

UNIVERSIDADE FEDERAL DE ALFENAS

AFONSO DE PAULA FELICIANO

**APLICAÇÃO DE UMA REDE NEURAL ARTIFICIAL LSTM NO ÍNDICE BOVESPA
E O IMPACTO DE NOTÍCIAS E ACONTECIMENTOS POLÊMICOS NA
PREVISÃO DOS DADOS**

Varginha/MG

2020

AFONSO DE PAULA FELICIANO

**APLICAÇÃO DE UMA REDE NEURAL ARTIFICIAL LSTM NO ÍNDICE BOVESPA
E O IMPACTO DE NOTÍCIAS E ACONTECIMENTOS POLÊMICOS NA
PREVISÃO DOS DADOS**

Trabalho de conclusão de curso apresentado ao Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas, como requisito parcial à obtenção do título de Especialista em Modelagem Científica e Tecnológica.

Orientador: Prof. Dr. Lázaro Eduardo da Silva

Coorientador: Prof. Dr. Leandro Ferreira

Varginha/MG

2020

AFONSO DE PAULA FELICIANO

**APLICAÇÃO DE UMA REDE NEURAL ARTIFICIAL LSTM NO ÍNDICE BOVESPA
E O IMPACTO DE NOTÍCIAS E ACONTECIMENTOS POLÊMICOS NA
PREVISÃO DOS DADOS**

A Banca examinadora abaixo-assinada, aprova a monografia apresentada como parte dos requisitos para obtenção do título de Especialista em Modelagem Científica e Tecnológica da Universidade Federal de Alfenas em parceria com o Centro Federal de Educação Tecnológica de Minas Gerais.

Aprovada em: ____ / ____ / ____

Profº. Dr. Lázaro Eduardo da Silva (Orientador) Assinatura:
Instituição: Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

Profº. Dr. Leandro Ferreira (Coorientador) Assinatura:
Instituição: Universidade Federal de Alfenas – UNIFAL/MG

Profº. Dr. André Rodrigues Monticeli Assinatura:
Instituição: Centro Federal de Educação Tecnológica de Minas Gerais – CEFET/MG

Profª. Drª. Cláudia Adam Ramos Assinatura:
Instituição: Universidade Federal de Alfenas – UNIFAL/MG

AGRADECIMENTOS

A Deus por sempre guiar os meus passos, concedendo-me sabedoria e determinação, e principalmente por colocar pessoas muito especiais no decorrer desta caminhada.

A minha família, que me apoiou em todos os momentos, acreditando no meu potencial e estando sempre presente em minhas batalhas, em especial ao meu pai Divino e minha mãe Sandra.

A minha companheira para todos os momentos, Alessandra, que sempre se mostrou paciente, compreendeu meus momentos de ausência e me serviu como fonte de inspiração durante a realização deste trabalho.

Ao Professor Dr. Lázaro Eduardo da Silva do Departamento de Computação e Engenharia Civil do Centro Federal de Educação Tecnológica de Minas Gerais – CEFET-MG que além de se manter sempre disponível e paciente, me proporcionou desafios, conhecimentos e me incentivou a explorar mais sobre Redes Neurais. Agradeço também por acreditar no meu trabalho e no meu potencial, desde a entrevista para ingresso nesta Pós-Graduação e principalmente pelo exemplo de orientador que se mostrou ser. Competente, responsável e amigo.

Ao Professor Dr. Leandro Ferreira do Instituto de Ciências Sociais Aplicadas da Universidade Federal de Alfenas - UNIFAL pelas aulas dinâmicas de Estatística e incrível coorientação transmitindo muitos conhecimentos e sugestões de leituras, paciência e disponibilidade.

Aos professores Dr. Ariovaldo Francisco da Silva e Dr. Rodrigo Franklin Frogeri do Centro Universitário do Sul de Minas – UNIS/MG por opinar durante a construção do meu trabalho, fornecendo ricas sugestões, ideias e pelos livros em que gentilmente me sugeriram e emprestaram para leituras.

As amigas construídas durante esta jornada: Humberto, Maurício, Leopoldina e Marcus Vinícius, no qual agradeço imensamente pelo apoio, união e superação. Amigas nas quais considero a maior das conquistas que um ser humano pode obter.

Agradeço também aos amigos de longa data Rafael Kerner e Fernando Martins, no qual foi uma honra aprender com vocês e conseguir explorar o universo da tecnologia e matemática.

Ao Professores Dr. André Rodrigues Monticeli, Dra. Cláudia Adam Ramos, Dr. Nilton César da Silva, Dr. Pablo Javier Grunmann, Dr. Paulo César Mappa e Dr. Silvio Antônio

Bueno Salgado que se esforçaram o máximo possível para transmissão de conhecimento e formação de todos os discentes do curso.

Aos amigos Pedro, Carol, Alex, Janderson, Paulo, Helena, Nivânia, Lucas, Alexandre, Edvaldo, Júlio, Guilherme e Ronald por todo o apoio e principalmente por acreditarem em meu potencial.

A todas as pessoas que estiveram e estão presentes na minha vida nesse período de muito aprendizado e crescimento, por fazerem parte deste momento tão importante.

RESUMO

Diante de uma ressignificação no uso de tecnologias e mudanças no cenário bancário, o número de investidores vem crescendo cada vez mais. Em 2019 esse número aumentou consideravelmente no Brasil, atingindo uma alta de 96% em relação a 2018, observando a preferência do investidor por investimentos com maior risco e que possuem como referência o Índice Bovespa – IBOVESPA. Tais investimentos, apesar de carecer de estratégias mais elaboradas, tem a possibilidade de maiores retornos financeiros. Nesse contexto, os investimentos de risco elevado possuem uma grande oscilação causada por notícias e acontecimentos polêmicos, despertando ao investidor os sentimentos de insegurança e dúvida para decidir se é o momento de aportar ou retirar seus recursos de um determinado investimento. Desse modo, o objetivo geral deste estudo foi aplicar uma RNA LSTM, isto é, uma rede neural artificial com memória de longo prazo (do inglês, *Long Short Term Memory*) para realizar a previsão dos dados do IBOVESPA para o primeiro semestre de 2019 e analisar se notícias e fatos polêmicos impactaram a previsão. Este intento foi conseguido mediante a coleta de dados através da plataforma Yahoo Finanças, o mapeamento de notícias que ocorreram no primeiro semestre de 2019 e uma série de treinamentos para encontrar a melhor arquitetura da RNA. As análises foram realizadas através de medidas de desempenho estatísticas validando os dados reais *versus* os dados previstos e comparando-os com os fatos polêmicos. Os resultados demonstraram que com a aplicação de uma margem de erro de 1% a 5%, a RNA apresentou uma taxa de acerto de 65,58% para a margem de 1% e atingiu mais de 90% de acerto para as margens de 2% a 5%. Concluiu-se que a RNA LSTM possui uma grande capacidade de aprendizado e acerto em suas previsões se utilizada com um grande volume de dados.

ABSTRACT

Around a resignation about the use of technologies and changes in the banking landscape, the number of investors has been growing significantly. In 2019 there was a considerable increase in Brazil, reaching a high of 96% percent compared to 2018, emphasizing the investor's preference for investments with higher risk and that they themselves as reference the IBOVESPA Index. Such investments despite carries of elaborate strategies has the possibility of greater financial returns. Starting from that context the high-stakes investments have a large oscillation caused by news and controversial events awakening to the investor the feelings of insecurity and doubt about how to act in the moments of entry and withdrawal of their resources in one determining investment. In this way, the overall aim of this study was to apply an RNA LSTM, i.e., an artificial neural network with long-term memory (from English, Long Short-Term Memory) to carry out the prediction of IBOVESPA data for the first half of 2019 and to analyze whether news and polemical facts impacted the forecast. This intent was achieved by collecting data through the Yahoo Finance platform, the mapping of news that took place in the first half of 2019 and a series of trainings to find the best architecture of RNA. The analyses were carried out through statistical performance measures validating the actual data versus the predicted data and comparing them with the controversial facts. The results demonstrated that with the application of a margin of error of 1% to 5%, RNA presented a hit rate of 65.58% to the margin of 1% and reached more than 90% of the hit to the margin of 2% to 5%. It was concluded that the RNA LSTM has a great learning ability and hit in its predictions if used with a large volume of data.

SUMÁRIO

1 INTRODUÇÃO	8
1.1 Motivação	9
1.2 Objetivos	9
1.1.1 Objetivo geral	9
1.1.2 Objetivos específicos.....	10
1.3 Organização do texto	10
2 REFERENCIAL TEÓRICO	11
2.1 Estudos Correlatos	11
2.2 Índice Bovespa.....	15
2.3 O impacto de notícias no mercado de ações	16
2.4 Redes Neurais Artificiais	17
2.4.1 LSTM	20
2.5 Medidas de desempenho	24
3 MATERIAIS E MÉTODOS	25
4 RESULTADOS E DISCUSSÃO	40
5 CONSIDERAÇÕES FINAIS.....	53
REFERÊNCIAS	56
ANEXOS	62

1 INTRODUÇÃO

O segmento acionário é movimentado por informações, sejam elas quantitativas, qualitativas ou quali-quantitativas, relacionadas as empresas, índices, dentre outros. Essas informações são utilizadas para averiguar o desempenho de uma ação ou do mercado como um todo além de possuírem diversas interpretações, resultando em comportamentos distintos do investidor (SILVA, CARVALHO E NUNES, 2012).

Em 2019 o número de investidores brasileiros contabilizados pela B3 obteve uma alta de 96% se comparado a 2018, totalizando 1.590.652 de pessoas, sendo 94,41% pessoas físicas e apenas 1,59% pessoas jurídicas (B3, 2019a).

Um investidor cuidadoso, de acordo com o princípio da dominância baseado em Sharpe, Alexander e Bailey (1995), deverá investir seu capital em ativos que possam lhe garantir maiores retornos mediante um determinado risco, retorno financeiro e tempo de aplicação. Dentre os investidores, observam-se três perfis considerados relevantes no ato de um aporte, sendo eles o perfil conservador no qual prioriza a segurança em seus investimentos. O moderado, no qual deseja segurança, mas tem tolerância para aceitar riscos de longo prazo nos seus ativos, desse modo, esse perfil de investidor tende a arriscar mais, de maneira flexível, buscando um lucro acima da média que supera os ganhos do perfil conservador e, por fim, o perfil agressivo, com capacidade de sofrer perdas a curto prazo desde que façam parte de uma estratégia para obter lucros maiores a médio e longo prazo.

Segundo a corretora Rico (2019) o perfil agressivo possui conhecimentos mercadológicos, almejando grandes rentabilidades e exposição a riscos para obter ganhos maiores que a inflação no médio e longo prazo. Esse investidor utiliza estratégias, ferramentas e possui muita preparação técnica e emocional para acompanhar as oscilações do mercado acionário. Sabe-se que notícias e acontecimentos tendem a impactar esse mercado, amplificando as oscilações e, muitas vezes, causando desconforto nos acionistas, e mesmo com a utilização de ferramentas e técnicas de análise, prever o comportamento de um índice ou ação.

Nesse contexto, estabelece-se como pergunta de pesquisa: uma RNA é capaz de prever o IBOVESPA em dias que foram observadas notícias e acontecimentos polêmicos? O objetivo geral deste estudo foi aplicar uma RNA com memória de longo prazo para realizar a previsão dos dados do IBOVESPA para o primeiro semestre de 2019 e analisar se notícias e

fatos polêmicos impactaram a previsão. Este intento foi conseguido mediante a coleta de dados através da plataforma Yahoo Finanças, o mapeamento de notícias que ocorreram no primeiro semestre de 2019 e uma série de treinamentos para encontrar a melhor arquitetura da RNA. As análises foram realizadas através de medidas de desempenho estatísticas validando os dados reais *versus* os dados previstos e comparando-os com os fatos polêmicos.

Tal abordagem se justifica pelos altos investimentos e crescimento do número de investidores no Brasil, além da popularização da utilização das redes neurais como objeto de estudos pela literatura (B3, 2019a; HAYKIN, 2007).

1.1 Motivação

A motivação deste trabalho se concentra em mapear o impacto de notícias e acontecimentos polêmicos no IBOVESPA. Existem vários estudos que buscam mapear e determinar a consequência desses fatos tais como: (DACACHE E LETOUZÉ, 2014; FAUSTINO, 2013; SILVA, 2015; SILVA, CARVALHO E NUNES, 2012).

Além disso, em outros trabalhos, são aplicadas as rede neurais, que se mostram promissoras na previsão de dados, tornando-se uma forte ferramenta de auxílio na tomada de decisão (CHEN et al., 2018; FISCHER; KRAUSS, 2018; HAFEZI, SHAHRABI E HADAVANDI, 2015; PATEL et al., 2015; QIU E SONG, 2016; SHAH et al., 2018; WANG et al., 2016; YANG, PAN E TAO, 2017). Desse modo, nessa pesquisa, realizou-se a implementação de uma RNA e avaliou-se o seu desempenho, comprovando que a modelagem realizada neste estudo pode ser utilizada nesta previsão de valores, apesar do Ibovespa ser afetado por notícias e fatos polêmicos.

1.2 Objetivos

Este trabalho de conclusão de curso foi dividido nos objetivos geral e específicos detalhados a seguir.

1.1.1 Objetivo geral

O objetivo geral deste estudo foi aplicar uma RNA LSTM, isto é, uma rede neural artificial com memória de longo prazo (do inglês, *Long Short Term Memory*) para realizar a

previsão dos dados do IBOVESPA para o primeiro semestre de 2019, e analisar se notícias e fatos polêmicos impactaram a previsão.

1.1.2 Objetivos específicos

- Coletar os dados históricos do IBOVESPA.
- Coletar notícias que impactaram o IBOVESPA.
- Normalizar os dados históricos do IBOVESPA.
- Normalizar as notícias que impactaram o índice utilizando o princípio da falseabilidade.
- Aplicar o modelo de RNA LSTM para os dados históricos do IBOVESPA.
- Realizar o treinamento da rede, verificando o nível de acerto desses testes com o objetivo de encontrar a melhor arquitetura para este problema.
- Realizar a aplicação de um *cross validation* para avaliar a capacidade de generalização do modelo, utilizando 75% de dados para treinamento e 25% dos dados para teste.
- Utilizar o aprendizado supervisionado, isto é, comparar os dados retornados pela RNA com os dados reais do índice e verificar se houve impacto das notícias mapeadas nos dados da RNA.
- Refletir sobre o custo computacional gerado, ao utilizar as redes neurais.

1.3 Organização do texto

Para atingir os objetivos propostos neste estudo, foi realizada uma pesquisa na literatura em cada um dos capítulos para um melhor fundamento teórico sobre o assunto.

O Capítulo 1, a Introdução, descreve o cenário em que o trabalho está inserido, seus objetivos, as lacunas encontradas na literatura que podem contribuir para um investidor durante a realização de um investimento, as ferramentas utilizadas para o desenvolvimento desta pesquisa e a estrutura organizacional do texto.

O Capítulo 2 aborda cinco seções a saber: a primeira seção consiste na investigação de estudos correlatos ao tema proposto. Essa seção foi estruturada de forma a apresentar os estudos mais recentes que abordam RNAs para resolução e melhoria de situações em que índices de bolsas de valores são utilizados como objetos de análise e tomada de decisão, onde valores e comportamentos futuros são incertos, sendo possível observar diversas categorias de RNAs, sendo a mais utilizada a LSTM. A segunda seção, apresentou as principais informações do IBOVESPA, o índice da bolsa de valores considerado o mais importante

indicador do desempenho médio das cotações das ações negociadas na bolsa de valores do Brasil, evidenciando sua relevância no cenário econômico do país. A terceira seção buscou caracterizar as categorias de notícias e acontecimentos polêmicos que causam impactos durante um investimento. A quarta seção, apresentou, em detalhes, as características de uma RNA desde a sua constituição até o seu funcionamento, evidenciando conceitos como camadas de entrada, camadas escondidas, neurônios, entre outros. Ainda nesta seção, foi exposta a RNA LSTM e sua arquitetura, além de descrever suas principais funções e métodos. Por fim, na quinta seção, realizou-se uma busca na literatura pelas principais medidas de desempenho capazes de avaliar uma previsão, utilizando cálculos que retornam a raiz do erro quadrático médio, o erro percentual absoluto médio e o coeficiente de determinação.

O Capítulo 3 demonstra a utilização dos materiais e métodos adotados neste estudo, desde a coleta de dados, o treinamento da RNA modificando sua arquitetura com o intento de obter os menores valores de erro quadrático médio e posteriormente a realização dos testes para validação da RNA.

O Capítulo 4 contém os resultados e discussões realizando uma confrontação com os principais fatos polêmicos mapeados no primeiro semestre 2019, averiguando se os dados previstos foram impactados por esses fatos.

Por fim, o Capítulo 5 apresenta as principais considerações deste estudo, as limitações, contribuições e apontamentos para pesquisas futuras, concluindo este estudo.

2 REFERENCIAL TEÓRICO

2.1 Estudos Correlatos

Fischer e Krauss (2018) desenvolveram um trabalho no qual teve como objetivo realizar a previsão das tendências financeiras da base de dados S&P 500, no qual é responsável por fornecer um índice de valores dos ativos das bolsas de valores de Nova York, que contém mais de 2800 ações. Os dados foram coletados para o período de 1992 a 2015, e, para avaliar a melhor precisão, foram utilizadas uma RNA recorrente com arquitetura LSTM conhecida como rede de memória de longo prazo, uma RNA profunda, a floresta aleatória (RF - *random forest*) e um classificador de regressão logística. Observou-se que para o período de 1992 a 2009, o desempenho da LSTM foi completamente superior a aplicação da RNA profunda, floresta aleatória e classificador de regressão logística. Após isso, os retornos de possível lucratividade foram atenuados devido ao aumento das taxas de transferência ao

realizar ou interromper um investimento. Além disso, percebeu-se a precisão da rede LSTM utilizando uma previsão de 240 dias futuros. Por fim, os autores investiram em ações comprando-as e vendendo-as conforme as tendências indicadas pela LSTM, obtendo retornos aproximados de 0,23% ao dia, concluindo que a LSTM foi capaz de extrair efetivamente informações precisas dos dados utilizados no treinamento.

Chen et al. (2018) buscaram verificar qual algoritmo de inteligência artificial (IA) possui a melhor capacidade de prever o mercado chinês de ações financeiras. Para isso, foi sugerida a combinação de um modelo de previsão baseado em aprendizado profundo (*deep learning*), um *autoencoder* utilizando aprendizado não supervisionado e uma máquina Boltzmann considerada um tipo de RNA recorrente estocástica com o campo aleatório de Markov. Após isso, a combinação criada foi comparada com três RNAs clássicas, sendo elas a RNA com retro propagação (do inglês, *back propagation*), a máquina de aprendizado extremo (do inglês, *extreme learning machine* - ELM) e uma função de base radial (RBF). Os dados utilizados para o treinamento, foram extraídos do CSI 300, no qual é conhecido como o índice do mercado de ações, com indicadores das 300 principais ações negociadas na bolsa de valor de Xangai e Shenzhen. Descobriu-se que o experimento criado obteve um índice de previsão superior aos demais métodos utilizados para comparação (RNA com retro propagação, máquina de aprendizado extremo e a função de base radial) além disso, quanto maior a quantidade de dados utilizada, maior foi o desempenho preditivo, indicando que o aprendizado foi capaz de capturar recursos não lineares da base utilizada.

