

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

VINÍCIUS GUILHERME MÜLLER

PREVISÃO DO ÍNDICE BOVESPA UTILIZANDO REDES NEURASIS

Porto Alegre

2018

VINÍCIUS GUILHERME MÜLLER

PREVISÃO DO ÍNDICE BOVESPA UTILIZANDO REDES NEURAIAS

Monografia apresentada para a
obtenção do grau de Bacharel em
Engenharia Elétrica na Universidade
Federal do Rio Grande do Sul.

ORIENTADOR: PROF. DR. ALEXANDRE BALBINOT

Porto Alegre

2018

VINÍCIUS GUILHERME MÜLLER

PREVISÃO DO ÍNDICE BOVESPA UTILIZANDO REDES NEURAIAS

Esta monografia foi analisada e julgada adequada para a obtenção do título de Bacharel em Engenharia Elétrica e aprovada em sua forma final pelo Orientador e pela Banca Examinadora proposta pelo Orientador.

Orientador: Prof. Dr. Alexandre Balbinot

Coordenador: Prof. Dr. Alberto Bastos do Canto Filho

Aprovado em: ___/___/___

BANCA EXAMINADORA:

Prof. Dr. Alexandre Balbinot – UFRGS – DELET _____

Profa. Dra. Leia B. Bagesteiro – UFRGS – DELET _____

Me. Vinícius Cene – UFRGS – PPGEE _____

RESUMO

A previsão do mercado de ações tem sido um tema de grande interesse no campo das finanças, da engenharia e da matemática devido ao seu potencial de ganho financeiro. No entanto, a previsão do preço das ações é altamente complexa, pois as séries financeiras possuem altos níveis de ruído, são dinâmicas, não-lineares e caóticas por natureza. Além disso existem muitos fatores como eventos políticos, condições econômicas, expectativas dos investidores e outros fatores ambientais que podem influenciar no preço das ações. O objetivo deste estudo foi desenvolver uma Rede Neural *Feedforward* para prever as tendências do Índice Bovespa em um horizonte de tempo de uma semana. Foram calculadas 220 diferentes características consistindo em indicadores técnicos, volatilidade e nas cotações do dólar, do petróleo, do ouro, assim como das ações que compõem o índice. As características foram ranqueadas utilizando-se como modelo um *Gradient Boosting* de Árvores de Decisão em um algoritmo *Recursive Feature Elimination*. Foram implementadas diferentes redes possuindo diferentes números de neurônios e quantidades variadas das características mais bem classificadas. Foi possível obter uma taxa de acertos de 58,63% para o conjunto de validação e 58,13% para o conjunto de teste para a rede utilizando as 40 características mais bem classificadas pelo *ranking*. Utilizando-se todas as 220 entradas pré-selecionadas a rede obteve uma taxa de acertos de 51,39%, evidenciando a importância da seleção de características para esse problema.

Palavras-chave: Previsão do Mercado Financeiro. Redes Neurais. Séries Financeiras. *Recursive Feature Elimination*. *Gradient Boosting*.

ABSTRACT

Stock market forecasting has been a subject of great interest in the field of finance, engineering and mathematics because of its potential for financial gain. However, stock price prediction is highly complex since financial series have high levels of noise, are dynamic, non-linear and chaotic by nature. In addition, there are many factors such as political events, economic conditions, investor expectations and other environmental factors that may influence stock prices. The objective of this study was to develop a Feedforward Neural Network to predict trends of the Bovespa Index over a one-week time horizon. Different characteristics were calculated, consisting of technical indicators, volatility and the dollar, oil and gold prices, as well as the stocks that make up the index. The features were ranked using a Gradient Boosting of Decision Trees in a Recursive Feature Elimination algorithm. Different networks with different numbers of neurons and varying amounts of the best classified features were implemented. It was possible to obtain an accuracy rate of 58.63% for the validation set and 58.13% for the test set for the network using the 40 best classified features in the ranking. Using all the 220 pre-selected features, the network obtained an accuracy rate of 51.39%, evidencing the importance of the feature selection on this problem.

Keywords: Stock Market forecasting. Neural Networks. Financial Time Series. Recursive Feature Elimination. Gradient Boosting.

LISTA DE ILUSTRAÇÕES

Figura 1 - Bandas de Bollinger	14
Figura 2 - Estrutura do neurônio artificial	15
Figura 3 - Representação de uma rede neural MLP	17
Figura 4 - Tipos de tarefas	20
Figura 5 - Árvores de Decisão.....	21
Figura 6 - Índice Bovespa e taxa de retorno de 5 dias	29
Figura 7 - Histograma da taxa de retorno.....	29
Figura 8 - Cálculo das características	31
Figura 9 - Média Móvel Simples (20 dias) do Índice Bovespa.....	32
Figura 10 - Média Móvel Simples (20 dias) do Índice Bovespa.....	33
Figura 11 - Taxa de Variação (10 dias)	33
Figura 12 - Volume e razão entre médias móveis	34
Figura 13 - Desvio padrão (20 dias)	35
Figura 14 - Relação entre os preços máximo e mínimo (10 dias)	36
Figura 15 - Relação entre os preços de fechamento e de abertura (10 dias)	37
Figura 16 - Relação entre os preços de abertura e de fechamento do período anterior (10 dias)	37
Figura 17 - Taxa de Variação (10 dias) da cotação do Dólar/Real.....	38
Figura 18 - Taxa de Variação (10 dias) da cotação do Petróleo	38
Figura 19 - Taxa de Variação (10 dias) da cotação do Ouro.....	39
Figura 20 - Algoritmo de ranqueamento das características	40
Figura 21 - Número de neurônios dos modelos em cada camada	42
Figura 22 - Validação cruzada k folds	43
Figura 23 - Taxa de acertos dos conjuntos de treinamento e de validação	44
Figura 24 - Custo dos conjuntos de treinamento e de validação.....	44
Figura 25 - Taxa de acertos do conjunto de validação.....	48
Figura 26 - Taxa de acertos do conjunto de testes	50
Figura 27 - Taxa de acertos para os dados de teste do algoritmo XGBoost	51

LISTA DE TABELAS

Tabela 1 – Funções de ativação e suas características.....	17
Tabela 2 – Divisão de classes.....	30
Tabela 3 – Ranking de características.....	46
Tabela 4 – Taxa de acertos de validação com 10 características.....	47
Tabela 5 - Taxa de acertos de validação com 20 características.....	47
Tabela 6 - Taxa de acertos de validação com 30 características.....	47
Tabela 7 – Taxa de acertos de validação com 40 características.....	48
Tabela 8 – Taxa de acertos de validação com 50 características.....	48
Tabela 9 – Taxa de acertos de validação com 60 características.....	48
Tabela 10 – Taxa de acertos de teste com 10 características.....	49
Tabela 11 – Taxa de acertos de teste com 20 características.....	49
Tabela 12 – Taxa de acertos de teste com 30 características.....	49
Tabela 13 – Taxa de acertos de teste com 40 características.....	49
Tabela 14 – Taxa de acertos de teste com 50 características.....	49
Tabela 15 – Taxa de acertos de teste com 60 características.....	49
Tabela 16 – Matriz de confusão da rede.....	50
Tabela 17 – Taxa de acertos de teste com 60 características.....	51

SUMÁRIO

1 INTRODUÇÃO	7
2 FUNDAMENTAÇÃO TEÓRICA	10
2.1 O MERCADO DE AÇÕES.....	10
2.1.1 A Hipótese do Mercado Eficiente (HME)	11
2.1.2 A Análise Técnica	12
2.2 AS REDES NEURAS ARTIFICIAIS	15
2.3 ÁRVORES DE DECISÃO E O <i>GRADIENT BOOSTING</i>	20
2.4 MÉTRICAS DE DESEMPENHO	23
2.5 TRABALHOS RELACIONADOS.....	24
3 METODOLOGIA	28
3.1 BASE DE DADOS E SUA DIVISÃO.....	28
3.2 VARIÁVEL DE PREVISÃO.....	28
3.3 CARACTERÍSTICAS DE PREVISÃO	30
3.3.1 Cálculo das características	31
3.3.2 Ranqueamento das características	39
3.4 MODELO PROPOSTO.....	41
4 RESULTADOS E DISCUSSÕES	46
5 CONCLUSÃO	52
6 PROPOSTAS PARA TRABALHOS FUTUROS	54
REFERÊNCIAS BIBLIOGRÁFICAS	55
ANEXO - Ações incluídas nas características	59

1 INTRODUÇÃO

A previsão do mercado financeiro tem sido um tema de grande interesse no campo das finanças, da engenharia e da matemática devido ao seu potencial de ganho financeiro (YOO *et al.*, 2005). No entanto, a previsão do preço das ações é altamente complexa, pois existem muitos fatores como eventos políticos, condições econômicas, expectativas dos investidores e outros fatores ambientais que podem influenciar nos preços. Além disso, as séries históricas de preços possuem altos níveis de ruído, são dinâmicas, não-lineares, não-paramétricas e caóticas por natureza (ZHANG & WU, 2009).

De acordo com Box *et al.* (2008), a previsão baseada em séries temporais representa um meio de prover informação e conhecimento para apoiar uma decisão subsequente. Portanto, a análise de séries temporais tem por objetivo obter relações de dependência entre dados históricos, ou seja, determinar estruturas e padrões nesses dados e, desta análise, desenvolver um modelo que preveja seu comportamento.

Em geral, as abordagens existentes para prever os preços do mercado financeiro podem ser amplamente classificadas em dois tipos, a Análise Fundamental e a Análise Técnica (BLACK, 1982). A Análise Fundamental baseia-se em dados macroeconômicos, como exportações e importações, oferta monetária, taxas de juros, taxas de inflação, taxas de câmbio, taxa de desemprego, assim como no perfil financeiro específico da empresa, que inclui rendimentos, dividendos, valor de mercado, dívidas, entre outros (BASU, 1977; DECHOW *et al.*, 2001).

A Análise Técnica baseia-se na lógica de que a história se repetirá e que a correlação entre preço e volume revela o comportamento do mercado. A previsão ocorre ao se explorar os históricos anteriores, analisando padrões e tendências mostrados nos gráficos de preço e volume. A principal ferramenta da Análise Técnica são os indicadores técnicos, que são transformações matemáticas realizadas nos dados históricos (PRING, 2016).

As análises Fundamental e Técnica de previsão de preços do mercado de ações tem uma importância diferente quando examinada sob o ponto de vista do horizonte de previsão. De acordo com Vanstone & Finnie (2008), em sistemas de previsão de longo termo, que possuem um horizonte temporal maior do que um ano,

a Análise Fundamental é preferida. Caso contrário, se o horizonte for inferior a um ano, a Análise Técnica é preferida.

Desde a introdução do mercado financeiro, técnicas computacionais foram amplamente aplicadas para sua previsão. Elas oferecem ferramentas úteis na previsão de ambientes ruidosos como o mercado de ações, capturando seu comportamento não-linear (ATSALAKIS & VALAVANIS, 2008). Dentre as diversas técnicas destaca-se a utilização de regressão linear e de sistemas inteligentes como as Redes Neurais, Sistemas Fuzzy, Algoritmos Genéticos, Máquinas de Vetores de Suporte, entre outros.

As Redes Neurais possuem uma ótima capacidade de prever as direções do mercado, pois sua habilidade em descobrir relações não-lineares entre as entradas e saídas possibilita a modelagem de sistemas não-lineares dinâmicos como, por exemplo, o mercado de ações (QIU et al, 2016; LAWRENCE, 1997). Além disso, diferentemente dos métodos estatísticos clássicos, técnicas como as Redes Neurais possuem maior tolerância diante de dados incertos, incompletos e imprecisos, sendo, assim, uma ferramenta atrativa para previsão de ambientes ruidosos como o mercado de ações (CHANG *et al*, 2008).

No entanto, embora haja muito estudo da aplicação de inteligência computacional na previsão do mercado financeiro, existe pouco estudo a respeito da performance de sistemas inteligentes na previsão de suas tendências semanais, por exemplo. Esses estudos, em sua maioria, ficam restritos à previsão do valor do dia seguinte dos ativos financeiros (NAEINI *et al.*, 2010; ATSALAKIS & VALAVANIS, 2009; KIM, 2003)

O cenário escolhido para este trabalho foi a BM&F Bovespa, pois é um mercado pouco estudado quando comparado ao mercado internacional. Dos 100 artigos avaliados por Atsalakis & Valavanis (2008), somente um se refere à Bolsa de Valores de São Paulo (BOVESPA). Esse é um número muito pequeno, levando em conta que em 2008, após a fusão com a BM&F, a nova BM&F BOVESPA se tornou a terceira maior Bolsa de Valores do mundo de acordo com seu valor de mercado (CPC, 2018).

Este trabalho tem por objetivo a aplicação de uma Rede Neural *Feedforward* na previsão do Índice Bovespa. A variável de previsão é um classificador que irá indicar se a cotação do índice irá subir ou descer em um horizonte de tempo de uma semana (5 dias úteis). O sistema de previsão desenvolvido possui como entradas:

- Indicadores Técnicos: Média Móvel e Taxa de Variação;
- Características estatísticas: desvio padrão;
- Relações entre os preços de abertura, fechamento, máximo e mínimo;
- Séries exógenas: preços das ações que compõem o Índice Bovespa, preço do dólar/real, preço do petróleo e preço do ouro;

O objetivo secundário deste trabalho é fazer a seleção das características mais significativas para a previsão através do algoritmo *Recursive Feature Elimination* utilizando como modelo um *Gradient Boosting* de árvores de decisão.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 O MERCADO DE AÇÕES

O papel fundamental da Economia é a alocação eficiente de capital. Para que isso seja obtido, o capital é investido em setores onde há grandes expectativas de retorno, enquanto que setores com perspectivas ruins são evitados. Baseado nisso, os mercados financeiros são encarregados de assegurar uma distribuição adequada de investimentos (WURGLER, 2000). O termo Mercado Financeiro se refere a qualquer mercado onde ativos como ações, obrigações, moedas e derivativos são negociados. Os dois maiores mercados são o mercado de ações e o mercado monetário nos quais são negociados trilhões de dólares diariamente (LEVINSON, 2005).

O mercado de ações é um mercado público que reúne investidores para a negociação de ações e derivativos de empresas a um preço acordado. O mercado de ações estabelece preços de acordo com a oferta e a demanda. Assim, uma ação com alta demanda aumentará de preço, enquanto uma ação que está sendo fortemente vendida diminuirá de preço. O mercado primário lida com as novas emissões de títulos diretamente da empresa. Um prospecto oficial é publicado sob a lei da corporação que contém todas as informações sobre a empresa, e é exigido pelos investidores para auxiliar na tomada de decisões de investimento. Os títulos existentes são comprados e vendidos no mercado secundário entre os comerciantes (CORPORATE FINANCE INSTITUTE, 2018).

Uma ação é um documento emitido por uma empresa, que autoriza seu titular a ser um dos proprietários da empresa. Ao possuir uma ação, pode-se ganhar uma parte do lucro da empresa chamado de dividendo. Além disso, ao vender as ações, obtém-se o ganho de capital. No entanto, existe o risco de uma perda de capital se o preço de venda da ação for inferior ao preço de compra (INVESTOPEDIA, 2018).

2.1.1 A Hipótese do Mercado Eficiente (HME)

Uma grande parte dos estudos existentes associados à previsão do mercado financeiro suporta a Hipótese do Mercado Eficiente (HME) segundo a qual o preço atual de uma ação reflete integralmente, a qualquer momento, as informações disponíveis assimiladas pelos investidores. À medida que novas informações estiverem disponíveis, qualquer desequilíbrio é imediatamente detectado e contabilizado por uma mudança contrária em seu preço (FAMA, 1965). Sendo assim, de acordo com Hawawini & Keim (1995) a previsão dos preços do mercado de ações é inútil e a correlação das séries cronológicas é economicamente e estatisticamente insignificante porque os preços do mercado financeiro seguem uma tendência aleatória.

Com base no grau de eficiência do mercado financeiro, existem três versões da HME. A versão fraca da HME afirma que qualquer informação resultante da análise do histórico passado da ação se reflete em seu preço. A versão semi-forte da HME afirma que toda nova informação, uma vez que se torna pública, é rapidamente refletida nos preços das ações. No entanto, se tal informação for, de alguma forma, conhecida previamente, ela pode ser usada para a obtenção de lucros extras. A versão forte da HME afirma que toda informação referente ao preço de uma ação, seja ela pública ou privada, é imediatamente contabilizada em seu preço (FAMA, 1991; HAUGEN, 1997).

De acordo com a forma fraca da HME, os preços das ações não podem ser previstos baseando-se apenas em seus valores passados devido ao seu comportamento aleatório. No entanto, apesar da popularidade da HME, há uma parcela significativa de investidores que não acreditam nela em qualquer uma de suas versões. Diversos estudos fornecem evidências de que a Análise Técnica pode consistentemente produzir lucros e que o mercado de ações não é completamente eficiente e nem os preços das ações seguem uma caminhada aleatória (BORODIN *et al.*, 2004; LO & MACKINLAY, 1988).

2.1.2 A Análise Técnica

A Análise Técnica é o estudo do comportamento do mercado, principalmente através do uso de gráficos, com o propósito de prever tendências futuras dos preços (MURPHY, 1999). A Análise Técnica assume que o preço dos ativos financeiros se move de acordo com tendências e padrões que se mantêm em períodos até que alguma mudança nas condições do mercado altere essa tendência. Tipicamente são utilizadas informações históricas de preço e de volume para extrair indicadores técnicos, que tem o propósito de facilitar o entendimento das movimentações do mercado. O principal objetivo da Análise Técnica é prever o preço de um ativo em algum horizonte de tempo a fim de auxiliar os investidores do mercado a realizar operações lucrativas (ROCKEFELLER, 2011).

