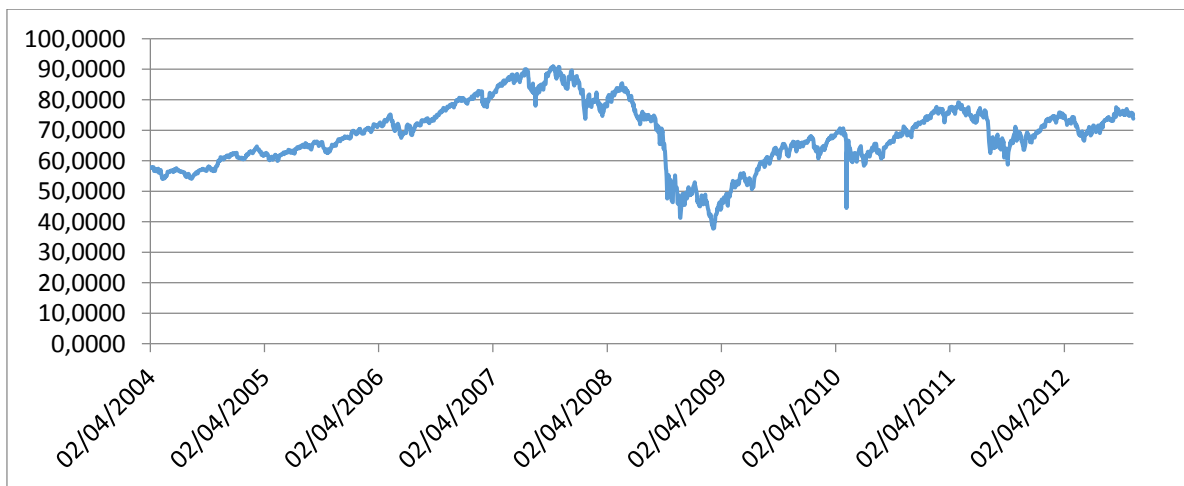


Figura 38 - Gráfico da evolução da série histórica do valor mínimo do índice NYSE Composite U.S. 100 no período entre 05/04/2004 e 08/11/2012.
Fonte: Elaborada pelo autor

Valor máximo do NYSE Composite U.S. 100

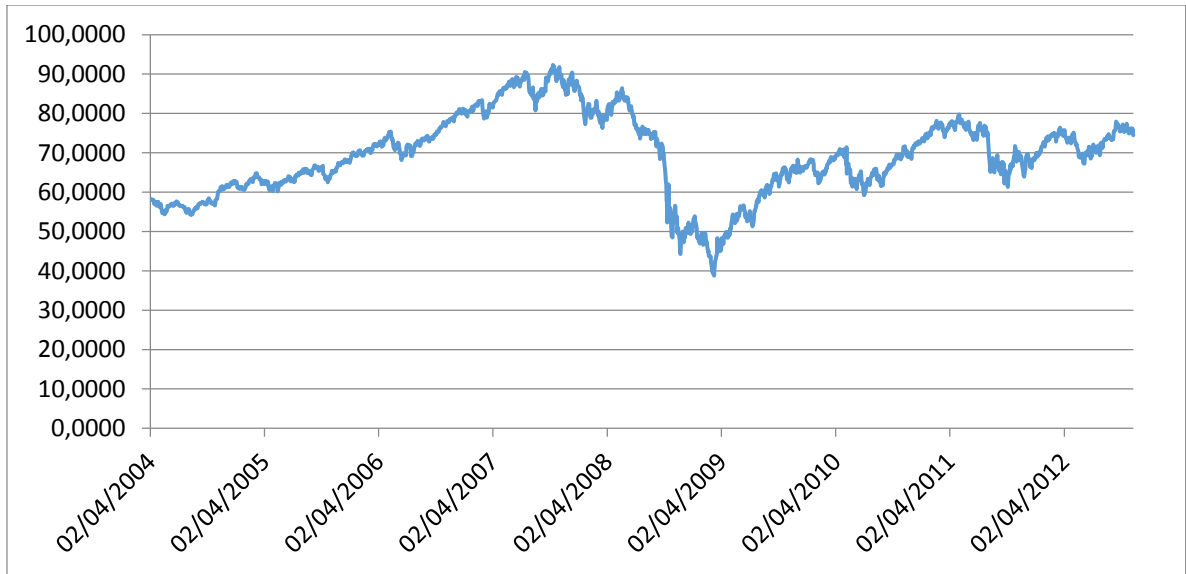


Figura 39 - Gráfico da evolução da série histórica do valor máximo do índice NYSE Composite U.S. 100 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