Shah et al. (2018) utilizaram uma bolsa de valores da Arábia Saudita, conhecida como *Saudi Stock Exchange*, com o objetivo de apresentar um novo algoritmo híbrido que fosse capaz de realizar os treinamentos e testes, adotando uma RNA, através de um algoritmo classificador conhecido como Curva ABC. Para os testes realizados, o período compreendido foi entre 2015 a 2016, utilizando um total de 250 dias, dos quais 175 dias (75%) foram para treinamento e 75 dias (25%) para teste. Para avaliar o desempenho do algoritmo, utilizou-se o erro quadrático médio, desvio padrão, erro médio quadrático normalizado, taxa de sucesso e precisão. Como conclusões gerais, observou-se superioridade da união da RNA com o algoritmo classificador ABC, visto que essa combinação obteve resultados superiores aos métodos utilizados de forma isolada, abrindo uma gama de possibilidades de implementação e exploração desse estudo em trabalhos futuros.

No estudo de Yang, Pan e Tao (2017), os autores buscaram resolver o problema de bases que possuem valores discrepantes quando a duração da série temporal é aumentada. Esses valores que fogem do padrão, normalmente enganam o modelo utilizado para realização

do treinamento e aprendizado da rede. Para corrigir isso, foi aplicado um método de aprendizado gradiente denominado RoAdam (Robust Adam) apropriado para redes LSTM. Com esse método, o intento foi ajustar a taxa de aprendizado do algoritmo de forma que fosse possível a redução de *outliers* (valores que estão fora da tendência encontrada). Foi possível, com a utilização de uma média ponderada rastrear o erro relativo de previsão da função de perda. Os experimentos realizados demonstraram que, com a aplicação do método citado, foi possível diminuir a taxa de erro durante uma previsão, tornando-o uma extensão que pode ser adotada em trabalhos que utilizam a rede LSTM.

Qiu e Song (2016) realizaram um trabalho com o objetivo de aumentar a precisão na previsão das tendências financeiras dos valores das ações no mercado japonês. Para isso, foi utilizada uma combinação de RNA com a implementação de algoritmos genéticos. A base de dados adotada para os experimentos foi o índice de Nikkei 225, conhecido como o principal índice econômico da bolsa de valores de Tóquio, e o período analisado foi de janeiro de 2007 a dezembro de 2013. Os treinamentos foram realizados submetendo os dados a um processo de normalização com o objetivo de diminuir possíveis inconsistências. Após isso, os dados foram inseridos no modelo da RNA, otimizando os seus pesos com a aplicação do algoritmo genético e, dessa forma, criou-se o modelo híbrido GA-RNA. Após a finalização dos experimentos, foi utilizado o aprendizado supervisionado, comparando os achados com dados históricos. Os resultados também foram comparados com quatro estudos semelhantes sendo eles: Chauvin e Rumelhart (1995); Huang, Nakamori e Wang (2005); Kara, Boyacioglu e Baykan (2011); Leung, Daoukb e Chen (2000) donde foi possível concluir que no trabalho dos autores Mingyue Qiu e Yu Song a maior precisão na previsão foi observada.

Os autores afirmaram ainda que os resultados deste estudo podem ser melhorados, se utilizado um nível de validação de três camadas, sendo possível combinar indicadores do índice da bolsa, incluindo mais variáveis que possam afetar a precisão da RNA, bem como a utilização de modelos de redes neurais probabilísticas e, por fim, a realização de investimentos com base nos resultados preditivos para validação prática dos experimentos e quantificação de possíveis lucros.

No trabalho de Wang et al. (2016) foi realizada a combinação de uma RNA recorrente baseada em Elman utilizando uma função estocástica de tempo. O artefato desenvolvido pelos autores foi submetido aos métodos clássicos de análise como a regressão linear e a distância invariante de complexidade e, posteriormente, comparado com a técnica de retro propagação. Os autores utilizaram para treinamento e previsão os dados extraídos do índice de Nikkei 225 e após os resultados, os dados foram submetidos a outras medidas de previsão, tais como o

MAE (erro absoluto médio, do inglês, *mean absolute error*) retornando o valor de 68,54. O RMSE (erro quadrático médio, do inglês, *root mean squared error*) com valor de 89,03 e o MAPE (erro percentual absoluto médio, do inglês, *mean absolute percentage error*) indicando 0,60. De acordo com os autores, esses dados demonstram que o modelo proposto por eles obteve vantagem sobre a técnica de retro propagação, sinalizando superioridade ao melhorar a precisão da previsão, uma vez que os dados previstos estavam mais próximos dos valores reais.

Os dados do mercado de ações DAX, nos quais qual são constituídos de uma relação das 30 companhias abertas de melhor performance financeira da Alemanha, foram aplicados no estudo de Hafezi, Shahrabi e Hadavandi (2015). Neste estudo, os autores apresentaram um modelo inteligente em uma estrutura multiagente, chamada sistema multiagente de rede bat-neural. O modelo foi configurado para processar os dados utilizando quatro camadas, buscando prever oito anos do preço das ações DAX em períodos trimestrais. Sua capacidade foi avaliada ao aplicar os dados e realizar os treinamentos, analisando seus resultados e comparando-os com outras técnicas, como a RNA com algoritmo genético e a RNA de regressão generalizada. Os autores mencionam que o modelo se destacou quando utilizada a validação estatística MAPE, no qual demonstrou valores de precisão maiores que os demais modelos, mesmo com a crise financeira enfrentada no período adotado para o trabalho.

Patel et al. (2015) buscaram comparar quatro modelos de previsão de dados, sendo eles a RNA, a máquina de vetor de suporte (SVM), a floresta aleatória (RF) e o Naive Bayes. Para realização dos experimentos, foi utilizada duas abordagens para o *input* (entrada dos dados). A primeira buscou a utilização dos dados de negociação de ações (valores de abertura, valor mais alto, valor mais baixo e valor de fechamento, entre outros) e a segunda abordagem se concentrou na representação desses dados como valores determinísticos de tendência. Para comparação da precisão dos quatro modelos, os dados de duas ações financeiras foram utilizados sendo elas a Reliance Industries e a Infosys Ltd, e dois índices de valores de ações, sendo eles o CNX Nifty e o S&P Bombay Stock Exchange (BSE), todos para o período entre 2003 a 2012. Os resultados para a primeira abordagem de entrada indicaram superioridade na utilização da RF com 83,56% (maior taxa de precisão) e o algoritmo de Naive Bayes apresentou 73,3% (menor taxa de precisão). Na segunda abordagem, ao submeter os valores determinísticos, todos os modelos foram aprimorados, sendo as taxas de precisão 86,69% (RNA), 89,33% (SVM), 89,98% (RF) e 90,19% (Naive Bayes). Dessa forma, foram observadas grandes melhorias nas previsões quando os dados foram transformados em valores determinísticos, tornando possível a ampliação deste estudo. Segundo os autores, as técnicas

abordadas poderiam ser utilizadas em tempo real, para a previsão de tendência das ações, o que poderia tornar os investimentos mais rentáveis e seguros.

Diante dos estudos correlatos, observou-se que já existe há algum tempo o interesse em estudos que envolvem a aplicação de uma ferramenta capaz de prever o comportamento de um índice de relevância nacional e internacional. Isso motivou o desenvolvimento deste estudo, alinhando-o com outros trabalhos deste campo de pesquisa.

A seguir, é apresentado o índice de maior relevância no Brasil e as principais empresas que o compõem.

2.2 Índice Bovespa

O IBOVESPA é a principal referência do desempenho das ações das empresas de maior importância no cenário financeiro do Brasil. Esse índice foi elaborado em 1968 e, com o passar dos anos, obteve grande crescimento e reconhecimento, sendo hoje o maior da América Latina (CASTRO et al., 2019). O seu objetivo é ser um parâmetro capaz de exibir a performance média dos valores das cotações dos ativos com maior movimentação de negociação no mercado brasileiro. Sabe-se que esse índice é submetido a uma reavaliação trimestral, para que considerem as ações das empresas que respeitam as normas presentes em sua metodologia (B3, 2019b).

No primeiro semestre de 2019, a composição da carteira teórica do IBOVESPA era constituída por 65 empresas nos mais diversos segmentos, ocupando os primeiros lugares na composição¹. Algumas destas empresas eram: Itaú Unibanco - 10,80%², a mineradora multinacional brasileira Vale - 10,77% e o banco Bradesco - 8,57%. Quando uma empresa compõe o índice, sua visibilidade aumenta, ganhando a atenção dos investidores (B3, 2019c).

Com as variações das ações da bolsa de valores implicando em valorização e desvalorização das empresas que compõem o IBOVESPA, o seu valor do Índice é ajustado em tempo real. O valor de fechamento é utilizado como referência para saber como foi o desempenho de um índice ou ação em um determinado dia. Durante a pesquisa, foi possível observar que essa medida foi adotada em diversos trabalhos como Caramico (2012); Colombo, Bampi e Camargo (2010); Locatelli, Silva e Alvim (2019) e além disso, foi possível constatar também através dos blogs oficiais das principais corretoras de investimento do Brasil, que os investidores que aplicam seu capital em ações e fundos de investimento,

¹ A composição completa da carteira está localizada no Anexo 1 deste estudo.

² Participação relativa das ações da carteira.

(considerados os principais investimentos de renda variável no Brasil), também utilizam o valor de fechamento do IBOVESPA para acompanhar o seu desempenho também dos seus ativos (CLEAR, 2020; RICO, 2020; XP INVESTIMENTOS, 2020).

De acordo com Gusmão (2018), no último relatório divulgado pela *StockstoTrade* o IBOVESPA ocupava a vigésima posição no ranking mundial das 20 maiores bolsas de valores. Para fazer parte desta cesta, o IBOVESPA precisou acumular pontos de acordo com o tempo de fundação, total de empresas participantes, capitalização de mercado e volume de negociação. Esse reconhecimento é importante pois quanto maior o nível do índice, mais visibilidade o país possui, além de atrair investidores de todo o mundo.

Dada a relevância do IBOVESPA, justifica-se entender como ele pode ser impactado por notícias e fatos polêmicos, o que torna seu comportamento imprevisível. A seguir, observa-se em detalhes como as notícias e fatos polêmicos podem influenciar o comportamento do índice.

2.3 O impacto de notícias no mercado de ações

Para Cutler, Poterba e Summers (1989) a instabilidade do mercado acionário espelha os fatos que acontecem no mundo que possuem relação com cenário financeiro e econômico. Dessa forma, quando uma empresa revela um resultado imprevisível de seus ativos, o resultado poderá ser observado como uma variação acima do normal dos preços das suas ações. Do mesmo modo, uma notícia econômica relevante, ou até mesmo novas ações realizadas por parte do governo, também podem afetar o comportamento do mercado acionário de maneira positiva ou negativa.

De acordo com Antunes (2019) existem três tipos de notícias que podem impactar os valores de uma ação, índice, ou até mesmo cotação do dólar, sendo elas:

Notícias pré-agendadas: são notícias com data e horário conhecidos pelos investidores. Essas notícias normalmente estão relacionadas aos indicadores de vendas, produto interno bruto (PIB), inflação, taxa de juros, informações sobre desemprego e balanços divulgados pelas próprias empresas.

Opiniões: nesta segunda categoria, estão centradas as convicções de especialistas que possuem uma grande visibilidade pública, tais como jornalistas, economistas, analistas financeiros, no qual expõem seu ponto de vista sobre os eventos que poderão acontecer.

Breaking News: é a categoria mais alarmante uma vez que ninguém consegue prever um fato que não aconteceu ou tem hora marcada para ser exposto. Nesse tipo de fato,

acontecimentos, os principais episódios, são por exemplo atentados, rompimento da barreira de Brumadinho, e qualquer tipo de incidente, que podem causar insegurança nos investidores, ocasionando uma instabilidade econômica.

Após entender como as notícias podem impactar o mercado acionário, observa-se um maior abalo das opiniões e *breaking news* visto que em muitos momentos o investidor não é capaz de prever os fatos informados por essas categorias de notícias. Desse modo, esses dois conjuntos podem impactar o IBOVESPA ocasionando em quedas que não podiam ser previstas. Para Taleb (2015) esses eventos são marcados pela imprevisibilidade e, além disso, o seu impacto é intenso e árduo. Nesse contexto, o investidor que não possui uma ferramenta de previsão em que possa se apoiar, para tomar a decisão de manter seus recursos ou retirá-los, é abalado pela dúvida e não é capaz de determinar se uma queda inesperada possuirá uma rápida recuperação ou não.

2.4 Redes Neurais Artificiais

Segundo Rezende (2003), a RNA pode ser representada por modelos matemáticos que buscam assimilar às estruturas neurais biológicas com capacidade computacional adquirida por meio do aprendizado e generalização. Para Haykin (2007) a RNA pode ser considerada como um processador constituído de unidades simples, com a capacidade de acumular conhecimento temporário e torná-lo disponível para o uso. Sua semelhança ao cérebro pode ser observada por duas características: i) o conhecimento é obtido pela rede através de um método de aprendizagem; ii) interconexão entre os neurônios, conceituadas como pesos sinápticos, são aplicadas para coletar e armazenar todo o conhecimento assimilado.

A maneira que um neurônio biológico funciona pode ser mapeada, utilizando um modelo que o descreve como um circuito binário, que possui dados de entrada com valores de 0 a 1, e que, quando combinados por uma soma ponderada, pode-se observar a criação de uma entrada efetiva do neurônio, determinando a continuação desse estímulo (TAFNER, FILHO E XEREZ, 1995). Desse modo, pode-se observar conforme Figura 1, a representação matemática de maneira simplificada de um neurônio.

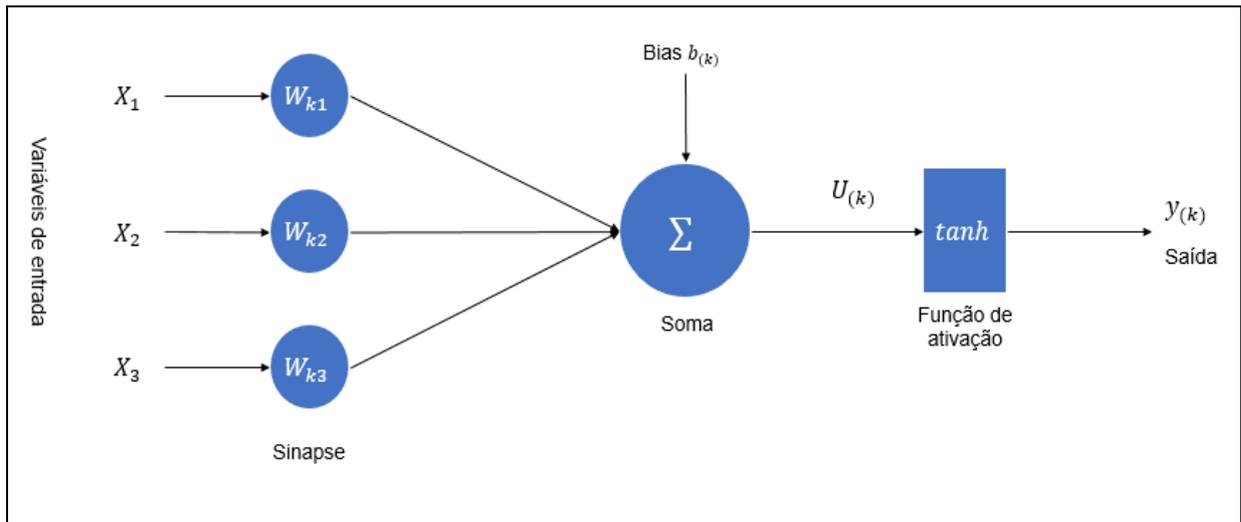


Figura 1 – Representação Simplificada de um Neurônio

Fonte: (DATA SCIENCE ACADEMY, 2019).

Nesta arquitetura que representa neurônio matemático, os impulsos elétricos originários de outros neurônios são exibidos pelos sinais de entrada (a variável x é utilizada para caracterizar os dados que sustentam seu modelo de RNA). Observa-se que, diante dos estímulos recebidos, alguns são capazes de excitar mais, e outros menos, o neurônio receptor e, essa medida de quão excitatório é o estímulo, pode ser representada no modelo de Warren McCulloch e Walter Pitts através dos pesos sinápticos. Quanto maior o valor do peso, mais excitatório é o estímulo. Os pesos sinápticos são representados por w_{kn} , onde k representa o índice do neurônio em questão e n se refere ao terminal de entrada da sinapse a qual o peso sináptico se refere. Já a função de soma da célula é caracterizada por uma constituição de dois parâmetros, sendo o primeiro uma junção aditiva, utilizada para realizar somatório dos estímulos (sinais de entrada) multiplicado pelo seu fator excitatório (pesos sinápticos), e posteriormente uma função de ativação representa por \tanh^3 , que define com base nas entradas e pesos sinápticos, qual será a saída do neurônio (DATA SCIENCE ACADEMY, 2019).

As entradas de um neurônio artificial podem ser comparadas exatamente como estímulos do neurônio biológico. Uma RNA é formada por vários neurônios divididos em camadas. A primeira é a camada de entrada e tem a função de armazenar os dados de entrada na rede. As camadas intermediárias são as camadas escondidas, nelas ocorrem os

³ A função \tanh é utilizada para retornar valores de -1 a +1 utilizada em diversos momentos na aplicação de um *back propagation*. Sua representação pode ser escrita diretamente como: $\tanh(x) = 2 / (1 + e^{(-2x)}) - 1$ (DATA SCIENCE ACADEMY, 2019).

processamentos da rede. A última camada, é definida como camada de saída, recebe os valores processados pela rede (RIBEIRO, 2003).

Para Haykin (2007), o processo de aprendizagem tem seus parâmetros adaptados através de um processo de fomentação no ambiente no qual a rede está introduzida. O método de aprendizagem é um dos elementos mais importantes do sistema de RNA e é o que diferencia dos algoritmos convencionais para classificação.

De maneira computacional, o aprendizado pode ser observado através de um algoritmo de aprendizagem que realiza uma série de cálculos em toda a cadeia de neurônios para o devido treinamento da RNA. A quantidade de neurônios por camadas, o número de camadas por rede e os pesos ajustados após o aprendizado é chamada de arquitetura da RNA (ITO et al., 2000), representada na Figura 2, a seguir.