A Análise Técnica é baseada em 3 premissas:

- Os movimentos do mercados já descontam tudo;
- Os preços se movem em tendências;
- A história se repete.

A primeira premissa é considerada o fundamento da Análise Técnica. Ela afirma que os preços do mercado refletem não somente a informação sobre fatores econômicos, mas que também os fatores psicológicos, políticos e geográficos já estão descontados e refletidos no preço atual do ativo. Dessa forma, o preço indiretamente fornece uma perspectiva dos aspectos fundamentais e, portanto, o estudo do preço é suficiente para realizar previsões (MURPHY, 1999).

A segunda premissa sugere que os movimentos do ativo financeiro seguem tendências. Isso significa que após o estabelecimento de uma tendência, é mais provável que o futuro movimento do preço seja na mesma direção da tendência do que contra ela, embora variações de preço possam ocorrer. Portanto, a Análise Técnica busca por padrões no preço que sinalizam uma continuação ou uma reversão nas tendências (MURPHY, 1999).

A terceira premissa se baseia na suposição de que a história tende a se repetir. Em outras palavras, padrões de comportamento do mercado que ocorreram no passado vão se repetir no futuro e, portanto, podem ser utilizados para propósitos de

previsão. Na terminologia estatística, as previsões dependem da dependência de sucessivas variações do preço (LEVY, 1966).

Os Indicadores Técnicos são uma ferramenta da Análise Técnica e são obtidos através de uma expressão matemática aplicada aos preços. Utilizados isoladamente eles não são capazes de prever o comportamento futuro de um ativo, no entanto, quando analisados em conjunto, podem fornecer informações importantes sobre as tendências do mercado (PRING, 2016).

Existem centenas de Indicadores Técnicos, sendo que muito investidores, inclusive, criam seus próprios indicadores. A seguir são listados os principais indicadores técnicos consolidados na literatura (PRING, 2016; WILDER, 1978):

- **Média Móvel Simples (MMS):** esse indicador é calculado pela média aritmética do preço dos últimos n dias. Como as séries temporais possuem muito ruído, ele é comumente usado para suavizar flutuações de curto prazo nos preços e ressaltar tendências de mais longo prazo. De forma geral, se a MMS é crescente, significa que a tendência é de alta, e se for decrescente que a tendência é de baixa. Uma das estratégias utilizadas é de que quando o preço da ação cruza a Média Móvel por baixo, o preço tende a subir nos próximos dias, já se o cruzamento for por cima, o preço tende a cair;

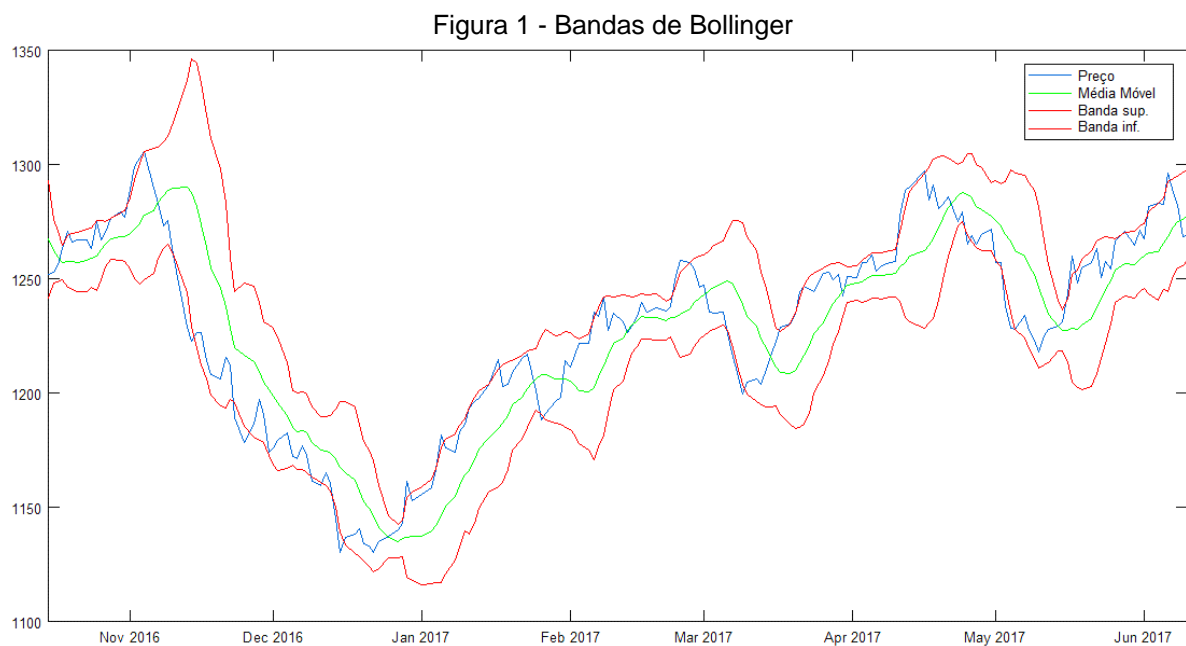
- **Taxa de Variação:** esse indicador mede (ver Equação (1)) a taxa com a qual o preço de uma ação variou em relação a n períodos atrás:

$$\text{Taxa de Variação} = \frac{\text{Preço} - \text{Preço } n \text{ períodos atrás}}{\text{Preço } n \text{ períodos atrás}} \times 100 \quad (1)$$

Quando o indicador assume valores maiores que zero, o preço da ação que ele está medindo é maior que seu preço n períodos atrás. Além disso, se o indicador também estiver crescendo, a diferença entre o preço atual e seu valor n períodos atrás também está crescendo. Se o indicador é positivo mas decrescente, o preço continua acima do seu valor n períodos atrás, porém está decrescendo. A mesma lógica é usada quando o indicador assumir valores negativos.

- **Bandas de Bollinger:** possibilitam avaliar a volatilidade e uma provável evolução dos preços. As bandas são constituídas de 3 curvas, uma delas calculada pela média móvel dos dados em n períodos, e outras duas curvas situadas cada uma

a uma distância de duas vezes o desvio-padrão, acima e abaixo, dos n períodos sobre os quais a média-móvel foi calculada. A largura da banda é uma indicação direta da volatilidade do ativo considerado. Teoricamente, há mais de 95% de chance de que a evolução do valor seja estabelecida dentro do quadro das bandas configuradas com uma distância de dois desvios-padrão. Uma das interpretações das Bandas de Bollinger é de que quando as bandas se estreitam, existe uma grande chance de que ocorrerão grandes mudanças no preço da ação em um futuro próximo. Uma das estratégias utilizadas é de que quando o preço do ativo ultrapassa uma das bandas, a tendência é de que irá continuar naquela direção. A Figura 1 ilustra o comportamento das bandas de Bollinger, onde o preço do ativo está em azul, a média móvel em verde e as bandas em vermelho.



- **Índice de Força Relativa (IFR):** esse indicador de momento mede a força ou o enfraquecimento de uma tendência baseado nos preços de fechamento de uma determinada ação. O seu valor varia entre 0 e 100 e é calculado de acordo com a Equação (2).

$$RSI = 100 - \frac{100}{1 + RS} \quad (2)$$

onde RS é a Força Relativa (*Relative Strength*), determinada pela Equação (3).

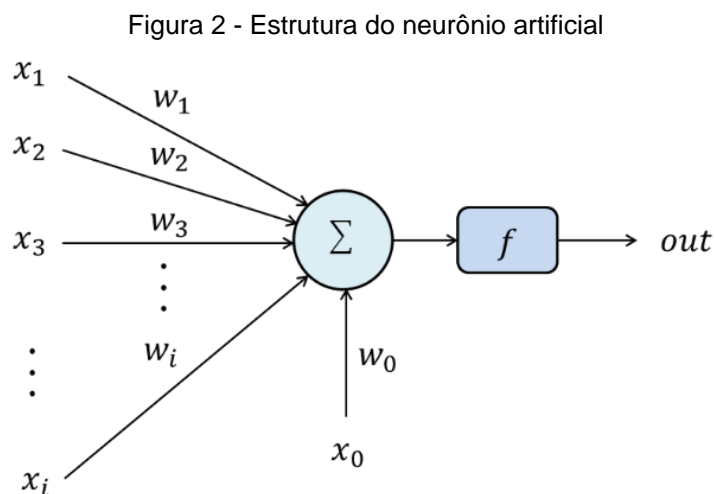
$$RS = \frac{Média_ganhos(n)}{Média_perdas(n)} \quad (3)$$

onde ambas as médias são as obtidas nos últimos n períodos.

O RSI compara os ganhos (tendência de alta) e as perdas (tendência de baixa) dos últimos n períodos. Wilder recomenda um número de períodos, n , igual a 14. Uma interpretação comum do RSI é que ele sugere um estado de sobrevenda do ativo para valores menores que 30, e de sobrecompra para valores acima de 70 ou 80.

2.2 AS REDES NEURAIS ARTIFICIAIS

Uma rede neural artificial é um modelo matemático inspirado no comportamento do cérebro humano (MCCULLOCH; PITTS, 1943). A Rede Neural é representada por um grafo orientado, onde cada nó deste grafo representa um neurônio e cada aresta representa uma sinapse, ou seja, uma conexão entre dois neurônios. A menor unidade de uma rede neural é o próprio neurônio, também conhecida por perceptron, apresentado na Figura 2.



Fonte: autoria própria

O neurônio recebe sinais de entradas representados pelo vetor coluna x dado pela Equação (4):

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \quad (4)$$

A cada uma destas entradas, é atribuído um peso sináptico w_{ij} , onde j representa o sinal de entrada x_j e i representa o neurônio analisado. Desta forma, pode-se representar o vetor dos pesos sinápticos pela Equação (5):

$$w_i = [w_{i1} \ w_{i2} \ w_{i3} \ \dots \ w_{in}] \quad (5)$$

Em seguida, a junção aditiva é aplicada aos sinais de entrada e seus respectivos pesos sinápticos. Esta junção aditiva é uma combinação linear e pode ser dada pela Equação (6):

$$a_i = w_{i1}x_1 + w_{i2}x_2 + \dots + w_{in}x_n + b_i = w_i * x_i + b_i \quad (6)$$

onde b_i indica o termo bias aplicado ao neurônio i .

As quantidades a_i são conhecidas como ativações. Cada uma delas é transformada utilizando uma função de ativação φ para gerar o sinal de saída do neurônio como mostrado pela Equação (7):

$$z_i = \varphi(a_i) \quad (7)$$

onde z_i é a saída do neurônio.

Ao processo desde a entrada dos sinais x até a geração do sinal de saída por meio da utilização da função de ativação, é dado o nome de *feedforward propagation*, ou propagação em frente (HAYKIN, 2008).

Em redes neurais artificiais, a função de ativação $\varphi(\cdot)$ de um neurônio representa a taxa com a qual determinado neurônio está sendo ativado ou utilizado (COPPIN, 2004). As funções de ativação têm um papel fundamental no treinamento de redes neurais e é importante que elas possuam algumas características a fim de melhorar o desempenho do treinamento das redes neurais. Segundo HAYKIN (2008), a função de ativação deve ser diferenciável para que se possa utilizar um algoritmo que visa a minimizar o erro das tarefas realizadas pelas redes neurais. Além disto, se a função de ativação for diferenciável mais de uma vez, é possível desenvolver algoritmos que extraíam informações de segunda ordem acerca do erro cometido

pelas redes neurais, o que possibilita um treinamento mais rápido e eficiente. As funções de ativação mais comuns e suas derivadas estão listadas na Tabela 1.

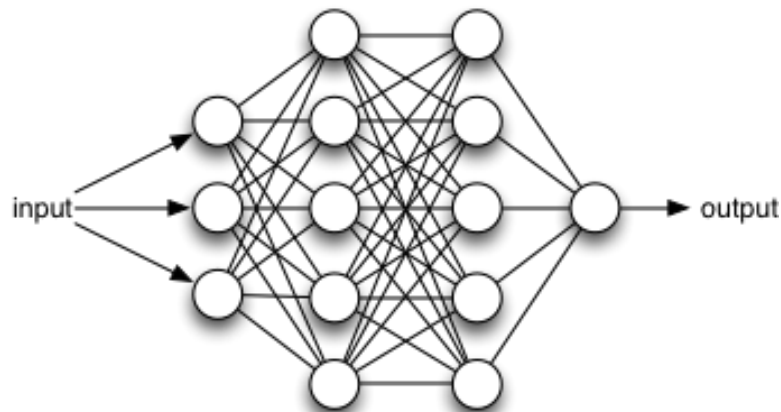
Tabela 1 – Funções de ativação e suas características.

Função de ativação	Equação	Derivada	Intervalo de valores	Ordem de continuidade
Passo binário	$\varphi(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases}$	$\varphi'(x) = 0, \quad x \neq 0$	[0,1]	C^{-1}
Identidade	$\varphi(x) = x$	$\varphi'(x) = 1$	$] -\infty, +\infty[$	C^∞
Sigmoide	$\varphi(x) = \frac{1}{1 + e^{-x}}$	$\varphi'(x) = \varphi(x) * (1 - \varphi(x))$	[0,1]	C^∞
Softplus	$\varphi(x) = \ln(1 + e^x)$	$\varphi'(x) = 1 - \varphi(x)^2$	$[0, +\infty[$	C^∞
Tangente hiperbólica	$\varphi(x) = \tanh(x)$	$\varphi'(x) = \frac{1}{1 + e^{-x}}$	$[-1,1]$	C^∞

Fonte: elaborada pelo autor

Um modelo de redes neurais de múltiplas camadas, ou *multilayer perceptron (MLP)*, é formado a partir da união de vários neurônios ligados entre si por meio de elos sinápticos e *bias*. A Figura 3 mostra um exemplo de uma rede neural com uma camada de entrada, uma camada oculta e uma camada de saída.

Figura 3 - Representação de uma rede neural MLP



Fonte: DOLHANSKY, 2018

O fluxo de sinal percorre a rede neural da esquerda para a direita através do mecanismo de *forward propagation*. Deste modo, pode-se realizar a modelagem matemática da rede neural como uma cadeia de perceptrons conectados uns aos outros, onde o sinal de saída de uma camada anterior representa o sinal de entrada da camada seguinte até que o sinal resultante saia da última camada de neurônios.

Segundo Haykin (2001), os perceptrons de múltiplas camadas têm sido aplicados com sucesso para resolver problemas difíceis, através do seu treinamento de forma supervisionada com um algoritmo muito popular conhecido como algoritmo de retropropagação de erro (*error backpropagation*).

O desenvolvimento do algoritmo de retropropagação representa um marco nas redes neurais, pois fornece um método computacional eficiente para o treinamento de perceptrons de múltiplas camadas. Apesar de não podermos afirmar que o algoritmo de retropropagação forneça uma solução ótima para todos os problemas resolúveis, ele acabou com o pessimismo sobre a aprendizagem em máquinas de múltiplas camadas (HAYKIN, 2001).

O algoritmo de retropropagação do erro visa a reajustar os pesos sinápticos e bias das conexões dentro da rede neural por meio de uma avaliação do erro obtido quando se apresenta um conjunto de sinais de entrada com sinais de saída conhecidos. Desta forma, é possível realizar um treinamento de uma rede neural num conjunto de dados disponíveis para que a mesma possa ser utilizada na previsão de sinais de saída desconhecidos.

Assumindo um conjunto de N exemplos constituídos de sinais de entrada x e seus respectivos sinais de saída y , pode-se definir o erro de um sinal de saída como na Equação (8):

$$e(n) = \hat{y}(n) - y(n) \quad (8)$$

Define-se a função custo quadrático médio para um neurônio pela Equação (9):

$$J(w, bias) = \frac{1}{2N} \sum_1^N e^2(n) \quad (9)$$

A função custo representa uma medida do desempenho do treinamento da rede neural. O objetivo do algoritmo de retropropagação do erro é encontrar os parâmetros \tilde{w} e \tilde{b} que minimizam a função custo. Uma das condições para que um par (\tilde{w}, \tilde{b}) indique um mínimo da função custo é a de que o gradiente em torno deste ponto seja aproximadamente zero. O gradiente da função custo resulta na Equação (10):

$$\frac{\partial J(w, b)}{\partial (w, b)} = \frac{1}{N} * \sum_1^N e(n) \times \varphi'(z) * x_i \quad (10)$$

A correção aplicada aos pesos sinápticos e bias é definida pela regra delta, dada pela Equação (11):

$$\Delta w_{ji}(n) = -\alpha \times \frac{\partial J(w, b)}{\partial (w, b)} \quad (11)$$

Ou seja, a correção é feita no sentido de descréscimo da função de custo a fim de se atingir o mínimo. O termo α é o parâmetro da taxa de aprendizagem do algoritmo de retropropagação. Ele é responsável por determinar a magnitude da mudança dos pesos sinápticos no sentido de descida do gradiente. O uso da Equação (10) em (11), resulta em:

$$\Delta w_{ji}(n) = -\gamma * \frac{1}{N} \sum_1^N e(n) * \varphi'(z) * x_i \quad (12)$$

Define-se o gradiente local por:

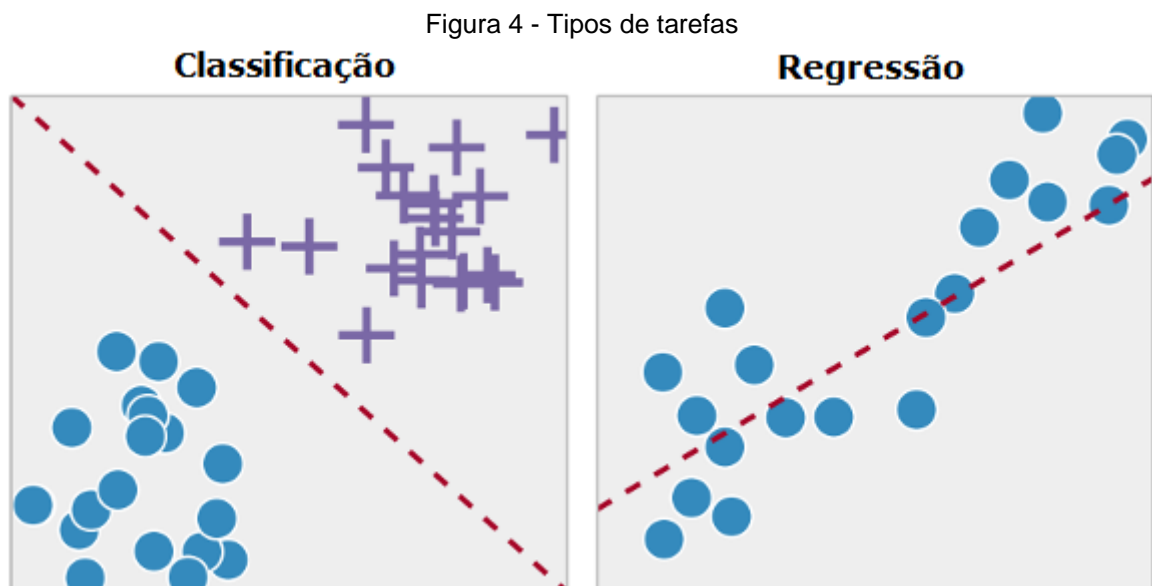
$$\delta_j(n) = -\frac{1}{N} \sum_1^N e(n) * \varphi'(z) \quad (13)$$

Logo:

$$\Delta w_{ji}(n) = \gamma * \delta_j(n) * x_i \quad (14)$$

Desta forma, o gradiente local aponta no sentido das modificações necessárias nos pesos sinápticos.