APÊNDICE B – GRÁFICOS DE EVOLUÇÃO DOS ÍNDICES FINANCEIROS TÉCNICOS CALCULADOS COM BASE NO NYSE COMPOSITE U.S. 100

MA_10

Gráfico da evolução do índice técnico MA_10 no período entre 05/04/2004 e 08/11/2012:

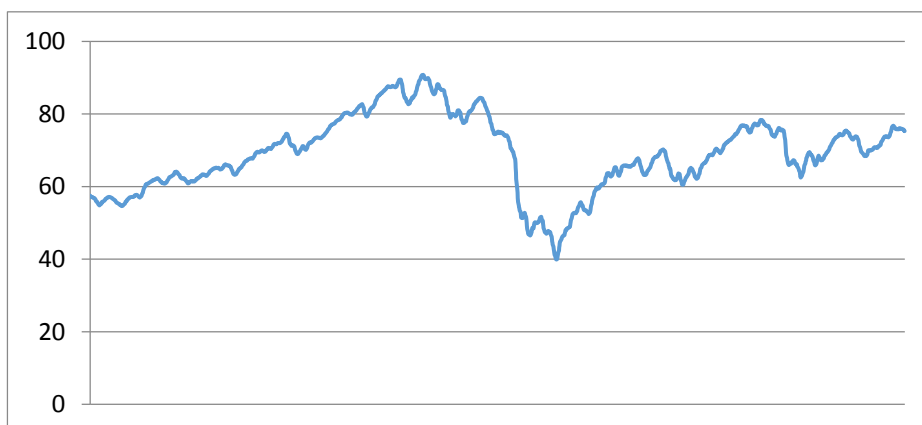


Figura 40 - Gráfico da evolução da série histórica do índice MA_10 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

BIAS_20

Gráfico da evolução do índice técnico BIAS_20 no período entre 05/04/2004 e 08/11/2012:

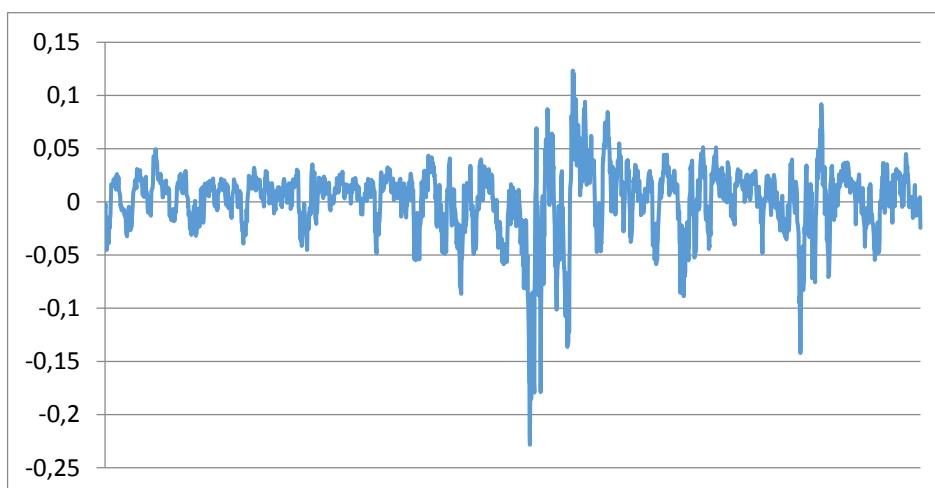


Figura 41 - Gráfico da evolução da série histórica do índice BIAS_20 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

WMS%R_9

Gráfico da evolução do índice técnico WMS%R_9 no período entre 05/04/2004 e 08/11/2012:

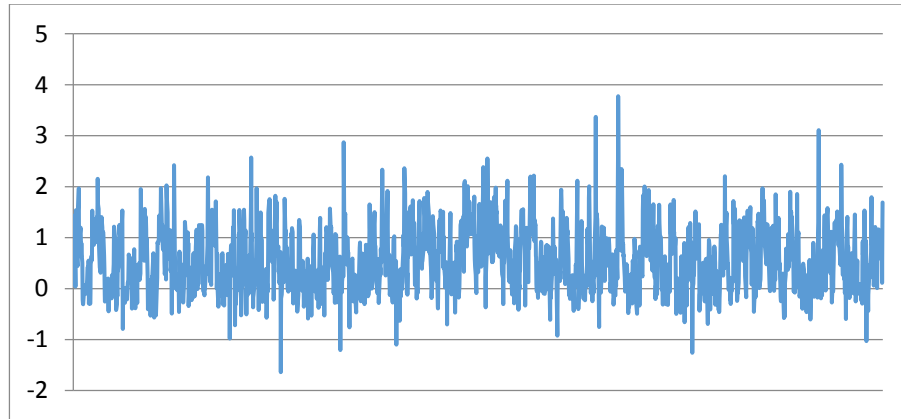


Figura 42 - Gráfico da evolução da série histórica do índice WMS%R_9 no período entre 05/04/2004 e 08/11/2012.
Fonte: Elaborada pelo autor

K_9

Gráfico da evolução do índice técnico K_9 no período entre 05/04/2004 e 08/11/2012:

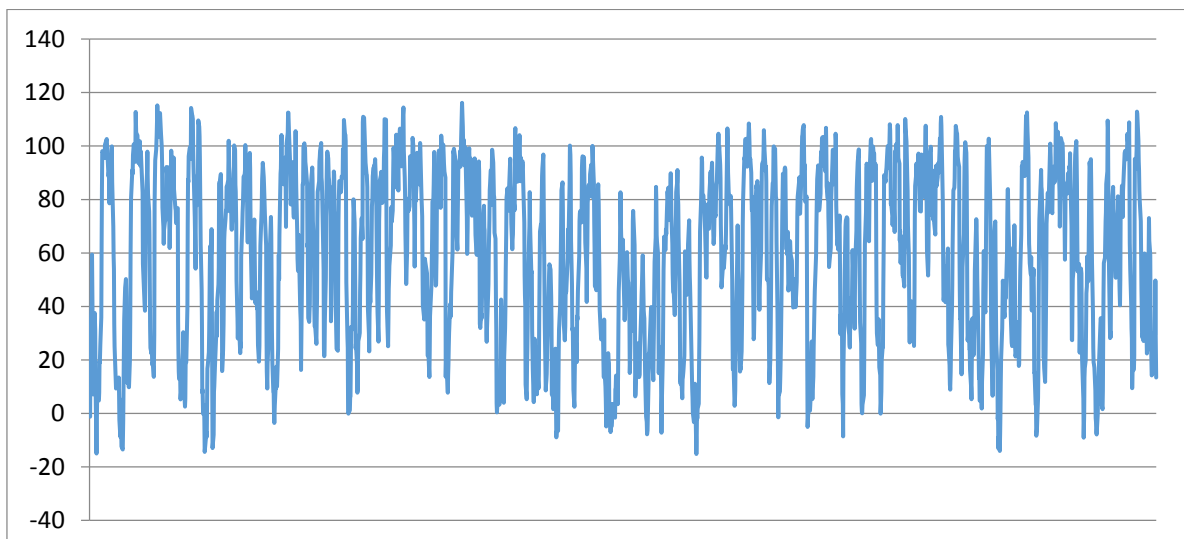


Figura 43 - Gráfico da evolução da série histórica do índice K_9 no período entre 05/04/2004 e 08/11/2012.
Fonte: Elaborada pelo autor

D_9

Gráfico da evolução do índice técnico D_9 no período entre 05/04/2004 e 08/11/2012:

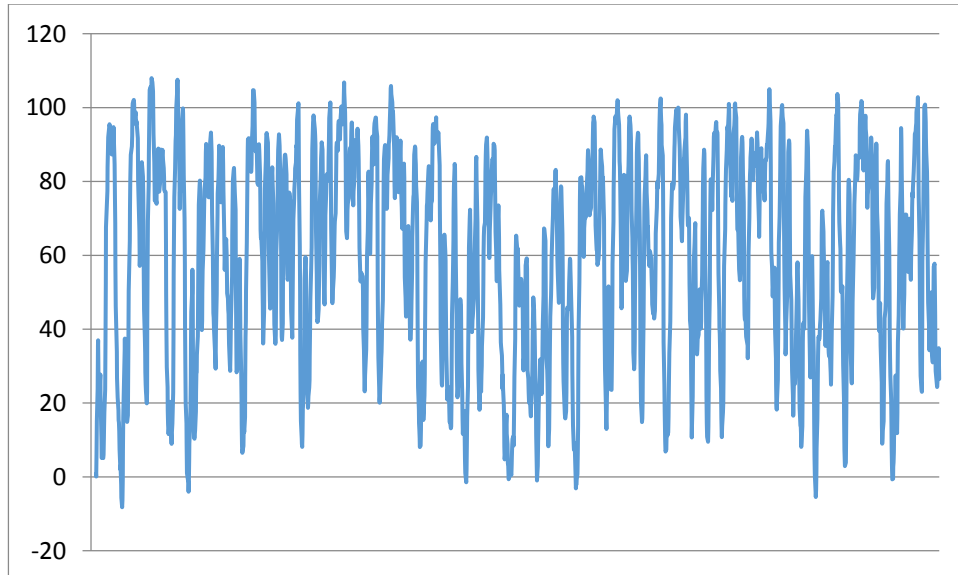


Figura 44 - Gráfico da evolução da série histórica do índice D_9 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

MTM_10

Gráfico da evolução do índice técnico MTM_10 no período entre 05/04/2004 e 08/11/2012:

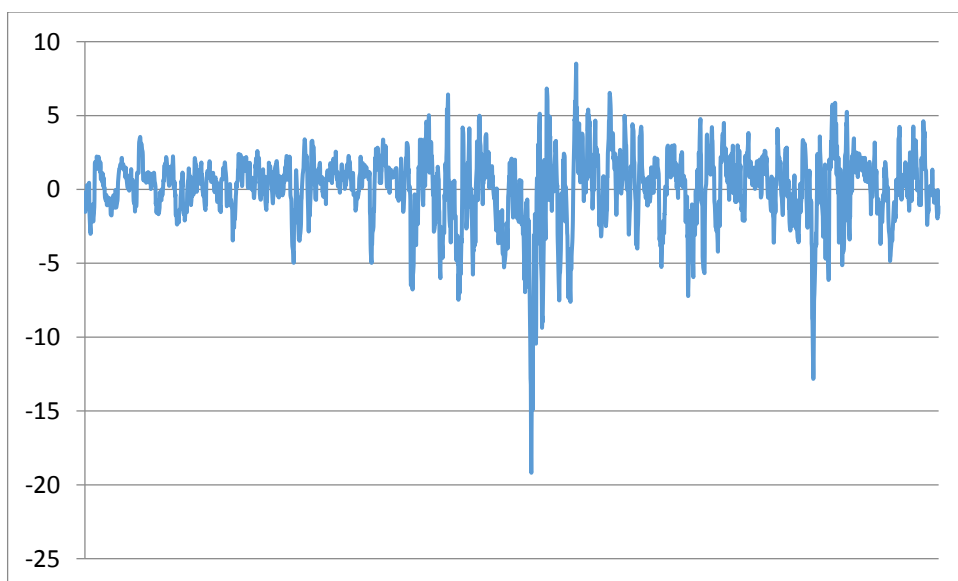


Figura 45 - Gráfico da evolução da série histórica do índice MTM_10 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

ROC_10

Gráfico da evolução do índice técnico ROC_10 no período entre 05/04/2004 e 08/11/2012:

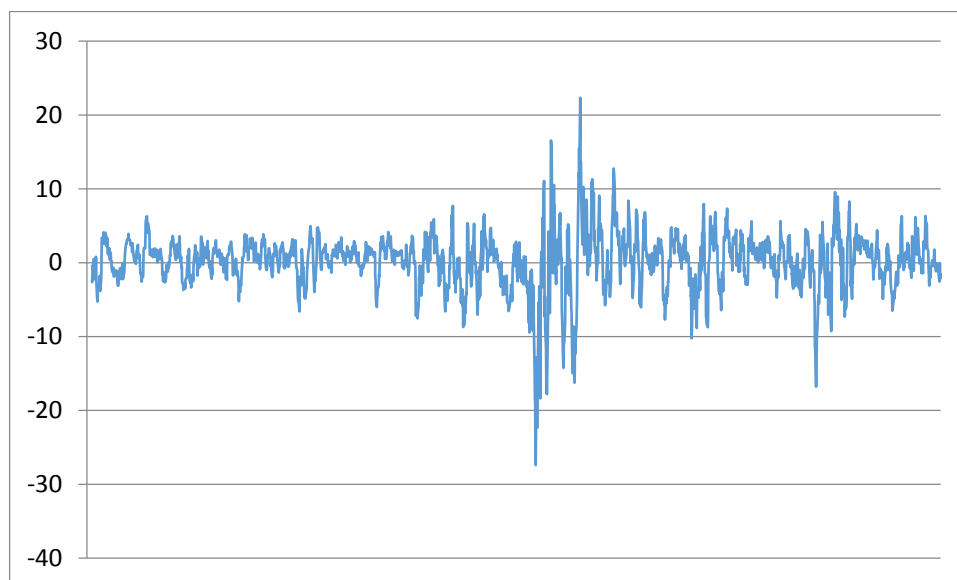


Figura 46 - Gráfico da evolução da série histórica do índice ROC_10 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

CCI_24

Gráfico da evolução do índice técnico CCI_24 no período entre 05/04/2004 e 08/11/2012:

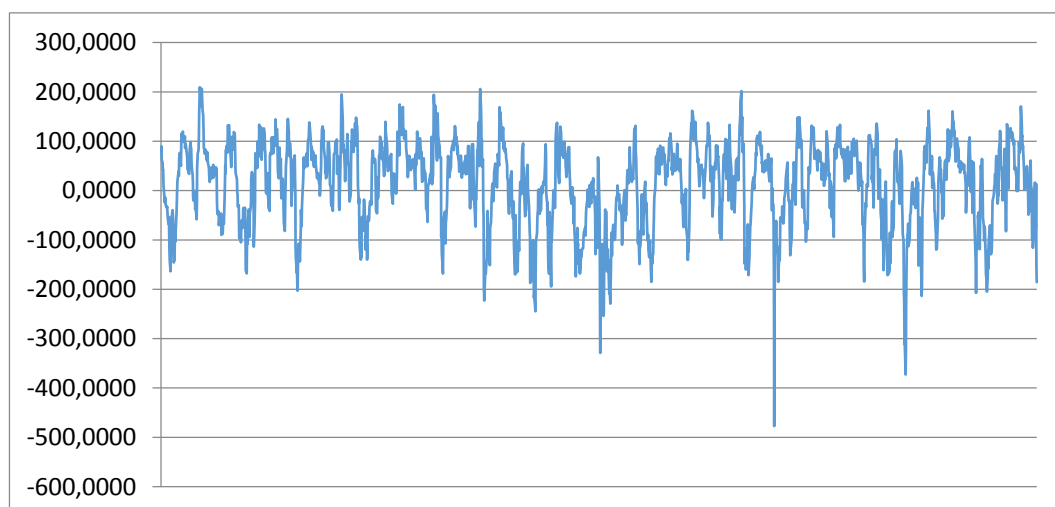


Figura 47 - Gráfico da evolução da série histórica do índice CCI_24 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

AR_26

Gráfico da evolução do índice técnico AR_26 no período entre 05/04/2004 e 08/11/2012:

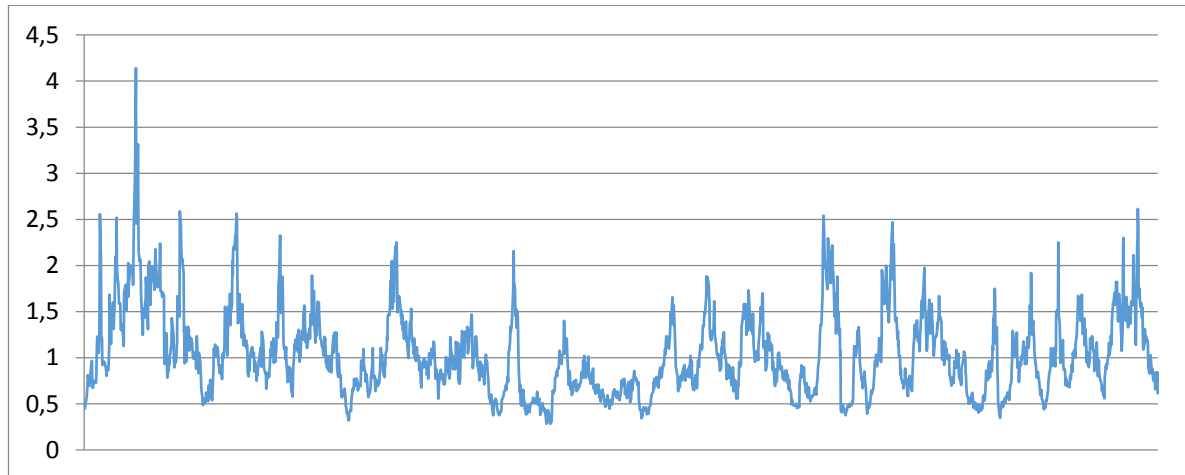


Figura 48 - Gráfico da evolução da série histórica do índice AR_26 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor

BR_26

Gráfico da evolução do índice técnico BR_26 no período entre 05/04/2004 e 08/11/2012:

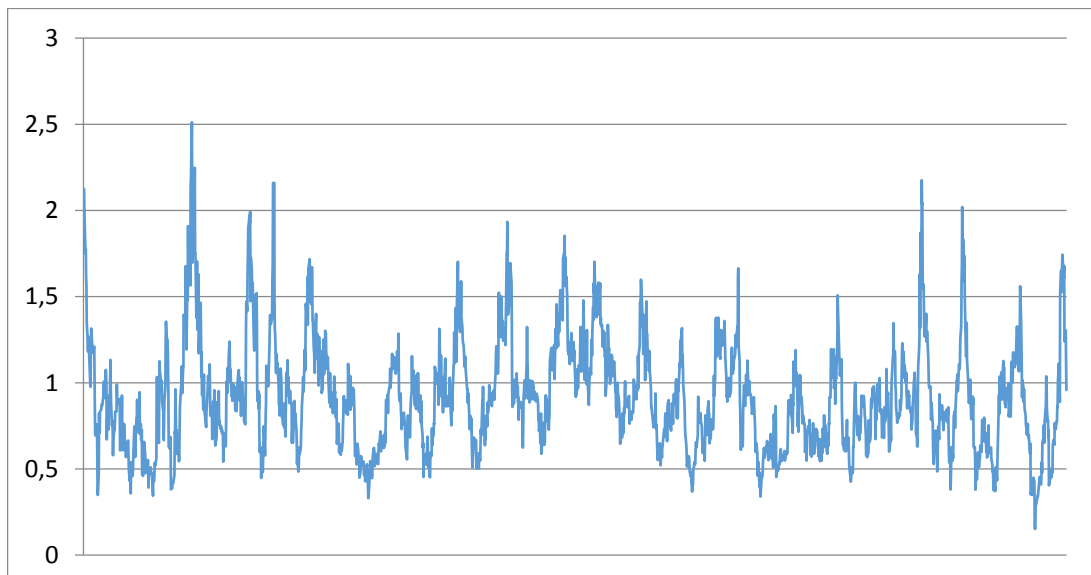


Figura 49 - Gráfico da evolução da série histórica do índice BR_26 no período entre 05/04/2004 e 08/11/2012.

Fonte: Elaborada pelo autor