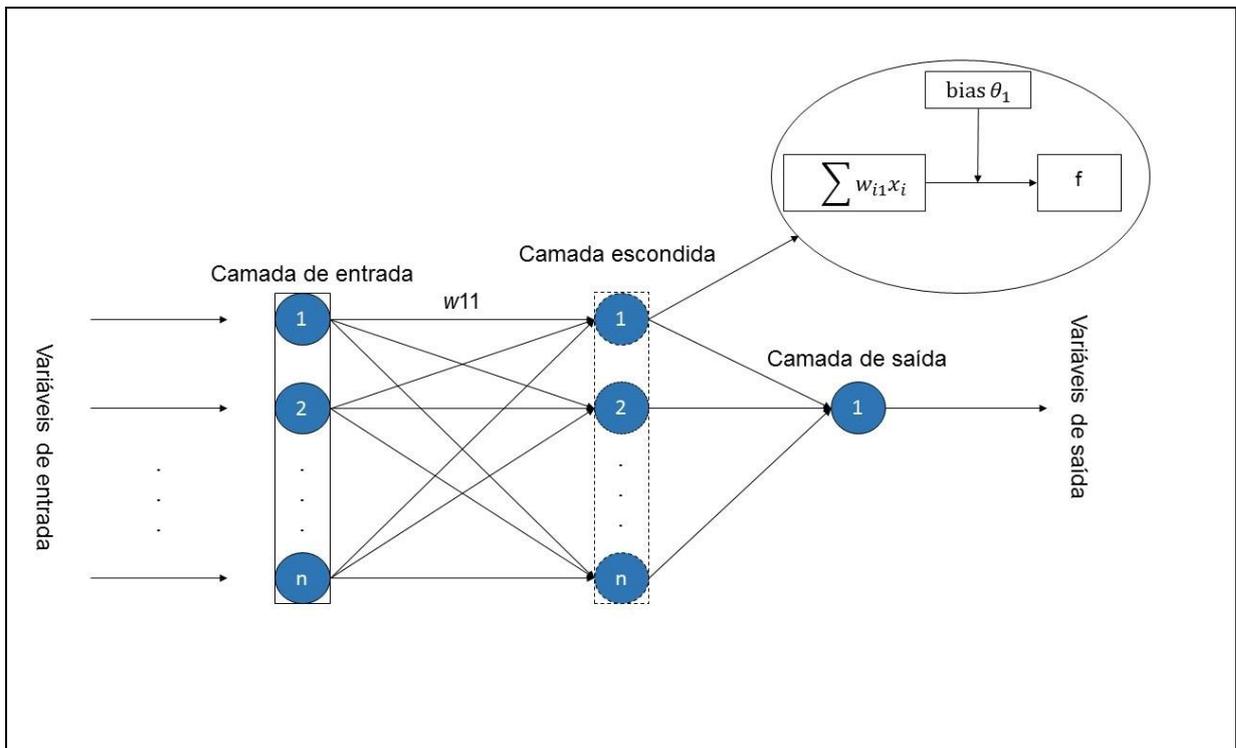


Figura 2 - Arquitetura de uma RNA

Fonte: (Qiu e Song 2016).

Na Figura 2, pode-se visualizar uma RNA com apenas uma camada escondida, responsável pelo grau de precisão nos dados de saída da rede. No entanto, por possuir uma maior flexibilidade e precisão, pode-se adicionar mais camadas escondidas, sendo possível aumentar a sua precisão e dessa forma, o custo computacional se tornará a maior barreira neste tipo de rede (RUCK et al., 1990).

Para melhor entendimento, pode observar as variáveis, onde: w_{11} é o peso do nó 1 da camada de entrada conectado com o nó 1 da camada escondida. O bias θ_1 representa a polarização do nó 1 na camada escondida. E por fim, os pesos e desvios da rede recebem os valores iniciais, o erro entre os valores de saída previstos e reais sendo propagado de volta pela rede para atualizar os pesos e desvios repetidamente até a conclusão do treinamento (WANG, 2009). Quando o erro é menor que um valor especificado ou quando o critério de término é satisfeito, o treinamento é considerado concluído e os valores de ponderação e polarização da rede são armazenados para utilização nos cálculos e retorno dos dados na camada de saída.

2.4.1 LSTM

A RNA denominada *Long Short-Term Memory* (LSTM) foi criada por Hochreiter e Schmidhuber (1997), onde observou-se uma popularização em sua utilização visto que pode ser adaptada em uma diversidade de problemas, além de fazer parte da arquitetura de rede recorrente (MEDSKER; JAIN, 2001). Essas redes, possuem o objetivo de aprender padrões complexos em estruturas com dependência temporal e vários estágios de processamento, no qual, muita informação é submetida a rede e pode se perder no começo de uma série de treinamentos até suas fases intermediárias após várias interações nos treinamentos em suas camadas ocultas (GREFF et al., 2017). Sua principal característica é a concentração em evitar o problema de dependência de longo prazo onde recordar a informação por longos períodos é seu comportamento padrão.

A seguir conforme, Figura 3, observa-se a estrutura de uma RNA LSTM.

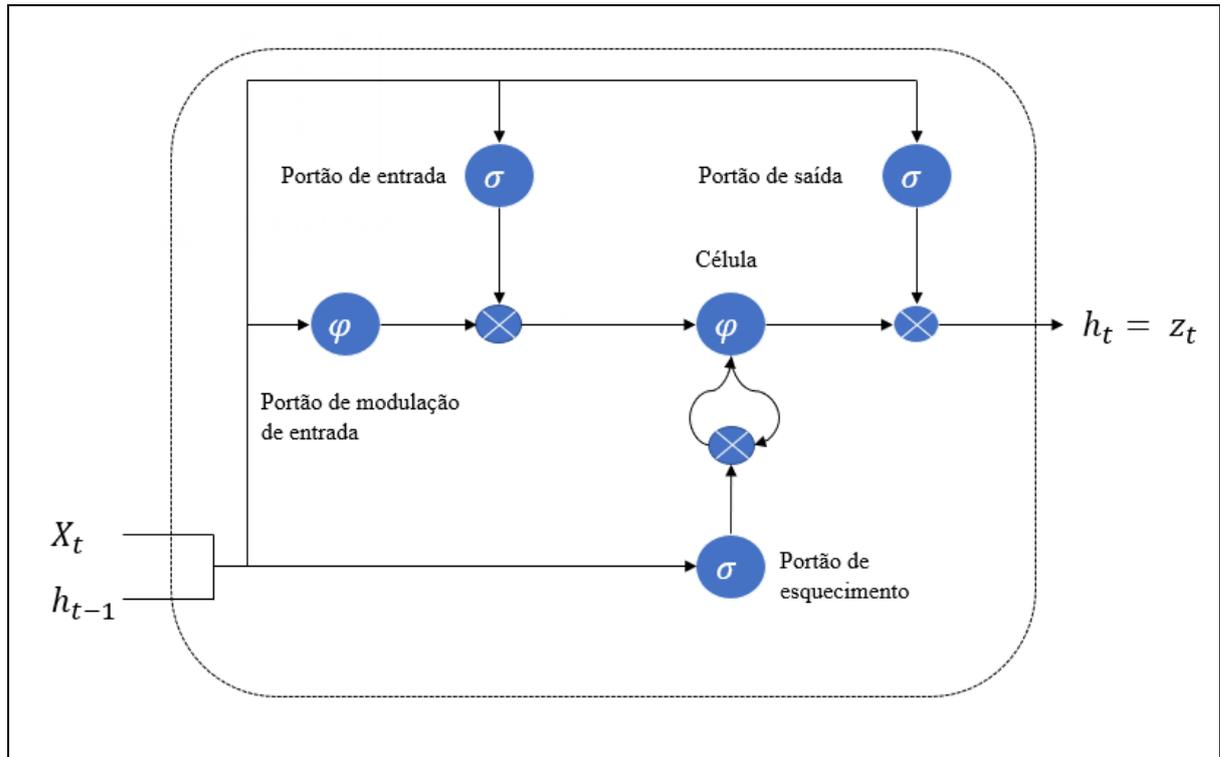


Figura 3 - Arquitetura de uma RNA LSTM

Fonte: (DATA SCIENCE ACADEMY, 2019).

Ao abordar uma RNA LSTM, pode-se mencionar que suas operações são realizadas pelos portões (do inglês, *gates*). Sendo eles, o *forget gate*, *input gate* e *output gate*. A seguir, são explanadas suas funcionalidades conforme o Data Science Academy (2019):

Portão de esquecimento: Realiza a tarefa de reciclar as informações que possuem, ou não, utilidade para a RNA. As informações que não são mais relevantes, são removidas. Para isso, são utilizadas duas entradas, o X_t e h_{t-1} nas quais são multiplicadas por matrizes que contém os pesos das células e adição do bias. Seu resultado é processado através da função de ativação, fornecendo um resultado binário (0 ou 1). Quando o resultado for 0, a informação é esquecida/deletada, e quando o resultado for 1, a informação é armazenada para reutilização.

Portão de entrada: As informações relevantes são adicionadas através do portão de entrada. Para isso, a informação é tratada utilizando a função sigmoide de maneira equivalente ao portão de esquecimento utilizando as entradas h_{t-1} e X_t . Com isso, um vetor baseado na função *tanh*, que é utilizada para compor o vetor com valores entre -1 a +1. Os valores armazenados no vetor e os valores regulados são multiplicados entre si com o objetivo de obter as informações necessárias.

Portão de saída: Seu objetivo é coletar as informações relevantes para serem expostas com uma saída da rede. Para isso, um vetor também é gerado com a aplicação da função *tanh*. Após isso, a informação é regulada utilizando a função de ativação sigmoide filtrando os valores que serão lembrados, desse modo, os valores do vetor e os valores regulados são multiplicados para serem enviados como uma saída e entrada. A célula recebe dois parâmetros, sendo eles o retorno do último estado oculto e a observação no tempo = t. Além do estado oculto, não há informações sobre o passado para se lembrar. As setas demonstram um estado de *loop*, indicando que a rede trabalha de maneira recursiva, permitindo que os dados dos intervalos anteriores sejam memorizados na célula LSTM. O estado da célula é modificado pelo portão de esquecimento inserido sob o estado da célula e ajustado pelo *gate* de modulação de entrada. De maneira análoga, observa-se que o estado da célula anterior é esquecido, multiplicado com a porta do esquecimento e adicionados novos dados através da saída das portas de entrada.

Após conhecer cada um dos portões que constituem a estrutura da RNA LSTM, foi elaborado um resumo com os principais métodos, funções e parâmetros da LSTM para o seu correto funcionamento ao realizar a codificação. Essas informações podem ser observadas a seguir conforme Quadro 1.

Quadro 1 – Funcionalidades para Realizar a Codificação da RNA LSTM

Função/Método/Parâmetro	Descrição	Pacote
<i>Read_csv()</i>	Função utilizada para importar uma base de dados no formato csv.	Pandas
<i>Array()</i>	Recebe dados que são transformados em uma matriz.	Numpy
<i>Dropna()</i>	Esse método permite analisar e remover linhas e colunas com valores nulos.	Pandas
<i>MinMaxScaler()</i>	Esse estimador transforma os dados em uma escala entre zero e um, de modo que os dados sejam normalizados para melhor desempenho no treinamento da RNA.	Sklearn.preprocessing
<i>Sequential()</i>	Representa a camada de entrada da RNA. Por se tratar de uma previsão, utiliza-se esse método uma vez que receberá os dados de	Keras

	entrada para o treinamento da série temporal.	
<i>units</i>	Representa a quantidade de neurônios na camada de entrada, escondida ou de saída.	Keras
<i>Return_sequences</i>	Quando este parâmetro possui o valor “True” as informações são enviadas para a próxima camada da RNA.	
<i>Dropout()</i>	Descarta de maneira aleatória conjuntos de neurônios para evitar com que as redes se adaptem demais, desse modo, a utilização da técnica de <i>Dropout</i> auxilia na prevenção do <i>overfitting</i> .	Keras
<i>Dense</i>	Camada de saída densa na qual possui 1 neurônio e uma função de ativação.	Keras
<i>Compile()</i>	Esse método é responsável por iniciar o treinamento da RNA e recebe como parâmetros a quantidade de épocas, uma função <i>loss</i> e uma métrica.	Keras
<i>Epochs</i>	Quantidade de vezes que o treinamento irá ser realizado.	Keras
<i>Loss</i>	Determina uma métrica no qual a RNA terá como base para melhorias em cada treinamento. Exemplo: erro quadrático médio.	Keras
<i>Metrics</i>	Retorna uma métrica para avaliar o treinamento da RNA. Exemplo: erro médio absoluto.	Keras
<i>Batch_size</i>	Quantidade de registros por atualização do gradiente da RNA. Possui o valor padrão de 32.	Keras
<i>Predict()</i>	Recebe a base de testes que será realizada a previsão.	Keras
<i>Input_dim</i>	Quantidade de atributos previsores. Para este estudo, o valor padrão será 1, visto que a	Keras

	previsão será realizada apenas para a coluna de valor de fechamento	
<i>Inverse_transform</i>	Desnormaliza os dados que foram normalizados através da função MinMaxScaler.	Sklearn.preprocessing

Fonte: (KERAS, 2020; NUMPY, 2020; SCIKIT-LEARN, 2020).

2.5 Medidas de desempenho

Para analisar o desempenho de uma RNA, pode-se utilizar medidas estatísticas capazes de analisar a diferença entre os valores previstos e os valores reais. Desse modo, observa-se a seguir as principais métricas para realização do estudo.

Erro Quadrático Médio: do inglês *Mean Squared Error* - MSE é considerado a métrica mais utilizada para avaliar a performance do modelo de previsão. Sua equação está representada a seguir:

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

onde: N é o número de observações, no caso de uma série temporal o N é interpretado como o número total de dias; as variáveis $R_{real,i}$ e $P_{previsto,i}$ são respectivamente o valor real e o valor previsto da série para um determinado dia i . Dessa forma, pode-se calcular a diferença entre a série projetada e a real ao quadrado, tirando uma média simples dos seus valores.

Raiz Quadrada do Erro Quadrático Médio: do inglês *Root Mean Squared Error* – RMSE, possui o objetivo de indicar a dimensão do erro médio obtido, e quanto mais próximo de zero for o erro, melhores são os resultados encontrados.

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

em que N é o número total de dias da série temporal, $R_{real,i}$ e $P_{previsto,i}$ são respectivamente o valor real e o valor previsto da série para um determinado dia i .

Erro Percentual Absoluto Médio: do inglês, *Mean Absolute Percentage Error* – MAPE, tem como objetivo obter as diferenças percentuais entre os valores reais e previstos, realizando uma média simples dos valores. Observa-se que se $R_{real,i} = P_{previsto,i}$, obtém-se o MAPE = 0, ou seja, quanto menor for o valor do MAPE, melhor é a previsão. Sua equação pode ser observada a seguir:

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

em que N é o número total de dias da série temporal e $R_{real,i}$ e $P_{previsto,i}$ são respectivamente o valor real e o valor previsto da série para um determinado dia i .

Coefficiente de determinação: é representado pelo R^2 buscando validar o ajustamento de um modelo, fornecendo a porcentagem da variância da variável real com relação a variável prevista.

$$R^2 = \frac{\sum(R_{real} - \overline{R_{real}})(P_{previsto} - \overline{P_{previsto}})}{\sqrt{\sum(R_{real} - \overline{R_{real}})^2 \sum(P_{previsto} - \overline{P_{previsto}})^2}}$$

em que $R_{real} - \overline{R_{real}}$ e $P_{previsto} - \overline{P_{previsto}}$ representam respectivamente a diferença dos valores reais e a sua média, e, a diferença dos valores previstos com a sua média.

Seu valor sempre estará entre 0 e 1, e quanto mais próximo de 1, mais preciso está o modelo. Desse modo, para se avaliar a robustez dos modelos, foram aplicadas estas medidas de desempenho com o objetivo de encontrar o melhor resultado, respeitando suas características, visto que, para algumas medidas, o retorno mais próximo de um (Coeficiente de Determinação) é o ideal, enquanto que para outras, quanto mais próximo de zero (MSE, RMSE, MAPE), mais preciso está o modelo.

De acordo com Kim e Kim (2016) a medida de acurácia mais popular dentre as existentes é o MAPE e, além desta, o uso das medidas de escalas como o RMSE e o coeficiente de determinação também podem ser utilizados como critério de avaliação. A seguir, serão descritos os aspectos metodológicos que sustentam este trabalho.

3 MATERIAIS E MÉTODOS

A coleta dos dados históricos foi realizada considerando o período de 03 de janeiro de 2000 a 28 de dezembro de 2018 totalizando 18 anos. Para isso, foi utilizada a plataforma Yahoo Finanças (YAHOO, 2019), na qual permite uma variedade de consultas, além do download em um arquivo de formato *csv* (valores separados por vírgulas, do inglês: *comma-separated-values*). A Figura 4 exibe os dados de fechamento do IBOVESPA em formato gráfico.



Figura 4 – Dados de fechamento reais para treinamento e teste

Fonte básica dos dados: (YAHOO, 2019).

Os dados foram segmentados em dois grupos, sendo o primeiro grupo com 75% dos dados, compreendidos no período de 03 de janeiro de 2000 a 07 de março de 2014, utilizados para o treinamento da rede, e o segundo grupo com 25%, no qual estão situados no período de 10 de março de 2014 a 28 de dezembro de 2018, para realização dos testes. Esse agrupamento é considerado clássico e um dos mais utilizados pela literatura para realização de previsões (HUANG, NAKAMORI E WANG, 2005; PAI E LIN, 2005).

A Tabela 1 traz um resumo da base de dados e o significado de cada coluna.

Tabela 1 - Dados do Ibovespa

Data	Abertura	Maior valor	Menor valor	Fechamento	Fechamento ajustado	Volume
03/01/2000	17.098,00	17.408,00	16.719,00	16.930,00	16.930,00	-
⋮	⋮	⋮	⋮	⋮	⋮	⋮
07/03/2014	47.093,00	47.218,00	46.086,00	46.244,00	46.244,00	3.982.800,00
10/03/2014	46.242,00	46.242,00	45.203,00	45.533,00	45.533,00	3.673.600,00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28/12/2018	85.469,00	88.044,00	85.469,00	87.887,00	87.887,00	3.649.100,00

Fonte: (YAHOO, 2019).

Conforme Tabela 1, os dados coletados possuem sete colunas, sendo elas:

Data: responsável por identificar o dia das movimentações.

Abertura: é o valor da ação no instante da abertura do pregão. Em muitos casos, o valor da abertura pode ser diferente do valor de fechamento do dia anterior visto que uma ação pode iniciar o dia já em alta ou em queda.

Maior valor: essa métrica é responsável por demonstrar o valor máximo atingido no decorrer do dia.

Menor valor: essa métrica, similar ao maior valor, é responsável por exibir o valor mínimo atingido no decorrer do dia.

Fechamento: o valor de fechamento é o valor registrado no término do dia, utilizado para fechamento dos desdobramentos.

Fechamento ajustado: similar ao fechamento, o fechamento ajustado pode ainda ser utilizado nas operações de dividendos.

Volume: é a soma do valor utilizado em ações de compras e vendas no decorrer do dia.

Neste estudo, a coluna de fechamento será utilizada como base dos treinamentos e testes, uma vez que esse valor é utilizado para estimar a variação percentual em relação ao dia anterior indicando movimentos de alta (ou queda) e percentual acumulado. Além disso, Lo e Hasanhodzic (2010) afirmam que o fechamento de uma ação ou índice possui maior representatividade informacional para os investidores uma vez que estratégias de investimentos são baseadas nesse valor.