Segundo Faceli *et al.* (2011), um algoritmo de aprendizado de máquina supervisionado é uma função que, dado um conjunto de exemplos rotulados, constrói um estimador. O rótulo ou etiqueta toma valores num domínio conhecido. Se esse domínio for um conjunto de valores nominais, tem-se um problema de classificação, também conhecido como aprendizado de conceitos, e o estimador gerado é um classificador. Se o domínio for um conjunto infinito e ordenado de valores, tem-se um problema de regressão, que induz um regressor. A Figura 4 ilustra os dois tipos de problemas.



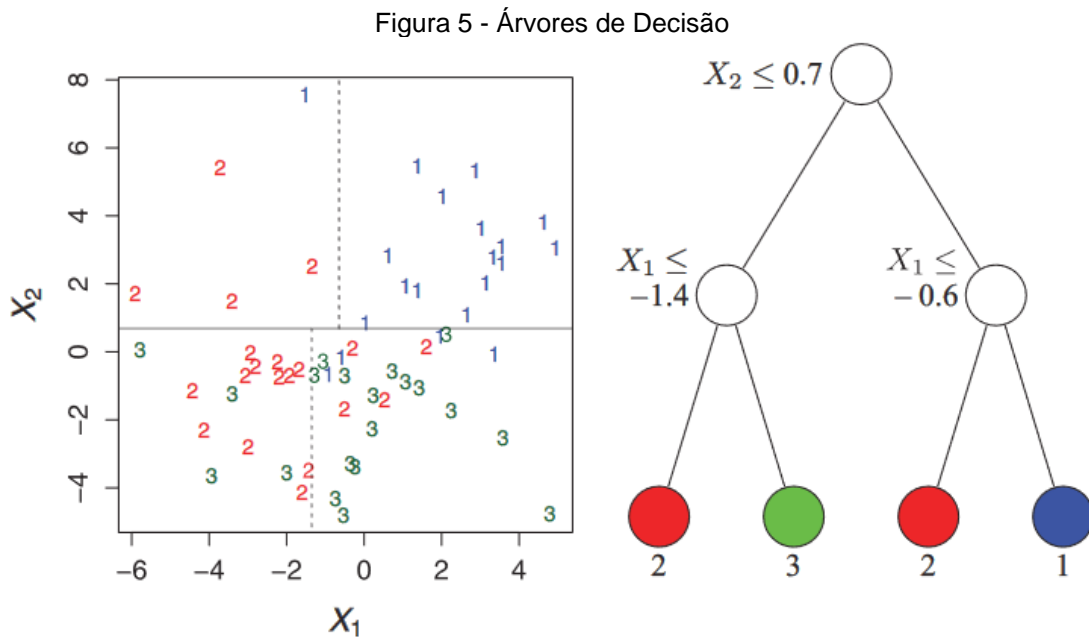
Fonte: MEDIUM, 2018

Em outras palavras, é possível realizar os dois tipos de tarefas por meio de redes neurais. Ao mostrar exemplos cujo resultado é conhecido, a rede neural aprende a prever um valor ou uma classificação quando encontra um novo conjunto de entradas.

2.3 ÁRVORES DE DECISÃO E O *GRADIENT BOOSTING*

Árvores de Decisão são modelos obtidos particionando-se recursivamente o espaço de dados, tornando os novos subconjuntos de dados mais fáceis de interpretar. Assim, dentro de cada partição é ajustado um modelo de previsão simples

(LOH, 2011). A Figura 5 ilustra as partições (na esquerda) e a estrutura de uma Árvore de Decisão (na direita) que possui três classes e duas variáveis de entrada. A classe de previsão é dada em cada folha (nó terminal) da árvore.



Fonte: LOH, 2011

O *Gradient Boosting* é uma técnica de Aprendizagem de Máquina utilizada para problemas de regressão e classificação desenvolvida por Friedman (2001). Os métodos chamados de *boosting* se baseiam no princípio da minimização de uma função custo através da agregação de múltiplos modelos fracos (*weak learners*) para a construção de um modelo final mais forte utilizando o método do gradiente como procedimento.

De forma semelhante a outros métodos de Aprendizagem de Máquina, dado um conjunto de treinamento $\{y_i, x_i\}_1^N$ de valores conhecidos (y, x) , o objetivo é encontrar uma função $F^*(x)$ que mapeie x em y , de forma que sobre a distribuição conjunta de todos os valores de (y, x) , o valor esperado de uma função custo especificada $\Psi(y, F(x))$ é minimizada, conforme mostra a Equação (15).

$$F^*(x) = \arg \min_{F(x)} E_{y,x} [\Psi(y, F(x))] \quad (15)$$

O método *Boosting* faz uma aproximação da função $F^*(\mathbf{x})$ através de uma expansão aditiva, conforme a Equação (16).

$$F(\mathbf{x}) = \sum_{m=0}^M \beta_m h(\mathbf{x}; \mathbf{a}_m) \quad (16)$$

onde as funções genéricas $h(\mathbf{x}; \mathbf{a})$, chamadas de “modelos fracos”, normalmente são funções simples de \mathbf{x} com parâmetros $\mathbf{a} = \{a_1, a_2, \dots\}$, e M é o de funções utilizadas.

Os coeficientes de expansão $\{\beta_m\}_0^M$ e os parâmetros $\{\mathbf{a}_m\}_0^M$ são determinados em etapas para os dados de treinamento. Inicia-se com uma estimativa inicial $F_0(\mathbf{x})$, e para $m = 1, 2, \dots, M$, os parâmetros são determinados de forma a minimizar a função custo da soma entre o modelo anterior e a nova função aditiva, conforme mostram as Equações (17) e (18).

$$(\beta_m, \mathbf{a}_m) = \arg \min_{\beta, \mathbf{a}} \sum_{i=1}^N \Psi(y_i, F_{m-1}(x_i) + \beta h(\mathbf{x}; \mathbf{a})) \quad (17)$$

$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \beta_m h(\mathbf{x}; \mathbf{a}_m) \quad (18)$$

O Gradient Boosting utilizando Árvores de Decisão especializa este método onde a função genérica $h(\mathbf{x}; \mathbf{a})$ é uma árvore com L folhas (ver Equação (19)). A árvore particiona o espaço de entradas em L regiões disjuntas R_{1m}, \dots, R_{Lm} e prevê um valor constante em cada região.

$$h(\mathbf{x}, \{R_{lm}\}_1^L) = \sum_{l=1}^L 1(\mathbf{x} \in R_{lm}) \quad (19)$$

onde \bar{y}_{lm} é a média de \tilde{y}_{im} em cada região R_{lm} . E \tilde{y}_{im} é chamada de pseudo-resíduo, calculado pela Equação (20).

$$\tilde{y}_{im} = - \left[\frac{\partial \Psi(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})} \quad (20)$$

Os parâmetros dessas funções genéricas são as variáveis de divisão dos nós e os correspondentes pontos de divisão que definem a árvore, que por sua vez definem as correspondentes regiões $\{R_{lm}\}_1^L$ da partição na m-ésima iteração.

O *Gradient Boosting* é tipicamente utilizado com Árvores de Decisão, especialmente árvores CART (*Classification and Regression Trees*). Um dos benefícios de se utilizar métodos de *ensemble* de Árvores de Decisão como o *Gradient Boosting* é que eles automaticamente provêm uma estimativa da importância das características do modelo preditivo treinado. A importância de uma característica é calculada pelo número de vezes que sua utilização melhora a medida de performance da árvore. Assim, quanto mais uma característica é utilizada para tomar decisões nas árvores, maior será sua relativa importância. A importância final de cada característica é, então, calculada pela média de sua importância em cada árvore que compõe o modelo. Essa importância é calculada explicitamente para cada característica, permitindo que elas sejam classificadas e comparadas umas às outras (CHEN & GUESTRIN, 2016).

2.4 MÉTRICAS DE DESEMPENHO

Para quantificar o desempenho de uma rede neural na previsão de séries temporais, costuma-se utilizar métricas que analisam as diferenças entre a série de valores previstos e a série de valores reais. A seguir, são mostradas algumas das métricas mais utilizadas para estes cálculos nos estudos pesquisados.

- **Erro Percentual Absoluto Médio:** o Erro Médio Percentual Absoluto (ou MAPE – *Mean Absolute Percentage Error*) obtém as diferenças percentuais entre todos os valores reais e previstos obtidos e faz uma média simples destes valores. Como todos os elementos da série temporal tem igual peso no resultado final, resultados isolados muito diferentes dos demais não fazem tanta diferença. Essa métrica é útil para se ter uma visão geral do erro médio gerado pelo algoritmo de previsão escolhido. O MAPE é definido de acordo com a Equação (21), onde N é o número de dias da série temporal sendo analisada, $P_{real,i}$ é o valor real da série no dia i e $P_{previsto,i}$ é o valor previsto para a série, também no dia i .

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{|P_{real,i} - P_{previsto,i}|}{P_{previsto,i}} \right) \times 100 \quad (21)$$

• **Raiz Quadrada do Erro Quadrático Médio:** essa métrica, também chamada de RMS ou RMSE (Root Mean Squared Error), é semelhante ao MAPE no sentido de que obtêm o erro de todos os valores calculados. A diferença está no fato do erro calculado não ser um percentual, mas sim um valor numérico que indica o tamanho do erro médio obtido. O quanto mais próximo de zero for o erro, melhor são os resultados obtidos. O RMS é calculado a partir da Equação (22), onde N é o número de dias da série temporal sendo analisada, $P_{previsto,i}$ é o valor previsto para a série no dia i e $P_{real,i}$ é o valor real da série também no dia i.

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_{real,i} - P_{previsto,i})^2} \quad (22)$$

• **Erro Médio Quadrático:** o erro médio quadrático, ou MSE (*Mean Squared Error*) é a métrica mais utilizada para calcular o desempenho de modelos de previsão. É calculado a partir da soma da variância e dos quadrados das diferenças obtidas entre os valores reais e previstos, e é dado pela Equação (23), onde N é o número de dias da série temporal sendo analisada, $P_{previsto,i}$ é o valor previsto para a série no dia i e $P_{real,i}$ é o valor real da série também no dia i.

$$MSE = \frac{1}{N} \sum_{i=1}^N (P_{real,i} - P_{previsto,i})^2 \quad (23)$$

2.5 TRABALHOS RELACIONADOS

Existe uma vasta quantidade de trabalhos na literatura que estuda o problema da previsão de tendências do mercado financeiro, e um grande número entre esses utiliza Redes Neurais Artificiais como modelo de previsão. Entre os estudos sobre o assunto, Atsalakis & Valavanis (2008) analisam 100 trabalhos relevantes que estudam

a utilização de técnicas computacionais para abordar o problema. Essas técnicas são classificadas de acordo com:

- O mercado do qual os dados são obtidos;
- As variáveis de entrada;
- A metodologia e os parâmetros utilizados (pré-processamento, tamanho do conjunto de dados, tipo da rede e método de treinamento);
- Os modelos utilizados como *benchmark*;
- As métricas de performance utilizadas para avaliar o método proposto.

Além disso, quanto à variável de previsão, esses problemas podem ser classificados em dois grandes grupos: aqueles que tem por objetivo prever os níveis, ou seja, os valores que o índice atingirá no futuro ou o retorno previsto em um determinado horizonte de tempo, os quais se enquadram em problemas de regressão. O outro grupo tem por objetivo prever a direção da variação dos preços, ou seja, se os valores futuros serão maiores ou menores do que o valor atual, os quais se enquadram nas técnicas de classificação.

Naeini *et al.* (2010) propuseram dois modelos de previsão do preços de ações cotadas na Bolsa de Valores de Teerã. O primeiro deles é uma Rede Neural *Feedforward* e o segundo uma Rede Recorrente de Elman. O objetivo deste estudo foi utilizar essas redes para fazer previsões baseadas nos dados históricos das ações. Foram utilizadas como entradas das redes os preços máximo, mínimo e a média entre o máximo e o mínimo dos últimos d dias, com d variando de 1 a 10. Foi avaliada a performance das redes na previsão da direção do preço do dia seguinte (problema de classificação) e do valor do preço do dia seguinte (problema de regressão). Os autores concluíram que para a previsão da direção do dia seguinte, uma simples regressão linear obteve a melhor taxa de acertos. Para a previsão do valor do dia seguinte, o método que obteve a melhor performance (medida através do Erro Quadrático Médio) foi a Rede Neural *Feedforward*.

Qiu *et al.* (2016) propuseram a utilização de uma Rede Neural treinada com o algoritmo *Backpropagation* na previsão do índice da bolsa de valores japonesa, Nikkei 225. Foram pré-selecionadas 71 variáveis que incluem indicadores financeiros e dados macroeconômicos como, por exemplo, valor da taxa de juros, taxa de câmbio,

índice de preços ao consumidor, índice de produção industrial e taxa de depósito. Através da implementação de superfícies Fuzzy, foram selecionadas 18 variáveis das 71 que demonstraram ter maior importância, através da medida da correlação entre as variáveis de entrada e a variável de saída. O conjunto de dados utilizados cobre o período de Novembro de 1993 até Julho de 2013, sendo que 70% dos dados foram utilizados para o treinamento e 30% para os testes. A arquitetura da rede possui 3 camadas, sendo que a camada de entrada possui 18 nós e a camada de saída possui um nó com o valor da previsão. Foi constatado que o menor Erro Médio Quadrático foi mínimo quando a rede possuía entre 10 e 30 neurônios na camada oculta. A fim de contornar o problema da convergência local das Redes Neurais, técnicas de procura de convergência global foram utilizadas com o auxílio de Algoritmos Genéticos e Arrefecimento Simulado. Foi demonstrado que o algoritmo híbrido proposto obteve uma melhor performance comparado ao *Backpropagation* comum. O erro quadrático médio do algoritmo híbrido foi de 0,0043 enquanto que do algoritmo *Backpropagation* comum foi de 0,1077.

Kim (2003) propõe a utilização de Máquinas de Vetores de Suporte na previsão da direção do dia seguinte do Índice composto do KOSPI (Índice de bolsa de valores sul-coreano). O modelo possui como entradas 12 indicadores técnicos, e a SVM é treinada com diferentes parâmetros C e δ^2 , onde C é o termo de regularização e δ^2 é o parâmetro da função de Kernel utilizada. O modelo SVM desenvolvido obteve uma taxa de acertos de 57,83% em sua melhor performance, enquanto uma Rede Neural *Feedforward* com *Backpropagation* obteve uma taxa de acertos de 54,73% na mesma série histórica.

Kara *et al.* (2010) propõem a utilização de Redes Neurais *Feedforward* e de Máquinas de Vetores de Suporte para a previsão dos movimentos diários do Índice da Bolsa de Valores de Istambul (*Istanbul Stock Exchange (ISE) National 100 Index*). Como entradas dos modelos foram utilizados 10 indicadores técnicos. O conjunto de dados é composto das cotações do índice no período de 2 de janeiro de 1997 a 31 de dezembro de 2007. Do conjunto total de dados, 20% foram utilizados para a determinação de parâmetros dos modelos. Em seguida, o conjunto total de dados foi dividido em dois conjuntos de mesmo tamanho para treinamento e para validação. Essa divisão foi realizada de forma que cada conjunto possuísse a mesma proporção de classes do conjunto original. Os modelos foram implementados utilizando-se

diferentes configurações de parâmetros, como número de neurônios e taxa de aprendizagem, e a taxa de acertos média de validação para a Rede Neural foi de 75,74% e para o modelo SVM foi de 71,52%.

Pesquisas recentes tendem a utilizar técnicas híbridas de Aprendizagem de Máquina. A proposta de Hiemstra (1995) foi de combinar Redes Neurais com Lógica Fuzzy, pois assim seria possível captar as complexidades do mapeamento funcional, sem ser necessária a especificação da função a ser aproximada. Kim e Shin (2007) propuseram um modelo híbrido utilizando Algoritmos Genéticos com Redes Neurais para a otimização do tempo de *delay* e fatores da arquitetura da rede, utilizando Algoritmos Genéticos para melhorar a eficácia do modelo na Rede Neural.