Após isso, foi realizada a coleta dos dados utilizados para o aprendizado supervisionado, que, de acordo com Russel e Norvig (2009), é caracterizado por possuir os

dados de saída conhecidos, ou seja, ao aplicar uma RNA, sua saída poderá ser comparada com dados concretos que podem determinar se essa saída possui bons resultados ou não. Os dados supervisionados foram tomados para o período de 02 de janeiro de 2019 a 28 de junho de 2019, totalizando 122 dias. Esses dados foram utilizados como critério de comparação com os valores retornados pela RNA, além de validarem outras informações, que poderão ser vistas no decorrer deste estudo. Observa-se na Figura 5 os dados supervisionados.

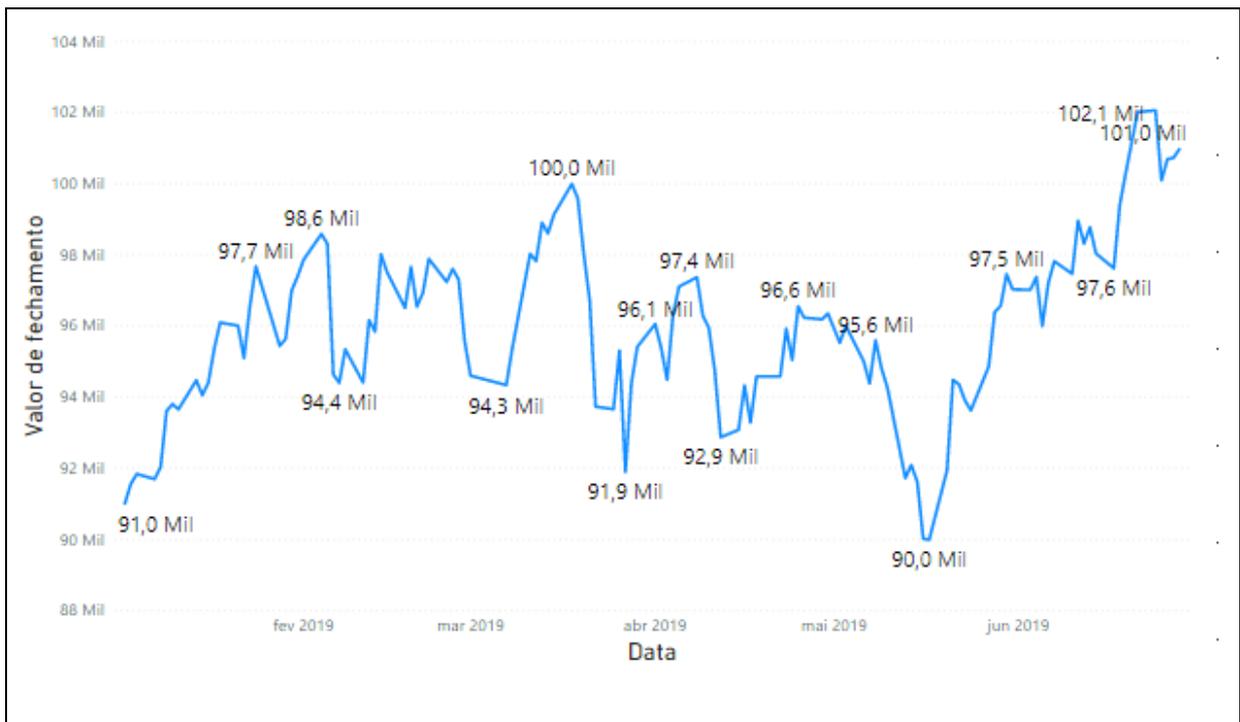


Figura 5 – Dados de Fechamento Supervisionados

Fonte básica dos dados: (YAHOO, 2019).

Quanto às notícias, realizou-se uma pesquisa documental o qual pode ser considerado um procedimento metodológico decisivo para pesquisa. Segundo Silva, Almeida e Guindani (2009) a pesquisa documental pode se caracterizar como o principal caminho de investigação, tendo como objetivo o acesso às fontes pertinentes, fazendo parte da heurística da investigação. Desse modo, possui o comprometimento de produzir novos conhecimentos, criando maneiras de entender os fenômenos e descrever a forma como foram desenvolvidos.

Foram mapeadas um total de 53 notícias com as respectivas datas que ocorreram no período dos dados supervisionados (02 de janeiro de 2019 a 28 de junho de 2019), considerando os três tipos mencionados no tópico 2.3 deste estudo.

Para maior confiabilidade, foi aplicado o raciocínio científico através do método hipotético-dedutivo, que, de acordo com Popper (2013), pode-se considerar que diante da

ausência de explicação para um fenômeno tem-se um problema, que pode ser expresso na forma de hipóteses. A dedução das hipóteses gera consequências passíveis de serem testadas ou falseadas. Por fim, Prodanov e Freitas (2013) afirmam que o método hipotético-dedutivo tem início com um problema de pesquisa ou a observação de uma lacuna na literatura científica, “seguindo-se pela formulação de hipóteses, e por um processo de inferência dedutiva, sendo este responsável por validar ou refutar a predição da ocorrência de fenômenos contidos nas referidas hipóteses”.

Observou-se que dos 122 dias que foram utilizados no aprendizado supervisionado, em 64 dias o índice exibiu um movimento de alta, 57 dias de baixa e um dia não se aplicou o movimento visto que foi o primeiro dia do índice. Para identificar o movimento, utilizou-se uma condicional que verifica dois estados:

Primeiro estado: se o valor de fechamento do dia atual for maior que o valor de fechamento do dia anterior, indica movimento de alta.

Segundo estado: se o valor de fechamento do dia atual for menor que o valor de fechamento do dia anterior, indica movimento queda.

Observa-se abaixo um exemplo para melhor entendimento, conforme Tabela 2.

Tabela 2 - Demonstração da Realização dos Movimentos de Queda e Alta

Dia	Valores	Movimento
02/01/2019	87887	Não se aplica
03/01/2019	91011	Alta
04/01/2019	91577	Alta
07/01/2019	91845	Alta
08/01/2019	91699	Queda

Fonte: Dados da pesquisa (2020).

Após essa classificação foi realizada a aplicação do método hipotético-dedutivo, e de 53 notícias, 46 foram utilizadas como conferência e análise dos dados, verificando se de fato houve impacto no índice ocasionando queda ou alta. Desse modo, os dados foram tabulados e organizados através do pacote Microsoft Office 365 *ProPlus* utilizando o software Microsoft Excel, relacionando as notícias com os respectivos dias dos índices. Os dados podem ser visualizados no Anexo 2 deste estudo.

Com todos os dados para realização da aplicação da RNA estruturados, foi configurado o ambiente, para isso foi utilizada a IDE – Ambiente de Desenvolvimento

Integrado (do inglês, *Integrated Development Environment*), Spyder na versão 3.3.6. Essa IDE é responsável pela codificação da RNA na linguagem de programação Python na versão 3.7.3. Realizou-se na sequência a configuração do pacote *Tensorflow* na versão 1.14.0. Esse pacote é responsável por detectar padrões durante o treinamento da RNA. Além deste, foi adotado o pacote *Keras* na versão 2.3.0 devido sua facilidade e adaptação em redes neurais profundas. Os pacotes *Sklearn*, *Numpy*, *Pandas* e *Matplotlib* também foram utilizados neste estudo, proporcionando grande flexibilidade em operações que necessitaram a utilização de vetores, matrizes, impressão de gráficos e modelos matemáticos em geral.

Ao iniciar a codificação da RNA, foi necessário definir o número de dias para aplicação da técnica de retro propagação (do inglês, *back propagation*) para otimizar os pesos durante o treinamento da RNA, tornando possível gerar um mapeamento mais preciso das entradas para o cálculo das saídas. A Figura 6 exhibe, com um exemplo fictício do funcionamento do *back propagation* para uma quantidade de 5 dias.

Data	Dia	Valor previsto	Valor real
02/01/2019	Dia 1	Não aplicável	87887
03/01/2019	Dia 2	Não aplicável	91011
04/01/2019	Dia 3	Não aplicável	91577
07/01/2019	Dia 4	Não aplicável	91845
08/01/2019	Dia 5	Não aplicável	91699
09/01/2019	Dia 6	92200	?
10/01/2019	Dia 7	93421	?
11/01/2019	Dia 8	92000	?
14/01/2019	Dia 9	94101	?

Definir o intervalo do tempo
Exemplo: 5 dias

Back Propagation (Dia 1, Dia 2, Dia 3, Dia 4, Dia 5) = valorPrevisto1
Back Propagation (Dia 2, Dia 3, Dia 4, Dia 5, valorPrevisto1) = valorPrevisto2
Back Propagation (87887, 91011, 91577, 91845, 91699) = 92200
Back Propagation (91011, 91577, 91845, 91699, 92200) = 93421
Back Propagation (91577, 91845, 91699, 92200, 93421) = 92000
Back Propagation (91845, 91699, 92200, 93421, 92000) = 94101

Figura 6 - Exemplificação do Movimento de *Back Propagation*

Fonte: Dados da pesquisa (2020).

Ao analisar a Figura 6, acima, é possível compreender que o modelo do *back propagation* incorporado na RNA LSTM possui um movimento de retro propagação, para isso, ele captura os dados de entrada de acordo com o parâmetro de dias, e depois, realiza os treinamentos e testes com esses dados. Nesse exemplo, com o parâmetro limitado a cinco dias, realizou-se uma exemplificação em que, são necessários os dias de 1 a 5 para realização da previsão do dia de número 6. Na seguinte iteração, o dia 2 é o primeiro dia a ser utilizado, e o primeiro valor previsto torna-se o último dia a ser inserido na RNA, uma vez que o valor do dia 6 é desconhecido.

Os primeiros treinamentos foram realizados com um *back propagation* de 30 dias. Isso significa que, para gerar uma saída, são necessários valores de 30 dias anteriores a ela para que a RNA possa realizar os cálculos.

Posteriormente, buscou-se definir a estrutura da RNA, como o número de camadas escondidas, a quantidade de neurônios, o número de épocas, a validação do valor do *back propagation* determinando o número de dias ideal para realização do movimento retro propagação para o ajustes dos pesos e o valor do *Dropout* que, de acordo com Srivastava et al. (2014), possui como principal objetivo, descartar de maneira aleatória conjuntos de neurônios para evitar com que as redes se adaptem demais, desse modo, a utilização da técnica de *Dropout* auxilia na prevenção do *overfitting*⁴.

Para definir a melhor arquitetura da RNA, foram realizados diversos treinamentos alterando os parâmetros da arquitetura da RNA, observando o valor do erro absoluto médio (do inglês, *mean absolute error*) em que, neste momento, não é calculado, uma vez que é automaticamente retornado através do método *compile* (Quadro 1) utilizado para compilar o treinamento da RNA, onde quanto menor for o valor, mais precisa está a estrutura. Observa-se abaixo os primeiros 8 treinamentos realizados conforme Tabela 3.

⁴ É considerado um termo estatístico utilizado para retratar modelos que possuem um ajuste muito preciso a um conjunto de dados utilizado anteriormente, porém, ao utilizar esse modelo para previsão de outros dados diferentes do conjunto utilizado, o modelo demonstra-se ineficaz.

Tabela 3 - Realização dos Primeiros Treinamentos da RNA

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
1	1 camada de entrada com 20 neurônios e peso 0,3 1 camada escondida com 20 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0379	00:41:32
2	1 camada de entrada com 25 neurônios e peso 0,3 1 camada escondida com 25 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0133	00:41:35
3	1 camada de entrada com 30 neurônios e peso 0,3 1 camada escondida com 30 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0133	00:41:40
4	1 camada de entrada com 35 neurônios e peso 0,3 1 camada escondida com 35 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0133	00:41:39
5	1 camada de entrada com 40 neurônios e	0,0131	00:41:41

	peso 0,3 1 camada escondida com 40 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias		
6	1 camada de entrada com 45 neurônios e peso 0,3 1 camada escondida com 45 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0125	00:42:01
7	1 camada de entrada com 50 neurônios e peso 0,3 1 camada escondida com 50 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0118	00:42:03
8	1 camada de entrada com 55 neurônios e peso 0,3 1 camada escondida com 55 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0252	00:42:40

Fonte: Dados da pesquisa (2020).

Observou-se nos primeiros 8 treinamentos, que o de número 7 foi o mais preciso, e após isso, o treinamento de número 8 possuiu um valor de *mean absolute error* alto, tendo um padrão diferente dos demais, visto que esse valor foi diminuindo a cada treinamento realizado.

Desse modo, a arquitetura do treinamento de número 7 foi a mais precisa até o momento, no entanto, foram efetuadas mais validações nos treinamentos para verificar se seria possível melhorar mais ainda a rede. Os treinamentos podem ser observados abaixo, de acordo com a Tabela 4.

Tabela 4 -Treinamentos 9 a 12

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
9	1 camada de entrada com 48 neurônios e peso 0,3 1 camada escondida com 48 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation de 30 dias</i>	0,0120	00:41:57
10	1 camada de entrada com 49 neurônios e peso 0,3 1 camada escondida com 49 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation de 30 dias</i>	0,0119	00:42:01
11	1 camada de entrada com 51 neurônios e peso 0,3 1 camada escondida com 51 neurônios e peso 0,3 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation de 30 dias</i>	0,0135	00:42:05
12	1 camada de entrada com 52 neurônios e peso 0,3 1 camada escondida com 52 neurônios e peso 0,3	0,0153	00:42:07

1 camada de saída com função de
ativação linear
20 épocas
Back propagation de 30 dias

Fonte: Dados da pesquisa (2020).

Nos treinamentos da Tabela 4, pode-se observar que o valor do *mean absolute error* também aumentou, indicando que o treinamento 7 continua sendo o mais preciso. Desse modo, foram realizados mais testes para validar o melhor treinamento, alterando os pesos da arquitetura da RNA conforme Tabela 5.

Tabela 5 - Treinamento 13 e 14

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
13	1 camada de entrada com 50 neurônios e peso 0,35 1 camada escondida com 50 neurônios e peso 0,35 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0136	00:42:13
14	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0116	00:42:10

Fonte: Dados da pesquisa (2020).

No término dos treinamentos 13 e 14, observou-se uma queda nos valores que medem a precisão, conforme treinamento 14, desse modo, ao utilizar um peso de valor 0,25, a rede começou a se adaptar de maneira mais precisa. Desse modo, buscou-se diminuir e aumentar o valor dos pesos, para avaliar se seria possível aprimorar ainda mais a RNA, conforme Tabela 6.

Tabela 6 - Treinamento 15 a 20

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
15	1 camada de saída com 50 neurônios e peso 0,20 1 camada escondida com 50 neurônios e peso 0,20 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0119	00:42:11
16	1 camada de entrada com 50 neurônios e peso 0,21 1 camada escondida com 50 neurônios e peso 0,21 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0118	00:42:05
17	1 camada de entrada com 50 neurônios e peso 0,22 1 camada escondida com 50 neurônios e peso 0,22 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0141	00:42:11
18	1 camada de entrada com 50 neurônios e peso 0,23 1 camada escondida com 50 neurônios e	0,0187	00:42:07

	peso 0,23 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias		
19	1 camada de entrada com 50 neurônios e peso 0,24 1 camada escondida com 50 neurônios e peso 0,24 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0137	00:42:21
20	1 camada de entrada com 50 neurônios e peso 0,26 1 camada escondida com 50 neurônios e peso 0,26 1 camada de saída com função de ativação linear 20 épocas <i>Back propagation</i> de 30 dias	0,0395	00:42:14

Fonte: Dados da pesquisa (2020).

Ao realizar a variação dos pesos, aumentando ou diminuindo, foi possível observar que o melhor treinamento prevaleceu no de número 14. Com o intento de observar uma possível evolução, realizou a adaptação do algoritmo para aumentar o número de épocas em que o treinamento é compilado através de uma função denominada *call-back*, no qual interrompe o treinamento quando não há mais melhorias. Observa-se, no entanto, que quanto maior o número de épocas, camadas e neurônios, maior é o tempo de processamento da RNA. Os dados estão registrados na Tabela 7.

Tabela 7 - Treinamento 21

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
21	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 57 épocas <i>Back propagation</i> de 30 dias	0,0110	00:42:11

Fonte: Dados da pesquisa (2020).

Ao finalizar o teste conforme Tabela 7, observou-se que os resultados do treinamento 21 superaram os do treinamento 14, diminuindo ainda mais o valor do *mean absolute error*. Uma outra alteração nos parâmetros da rede, foi aumentar o número de camadas escondidas, que pode ser observado, a seguir de acordo com a Tabela 8.

Tabela 8 - Treinamento 22

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
22	1 camada de entrada com 50 neurônios e peso 0,25 2 camadas escondidas com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 57 épocas <i>Back propagation</i> de 30 dias	0,0121	01:28:17

Fonte: Dados da pesquisa (2020).

Foi evidenciado um grande aumento no processamento durante o treinamento da rede que pode ser observado através do tempo de processamento visto que nos treinamentos anteriores o tempo era em média de 42 minutos, e após o aumento das camadas escondidas, a duração do treinamento praticamente dobrou, além disso, não foram observadas melhorias no treinamento para a configuração realizada na Tabela 8. Por fim, as próximas validações

retratam a combinação dos parâmetros anteriores aumentando e diminuindo o número dias na utilização do *back propagation*. Foram realizadas mais cinco tentativas conforme Tabela 9.

Tabela 9 - Treinamento 23 – 27

Treinamento	Estrutura	Mean Absolute Error	Tempo de Processamento
23	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 57 épocas <i>Back propagation</i> de 15 dias	0,0198	00:41:30
24	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 57 épocas <i>Back propagation</i> de 45 dias	0,0115	00:42:37
25	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear 57 épocas <i>Back propagation</i> de 60 dias	0,0109	00:42:41
26	1 camada de entrada com 50 neurônios e peso 0,25 1 camada escondida com 50 neurônios e peso 0,25 1 camada de saída com função de ativação linear	0,0103	00:42:51

	57 épocas		
	<i>Back propagation</i> de 75 dias		
27	1 camada de entrada com 50 neurônios e peso 0,25	0,0210	00:42:58
	1 camada escondida com 50 neurônios e peso 0,25		
	1 camada de saída com função de ativação linear		
	57 épocas		
	<i>Back propagation</i> de 90 dias		

Fonte: Dados da pesquisa (2020).

Observou-se que ao utilizar um *back propagation* de 60 e 75 dias, os resultados foram superiores ao do treinamento de número 21. Dessa forma, após a validação da arquitetura da RNA, os três melhores treinamentos (21, 25 e 26) foram separados para realização dos testes, tendo como retorno os valores previstos. Desse modo, na próxima seção, serão exibidos em detalhes as características das três melhores topologias que foram submetidas aos testes retornando os dados previstos e, em sequência, foram avaliados através das medidas de desempenho.

4 RESULTADOS E DISCUSSÃO

Inicialmente, as análises foram agrupadas pelo número do treinamento, ou seja, para cada treinamento foi realizado um teste com o retorno dos valores previstos. Observa-se, na Tabela 10, as métricas dos melhores testes.

Tabela 10 - Métricas de Desempenho dos Principais Testes da RNA

Teste	MSE	RMSE	MAPE	Coefficiente de determinação
25	1.536.281,9	1.239,46	1	0,7771
26	1.754.128,1	1.324,43	1,08	0,7812
21	3.548.793,5	1.883,82	1,64	0,6965

Fonte: Dados da pesquisa (2020).