Atsalakis & Psomatakis (2014) propõem a utilização de um modelo híbrido utilizando Redes Neurais com Algoritmos Genéticos na previsão das movimentações diárias das 5 principais ações cotadas na Bolsa de Valores de Atenas (*Athens Stock Exchange*). Foram utilizadas como entradas indicadores técnicos como o Índice de Força Relativa, a Taxa de Variação e a Média Móvel. Foi obtida com esse modelo uma taxa de acertos média de 57,5% na previsão da cotação das ações.

3 METODOLOGIA

3.1 BASE DE DADOS E SUA DIVISÃO

A base de dados utilizada para este estudo é composta pelos preços históricos diários do Índice BOVESPA, das ações que o compõem, do preço do dólar (em reais), assim como dos preços do petróleo e do ouro obtidos da base de dados do YAHOO FINANCE (2018). Cada dia possui as informações das cotações em seu fechamento e abertura, seus pontos máximo e mínimo e o volume de transações realizadas. O período utilizado foi de 27 de setembro de 2011 a 15 de março de 2018, resultando em um total de 1597 dias.

Os dados foram divididos inicialmente em 2 conjuntos, sendo os primeiros 70%, de 27 de setembro de 2011 a 12 de fevereiro de 2016, utilizados para o treinamento e validação da rede, e os últimos 30%, de 13 de fevereiro de 2016 a 15 de março de 2018, para teste. Essa divisão foi feita em ordem cronológica, pois é uma das metodologias mais utilizadas pela literatura para a previsão do mercado financeiro e que possui maior sentido prático (TAY & CAO, 2001; HUANG et al., 2004; PAI & LIN, 2004).

3.2 VARIÁVEL DE PREVISÃO

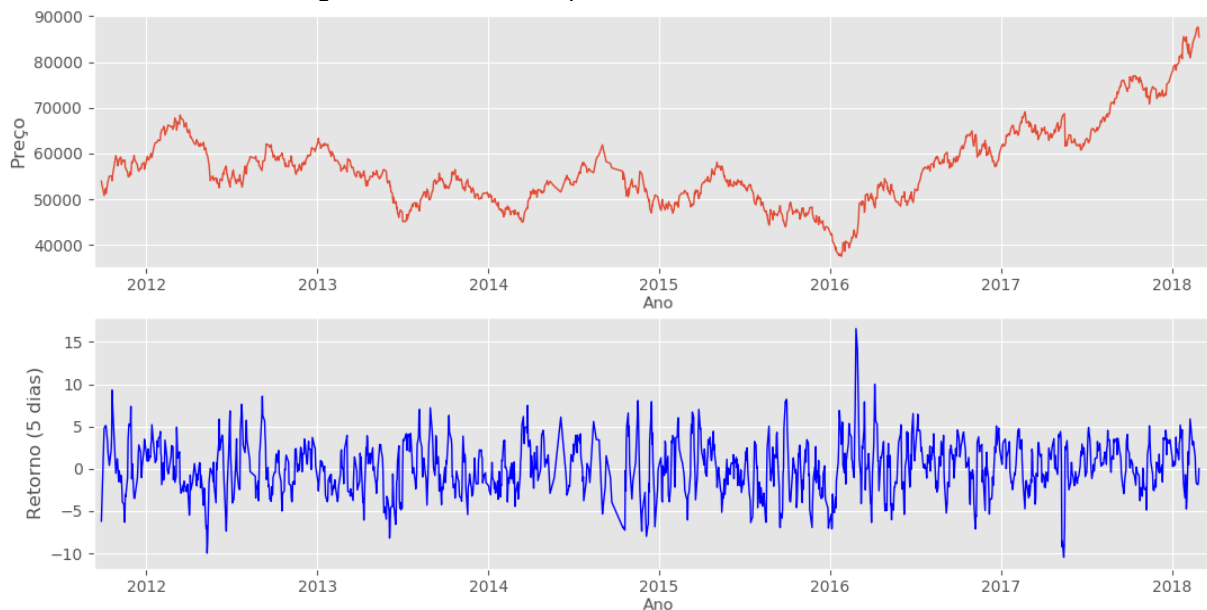
A variável de previsão está relacionada com a taxa de variação dos próximos 5 dias do Índice Bovespa, que significa o quanto se obteria de retorno se o índice fosse comprado no dia em questão e vendido 5 dias depois. A Taxa de Retorno é calculada de acordo com a Equação (24).

$$\text{Taxa de Retorno (\%)} = \frac{\text{Preço}_{d+5} - \text{Preço}_d}{\text{Preço}_d} \times 100 \quad (24)$$

onde Preço_d é o preço atual e Preço_{d+5} é o preço 5 dias adiante.

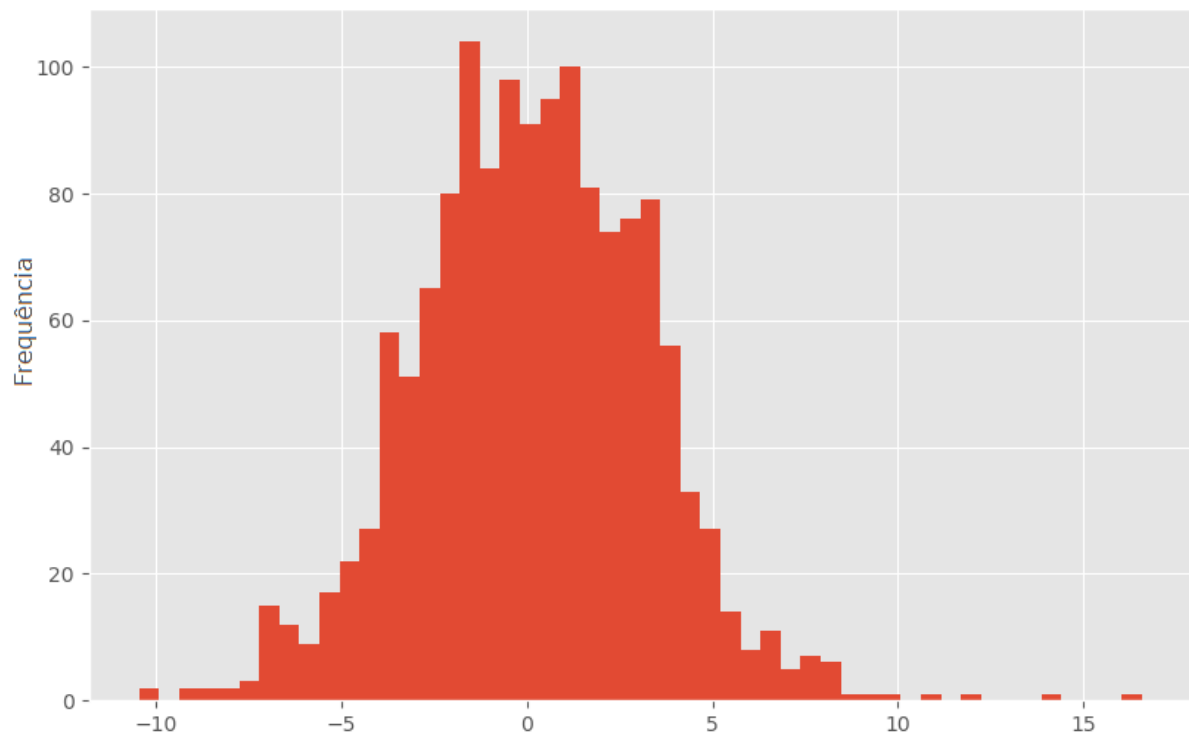
O preço do Índice Bovespa e a Taxa de Retorno de 5 dias no período de 27 de setembro de 2011 a 15 de março de 2018 são apresentados na Figura 6.

Figura 6 - Índice Bovespa e taxa de retorno de 5 dias



A fim de se obter um melhor entendimento da distribuição da Taxa de Retorno, foi criado um histograma dessa variável. Ele é apresentado na Figura 7.

Figura 7 - Histograma da taxa de retorno



Observa-se a que a distribuição dos retornos é aproximadamente normal, com uma média de 0,16 e desvio-padrão de 3,12.

Este problema de previsão foi tratado como um problema de classificação binária. Dessa forma, o classificador assumirá valores 0 e 1, de acordo com as seguintes condições:

$$y = \begin{cases} 0 & \text{se } r < 0 \\ 1 & \text{se } r \geq 0 \end{cases}$$

onde y é o classificador e r é a Taxa de Retorno.

A distribuição de classes dentro de cada um dos conjuntos de dados é apresentada na Tabela 2.

Tabela 2 – Divisão de classes.

Classe	Treinamento / Validação	Teste	Total
0	653 (58%)	229 (48%)	882 (55%)
1	464 (42%)	251 (52%)	715 (45%)

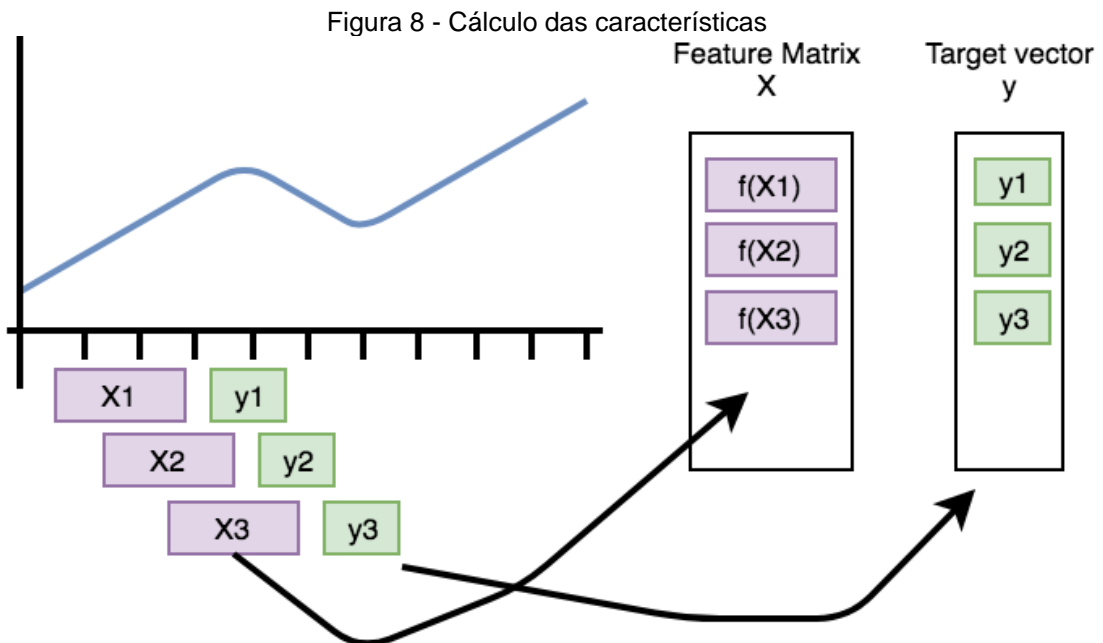
3.3 CARACTERÍSTICAS DE PREVISÃO

As características de previsão foram calculadas sobre a própria série a ser prevista, assim como sobre séries exógenas que possuem um possível caráter preditivo. As características de previsão consistem em:

- Indicadores Técnicos: Média Móvel, Taxa de Variação, relação entre médias móveis de volume;
- Volatilidade (desvio padrão);
- Relações entre os preços de abertura, fechamento, máximo e mínimo;
- Séries exógenas: cotações das ações que compõem o Índice Bovespa, o preço do dólar/real, o preço do petróleo e preço do ouro.

O processo de cálculo das características pode ser observado na Figura 8. As características são calculadas em uma sub-série (em lilás) que antecede o momento a ser realizada a previsão (em verde). É importante observar que a escolha do número de períodos para o cálculo desses indicadores influencia diretamente nas características extraídas do mercado. Por exemplo, a Média Móvel de 50 dias vai capturar tendências de mais longo termo que a Média Móvel de 10 dias. Por isso, cada

um dos indicadores foi calculado em 5 diferentes janelas de tempo, de 5, 10, 20, 50 e 100 períodos.



Fonte: TSFRESH, 2018

3.3.1 Cálculo das características

No campo das finanças, o fato da cotação de um ativo estar a R\$20 ou ter um volume de 10000 negociações, por exemplo, é de pouca importância quando se trata da previsão dos movimentos futuros dos preços. Ao invés dos valores brutos, é mais relevante que as entradas da Rede Neural forneçam valores relativos à outras variáveis. Por exemplo, em vez de informar o preço de uma cotação, ela deve indicar a taxa de variação do preço em relação a valores anteriores, e em vez de informar o volume de negociações, a característica deve indicar o quanto esse volume está abaixo ou acima da média. Isso irá prevenir a Rede Neural de focar se um ativo financeiro possui um valor alto ou baixo e permitirá uma maior generalização das características e de suas relações (VANSTONE & FINNIE, 2008).

Com exceção das características que são calculadas a partir dos preços de abertura, máximo e mínimo, optou-se por utilizar o valor de fechamento como referência para os cálculos. Tal decisão se justifica pelo fato de que esse valor é utilizado para estimar a variação percentual em relação ao dia anterior para os papéis

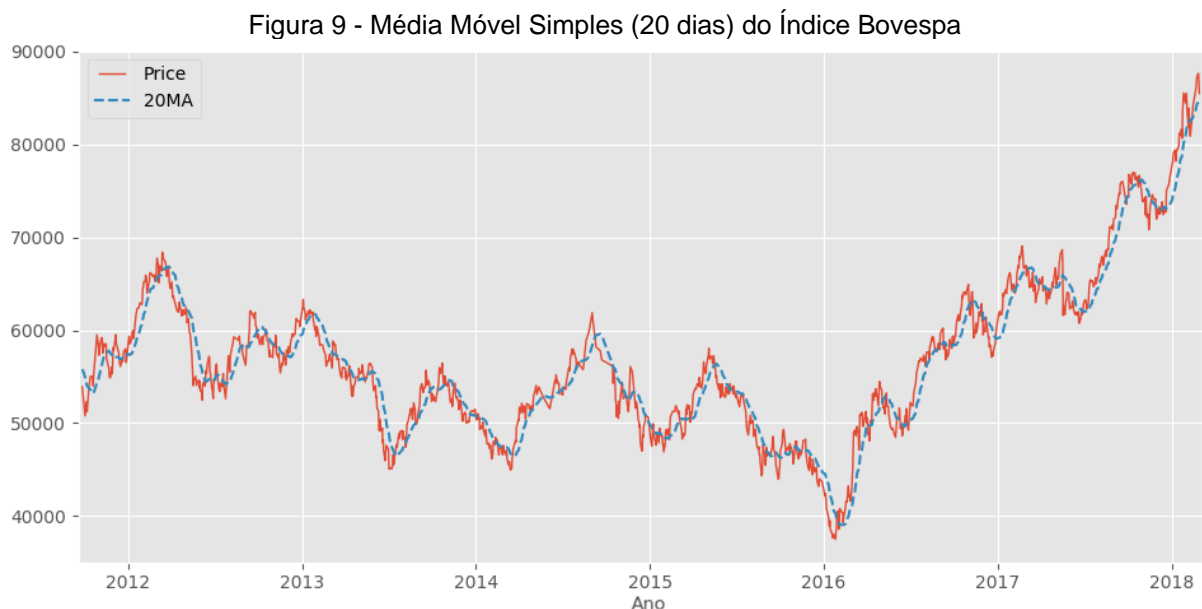
negociados em bolsa e, conforme apresentado por DOW C.H. (1902) seria a representação do valor a ser pago aceito pelos investidores que possuem o melhor nível informacional para formarem posições para o dia seguinte. As características foram calculadas como segue:

- **Média Móvel Simples:** é uma versão suavizada do preço do ativo financeiro. Para ser utilizada como entrada da Rede Neural, a Média Móvel foi calculada de acordo com a Equação (25).

$$\text{Média Móvel (pré - processada)} = \frac{\text{Preço}}{\text{Média Móvel}} - 1 \quad (25)$$

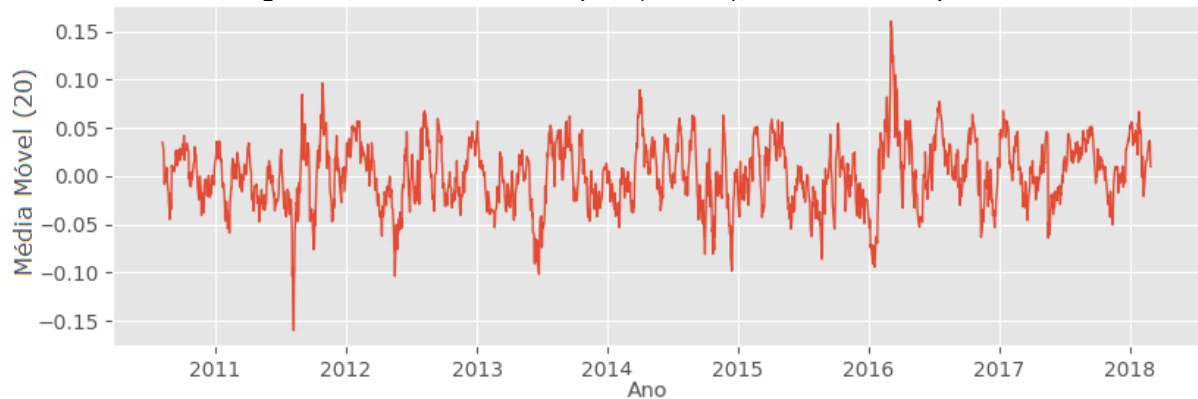
Dessa forma, a variável assumirá valores positivos se o preço for maior que sua Média Móvel, negativos se o preço for menor que a Média Móvel e iguais a zero no cruzamento entre as curvas. Assim é possível resumir o comportamento da Média Móvel em relação ao preço original em uma única variável de entrada.

Na Figura 9 é apresentado o aspecto da Média Móvel de 20 dias juntamente com as cotações do Índice Bovespa.