Observa-se na Tabela 10 os valores das métricas de desempenho dos melhores testes abordados na RNA. Para interpretação dos dados, pode-se levar em consideração as medidas pela capacidade de explicar a taxa de erro/acerto de um modelo. Assim sendo, a medida mais

relevante é o MAPE, seguida pelo RMSE, na sequência o Coeficiente de Determinação e por fim o MSE, no entanto as medidas de MSE e RMSE podem colocar mais peso em valores altos de erros não fornecendo uma análise precisa (KIM; KIM, 2016).

Ao analisar cada teste, foram realizados os cálculos para obtenção das métricas de desempenho. O teste 25 se destacou entre os demais por possuir os menores valores do MSE, RMSE e MAPE se comparados aos demais testes, indicando a melhor aproximação dos valores quando comparados aos valores reais. O coeficiente de determinação foi o segundo melhor, relacionando 77,71% os valores.

O teste 26 possuiu o maior valor do coeficiente de determinação, possuindo uma relação de 78,12% dos valores, além de ser muito similar ao teste 25, com diferenças mínimas nos valores de MSE, RMSE e MAPE. Já o teste 21 foi inferior em todas as medidas, se comparado aos testes 25 e 26, sendo descartado neste momento.

A seguir, são visualizadas cada métrica agrupadas por grupo de acordo com a Figura 7.

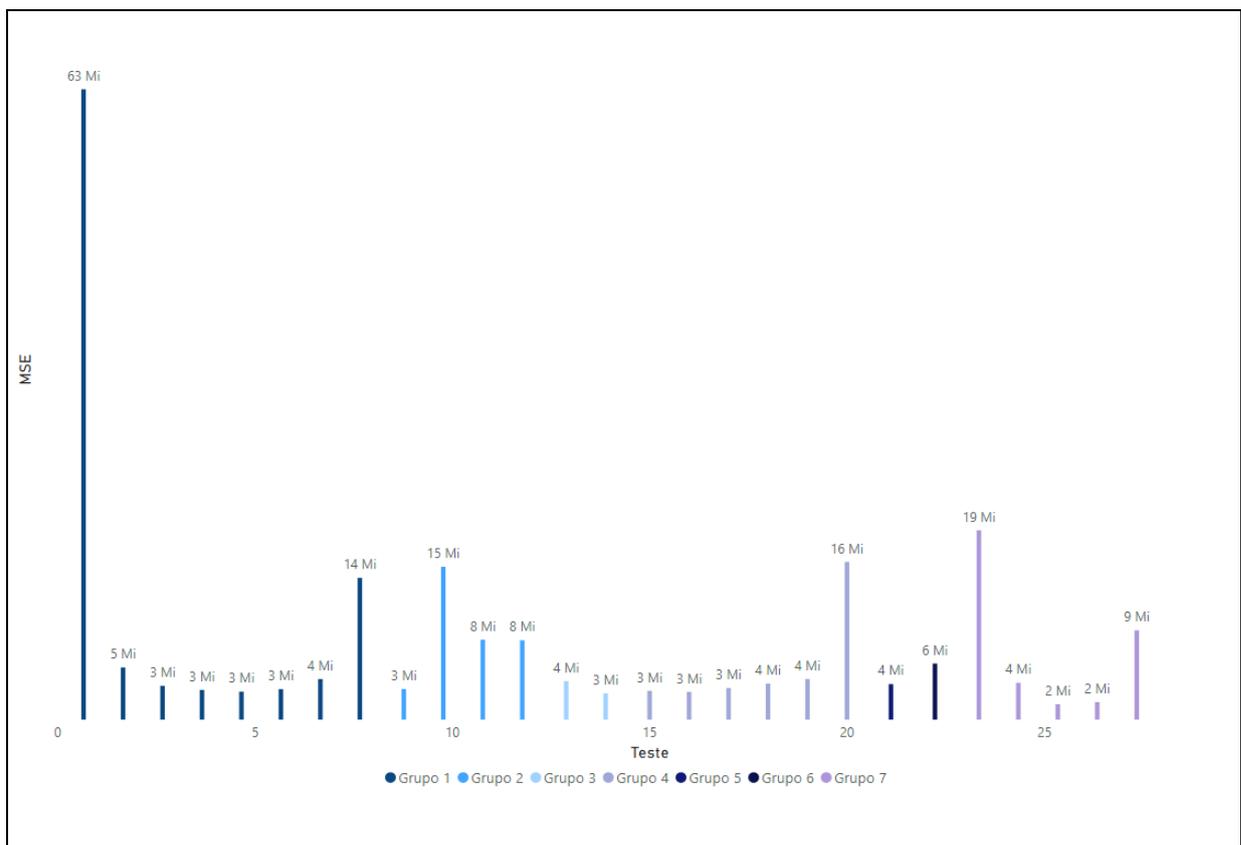


Figura 7 - MSE de Todos os Testes

Fonte: Dados da pesquisa (2020).

Os valores obtidos nos testes foram agrupados conforme as Tabelas de treinamentos visualizadas na seção anterior. Desse modo, O Grupo 1 contém os testes que correspondem aos treinamentos de 1 a 8, o Grupo 2 possui os testes de 9 a 12, o Grupo 3 os testes 13 e 14, o Grupo 4 os testes de 15 a 20, Grupo 5 possui o teste de número 21, o Grupo 6 retrata o teste 22 e, por fim, o Grupo 7 evidencia os testes de número 23 a 27

Inicialmente, pode-se observar um valor discrepante para o primeiro teste pertencente ao Grupo 1, e no decorrer dos testes, foi possível analisar os menores valores, presentes nos testes 25 e 26 correspondentes ao Grupo 7. A seguir, em detalhes, encontram-se as informações do RMSE conforme Figura 8.

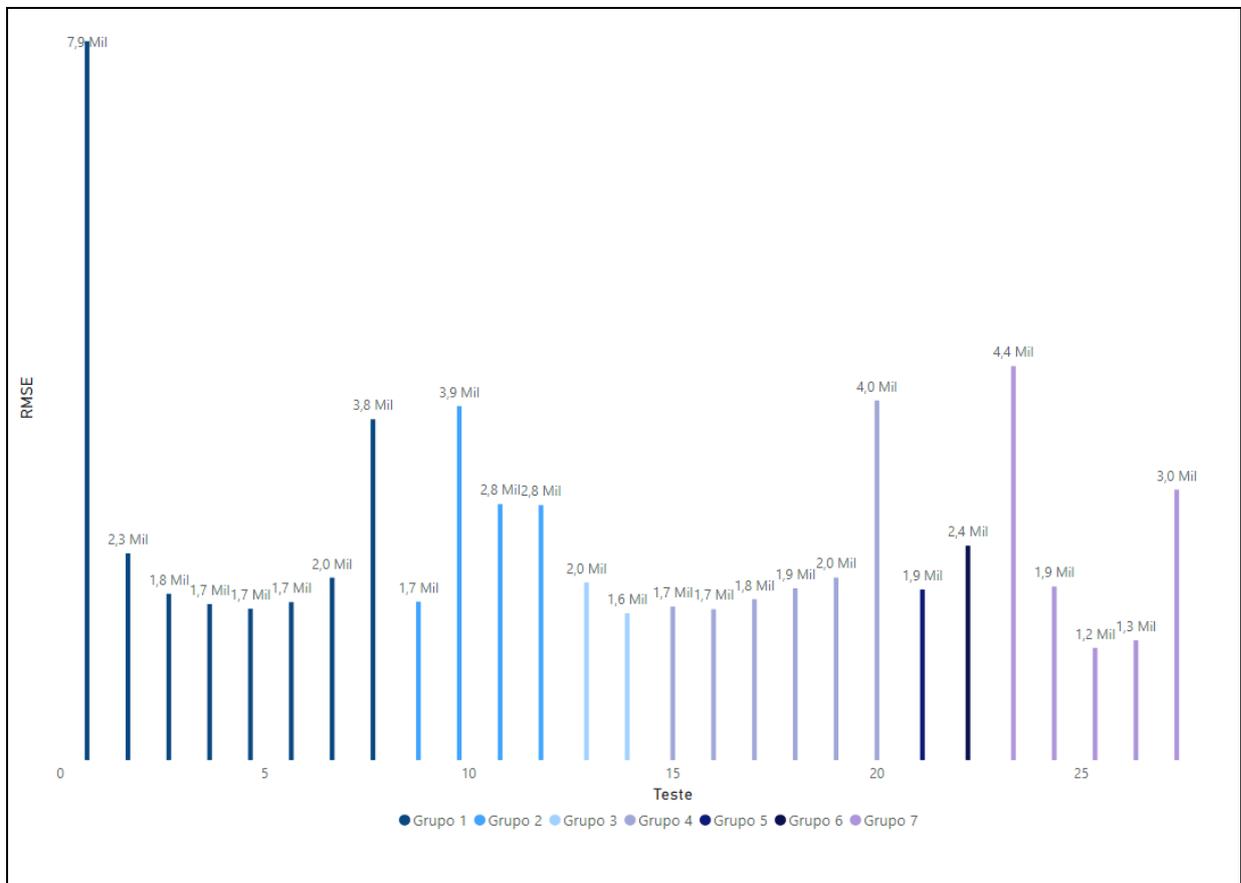


Figura 8 - RMSE de Todos os Testes

Fonte: Dados da pesquisa (2020).

O RMSE também possuiu valores discrepantes nos primeiros testes pertencentes ao Grupo 1, e à medida que os parâmetros da RNA foram modificados, o RMSE oscilou entre valores altos e baixos, porém, foi observado os melhores resultados (considerando os menores valores dentre todos os testes) nos testes 25 e 26 também pertencentes ao Grupo 7. No entanto, observou-se que as medidas de MSE e RMSE possuíram um valor muito alto, visto

que o cálculo concedeu um peso muito grande aos maiores erros. Na Figura 9, serão observados os valores do Erro Percentual Médio, no qual é considerado o mais popular dentre as medidas de desempenho em estudos de previsão, sendo capaz de descrever se o ajuste da previsão se encontra em um bom estado, isto é, quando o retorno dessa métrica está muito próximo de zero.

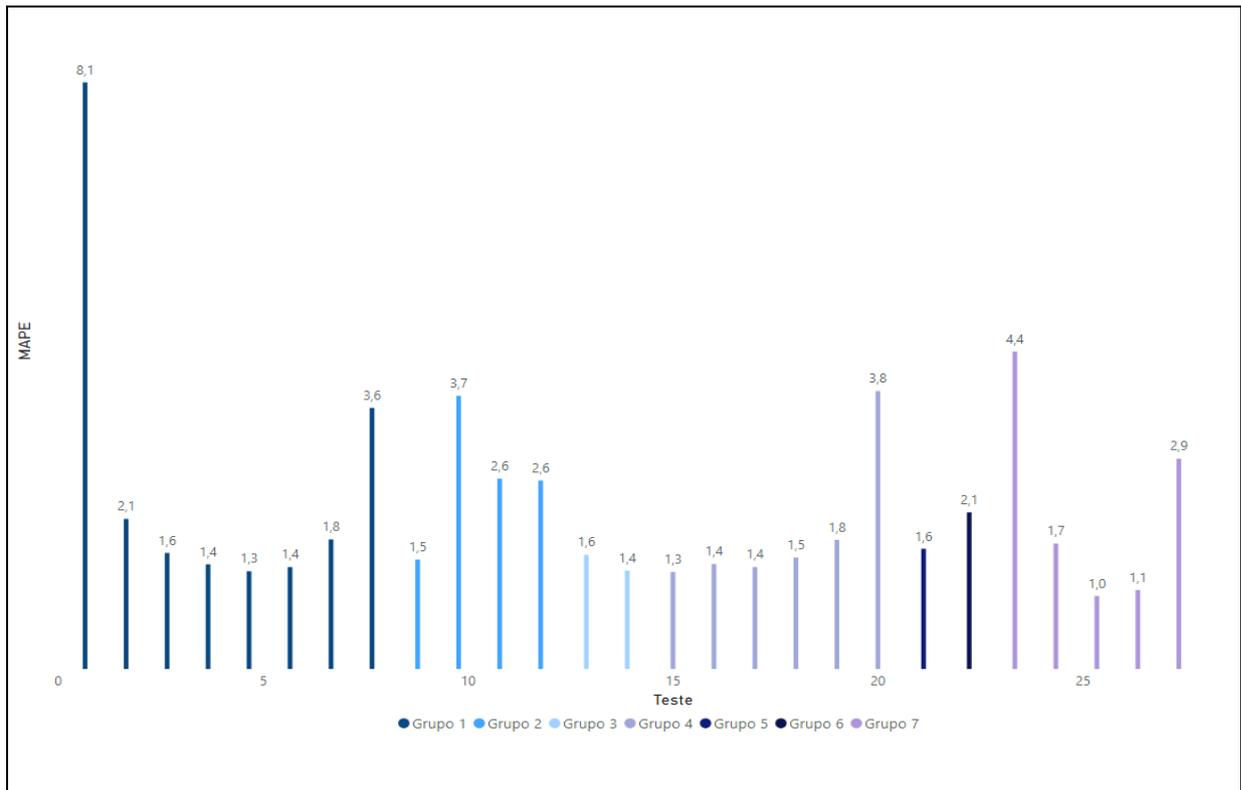


Figura 9 - MAPE de Todos os Testes

Fonte: Dados da pesquisa (2020).

Conforme mencionado, quanto mais próximo de zero o MAPE se encontra, mais precisos estão os resultados. Neste caso, pode-se observar também que os testes 25 e 26 demonstram os melhores valores, respectivamente, 1,0 e 1,1. Por fim, observa-se na Figura 10, os valores do coeficiente de determinação, que, diferentemente das demais métricas, possui um bom desempenho quando os valores estão mais próximos de 1.

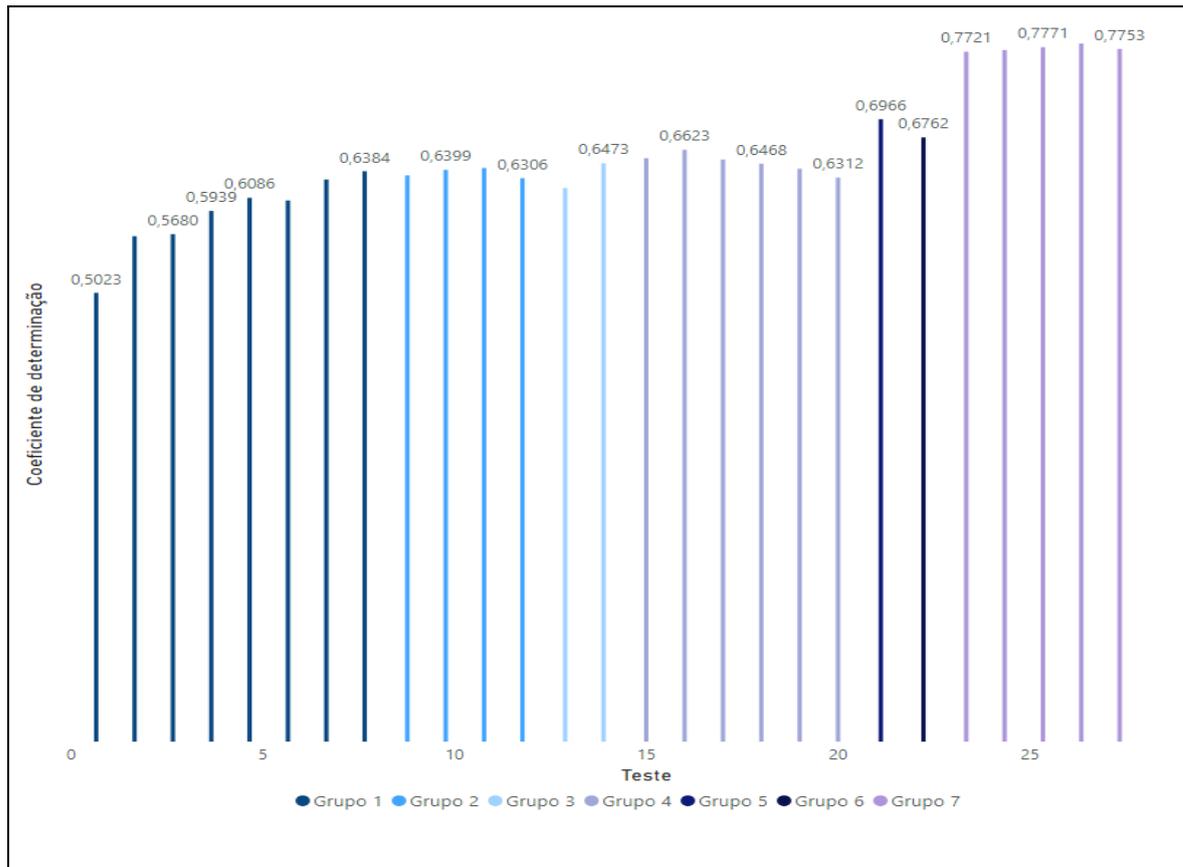


Figura 10 - Coeficiente de Determinação de Todos os Testes

Fonte: Dados da pesquisa (2020).

Pode-se observar que os testes do Grupo 7 demonstraram um grande percentual de correlação. Dessa forma, com os ajustes que foram realizados nos treinamentos, observou-se que os valores encontrados na previsão podem ser utilizados para uma análise mais detalhada.

Ao analisar os dados do Anexo 2, foi possível observar que o IBOVESPA possuiu movimento de alta em 16 dias, queda em 24 dias e no primeiro dia, o movimento não foi considerado, totalizando 41 dias com acontecimentos que impactaram o índice. A Tabela 11 descreve, em detalhes, os dias em alta e dias em queda.

Tabela 11 - Análises dos Testes *versus* Acontecimentos

Tipo de dados	Dias em alta	Dias em queda	Não se aplica	Acertos
Valores reais	16	24	1	-
Teste 25	20	20	1	19
Teste 26	21	19	1	18

Fonte: Dados da pesquisa (2020).

Observou-se que as notícias impactam diretamente na capacidade da RNA de acompanhar o movimento de alta ou baixa, devido o número de acertos nos testes 25 e 26. Por fim, observa-se uma análise geral dos dados reais *versus* os dados previstos conforme Tabela 12.

Tabela 12 - Análise Macro dos Dados Reais *versus* Dados Previstos

Tipo de dados	Dias em alta	Dias em queda	Não se aplica	Acertos
Valores reais	64	57	1	-
Teste 25	65	56	1	55
Teste 26	66	55	1	54

Fonte: Dados da pesquisa (2020).

Embora as taxas de acerto apresentadas na Tabela 12 sejam relativamente baixas, pode-se observar que essa diferença ocorreu devido um atraso na RNA onde os movimentos de alta e queda foram, em muitas vezes, ocorridos alguns dias após o verdadeiro movimento. Tal comportamento pode ser analisado abaixo, nas Figuras 11 e 12 que traçam uma comparação dos dados reais *versus* os dados previstos.

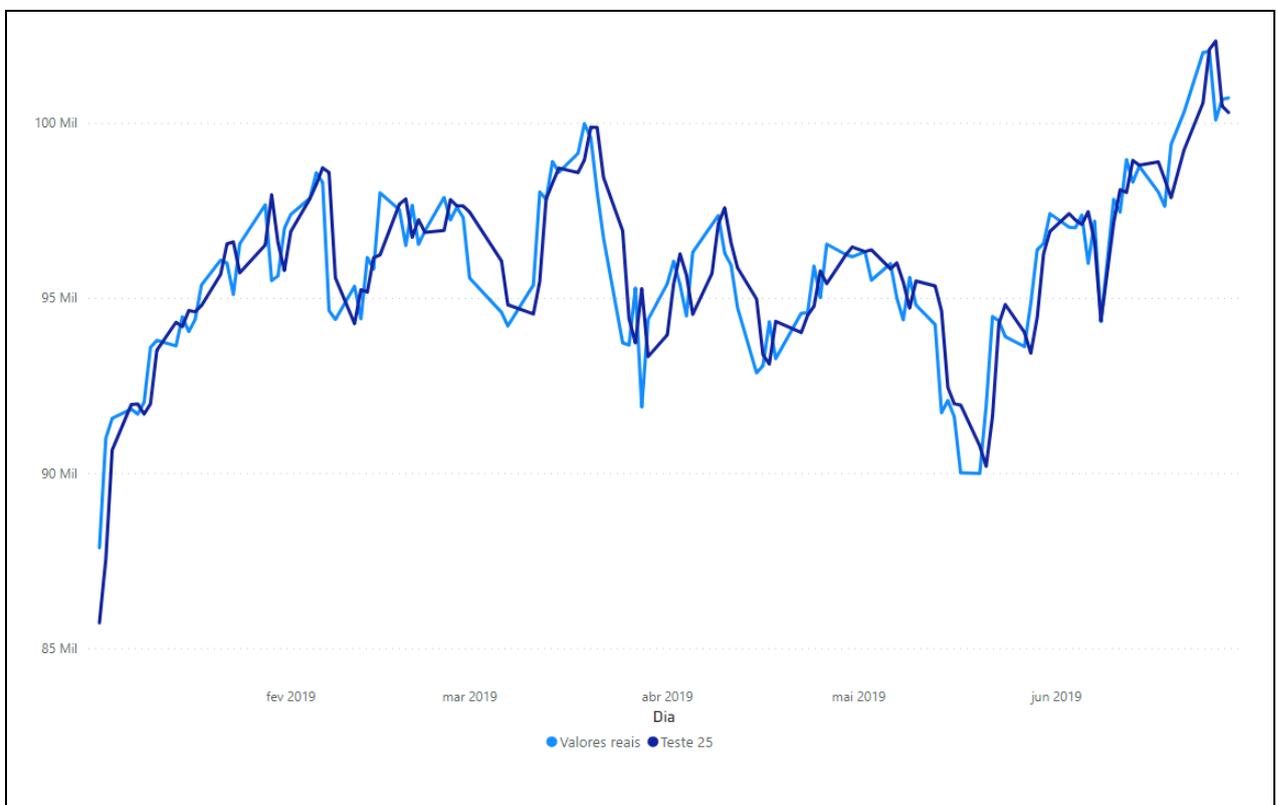


Figura 11 - Plotagem do Teste 25

Fonte: Dados da pesquisa (2020).

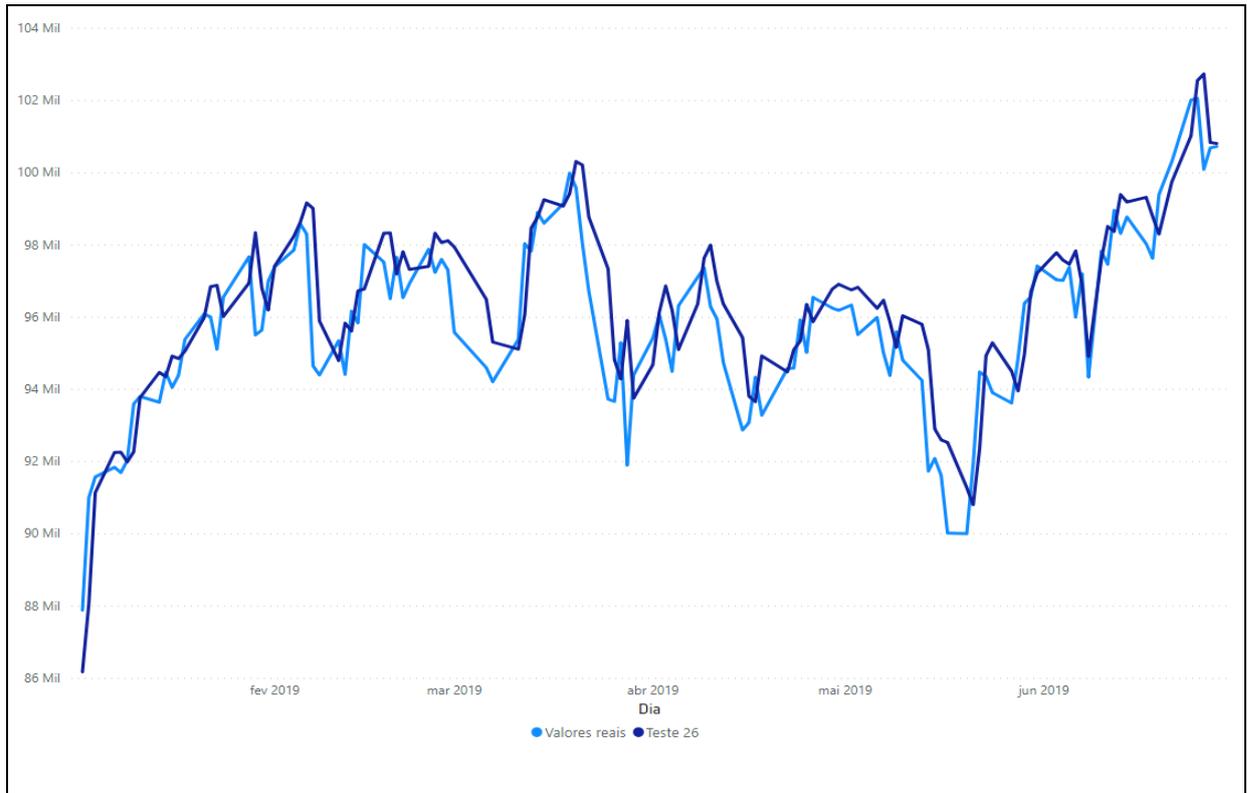


Figura 12 - Plotagem do Teste 26

Fonte: Dados da pesquisa (2020).

Ao realizar as análises gráficas, uma das possíveis interpretações é de que o IBOVESPA, apesar de suas oscilações no decorrer dos meses, demonstrou características de crescimento considerando um período semestral.

E após, as análises estatísticas e gráficas, pode-se observar que o Teste 25, que correspondente ao treinamento 25, possui mais proximidade com os valores reais que os demais. Desse modo, realizou-se uma relação entre dados reais *versus* dados previstos *versus* notícias, para tornar claro o impacto implicado pelos acontecimentos.

Observou-se com maior frequência esse impacto no final do mês de janeiro, em todo o mês de fevereiro e abril. As Figuras 13 e 14 demonstram a tendência real e prevista *versus* as notícias.

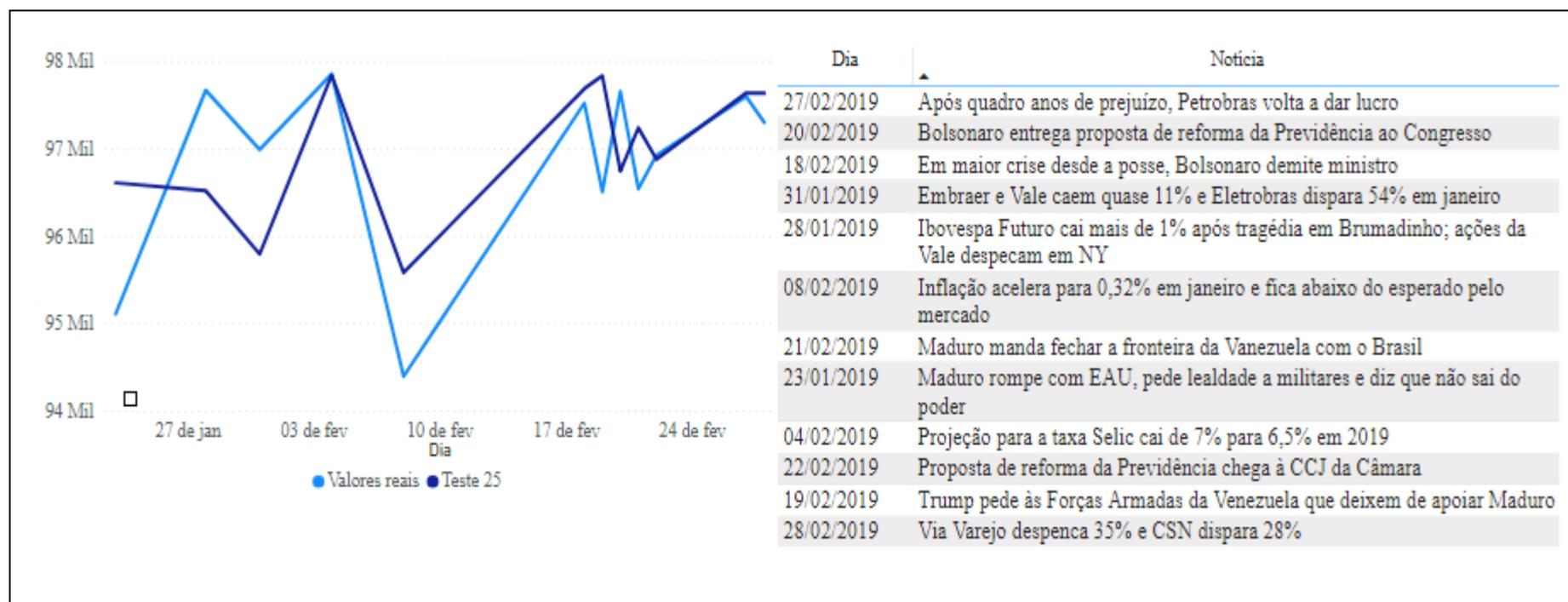


Figura 13 – Comparação dos Dados no Final do Mês de Janeiro e Todo o Mês de Fevereiro

Fonte: Dados da pesquisa (2020).

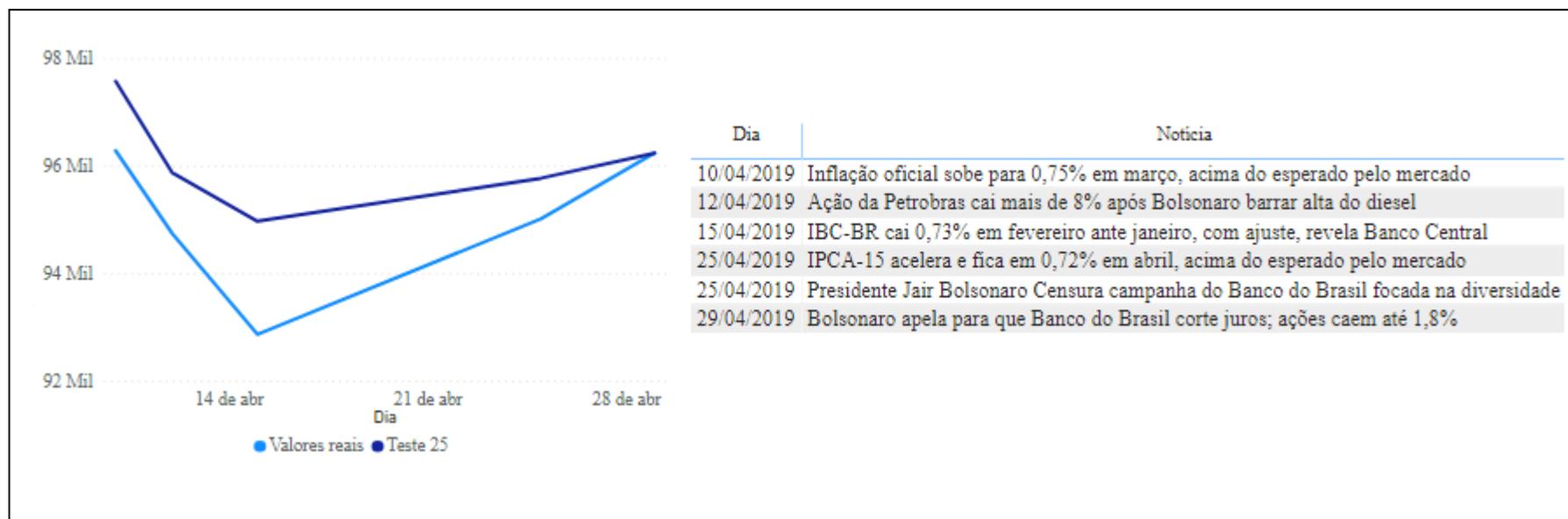


Figura 14 – Comparação dos Dados no Mês de Abril

Fonte: Dados da pesquisa (2020).

Ao observar as Figuras 13 e 14, pode-se interpretar que os maiores erros foram encontrados nos meses de fevereiro totalizando 6 dias consecutivos, que pode ser explicado devido o acontecimento em que a Barragem de Brumadinho foi rompida resultando em um dos maiores desastres de rejeitos de mineração no Brasil, instabilidade política e aumento da inflação. E 8 erros consecutivos no mês de abril, relacionados a um apelo do governo brasileiro para redução de juros, projeção de alta inflação e aumento no valor do Diesel no Brasil.

Por se tratar de uma previsão, os dados históricos são importantes para que a RNA consiga prever os passos seguintes, desse modo, após a validação dos melhores treinamentos e testes, foram realizados os testes mês a mês para uma melhor previsão, baseado na Topologia 25, importando os dados até o final de 2018 para a previsão do mês de janeiro de 2019 e, assim, subsequentemente até o mesmo de junho de 2019, tal fato corrobora a Ehlers (2007) que menciona que uma das características que definem uma série temporal, é que as observações vizinhas são dependentes, enquanto em modelos de regressão a ordem das observações é irrelevante, para as séries temporais cada dado depende dos valores anteriores. A seguir, pode-se observar os dados dos testes segmentados por mês conforme Tabela 13.

Tabela 13 - Métricas de Desempenho da Previsão da Arquitetura de Treinamento 25 Segmentada por Mês

Mês	MSE	RMSE	MAPE	Coefficiente de Determinação
Janeiro	683.491,9	826,73	0,7	0,8948
Fevereiro	803.350,8	896,29	0,8	0,8687
Março	1.256.760	1.121,05	0,9	0,7859
Abril	930.410,1	964,57	0,8	0,6107
Mai	1.971.827	1.404,21	1,21	0,7073
Junho	773.243,2	879,34	0,69	0,7458
Semestre	1.081.223	1.039,81	0,8	0,8258

Fonte: Dados da pesquisa (2020).

Os dados da Tabela 13 evidenciam que ao compará-los com os achados do teste 25 conforme Tabela 10, obteve-se valores superiores em todas as métricas. Desse modo, ao reduzir a previsão para um mês, a precisão aumenta. Devido a isso, foi realizada a comparação com as demais Tabelas, analisando a quantidade de acertos dessa nova previsão

para os movimentos de alta e queda. Na Tabela 14, são apresentados os novos números de queda, alta e acertos.

Tabela 14 – Análise do Teste 25 por mês *versus* Acontecimentos

Tipo de dados	Dias em alta	Dias em queda	Não se aplica	Acertos
Valores reais	16	24	1	-
Teste 25 por mês	19	22	1	24

Fonte: Dados da pesquisa (2020).

Observou-se uma evolução dos resultados, se comparadas as Tabelas 11 e 14, visto que o número de acertos nos dias que possuem notícias aumentou para 24 totalizando 60%. Desse modo, a RNA foi capaz de absorver melhor as informações, além de não ser impactada de maneira predominante por acontecimentos polêmicos. Além desta análise, os dados de maneira macro são apresentados a seguir na Tabela 15.

Tabela 15 - Análise Macro do Teste 25

Tipo de dados	Dias em alta	Dias em queda	Não se aplica	Acertos
Valores reais	64	57	1	-
Teste 25 por mês	62	59	1	69

Fonte: Dados da pesquisa (2020).

Inicialmente o Teste 25 possuiu uma taxa de acerto de 55 dias representando apenas 45,45% sobre o período analisado. Após diminuir o período da previsão, o Teste 25, com os dados mensais, obteve uma taxa de acerto de 69 dias que corresponde a aproximadamente 57% sobre o período. Desse modo, observa-se uma maior evolução na previsão em um contexto macro.

Dessa forma, realizou-se a plotagem gráfica do teste conforme Figura 15.

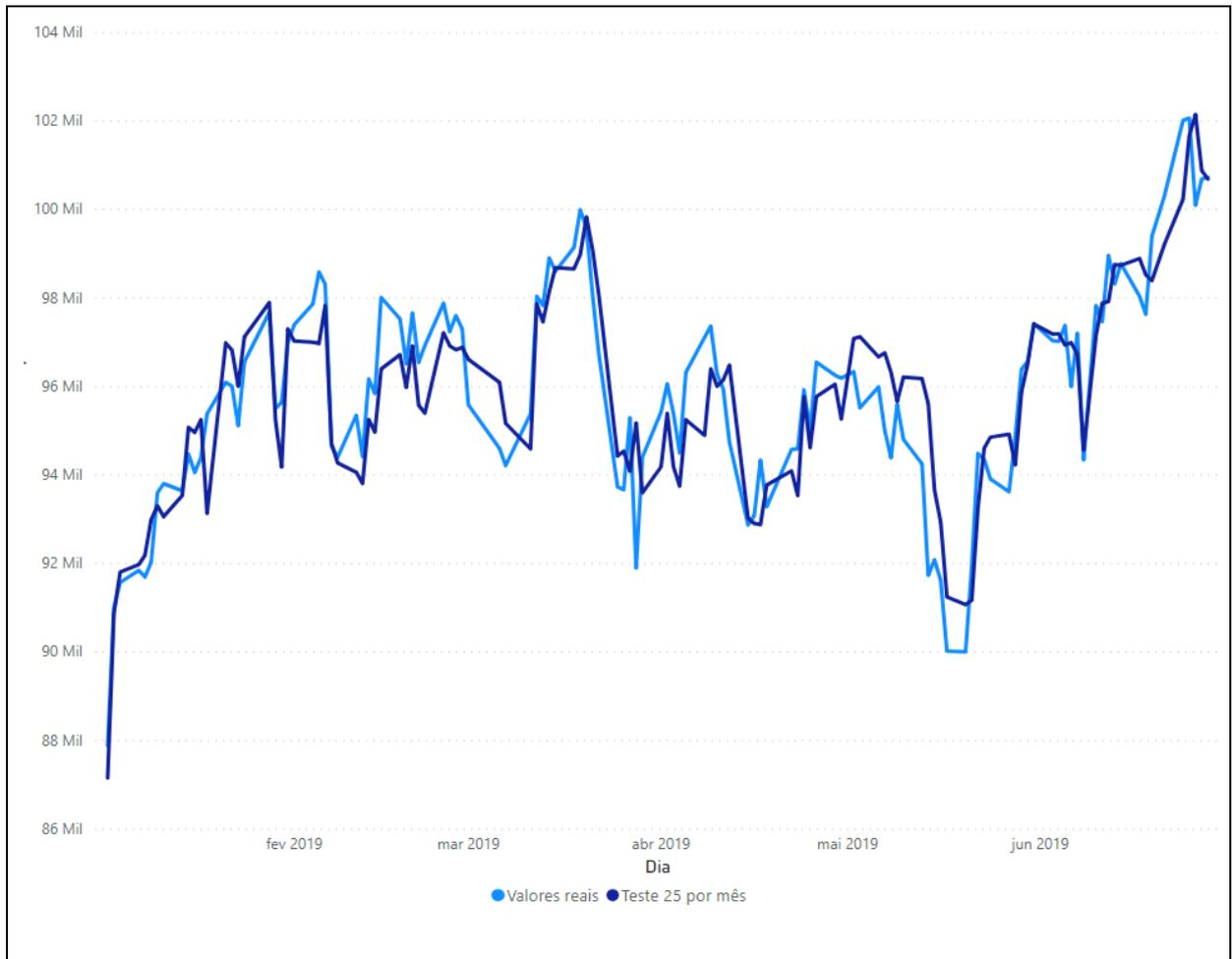


Figura 15 - Plotagem do Teste 25 Realizado por Mês

Fonte: Dados da pesquisa (2020).