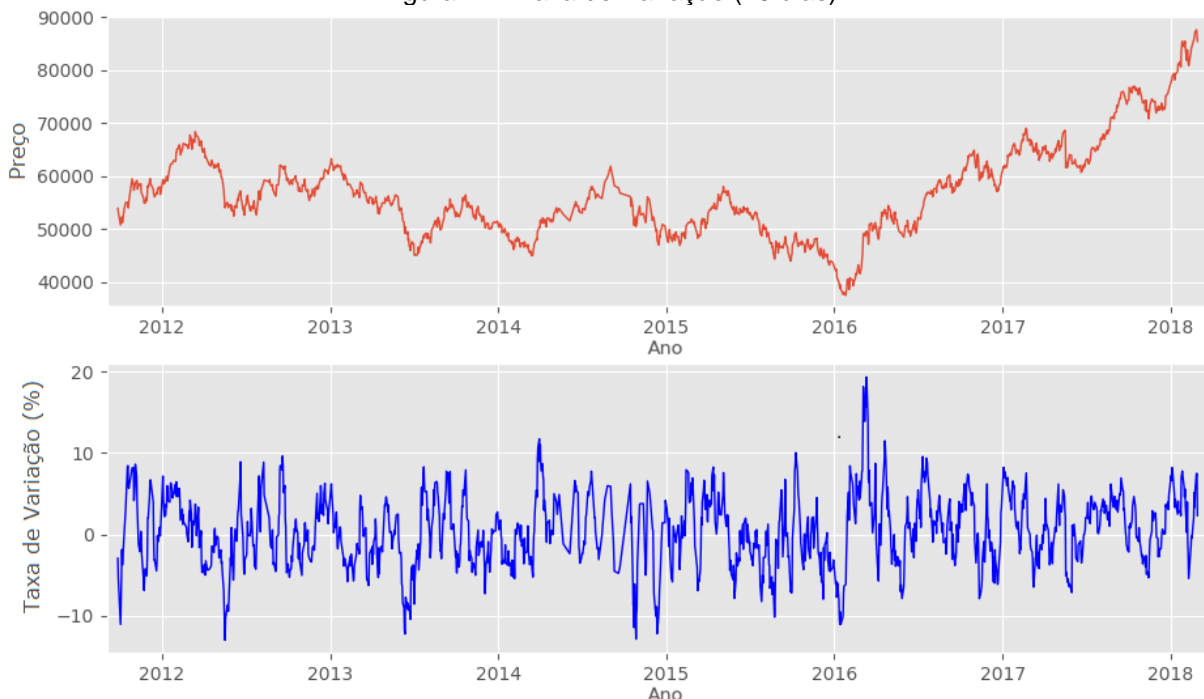
Na Figura 10 é apresentado o aspecto da Média Móvel de 20 dias após aplicado o pré-processamento da Equação (25).

Figura 10 - Média Móvel Simples (20 dias) do Índice Bovespa



- **Taxa de Variação:** foi calculada de acordo com a Equação (1). Não foi necessário aplicar cálculos suplementares à Taxa de Variação nesta etapa uma vez que ela já possui valores relativos a períodos anteriores. A cotação do Índice Bovespa e o aspecto da Taxa de Variação calculada em um período de 10 dias são apresentados na Figura 11, respectivamente.

Figura 11 - Taxa de Variação (10 dias)



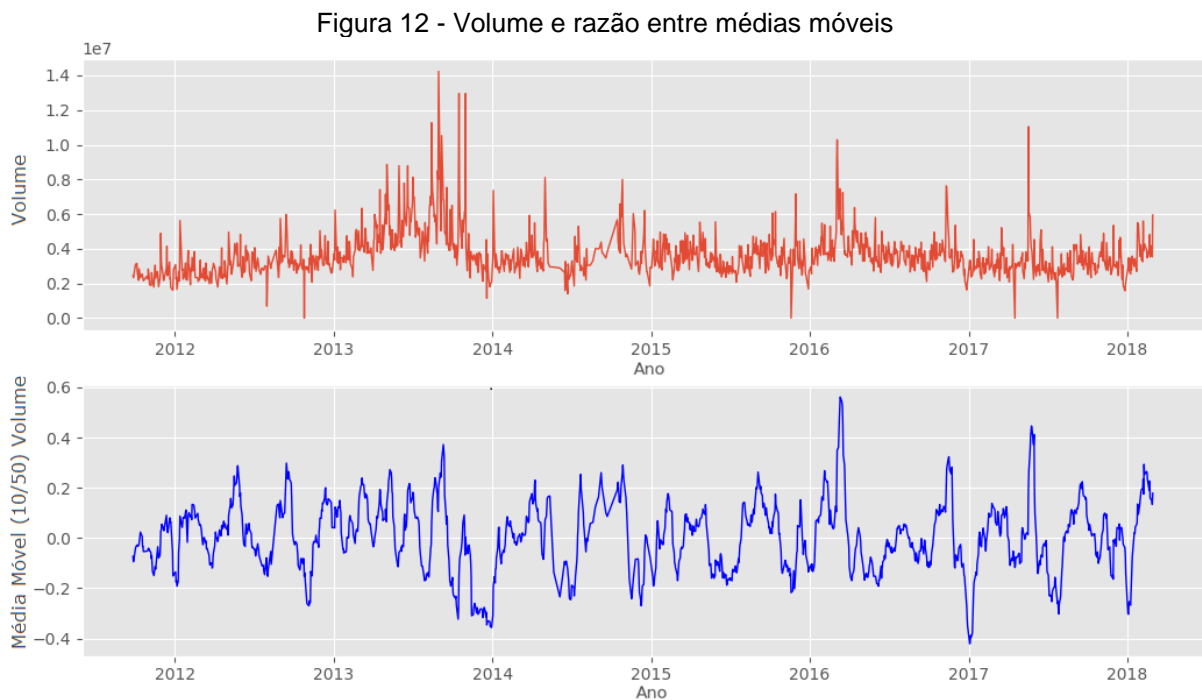
- **Volume de transações:** o volume de transações realizadas é um fator importante de séries temporais financeiras, pois assim como o preço, padrões de volume também existem nessas séries e, portanto podem ser utilizados na previsão

de preços. No entanto, assim como os preços, os volumes brutos possuem pouco significado comparados aos volumes relativos, portanto um bom indicador dos níveis de volume é a razão entre médias móveis, como mostra a Equação (26):

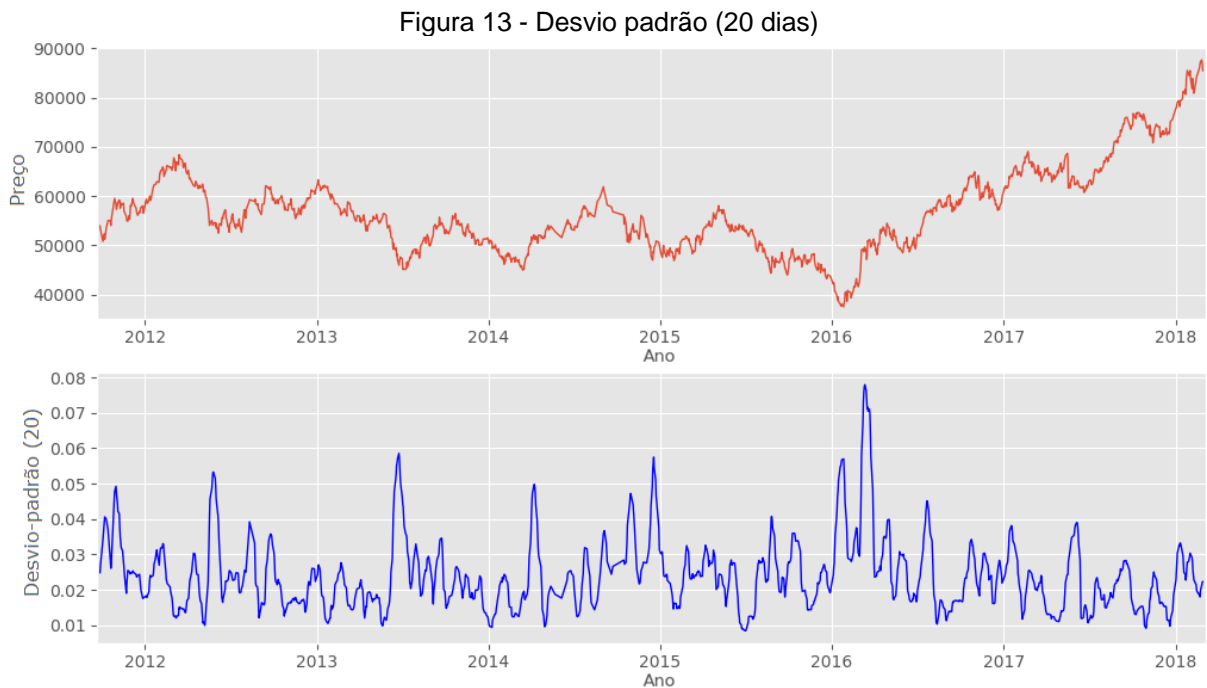
$$\text{Volume (pré - processado)} = \frac{\text{Média Móvel Volume 1}}{\text{Média Móvel Volume 2}} - 1 \quad (26)$$

onde a Média Móvel 1 é calculada em um período menor que a Média Móvel 2.

Na Figura 12 são apresentados o volume de transações do Índice Bovespa e a razão entre as médias móveis de 10 e 50 períodos.



- **Desvio padrão móvel:** o desvio padrão é uma característica importante pois ele estima a volatilidade do mercado. A fim de se desprezar os níveis de preço, seus valores foram divididos pela média móvel, resultando em um valor de desvio percentual em relação à média. Na Figura 13 são apresentados o Índice Bovespa e o comportamento do desvio padrão para uma janela de 20 dias, respectivamente.



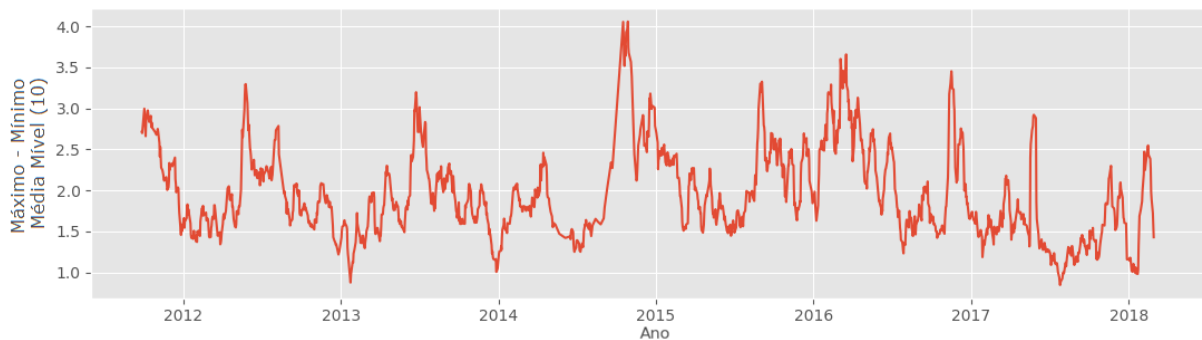
- **Preços de abertura, fechamento, máximo e mínimo:** o cálculo dos indicadores até aqui foram realizados com os preços de fechamento, no entanto os preços de abertura, de máximo e de mínimo podem conter informações importantes para a descoberta de padrões. Uma das características calculadas foi a volatilidade diária, calculada pela taxa percentual de variação entre o máximo e o mínimo do período, de acordo com a Equação (27).

$$Preço\left(\frac{m\acute{a}ximo}{m\acute{m}imo}\right) = \frac{Preço_{m\acute{a}ximo}(n) - Preço_{m\acute{m}imo}(n)}{Preço_{m\acute{m}imo}(n)} \times 100 \quad (27)$$

onde os preços máximo e mínimo são calculados no mesmo período n .

Partindo-se da hipótese de que um comportamento suavizado das características, assim como foi feito para os outros indicadores calculados até aqui, é mais significativo para previsões de múltiplos períodos adiante, foi calculada a média móvel da volatilidade diária nas mesmas janelas de tempo utilizadas para os outros indicadores. O comportamento da média móvel de 10 dias da volatilidade diária é apresentado na Figura 14.

Figura 14 - Relação entre os preços máximo e mínimo (10 dias)



Outra característica extraída foi a variação nos preços entre a abertura e o fechamento do período. Ela é calculada conforme a Equação (28).

$$\text{Preço}\left(\frac{\text{fechamento}}{\text{abertura}}\right) = \frac{\text{Preço}_{\text{fechamento}}(n) - \text{Preço}_{\text{abertura}}(n)}{\text{Preço}_{\text{abertura}}(n)} \times 100 \quad (28)$$

onde os preços de fechamento e de abertura são calculados no mesmo período n .

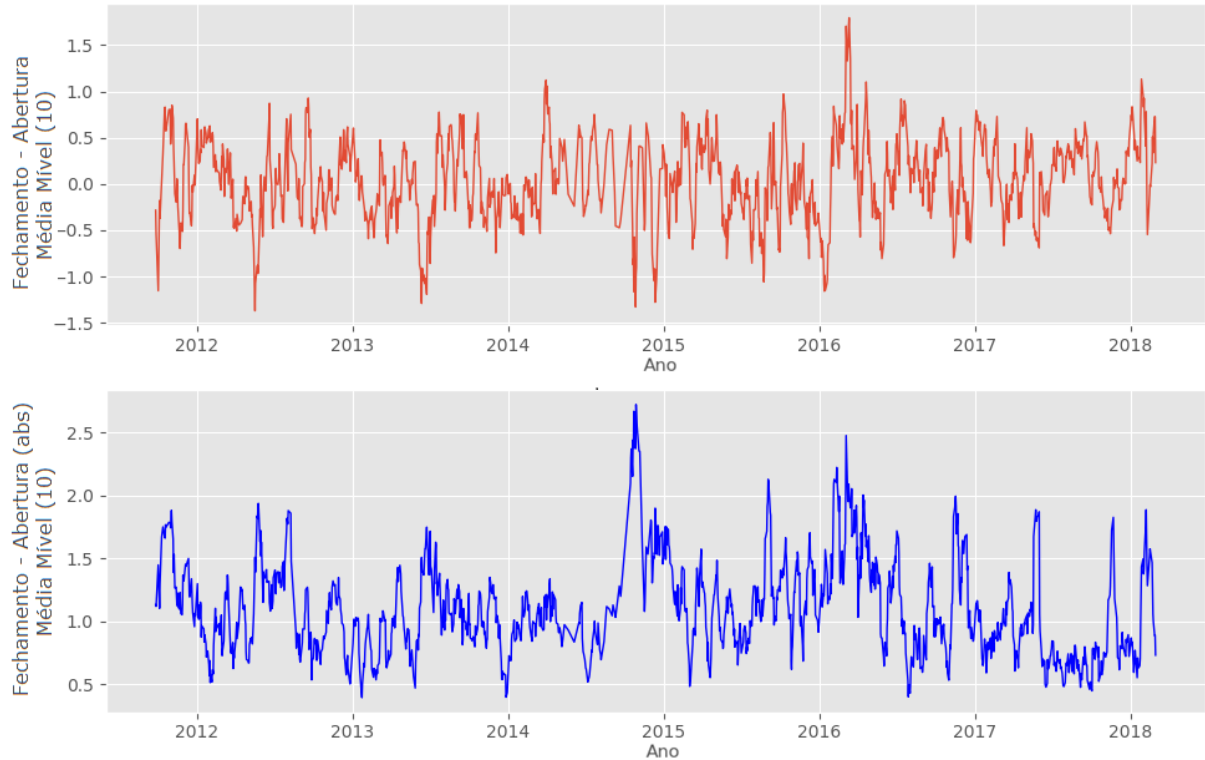
Diferentemente da volatilidade diária, essa característica pode apresentar valores positivos ou negativos. Dessa forma, o cálculo da média móvel foi realizado para os valores reais e também para os valores absolutos. A média móvel dos valores reais irá indicar o comportamento do equilíbrio entre as altas e as quedas, enquanto que a média móvel dos valores absolutos irá indicar a magnitude desses movimentos. A média móvel de 10 dias da relação entre os preços de fechamento e de abertura é apresentado na Figura 15 para seus valores reais e absolutos, respectivamente.