Após a plotagem do gráfico conforme Figura 15, buscou-se identificar quais os principais períodos sofreram impactos por notícias.

Para os meses de janeiro e maio, não foram detectados impactos. Os meses de fevereiro e abril melhoraram após a redução de período da RNA. Desse modo, realizou-se a plotagem dos principais acontecimentos do qual a RNA não conseguiu acompanhar a tendência de queda ou alta para os meses de fevereiro e abril. Os dados são visualizados a seguir nas Figuras 16 e 17.

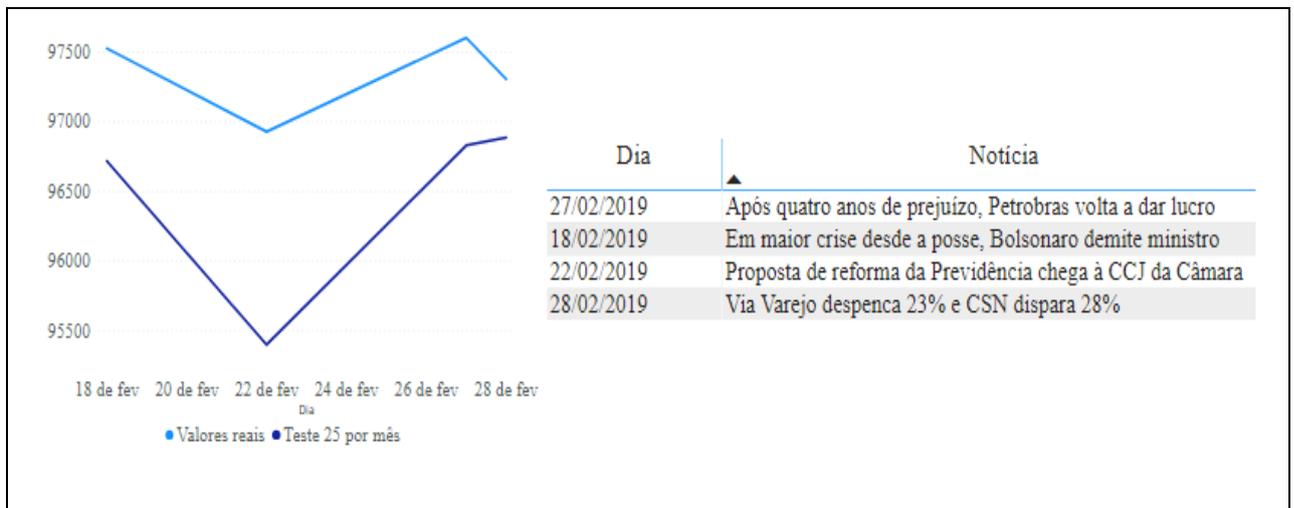


Figura 16 – Comparação dos Dados no Mês de Fevereiro Após o Teste Mensal

Fonte: Dados da pesquisa (2020).

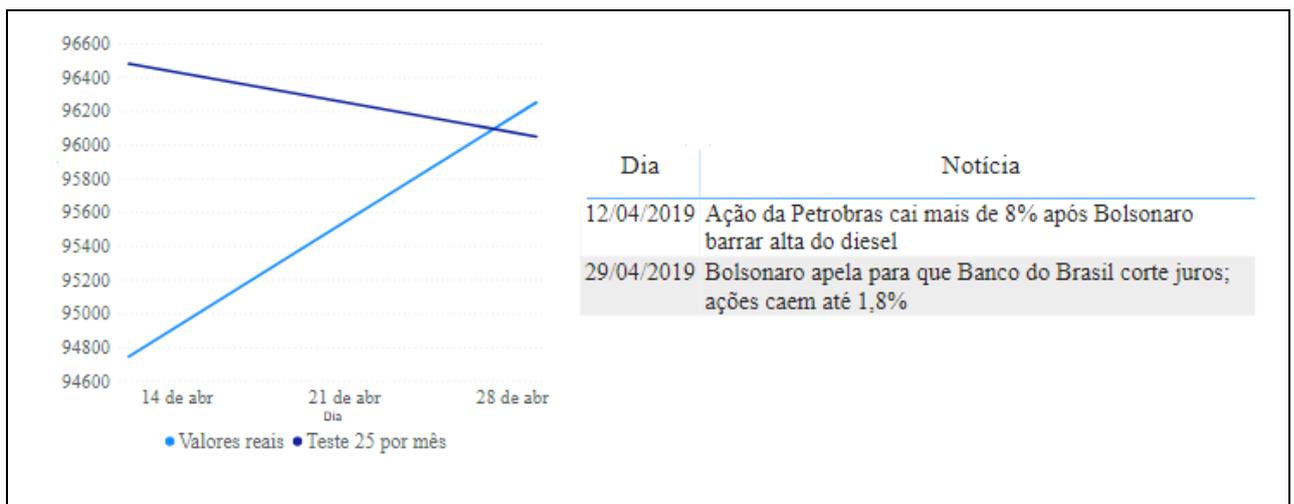


Figura 17 – Comparação dos Dados no Mês de Abril Após o Teste Mensal

Fonte: Dados da pesquisa (2020).

Ao observar as Figura 16 e 17, destaca-se que após a submissão dos testes mensais o número de notícias que impactaram o movimento de previsão da RNA diminuiu gradativamente. Tal comparação pode ser realizada ao analisar as informações da Figura 13 e compará-las com a Figura 16, do mesmo modo, a Figura 14 pode ser comparada com a Figura 17.

Por fim, para concluir os resultados desta seção, foi analisado o estudo de Roque (2009) para avaliar o desempenho de seu estudo, o autor utilizou o percentual de acurácia

considerando margens de erros. No estudo do autor, foram estipuladas as margens de 1% a 5% para mais ou para menos. O autor encontrou as seguintes taxas de acerto: 1% apenas 15,35% de acerto; 2% o total de 27,53% de acerto; 3% o valor de 43,07% de acerto; 4% uma taxa de 59,36%; 5% o valor de 72,85% de acerto.

Em sua análise o objetivo era superar 50% de acerto, no qual foi atingido com uma margem de erro de 4% para mais ou para menos. Desse modo, a análise foi replicada para este estudo conforme Figura 18.

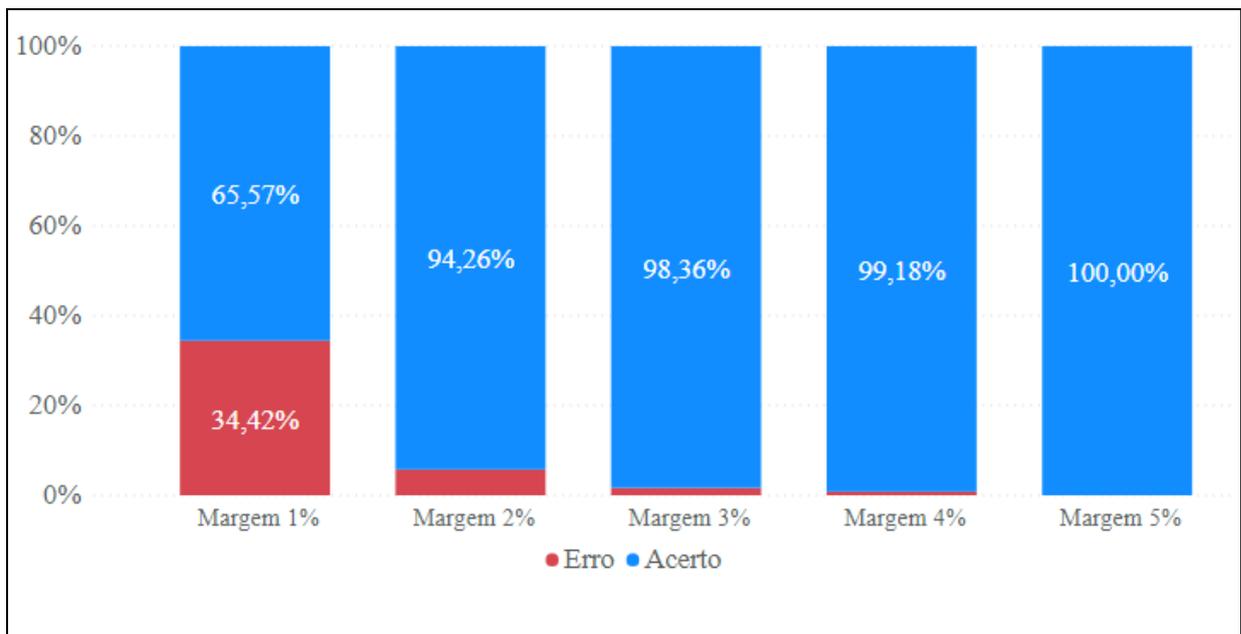


Figura 18 - Percentual de Acurácia da Previsão Considerando Margens de Erro

Fonte: Dados da pesquisa (2020).

Ao analisar os dados da previsão conforme Figura 18, pode-se observar uma taxa de erro de 34,42% considerando uma margem de erro de 1% para mais ou para menos. Isso demonstra que em 65,58% dos valores, os dados previstos se encontram próximos dos valores reais considerando essa margem de 1%. No entanto, ao aumentar a margem de erro, a previsão se torna mais precisa, com um potencial de acerto de mais de 90% para as margens de 2% até 5%.

5 CONSIDERAÇÕES FINAIS

Este estudo apresentou resultados e técnicas metodológicas que podem ser relevantes ao investidor brasileiro, desde o perfil agressivo e, conseqüentemente, aos demais perfis, realizando adaptações nas análises e objetivos. Destaca-se como maior contribuição a

previsão do IBOVESPA após várias alterações no parâmetro da RNA, onde foi possível obter um valor de erro percentual absoluto médio de 0,8 e 0,8258 no que se diz respeito ao coeficiente de determinação, sendo possível afirmar que os valores previstos neste estudo possuem uma boa estimativa de tendência. Com a adaptação da RNA, observou uma proximidade dos dados reais *versus* os dados previstos, concedendo a um investidor, por exemplo, a possibilidade de utilizar os dados históricos já existentes e traçar uma possível estratégia com a previsão dos dados em um período futuro.

Outro ponto a ser destacado pode ser observado durante a comparação das notícias com os dados reais e previstos do Teste 25, no qual foi realizado em um formato de longo prazo (semestral), sendo possível afirmar que houve um impacto desses acontecimentos na previsão, uma vez que houve uma taxa de acerto nos movimentos (alta ou queda) de 47,5% em dias que notícias foram mapeadas (40 dias). Em dias que não houve fatos relevantes, a taxa de acerto foi de 43,4% e, no contexto geral, considerando dias com e sem notícias, a previsão demonstrou uma taxa de acerto de aproximadamente 45,5%.

Após analisar esse impacto, ao diminuir o período para um formato de curto prazo (mensal), os dados previstos alcançaram um total de 60% de acerto em dias que houve acontecimentos e, no geral, a taxa de acerto foi de aproximadamente 57%, demonstrando a evolução da RNA. Desse modo, o objetivo deste estudo, que consiste em fornecer uma ferramenta de auxílio para o investidor de médio e/ou longo prazo que busca maiores rendimentos e aceita o risco de queda em curtos períodos de tempo, foi atingido. Além disso, observou-se que a RNA foi otimista em relação aos dados reais, visto que, para dias em que as quedas foram observadas, a previsão retornou um comportamento de alta, e, com a análise dos acontecimentos polêmicos, esse desempenho foi aceitável e reduzido no decorrer as alterações das arquiteturas de treinamento.

Concluiu-se que a RNA LSTM possui uma grande capacidade de aprendizado e acerto em suas previsões se utilizada com um grande volume de dados, mas para isso, torna-se necessária a realização de diversos treinamentos, alterando os parâmetros como neurônios, quantidade de camadas, número de épocas, entre outros, com o intento de obter melhores resultados, se tornando uma ferramenta que pode ser adotada por um investidor que busca traçar estratégias baseadas em médio e longo prazo, através de uma ferramenta de previsão.

Este estudo tem como principais limitações a maneira com que os treinamentos para melhoria da RNA foram realizados, uma vez que foi necessário a inspeção e, posteriormente, a alteração dos parâmetros da RNA. Para isso, torna-se necessário conhecimentos técnicos nos pacotes e linguagem de programação Python. Destaca-se ainda, além desta melhoria, a

possibilidade de criação de uma interface *Web* ou *Mobile*, onde o próprio investidor consiga realizar a importação de uma base de dados para treinamento, testes, geração de gráficos e retorno de informações estatísticas avaliando o desempenho da RNA.

REFERÊNCIAS

ANTUNES, A. **Qual é o impacto das notícias nos preços das Ações, Índice e Dólar?** 2019. Disponível em: <<https://scalpertrader.com.br/impacto-das-noticias-nas-acoes-indicie-e-dolar/>>. Acesso em: 6 jan. 2020.

B3. **Histórico pessoas físicas.** 2019a. Disponível em: <http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/historico-pessoas-fisicas/>. Acesso em: 23 dez. 2019.

B3. **Há 50 anos, com o mercado, para o futuro.** 2019b. Disponível em: <http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm>. Acesso em: 30 set. 2019.

B3. **Índice Bovespa (Ibovespa): composição da carteira.** 2019c. Disponível em: <http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/indice-ibovespa-ibovespa-composicao-da-carteira.htm>. Acesso em: 14 abr. 2019.

CARAMICO, A. Proposição de um modelo preditivo do Ibovespa utilizando redes neurais artificiais. **CAD**, [s. l.], v. 7, n. 1, p. 21–42, 2012.

CASTRO, F. H. et al. Fifty-year history of the Ibovespa. **Revista Brasileira de Finanças (Online)**, Rio de Janeiro, p. 47–65, 2019.

CHAUVIN, Y.; RUMELHART, D. E. **Back propagation: theory, architectures, and applications.** 1. ed. [s.l.] : Psychology Press, 1995.

CHEN, L. et al. Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market? **IEEE Access**, [s. l.], v. 6, p. 48625–48633, 2018.

CLEAR. **Blog Clear.** 2020. Disponível em: <<https://blog.clear.com.br/>>. Acesso em: 3 fev. 2020.

COLOMBO, J. A.; BAMPI, R. E.; CAMARGO, M. E. Redução de Risco na Formação de Carteiras: um Estudo da Correlação das Ações do IBOVESPA. **INGEPRO – Inovação, Gestão e Produção**, [s. l.], v. 2, n. 6, 2010.

CUTLER, D. M.; POTERBA, J. M.; SUMMERS, L. H. What moves stock prices? **Journal of Portfolio Management**, [s. l.], v. 15, n. 3, p. 4–12, 1989.

DACACHE, A. E.; LETOUZÉ, I. D. **Análise do impacto da divulgação de notícias na liquidez de ações do mercado brasileiro**. 2014. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

DATA SCIENCE ACADEMY. **Deep Learning Book**. 2019. Disponível em: <<http://www.deeplearningbook.com.br>>. Acesso em: 23 dez. 2019.

EHLERS, R. S. **Análise de séries temporais**. 4. ed. Paraná.

FAUSTINO, C. C. R. **O impacto das notícias no mercado financeiro brasileiro**. 2013. Fundação Getúlio Vargas, São Paulo, 2013.

FISCHER, T.; KRAUSS, C. Deep learning with long short-term memory networks for financial market predictions. **European Journal of Operational Research**, [s. l.], v. 270, n. 2, p. 654–669, 2018. Disponível em: <<https://doi.org/10.1016/j.ejor.2017.11.054>>

GREFF, K. et al. Lstm: a search space obyssey. **IEEE Transactions on Neural Networks and Learning Systems**, [s. l.], v. 28, n. 10, p. 2222–2232, 2017.

GUSMÃO, G. **As 20 maiores Bolsas de Valores do mundo**. 2018. Disponível em: <<https://www.bussoladoinvestidor.com.br/maiores-bolsas-de-valores-do-mundo/>>. Acesso em: 4 abr. 2020.

HAFEZI, R.; SHAHRABI, J.; HADAVANDI, E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. **Applied Soft Computing Journal**, [s. l.], v. 29, p. 196–210, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.asoc.2014.12.028>>

HAYKIN, S. **Redes Neurais: princípios e prática**. 2. ed. São Paulo: Bookman, 2007.

HOCHREITER, S.; SCHMIDHUBER, J. No Title. **Neural computation**, [s. l.], v. 9, n. 8, p. 1735–1780, 1997.

HUANG, W.; NAKAMORI, Y.; WANG, S.-Y. Forecasting stock market movement direction with support vector machine. **Computers & Operations**, [s. l.], v. 32, p. 2513–2522, 2005.

ITO, Y. et al. Extraction of damaged regions using sar data and neural networks. In: INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING 2000, Amsterdam. **Anais...** Amsterdam

KARA, Y.; BOYACIOGLU, M. A.; BAYKAN, Ö. 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**, [s. l.], v. 38, n. 5, p. 5311–5319, 2011.

KERAS. **Keras: The Python Deep Learning library**. 2020. Disponível em: <<https://keras.io/>>. Acesso em: 9 fev. 2020.

KIM, S.; KIM, H. A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, [s. l.], v. 32, n. 3, p. 669–679, 2016.

LEUNG, M. T.; DAOUKB, H.; CHEN, A.-S. Forecasting stock indices: a comparison of classification and level estimation models. **International Journal of Forecasting**, [s. l.], v. 16, n. 2, p. 173–190, 2000.

LO, A. W.; HASANHODZIC, J. **The Evolution of Technical Analysis: financial prediction from babylonian tablets to bloomberg terminals**. 1. ed. Berlim: Bloomberg Press, 2010.

LOCATELLI, A.; SILVA, C. E. L. e; ALVIM, A. M. O impacto dos ciclos políticos nos retornos e na volatilidade do Ibovespa. **Revista Estudo & Debate**, [s. l.], v. 26, n. 4, 2019.