Outra característica extraída é a relação entre o preço de abertura de um período e o preço de fechamento do período anterior. A disparidade entre esses valores reflete as expectativas e o interesse dos investidores causados por anúncios das corporações, desastres naturais, entre outras notícias (INVESTOPEDIA, 2018). Seu cálculo é feito conforme a Equação (29).

$$\text{Preço}\left(\frac{\text{abertura}(n)}{\text{fechamento}(n-1)}\right) = \frac{\text{Preço}_{\text{abertura}}(n) - \text{Preço}_{\text{fechamento}}(n-1)}{\text{Preço}_{\text{fechamento}}(n-1)} \times 100 \quad (29)$$

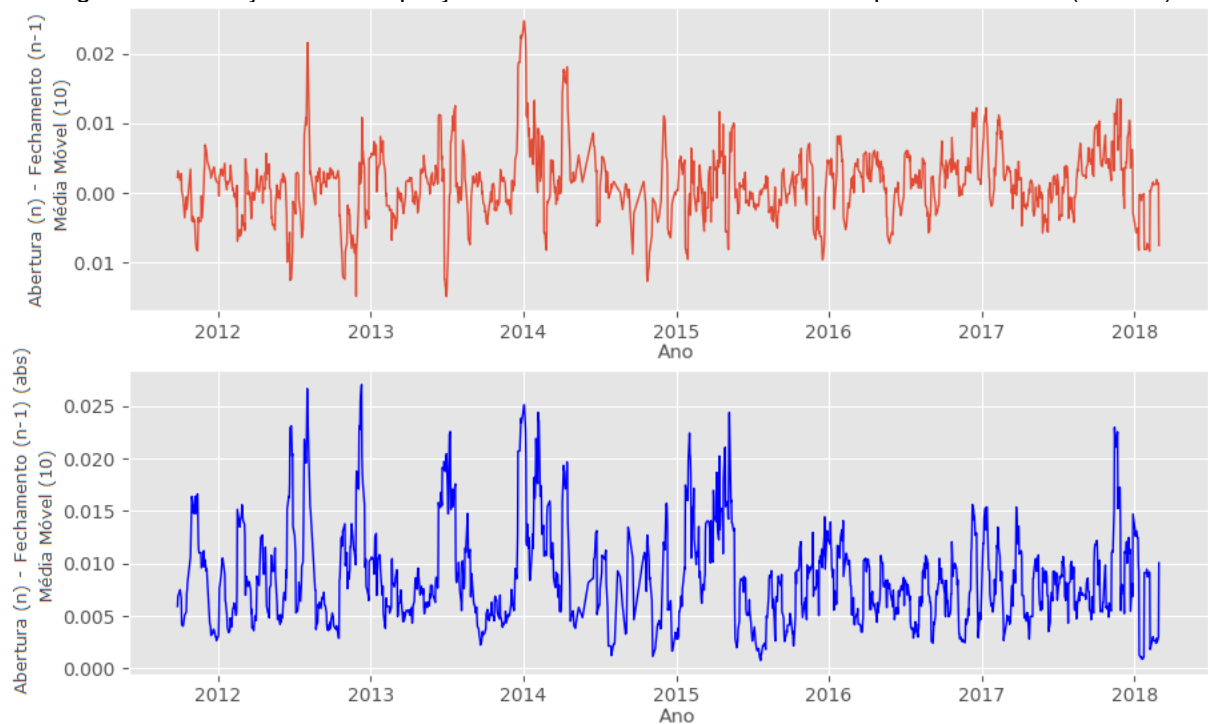
onde os preços com a notação (n) são calculados no período atual e os preços com a notação $(n-1)$ são calculados no período anterior.

Figura 15 - Relação entre os preços de fechamento e de abertura (10 dias)



O comportamento da média móvel de 10 dias da relação entre os preços de abertura e de fechamento do período anterior é apresentado na Figura 16 para seus valores reais e absolutos, respectivamente.

Figura 16 - Relação entre os preços de abertura e de fechamento do período anterior (10 dias)



- **Séries exógenas:** Além das características extraídas da própria série a ser prevista foram utilizadas as cotações de 52 dos 67 papéis que compunham o Índice Bovespa no último dia da base de dados (ANEXO). Algumas ações não foram utilizadas por falta de disponibilidade na internet ou por serem ações que existem há pouco tempo e, portanto, possuem somente dados históricos recentes.

Também foram incluídas as cotações do dólar/real, do petróleo e do ouro. Diversos estudos sugerem que essas variáveis tem influência sobre o mercado internacional embora poucos estudos tenham sido feitos para o mercado brasileiro. (LEE & ZENG, 2011; RAHMAN & MUSTAFA, 2011; BAHMANI-OSKOOEE & SOHRABIAN, 1992). Dessas séries foram utilizadas como características as taxas de variação dos preços, calculados pela Equação (1). O comportamento da taxa de variação de 10 dias do dólar/real, do petróleo e do ouro é apresentado nas Figuras 17, 18 e 19, respectivamente.

Figura 17 - Taxa de Variação (10 dias) da cotação do Dólar/Real

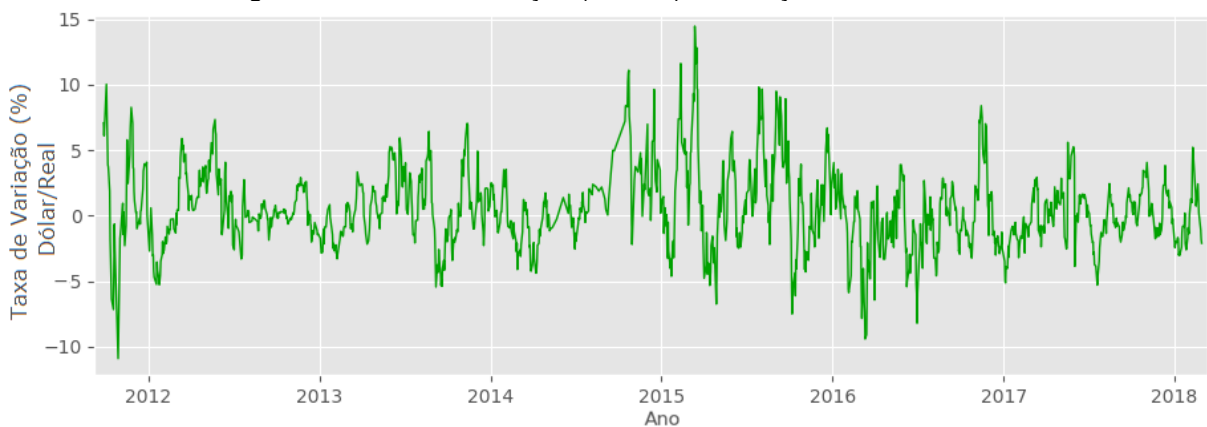
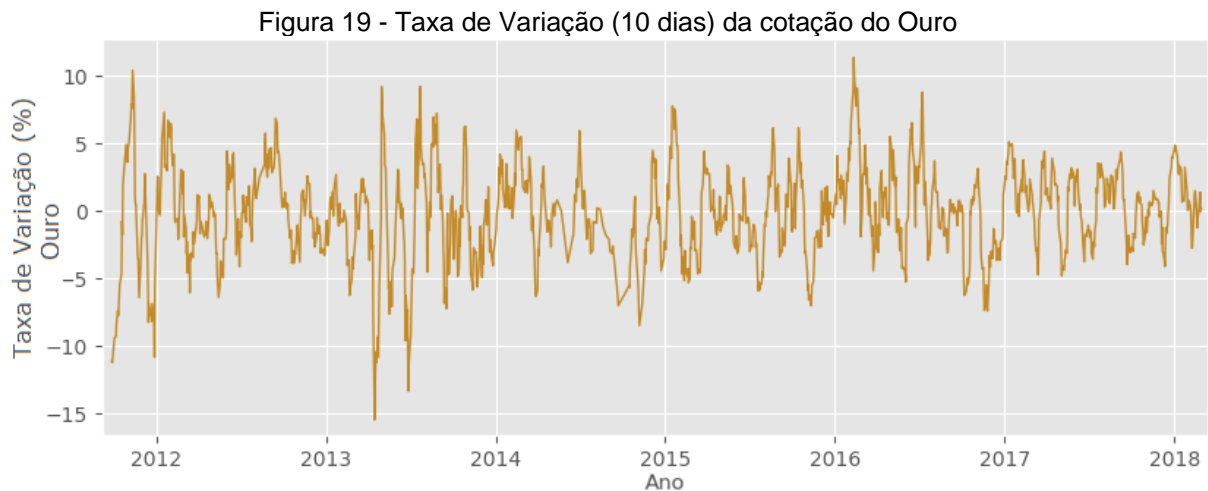


Figura 18 - Taxa de Variação (10 dias) da cotação do Petróleo





A fim de se dar a mesma importância para cada uma das características e acelerar o treinamento da rede, as variáveis foram normalizadas pela média e pelo desvio-padrão, conforme a Equação (30). Dessa forma, as variáveis assumirão valores na mesma ordem de grandeza, assumindo, na maior parte do tempo, valores entre -1 e 1.

$$z = \frac{x - \mu}{\sigma} \quad (30)$$

onde z é a variável normalizada, x o valor bruto da variável, μ a média da população e σ o desvio-padrão da população.

O cálculo da média e do desvio-padrão foram realizados sobre os dados de treinamento/validação, e os dados de teste foram normalizados baseados nesses valores. Dessa forma, as previsões no conjunto de teste não serão enviesadas pela contabilização de informações que estão no futuro, simulando, assim, uma situação real, onde os dados de teste são totalmente desconhecidos tanto no treinamento quanto durante a tomada de decisões.

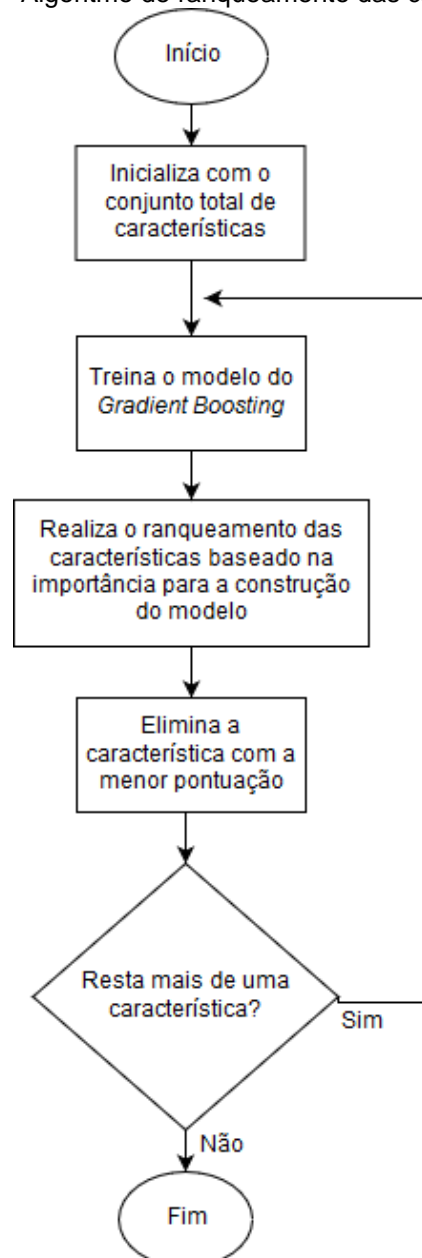
3.3.2 Ranqueamento das características

Após o cálculo e a normalização das variáveis, esta etapa tem por objetivo listar as características por ordem de importância. O ranqueamento das características foi realizado utilizando-se como modelo o método do *Gradient Boosting* em conjunto com o algoritmo de Eliminação Recursiva (*Recursive Feature Selection*).

O método *Gradient Boosting* foi implementado utilizando-se a biblioteca *xgboost 0.72* do Python. O algoritmo XGBoost fornece uma pontuação que indica o quão útil

cada variável foi na construção das árvores de decisão do modelo. A importância de uma característica é calculada pelo número de vezes que sua utilização melhora a medida de performance da árvore. Assim, quanto mais uma característica é utilizada para tomar decisões nas árvores, maior será sua relativa importância. A importância final de cada característica é, então, calculada pela média de sua importância em cada árvore que compõe o modelo. Assim, essa importância é calculada explicitamente para cada característica, permitindo que sejam ranqueadas e comparadas umas às outras. O algoritmo utilizado para a realização do ranqueamento das características é apresentado na Figura 20.

Figura 20 - Algoritmo de ranqueamento das características



Dessa forma, pelo algoritmo de Eliminação Recursiva, a característica que estiver com a menor pontuação é eliminada e, então, um novo subconjunto de características é novamente classificado pelo *Gradient Boosting*. Assim, a construção do *ranking* é feita até que todas as características sejam eliminadas, sendo a que for eliminada por último é a característica mais importante do *ranking*.

3.4 MODELO PROPOSTO

A Rede Neural proposta é uma *Feedforward* implementada utilizando-se a biblioteca Keras 2.1.4 do Python. A rede foi treinada com o algoritmo de otimização *Adam*, o qual tem apresentado resultados superiores a outros algoritmos como o método do gradiente (KINGMA & BA, 2014). A função custo utilizada foi a de Entropia Cruzada, calculada pela Equação (31), já que ela tende a minimizar o problema da desaceleração no aprendizado causado pelas derivadas parciais das funções de ativação.

$$C = -\frac{1}{n} \sum_{n=1}^N [y_n \cdot \ln(\hat{y}_n) + (1 - y_n) \cdot \ln(1 - \hat{y}_n)] \quad (31)$$

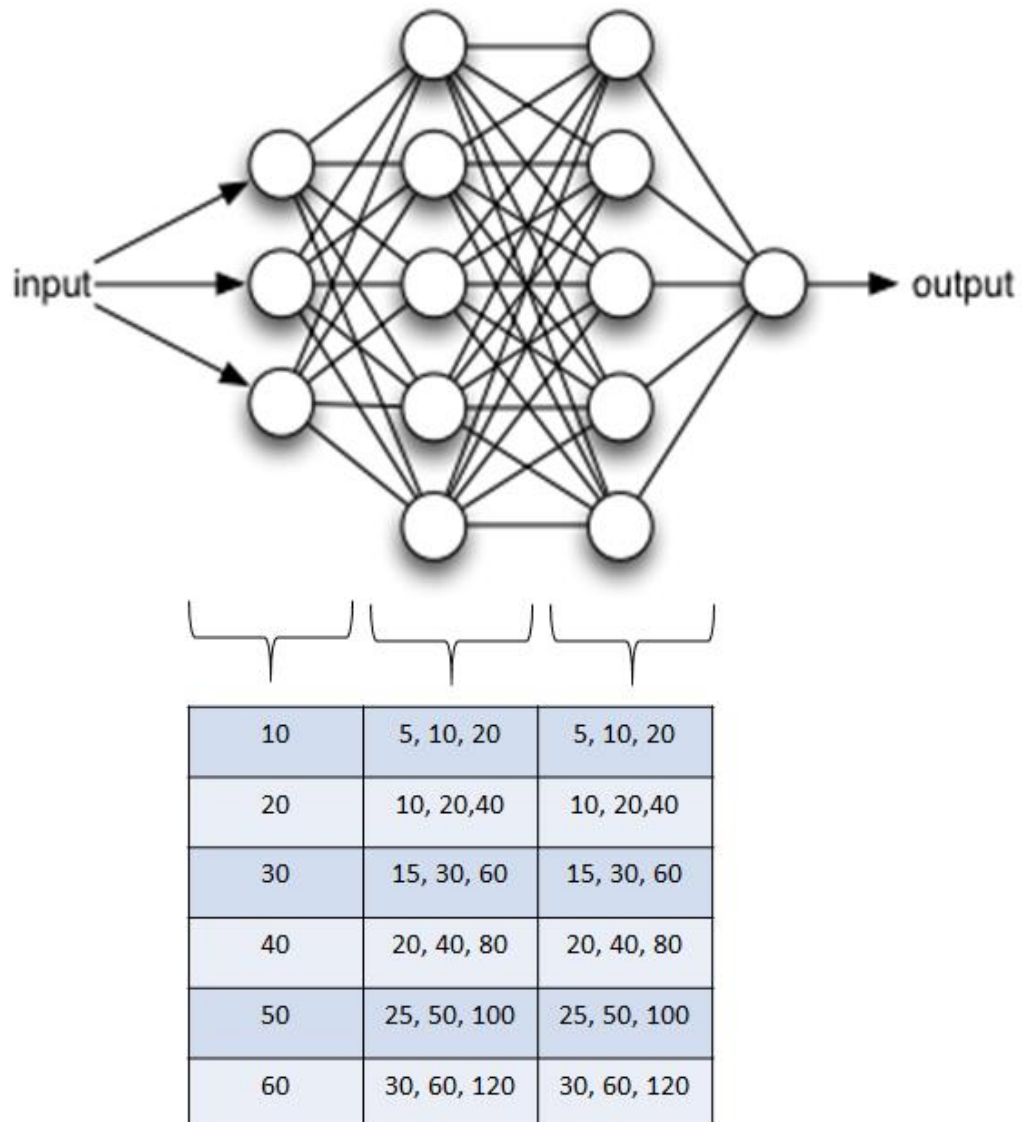
onde N é o número de itens de dados de treinamento, y_n é a saída desejada e \hat{y}_n é a ativação da unidade de saída do n-ésimo elemento dos dados de treinamento.

Foram realizados alguns testes preliminares nos quais foram utilizados diferentes números de camadas ocultas e diferentes funções de ativações para cada uma das camadas. Verificou-se que independente do número de características utilizadas, a rede obteve melhores taxas de acertos utilizando-se 2 camadas ocultas. Além disso, as funções de ativação que apresentaram melhores resultados foram a função tangente hiperbólica para os neurônios da primeira camada oculta, a função linear para os neurônios da segunda camada oculta e a função sigmoide para a camada de saída.

O ranqueamento das características não deixa explícita a quantidade de características que devem ser utilizadas. Portanto, como o seu número ideal não é conhecido, foram implementadas diferentes redes utilizando 10, 20, 30, 40, 50 e 60 das características mais bem classificadas no ranking. Para o número de neurônios das camadas ocultas utilizou-se 3 valores diferentes, sendo eles a quantidade de características, a metade da quantidade de características e o dobro da quantidade

de características, totalizando $6 \times 3 \times 3 = 54$ modelos. A Figura 21 ilustra o número de neurônios utilizados em cada camada oculta em função do número de características. A fim de se reduzir os efeitos da convergência a um mínimo local das Redes Neurais, cada modelo foi treinado 3 vezes e aquele com a maior taxa de acertos no conjunto de validação foi registrado.

Figura 21 - Número de neurônios dos modelos em cada camada

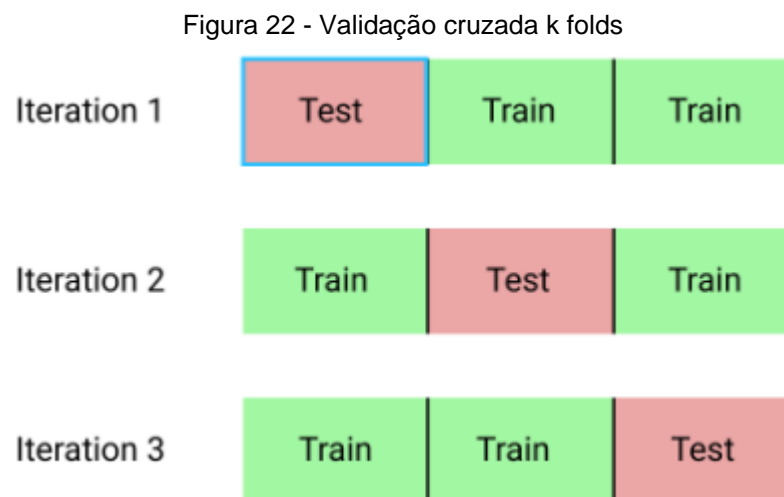


Fonte: Adaptado de DOLHANSKY, 2018

A fim de se avaliar a capacidade de generalização da rede para dados fora do conjunto de treinamento foi utilizada a técnica *stratified k folds* de validação cruzada. Na técnica *k folds*, a amostra original é dividida de forma aleatória em *k* sub-amostras. Uma entre elas é utilizada como conjunto de validação e as *k-1* restantes são

utilizadas para o treinamento. O processo de validação cruzada é repetido k vezes, utilizando-se cada um dos k conjuntos uma vez como conjunto de validação. A taxa de acertos do modelo é, então, obtida pela média dos k resultados.

Visto que a taxa de acertos é dependente da divisão aleatória dos conjuntos, foi feita uma estratificação proporcional dos dados, de forma que cada conjunto é construído de forma a conter aproximadamente a mesma distribuição de classes do conjunto original. Neste trabalho foi utilizado o valor de k igual a 3, pois este valor apresentou resultados melhores em alguns testes preliminares. A Figura 22 ilustra como é feita a divisão e o treinamento utilizando-se o método *k folds*.



Fonte: Adaptado de MEDIUM, 2018

Utilizando-se essa técnica de validação cruzada são construídos 3 modelos. Desses 3 modelos, é criado um modelo final através de uma técnica de *ensembling*, que tem por objetivo obter um modelo mais forte a partir dos outros modelos. O método utilizado é o método de votação, no qual a previsão final é aquela mais votada pela 3 redes. Para a avaliação do modelo final é utilizado um conjunto de dados de teste os quais não fazem parte dos conjuntos de treinamento e de validação e são os dados mais recentes do conjunto total.

A fim de prevenir um sobre-treinamento do modelo utilizou-se um ponto de verificação durante o treinamento que tem por objetivo salvar o modelo treinado a cada época somente se a taxa de acertos do conjunto de validação apresentar uma melhora em relação à época anterior. Caso mais de uma época apresentar a melhor taxa de acertos durante o treinamento, o desempate é feito pela função custo. Essa técnica tem por objetivo evitar um sobreajuste do modelo aos dados de treinamento,

melhorando, assim, sua generalização para dados desconhecidos. A Figuras 23 e 24 ilustram a taxa de acertos e o custo, respectivamente, do conjunto de treino (em vermelho) e de validação (em azul) durante o treinamento da rede com 30 características de entrada e possuindo 30 e 15 neurônios nas camadas ocultas.

Figura 23 - Taxa de acertos dos conjuntos de treinamento e de validação

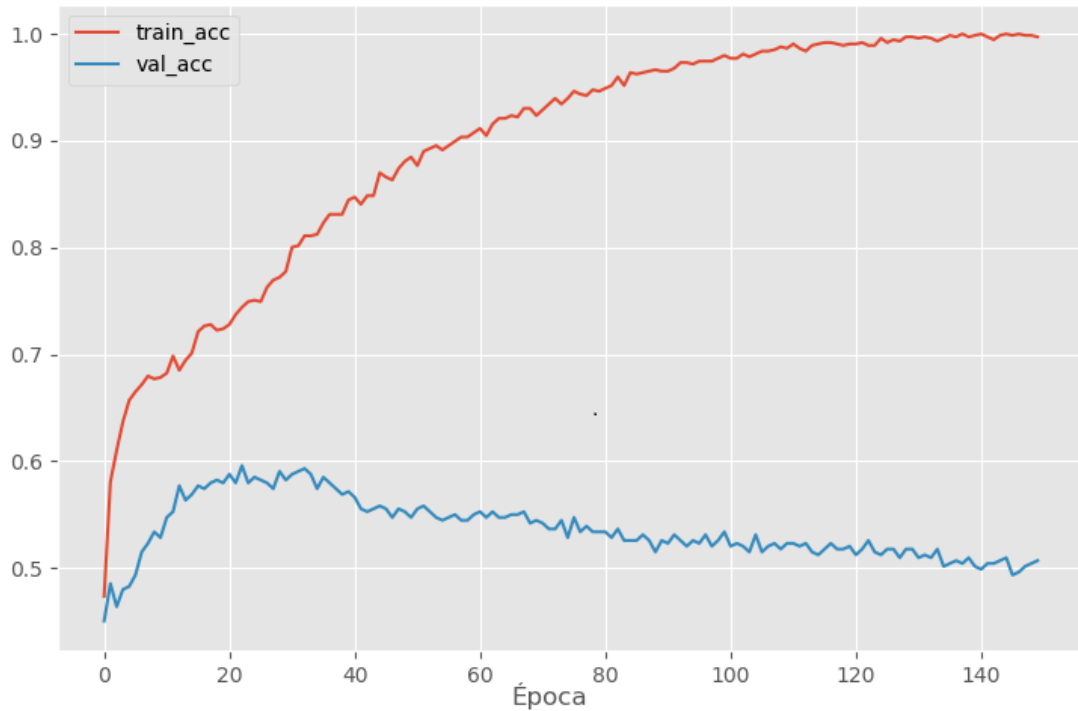
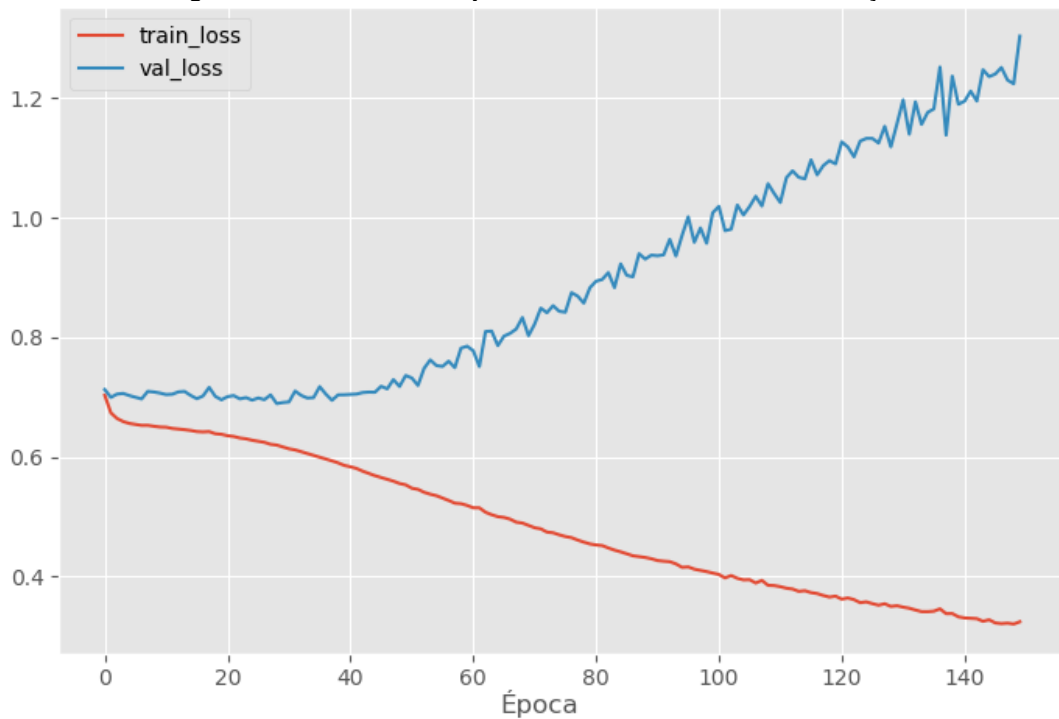


Figura 24 - Custo dos conjuntos de treinamento e de validação



Observa-se que apesar de a taxa de acertos e o custo dos dados de treinamento (em vermelho) continuarem melhorando, a partir da época 30, a taxa de acertos do conjunto de validação (em azul) volta a cair e o custo a aumentar, indicando que um sobreajuste está ocorrendo. Nesse caso o modelo é salvo em torno da época 30, onde o modelo obteve o melhor resultado.

4 RESULTADOS E DISCUSSÕES

A classificação das características utilizando-se o algoritmo *Gradient Boosting* com o método de Eliminação Recursiva obteve os resultados mostrados na Tabela 3, na qual são apresentadas as 60 características mais importantes identificadas pelo método. Os valores entre parênteses indicam o número de períodos para o cálculo das características.

Tabela 3 – Ranking de características.

Ranking	Característica
1	Desvio-padrão (50)
2	Desvio-padrão (100)
3	Abertura (n) / Fechamento (n-1) [Absoluto] (50)
4	Média móvel (100)
5	Média Móvel volume 10/50
6	Média Móvel volume 50/100
7	Petrobrás (5)
8	Telefônica Brasil (20)
9	Embraer (10)
10	Ouro (10)
11	Brasil Foods (5)
12	Equatorial Energia (10)
13	Companhia Paranaense de Energia (20)
14	Petróleo (10)
15	Cielo (5)
16	WEG (5)
17	Companhia Siderúrgica Nacional (10)
18	Braskem (20)
19	Companhia Paranaense de Energia (10)
20	Natura (20)
21	Cyrela (20)
22	Taxa de variação (50)
23	Desvio-padrão (20)
24	Abertura (n) / Fechamento (n-1) (1)
25	Eletrobrás (5)
26	Brasil Foods (20)
27	Máximo / Mínimo (50)
28	Companhia Brasileira de Distribuição (20)
29	Metalúrgica Gerdau (10)
30	Telefônica Brasil (10)
31	Companhia de Concessões Rodoviárias (5)
32	Eletrobrás (20)
33	Localiza (10)
34	MRV Engenharia (20)
35	Petrobrás (5)
36	Ambev (10)
37	Dólar / Real (20)
38	Banco Bradesco (10)

Tabela 3 – Ranking de características - conclusão

39	Abertura (n) / Fechamento (n-1) (100)
40	Abertura (n) / Fechamento (n-1) [Absoluto] (20)
41	Média Móvel volume 5/50
42	Magazine Luiza (5)
43	Gerdau S.A. (5)
44	Ultrapar (5)
45	Natura (10)
46	Equatorial Energia (20)
47	Abertura (n) / Fechamento (n-1) (20)
48	Marfrig Alimentos (20)
49	Média Móvel volume 10/20
50	Sabesp (20)
51	Multiplan (20)
52	Brasil Bolsa Balcão (10)
53	Embraer (5)
54	Cosan (10)
55	Máximo / Mínimo (5)
56	Ultrapar (10)
57	BR Malls (20)
58	Abertura (n) / Fechamento (n-1) (50)
59	Equatorial Energia (5)
60	Petróleo (5)

As taxas de acerto das redes para os dados de validação utilizando-se diferentes números de características são apresentadas nas Tabelas 4, 5, 6, 7, 8, 9, onde N1 representa o número de neurônios da primeira camada oculta e N2 o número de neurônios da segunda camada oculta.

Tabela 4 – Taxa de acertos de validação com 10 características.

N1 \ N2	5	10	20
5	59,53%	60,06%	58,64%
10	60,34%	60,61%	58,91%
20	58,81%	59,43%	59,89%

Tabela 5 - Taxa de acertos de validação com 20 características.

N1 \ N2	10	20	40
10	56,22%	57,47%	56,58%
20	56,04%	56,58%	55,68%
40	57,56%	57,57%	56,31%

Tabela 6 - Taxa de acertos de validação com 30 características.

N1 \ N2	15	30	60
15	57,83%	56,58%	56,95%
30	57,92%	59,37%	58,72%
60	57,39%	60,34%	59,00%

Tabela 7 – Taxa de acertos de validação com 40 características.

N1 \ N2	20	40	80
20	57,56%	57,38%	57,12%
40	58,63%	57,83%	58,55%
80	58,63%	59,08%	57,84%

Tabela 8 – Taxa de acertos de validação com 50 características.

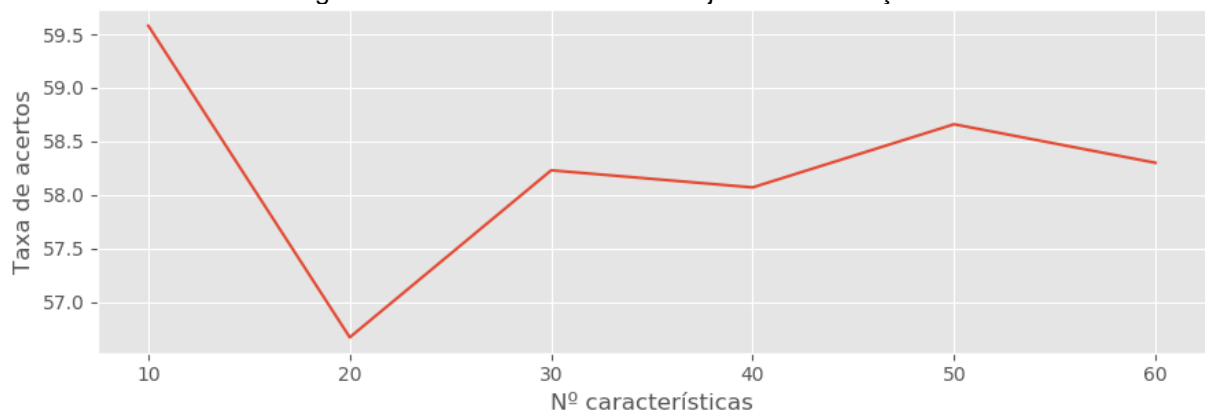
N1 \ N2	25	50	100
25	56,94%	59,63%	57,49%
50	59,71%	56,04%	60,97%
100	60,42%	57,92%	58,81%

Tabela 9 – Taxa de acertos de validação com 60 características.

N1 \ N2	30	60	120
30	57,75%	57,74%	57,74%
60	57,66%	57,38%	59,53%
120	60,06%	57,84%	59,00%

É possível observar que as taxas de acertos para o conjunto de validação ficou compreendida entre 56% e 61%, apresentando pouca variação em função da quantidade de características e de neurônios utilizados. A taxa de acertos média do conjunto de validação em função do número de características de entrada é apresentada na Figura 25.

Figura 25 - Taxa de acertos do conjunto de validação



Para o conjunto de testes as taxas de acerto das redes são apresentadas nas Tabelas 10, 11, 12, 13, 14, 15.

Tabela 10 – Taxa de acertos de teste com 10 características.

N1 \ N2	5	10	20
5	49,58%	51,66%	49,17%
10	52,29%	48,54%	50,00%
20	52,50%	47,50%	47,29%

Tabela 11 – Taxa de acertos de teste com 20 características.

N1 \ N2	10	20	40
10	53,95%	48,96%	48,75%
20	52,91%	51,87%	52,71%
40	54,16%	54,79%	52,08%

Tabela 12 – Taxa de acertos de teste com 30 características.

N1 \ N2	15	30	60
15	54,58%	56,04%	55,41%
30	57,50%	55,62%	54,58%
60	55,83%	54,17%	53,33%

Tabela 13 – Taxa de acertos de teste com 40 características.

N1 \ N2	20	40	80
20	57,45%	56,45%	55,41%
40	58,13%	57,75%	55,62%
80	56,62%	55,21%	55,21%

Tabela 14 – Taxa de acertos de teste com 50 características.

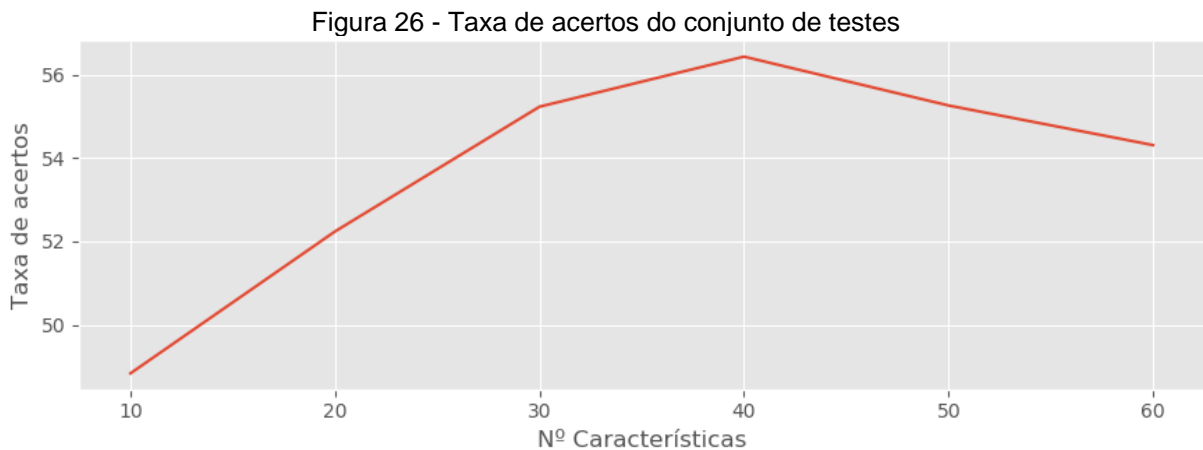
N1 \ N2	25	50	100
25	54,37%	54,79%	56,00%
50	55,83%	54,37%	54,91%
100	55,83%	56,87%	54,37%

Tabela 15 – Taxa de acertos de teste com 60 características.

N1 \ N2	30	60	120
30	55,21%	55,83%	56,67%
60	52,71%	52,91%	53,54%
120	54,79%	54,17%	52,92%

Observa-se que as taxas de acertos do conjunto de testes não tiveram uma variação significativa em função da quantidade de neurônios. As redes que obtiveram as maiores taxas de acertos foram aquelas que incluíram 30, 40 ou 50 características de entrada. Mais especificamente, as redes que obtiveram o melhor resultado foram aquelas incluindo 40 características de entrada, com uma taxa de acertos média de

56,43% e uma taxa de acertos máxima de 58,13% para a rede com 40 neurônios na primeira camada oculta e 20 neurônios na segunda camada oculta. A taxa de acertos média do conjunto de testes em função do número de características é apresentada na Figura 26.



A Matriz de Confusão do conjunto de teste da rede utilizando 40 características de entrada, 40 neurônios na primeira camada oculta e 20 neurônios na segunda camada oculta é apresentada na Tabela 16.

Tabela 16 – Matriz de confusão da rede

		Previsto		
		0	1	Total
Real	0	171	58	229
	1	143	108	251
	Total	314	166	480

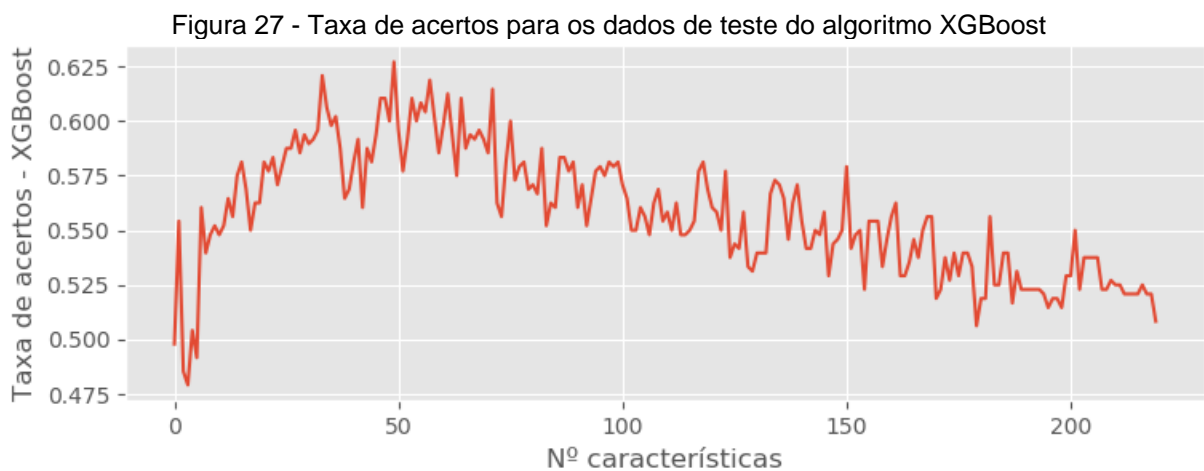
Sem a utilização da seleção de características as taxas de acerto das redes para o conjunto de teste, utilizando-se todas as 220 características, são apresentadas na Tabela 17.

Tabela 17 – Taxa de acertos de teste com 60 características.

N1 \ N2	110	220	440
110	50,14%	49,24%	50,28%
220	49,31%	51,60%	47,99%
440	51,39%	49,93%	48,54%

Observa-se que o resultado foi bastante inferior comparado às redes que utilizam a seleção de características. Além disso, muitas das taxas de acerto são inferiores à 50%, o que, pode-se dizer, é tão ruim quanto uma classificação aleatória.

Foi avaliada também as taxas de acerto para os dados de teste utilizando-se o algoritmo *Gradient Boosting* como modelo. A taxa de acerto em função do número de características utilizadas como entrada é apresentada na Figura 27.



Fonte: autoria própria

Observa-se que a taxa de acertos máxima foi de 62,71% com 49 características, o que é um resultado melhor do que o obtido com a Rede Neural. Além disso, o algoritmo foi executado com os parâmetros padrão, então é provável que ele ainda possa ser melhorado. No entanto, embora esse resultado indique o *Gradient Boosting* de Árvores de Decisão como um melhor modelo de previsão, a seleção de características foi implementada utilizando esse método, então é natural que ele tenha uma taxa de acertos mais alta.

5 CONCLUSÃO

Este trabalho teve por objetivo a aplicação de uma Rede Neural *Feedforward* na previsão das tendências do Índice Bovespa em um horizonte de tempo de uma semana. Foram calculadas 220 diferentes características consistindo em indicadores técnicos, volatilidade e nas cotações do dólar, do petróleo, do ouro, assim como das ações que compõem o índice. Foi realizado um ranqueamento das características utilizando-se como modelo o método *Gradient Boosting* em um algoritmo *Recursive Feature Elimination*. Foram implementadas diferentes redes utilizando diferentes números de neurônios e quantidades variadas das características mais bem classificadas.

Obteve-se uma taxa de acertos de 58,63% para o conjunto de validação e 58,13% para o conjunto de teste para a rede utilizando as 40 características mais bem classificadas pelo *ranking*. Utilizando-se todas as entradas pré-selecionadas a rede obteve uma taxa de acertos de 51,39%, mostrando a importância da seleção de características nesse problema. A taxa de acertos no conjunto de testes do algoritmo *Gradient Boosting* foi de 62,71% utilizando 49 características, o que é um resultado melhor que o obtido pela Rede Neural. Isso, no entanto, parece natural visto que a seleção de características foi implementada utilizando-se esse método.

Os resultados obtidos são similares aos de outros estudos da área, como Kim (2003), por exemplo, que obteve uma taxa de acertos de 57,83% na previsão do Índice Sul-Coreano, e Atsalakis & Psomatakis (2014) que obtiveram uma taxa de acertos média de 57,5% na previsão da cotação de ações. Outros trabalhos, como o de Kara *et al.* (2010), obtiveram uma taxa de acertos de 75,74% na previsão do Índice da Bolsa de Istambul para os dados de validação. No entanto, de acordo com Racine (2000), para que a validação cruzada funcione como uma ferramenta de seleção de modelos, é necessário que os conjuntos de treinamento e validação sejam independentes. O problema com séries temporais é que pontos adjacentes são altamente dependentes de modo que técnicas padrão de validação cruzada serão enviesadas e não cumprirão tão bem sua função. Portanto, como os autores não utilizam um conjunto de testes, é difícil saber a real capacidade de generalização da rede para dados fora dos conjuntos de treinamento e validação.

Prever o mercado financeiro é uma tarefa desafiadora, pois, por mais que existam padrões nas séries financeiras, ainda existe um grande fator psicológico e outros fatores que são muito complicados de modelar com precisão. No entanto, ainda que, na maior parte do tempo, o mercado se comporte de maneira imprevisível, a descoberta de padrões é possível, rejeitando, em parte, a Hipótese do Mercado Eficiente já que os preços das ações não se movimentam de forma completamente aleatória.

6 PROPOSTAS PARA TRABALHOS FUTUROS

Existem diversos caminhos que podem dar sequência a esse trabalho. Dentre eles está a inclusão de outras métricas ou séries que possam ter relação com o índice, assim como a utilização de outras técnicas de seleção de características. Podem ainda ser aplicados outros métodos de pré-processamento, como a discretização das variáveis e a aplicação de filtros que podem ajudar a minimizar os efeitos da aleatoriedade das séries financeiras.

Outra possibilidade é a utilização de outros métodos ou outras arquiteturas de redes como, por exemplo, as redes LSTM (*Long Short-Term Memory*). Esse tipo de rede é capaz de fazer previsões não só baseadas em um conjunto de entradas, mas também nas informações sequenciais armazenadas ao longo do tempo.

Este trabalho utilizou somente os valores diários do índice. É possível que possam existir informações nos dados de outros *timeframes* que podem ser úteis para as previsões. Uma possibilidade seria a construção de um *ensemble* de modelos que utilizem dados de diferentes bases de tempo, permitindo, assim, que o sistema de previsão extraia outras informações do comportamento do mercado ao longo do dia.

REFERÊNCIAS BIBLIOGRÁFICAS

ATSALAKIS, G. S. & PSOMATAKIS, M. A. Stock Market Trend Forecasting by Technical Analysis and Hybrid Genetically Evolved Neural Networks. *Journal of Computational Optimization in Economics and Finance*, Vol. 6, N. 3. 1941-3971, 2014.

ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques – part II: Soft computing methods. *Expert Systems with Applications*, 36(3), 5932–5941, 2008.

ATSALAKIS, G. S.; VALAVANIS, K. P. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications*, 36, 10696–10707, 2009.

BAHMANI-OSKOOEE M. & SOHRABIAN A. Stock prices and the effective exchange rate of the dollar, *Applied Economics Journal*, Volume 24, 1992.

BASU, S. The investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *Journal of Finance* 32: 663–682, 1977.

BLACK, F. The trouble with econometrics models. *Financial Analysis Journal* 4(5): 75–87, 1982.

BORODIN, A.; EL-YANIV, R; GOGAN, V. Can We Learn to Beat the Best Stock. *Journal of Artificial Intelligence Research* 21: 579-594, 2004.

BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. *Time Series Analysis, Forecasting and Control*, 4th ed., Wiley Series in Probability and Statistics, 2008.

CHANG R.P., RHEE S.G., STONE G.R., TANG N. How does the call market method affect price efficiency? Evidence from the Singapore Stock Market. *Journal of Banking & Finance*, Volume 32, Issue 10, 2008.

CHEN T. & GUESTRIN C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 978-1-4503-4232-2, 2016.

CHOUDHRY, R., and KUMUM, G. A hybrid machine learning system for stock market forecasting, *World Academy of Science, Engineering and Technology* 39.3: 315-318, 2008.

CORPORATE FINANCE INSTITUTE. Equity Capital Market. Disponível em: <<https://corporatefinanceinstitute.com/resources/knowledge/finance/equity-capital-market-ecm/>>. Visitado em Março de 2018.

CPC (Comitê de Pronunciamentos Contábeis). BM&F Bovespa. Disponível em: <<http://cpc.org.br/bovespa.htm/>>. Visitado em Março de 2018.

DECHOW, P. M. ; HUTTON, A. P.; MEULBROEK, L. ; SLOAN, R. G. Short-sellers, fundamental analysis, and stock returns *Journal of Financial Economics* 61: 77–106, 2001.

DOLHANSKY, B. Artificial Neural Networks: Mathematics of Backpropagation. Disponível em: <<http://briandolhansky.com/blog/2013/9/27/artificial-neural-networks-backpropagation-part-4>>. Visitado em Abril de 2018.

DOW C.H. Swings Within Swings. *The Wall Street Journal* of January, 1902.

FACELI K., LORENA A.C., GAMA J., CARVALHO A. Inteligência artificial: uma abordagem de aprendizado de máquina. 1ª Edição. Rio de Janeiro, LTC: 2011.

FAMA E.F. Short-Term Interest Rates as Predictors of Inflation. *The American Economic Review*, Vol. 65, No. 3, pp. 269-282, 1975.

FAMA E.F. Efficient Capital Markets: II. *The Journal of Finance*, Vol. 46, Issue 5, pp. 1575-1617, 1991.

FRIEDMAN J.H. Greedy Function Approximation: A Gradient Boosting Machine, *The Annals of Statistics*, Vol. 29, No. 5, pp. 1189-1232, 2001.

HAYKIN, Simon. Redes neurais: princípios e prática. 2ª Edição. Porto Alegre. Bookman: 2001.

HIEMSTRA, Y. Modeling structured nonlinear knowledge to predict stock market returns, *Chaos & Nonlinear Dynamics in the Financial Markets: Theory, Evidence and Applications*, Irwin, pp. 163-175, 1995.

HUANGA W., NAKAMORIA Y., WANGB S.Y., Forecasting stock market movement direction with support vector machine, *Computers & Operations Research* vol. 32, no. 10, 1820-6425, 2005.

INVESTOPEDIA. Financial Market. Disponível em: <<https://www.investopedia.com/terms/f/financial-market.asp>>. Visitado em Março de 2018.

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*, Volume 38, 5311-5319, 2010.

KIM K. J. Financial time series forecasting using support vector machines. *Neurocomputing* no. 55, 307-319, 2003.

KIM, H.; SHIN, K., A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets, *Applied Soft Computing*, Volume 7, Issue 2, Pages 569-576, 2007.

- KINGMA D.P. & BA J. Adam: A Method for Stochastic Optimization. Proceedings of the 3rd International Conference on Learning Representations (ICLR) : 1412.6980, 2014.
- LAWRENCE R. Using Neural Networks to Forecast Stock Market Prices, University of Manitoba, 1997.
- LEE C.C & ZENG J.H. The impact of oil price shocks on stock market activities: Asymmetric effect with quantile regression. Volume 81, 2011.
- LEVINSON M. Guide to Financial Markets. 4^a ed. London: The Economist, 2005.
- LEVY, R. Conceptual Foundations of Technical Analysis. Financial Analysts Journal. Vol. 22, No. 4: 83-89, 1966.
- LO A. W., MACKINLAY, A. C. Stock market prices do not follow random walks: evidence from a simple specification test, Rev. Financial Stud. 1: 41–66, 1988.
- LOH W.Y. Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Volume 1, Issue 1, 2011.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, Springer, v. 5, n. 4, p. 115–133, 1943.
- MEDIUM. Regression vs Classification Problems in Machine Learning. Disponível em: <<https://medium.com/@heyozramos/regression-vs-classification86d73c281c5e>>. Visitado em Abril de 2018.
- MURPHY, John J. Technical Analysis of the Financial Markets. 1^a ed. New York: New York Institute of Finance, 1999.
- NAEINI M.P., TAREMIAN H., HASHEMI H.B. Stock market value prediction using neural networks. International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 978-1-4244-7818-7, 2010.
- PRING, M. Technical Analysis Explained. 5^a Edição. McGraw-Hill, 2016.
- PAI P.F. & LIN C.S. A hybrid ARIMA and support vector machines model in stock price forecasting, Omega 33, 497–505, 2005.
- QIU, M; SONG, Y; AKAGI, F. Application of artificial neural network for the prediction of stock market returns: The case of the Japanese stock market. Chaos, Solitons and Fractals 85: 1–7, 2016.
- RACINE J. Consistent cross-validators model-selection for dependent data: hv-block cross-validation. Journal of Econometrics 99, 39-61, 2000.

RAHMAN M. & MUSTAFA M. Effects of Crude Oil and Gold Prices on US Stock Market: Evidence for USA from ARDL Bounds Testing. *Mathematics and Computers in Simulation* 81(9), 1910-1920, 2011.

ROCKEFELLER, Barbara. *Technical Analysis for Dummies*. 2ª ed. Indianapolis: Wiley Publishing, Inc., 2011.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

SAAD E. W., PROKHOROV D. V., WUNSCH D. C. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *Neural Networks, IEEE Transactions on*, IEEE, v. 9, n. 6, p. 1456–1470, 1998.

TAY F.E.H & CAO L. Application of support vector machines in financial time series forecasting, *Omega*, 29(4):309-317, 2001.

TSFRESH. Time Series Forecasting. Disponível em: <<http://tsfresh.readthedocs.io/en/latest/text/forecasting.html/>>. Visitado em Maio de 2018.

VANSTONE B. & FINNIE G. An empirical methodology for developing stock market trading systems using artificial neural networks, *Expert Systems with Applications* 36: 6668–6680, 2008.

WILDER, J. W. *New concepts in technical trading systems*. Hunter Publishing Company. Winston-Salem & Greensboro, NC: Trend Research. Library of Congress Card Catalog No. 78-60759, 1978.

WURGLER, Jeffrey. Financial markets and the allocation of capital. *Journal of Financial Economics*, n. 58, p. 187-214, 2000.

YAHOO FINANCE. Disponível em: <<https://finance.yahoo.com/>>. Visitado em Abril de 2018.

YOO, P. D.; KIM, M. H.; JAN, T. *Machine Learning Techniques and Use of Event Information for Stock, Market Prediction: A Survey and Evaluation*, 2005.

ZHANG, Y., & WU, L. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications*, 36(5), 8849–8854, 2009.

ANEXO - Ações incluídas nas características

CÓDIGO	AÇÃO
ABEV3	AMBEV S/A
BBAS3	BRASIL
BBDC3	BRADESCO
BBDC4	BRADESCO
BBSE3	BBSEGURIDADE
BRAP4	BRADESPAR
BRFS3	BRF SA
BRKM5	BRASKEM
BRML3	BR MALLS PAR
BTOW3	B2W DIGITAL
CCRO3	CCR SA
CIEL3	CIELO
CMIG4	CEMIG
CPFE3	CPFL ENERGIA
CPLE6	COPEL
CSAN3	COSAN
CSNA3	SID NACIONAL
CYRE3	CYRELA REALT
ECOR3	ECORODOVIAS
EGIE3	ENGIE BRASIL
ELET3	ELETROBRAS
ELET6	ELETROBRAS
EMBR3	EMBRAER
ENBR3	ENERGIAS BR
EQTL3	EQUATORIAL
FIBR3	FIBRIA
FLRY3	FLEURY
GBR4	GERDAU
GOAU4	GERDAU MET
GOLL4	GOL
HYPE3	HYPERA
IGTA3	IGUATEMI
ITSA4	ITAUSA
ITUB4	ITAUUNIBANCO
JBSS3	JBS
LREN3	LOJAS RENNER
MGLU3	MAGAZ LUIZA
MRFG3	MARFRIG
MRVE3	MRV
MULT3	MULTIPLAN
NATU3	NATURA

PCAR4	P.ACUCAR-CBD
PETR3	PETROBRAS
PETR4	PETROBRAS
QUAL3	QUALICORP
RADL3	RAIADROGASIL
RENT3	LOCALIZA
SBSP3	SABESP
TIMP3	TIM PART S/A
UGPA3	ULTRAPAR
USIM5	USIMINAS
VALE3	VALE
VIVT4	TELEF BRASIL
WEGE3	WEG