MEDSKER, L.; JAIN, L. C. **Recurrent neural networks: design and applications**. New York: CRC Press LLC, 2001.

NUMPY. **NumPy**. 2020. Disponível em: <<https://numpy.org/>>. Acesso em: 9 fev. 2020.

PAI, P.-F.; LIN, C.-S. A hybrid ARIMA and support vector machines model in stock price forecasting. **Omega**, [s. l.], v. 33, n. 6, p. 497–505, 2005. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0305048304001082>>

PATEL, J. et al. Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques. **Expert Systems with Applications**, [s. l.], v. 42, n. 1, p. 259–268, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.eswa.2014.07.040>>

POPPER, K. A. **A lógica da pesquisa científica**. 2. ed. São Paulo: Cultrix, 2013.

PRODANOV, C. C.; FREITAS, E. C. De. **Metodologia do Trabalho Científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. 2. ed. Novo Hamburgo: Universidade FEEVALE, 2013.

QIU, M.; SONG, Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. **PLoS ONE**, [s. l.], v. 11, n. 5, p. 1–11, 2016.

REZENDE, S. O. **Sistemas Inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005.

RIBEIRO, S. R. A. **Integração de Imagens Landsat TM e dados auxiliares para a delimitação automática de unidades fisionômicas (paisagem) e geológicas usando Redes Neurais Artificiais**. 2003. UFPR, Curitiba, 2003.

RICO. **Qual o Seu Perfil de Investidor - Conservador, Moderado ou Arrojado?** 2019. Disponível em: <https://blog.rico.com/vc/perfil-de-investidor?campaignid=316171546&adgroupid=55392294370&feeditemid=&targetid=dsa-19959388920&loc_interest_ms=&loc_physical_ms=9047743&matchtype=b&network=g&device=c&devicemodel=&ifmobile=&ifmobile=0&ifsearch=1&ifsearch=&if>.

RICO. **Blog Rico**. 2020. Disponível em: <<https://blog.rico.com.vc/>>. Acesso em: 4 fev. 2020.

ROQUE, R. do C. **Estudo sobre a empregabilidade da previsão do índice Bovespa usando redes neurais artificiais**. 2009. Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

RUCK, D. W. et al. The multilayer perceptron as an approximation to a Bayes optimal discriminant function. **IEEE Transactions on Neural Networks**, [s. l.], v. 1, n. 4, p. 296–298, 1990.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. [s.l.] : Prentice Hall, 2009.

SCIKIT-LEARN. **scikit-learn Machine Learning in Python**. 2020. Disponível em: <<https://scikit-learn.org/stable/>>. Acesso em: 9 fev. 2020.

SHAH, H. et al. A Quick Gbest Guided Artificial Bee Colony algorithm for stock market prices prediction. **Symmetry**, [s. l.], v. 10, n. 7, 2018.

SHARPE, W. F.; ALEXANDER, G. J.; BAILEY, J. V. **Investments**. 5. ed. New Jersey: Prentice Hall, 1995.

SILVA, A. M. L. **Finanças comportamentais e meios de comunicação: uma análise do reflexo das notícias nacionais na cotação das ações da petrobrás em 2014**. 2015. Universidade Federal do Ceará, Fortaleza, 2015.

SILVA, C. A. T.; CARVALHO, C. C. De; NUNES, D. M. S. O Que Move o Preço da Ação? Uma Abordagem sobre a Influência das Notícias no Mercado Acionário. **REUNIR – Revista de Administração, Contabilidade e Sustentabilidade**, [s. l.], v. 2, n. 3, p. 1–13, 2012.

SILVA, J. R. S. Da; ALMEIDA, C. D. De; GUINDANI, J. F. Pesquisa documental: pistas teóricas e metodológicas. **Revista brasileira de história & ciências sociais**, [s. l.], v. 1, n. 1, 2009.

SRIVASTAVA, N. et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. **Journal of Machine Learning Research**, [s. l.], v. 15, p. 30, 2014.

TAFNER, M. A.; FILHO, E. R.; XEREZ, M. De. **Redes Neurais Artificiais: introdução e princípios de neurocomputação**. Blumenau: Editora da FURB, 1995.

TALEB, N. N. **A lógica do Cisne Negro: o impacto do altamente improvável**. 9. ed. Rio de Janeiro: Best Seller, 2015.

WANG, J. et al. Financial Time Series Prediction Using Elman Recurrent Random Neural Networks. **Hindawi Publishing Corporation Computational Intelligence and Neuroscience**, [s. l.], v. 2016, p. 1–14, 2016. Disponível em: <<http://dx.doi.org/10.1155/2016/4742515>>

WANG, Y.-H. Nonlinear neural network forecasting model for stock index option price: Hybrid GJR–GARCH approach. **Expert Systems with Applications**, [s. l.], v. 36, n. 1, p. 564–570, 2009.

XP INVESTIMENTOS. **Blog XP**. 2020. Disponível em: <<https://blog.xpi.com.br/>>. Acesso em: 31 jan. 2020.

YAHOO. **Yahoo Finanças**. 2019. Disponível em: <<https://br.financas.yahoo.com/quote/%5EBvsp/>>. Acesso em: 26 dez. 2019.

YANG, H.; PAN, Z.; TAO, Q. Robust and adaptive online time series prediction with long short-term memory. **Computational Intelligence and Neuroscience**, [s. l.], v. 2017, 2017.

ANEXOS

Anexo 1 – Composição da Carteira Teórica IBOVESPA de Janeiro a Maio de 2019

POSICAO	COD.	PART. %
1	ITUB4	10,80
2	VALE3	10,77
3	BBDC4	8,57
4	PETR4	7,21
5	PETR3	5,02
6	ABEV3	4,84
7	BBAS3	4,37
8	ITSA4	3,85
9	B3SA3	3,85
10	LREN3	1,97
11	UGPA3	1,93
12	BBDC3	1,60
13	SUZB3	1,46
14	JBSS3	1,35
15	VIVT4	1,33
16	RAIL3	1,29
17	BBSE3	1,28
18	SANB11	1,18
19	BRFS3	1,17
20	EQTL3	1,04
21	GGBR4	1,03
22	EMBR3	1,03
23	RENT3	1,02
24	KROT3	0,93
25	WEGE3	0,93
26	CCRO3	0,92
27	CMIG4	0,90
28	LAME4	0,89
29	PCAR4	0,87
30	BRKM5	0,86
31	HYPE3	0,84
32	RADL3	0,82
33	MGLU3	0,81
34	SBSP3	0,78
35	CIEL3	0,76
36	BRML3	0,76
37	KLBN11	0,68
38	TIMP3	0,65
39	EGIE3	0,61
40	BRDT3	0,60
41	CVCB3	0,58
42	ELET3	0,55
43	NATU3	0,55
44	ELET6	0,52
45	ESTC3	0,51
46	BRAP4	0,48

47	BTOW3	0,47
48	MULT3	0,44
49	FLRY3	0,44
50	CSNA3	0,41
51	CSAN3	0,38
52	TAE11	0,37
53	USIM5	0,35
54	ENBR3	0,30
55	GOAU4	0,28
56	CYRE3	0,28
57	MRVE3	0,24
58	IGTA3	0,24
59	GOLL4	0,22
60	QUAL3	0,22
61	SMLS3	0,17
62	MRFG3	0,15
63	ECOR3	0,13
64	VVAR3	0,12
65	LOGG3	0,02
QUANTIDADE TEORICA TOTAL		100,00

Anexo 2 – Mapeamento de Notícias e Movimentos de Alta e Queda nos Valores de Fechamento do IBOVESPA

Dia	Valores reais	Movimento	Notícia	Fonte
02/01/2019	87887	Não se aplica	Dólar cai abaixo de R\$ 3,85 e Ibovespa se divide entre pessimismo com China e expectativas com Bolsonaro	https://www.infomoney.com.br/mercados/dolar-cai-abaixo-de-r-385-e-ibovespa-se-divide-entre-pessimismo-com-china-e-expectativas-com-bolsonaro/
03/01/2019	91011	Alta	BR Distribuidora entra no Ibovespa, enquanto Copel e Fibria deixam índice, aponta prévia	https://www.infomoney.com.br/mercados/br-distribuidora-entra-no-ibovespa-enquanto-copel-e-fibria-deixam-indice-aponta-previa/
15/01/2019	94474	Alta	Ibovespa cai abaixo de 94 mil pontos em dia de ajuste após recordes consecutivos	https://www.infomoney.com.br/mercados/ibovespa-cai-abaixo-de-94-mil-pontos-em-dia-de-ajuste-apos-records-consecutivos/
16/01/2019	94057	Queda	Países europeus aprovam barreiras contra aço brasileiro	https://www.infomoney.com.br/mercados/paises-europeus-aprovam-barreiras-contr-a-co-brasileiro/
23/01/2019	95116	Queda	Maduro rompe com EAU, pede lealdade a militares e diz que não sai do poder	https://www.bol.uol.com.br/noticias/2019/01/23/maduro-rompe-com-os-eua-e-diz-que-nao-deixara-presidencia-da-venezuela.htm
28/01/2019	97674	Alta	Ibovespa Futuro cai mais de 1% após tragédia em Brumadinho; ações da Vale despencam em NY	https://www.infomoney.com.br/mercados/ibovespa-futuro-cai-mais-de-1-apos-tragedia-em-brumadinho-acoes-da-vale-despencam-em-ny/
31/01/2019	96996	Alta	Embraer e Vale caem quase 11% e Eletrobras dispara 54% em janeiro; confira as maiores altas e baixas do mês	https://www.infomoney.com.br/mercados/embraer-e-vale-caem-quase-11-e-eletrobras-dispara-54-em-janeiro-confira-as-maiores-altas-e-baixas-do-mes/
04/02/2019	97861	Alta	Projeção para taxa Selic cai de 7% para 6,5% em 2019	https://www.infomoney.com.br/mercados/projecao-para-taxa-selic-cai-de-7-para-65-em-2019/

08/02/2019	94401	Queda	Inflação acelera para 0,32% em janeiro e fica abaixo do esperado pelo mercado	https://www.infomoney.com.br/mercados/inflacao-acelera-para-032-em-janeiro-e-fica-abaixo-do-esperado-pelo-mercado/
18/02/2019	97527	Queda	Em maior crise desde a posse, Bolsonaro demite ministro	https://www.bol.uol.com.br/noticias/2019/02/18/ministro-gustavo-bebianno-demissao.htm
19/02/2019	96513	Queda	Trump pede às Forças Armadas da Venezuela que deixem de apoiar Maduro	http://agenciabrasil.ebc.com.br/internacional/noticia/2019-02/trump-pede-forcas-armadas-da-venezuela-que-deixe-de-apoiar-maduro
20/02/2019	97661	Alta	Bolsonaro entrega proposta de reforma da Previdência ao Congresso	https://www.bol.uol.com.br/noticias/2019/02/20/reforma-da-previdencia-apresentacao-proposta.htm
21/02/2019	96546	Queda	Maduro manda fechar a fronteira da Venezuela com o Brasil	https://www.bol.uol.com.br/noticias/2019/02/21/maduro-manda-fechar-a-fronteira-da-venezuela-com-o-brasil.htm
22/02/2019	96929	Alta	Proposta de reforma da Previdência chega à CCJ da Câmara	http://agenciabrasil.ebc.com.br/politica/noticia/2019-02/proposta-de-reforma-da-previdencia-chega-ccj-da-camara
27/02/2019	97603	Alta	Após quatro anos de prejuízo, Petrobras volta a dar lucro	http://agenciabrasil.ebc.com.br/economia/noticia/2019-02/apos-quatro-anos-petrobras-volta-dar-lucro
28/02/2019	97307	Queda	PIB fecha 2018 com crescimento 1,1%, mostra IBGE	http://agenciabrasil.ebc.com.br/economia/noticia/2019-02/pib-fecha-2018-com-crescimento-de-11-mostra-ibge
28/02/2019	97307	Queda	Via Varejo despenca 23% e CSN dispara 28%: confirma as maiores altas e baixas de fevereiro	https://www.infomoney.com.br/mercados/via-varejo-despenca-23-e-csn-dispara-28-confirma-as-maiores-altas-e-baixas-de-fevereiro/
01/03/2019	95584	Queda	Os 2 fatores que fizeram o Ibovespa abrir março com queda de 1% e o dólar subir 0,7%	https://www.infomoney.com.br/mercados/os-2-fatores-que-fizeram-o-ibovespa-abrir-marco-com-queda-de-1-e-o-dolar-subir-07/

06/03/2019	94604	Queda	Vale trocou auditor que se recusou a certificar barragem, BRF perde mais de R\$ 200 mi com derivativos e mais destaques	https://www.infomoney.com.br/mercados/vale-trocou-auditor-que-se-recusou-a-certificar-barragem-brf-perde-mais-de-r-200-mi-com-derivativos-e-mais-destaques/
11/03/2019	95384	Alta	Selic no fim de 2019 permanece em 6,50% ao ano, aponta Boletim Focus	https://www.infomoney.com.br/mercados/selic-no-fim-de-2019-permanece-em-650-ao-ano-aponta-boletim-focus/
12/03/2019	98038	Alta	Banco Central cortará Selic até o meio do ano, diz ex-diretor da autoridade monetária	https://www.infomoney.com.br/mercados/banco-central-cortara-selic-ate-o-meio-do-ano-diz-ex-diretor-da-autoridade-monetaria/
14/03/2019	98905	Alta	STF decide que casos da Lava Jato podem ser julgados na Justiça Eleitoral	https://www.bol.uol.com.br/noticias/2019/03/14/stf-justica-eleitoral-lava-jato.htm
15/03/2019	98605	Queda	Estácio e Marisa caem 3% após balanço do 4º tri; JBS e BRF disparam com “efeito China” e mais destaques	https://www.infomoney.com.br/mercados/estacio-e-marisa-caem-3-apos-balanco-do-4o-tri-jbs-e-brf-disparam-com-efeito-china-e-mais-destaques/
25/03/2019	93735	Queda	Ex-presidente Michel Temer é solto no Rio de Janeiro após quatro dias de prisão	https://www.bol.uol.com.br/noticias/2019/03/25/apos-decisao-do-trf-2-michel-temer-deixa-a-prisao.htm
26/03/2019	93668	Queda	MP critica Bolsonaro por incentivar celebração do golpe: "enorme gravidade"	https://www.bol.uol.com.br/noticias/2019/03/26/mp-critica-bolsonaro-por-incentivar-celebracao-do-golpe-enorme-gravidade.htm
28/03/2019	91901	Queda	Três barragens da Vale em MG estão em nível máximo de alerta	http://agenciabrasil.ebc.com.br/geral/noticia/2019-03/tres-barragens-da-vale-em-mg-estao-em-nivel-maximo-de-alerta
29/03/2019	94402	Alta	Ibovespa cai em março, mas fecha trimestre com alta de 8,5%; dólar salta 4% no mês	https://www.infomoney.com.br/mercados/ibovespa-cai-em-marco-mas-fecha-trimestre-com-alta-de-85-dolar-salta-4-no-mes/

29/03/2019	94402	Alta	Os primeiros meses do ano foram de fortes emoções na bolsa, tendo como destaque positivo a CSN, enquanto Embraer teve a maior queda do período	https://www.infomoney.com.br/mercados/as-maiores-altas-e-baixas-do-ibovespa-no-1o-trimestre-de-2019/
10/04/2019	96292	Queda	Inflação oficial sobe para 0,75% em março, acima do esperado pelo mercado	https://www.infomoney.com.br/mercados/inflacao-oficial-sobe-para-075-em-marco-acima-do-esperado-pelo-mercado/
12/04/2019	94747	Queda	Ação da Petrobras cai mais de 8% após Bolsonaro barrar alta do diesel	https://economia.uol.com.br/cotacoes/noticias/redacao/2019/04/12/petrobras-acoes-desepencam-preco-diesel-bolsonaro.htm
15/04/2019	92875	Queda	IBC-Br cai 0,73% em fevereiro ante janeiro, com ajuste, revela BC	https://www.infomoney.com.br/mercados/ibc-br-cai-073-em-fevereiro-ante-janeiro-com-ajuste-revela-bc/
25/04/2019	95027	Queda	Presidente Jair Bolsonaro Censura campanha do Banco do Brasil focada na diversidade	https://www.b9.com.br/107113/presidente-jair-bolsonaro-censura-campanha-do-banco-do-brasil-focada-na-diversidade/
25/04/2019	95027	Queda	IPCA-15 acelera e fica em 0,72% em abril, acima do esperado pelo mercado	https://www.infomoney.com.br/mercados/ipca-15-acelera-e-fica-em-072-em-abril-acima-do-esperado-pelo-mercado/
29/04/2019	96254	Queda	Bolsonaro apela para que Banco do Brasil corte juros; ações caem até 1,8%	https://www.infomoney.com.br/bancodobrasil/noticia/8264691/bolsonaro-apela-para-que-banco-do-brasil-corte-juros-acoes-viram-para-queda
06/05/2019	95992	Alta	Bolsonaro enfrenta grandes testes no Congresso em maio, com MPs sob risco e Previdência em discussão	https://www.infomoney.com.br/politica/bolsonaro-enfrenta-grandes-testes-no-congresso-em-maio-com-mps-sob-risco-e-previdencia-em-discussao/
10/05/2019	94809	Queda	IPCA sobe 0,57% em abril, ante 0,75% em março; afirma IBGE	https://www.infomoney.com.br/mercados/ipca-sobe-057-em-abril-ante-075-em-marco-afirma-ibge/
13/05/2019	94252	Queda	Bitcoin dispara para US\$ 8.000 e já acumula alta de mais de 100% em 2019	https://www.infomoney.com.br/mercados/bitcoin-dispara-para-us-8-000-e-ja-acumula-alta-de-mais-de-100-em-2019/

20/05/2019	90006	Queda	Inflação: IGP-M desacelera a 0,58% na 2ª prévia de maio, aponta FGV	https://www.infomoney.com.br/mercados/inflacao-igp-m-desacelera-a-058-na-2a-previa-de-maio-aponta-fgv/
31/05/2019	97420	Alta	Suzano tinha tudo para “brilhar”, mas foi a pior ação do Ibovespa em maio	https://www.infomoney.com.br/negocios/suzano-tinha-tudo-para-brilhar-mas-foi-a-pior-acao-do-ibovespa-em-maio-o-que-aconteceu/
31/05/2019	97420	Alta	Conta de luz ficará mais barata em junho, diz Aneel	https://www.infomoney.com.br/consumo/conta-de-luz-ficara-mais-barata-em-junho-diz-aneel/
31/05/2019	97420	Alta	Ibovespa cai, mas não apaga alta no mês e tem melhor maio dos últimos 10 anos	https://www.infomoney.com.br/mercados/ibovespa-cai-mas-nao-apaga-alta-no-mes-e-tem-melhor-maio-dos-ultimos-10-anos/
13/06/2019	98321	Queda	Bancos caem com proposta de aumento da CSLL	https://www.infomoney.com.br/mercados/bancos-caem-com-proposta-de-aumento-da-csll-brf-sobe-quase-6-e-jbs-cai-com-recomendacoes/
14/06/2019	98776	Alta	IBC-Br cai 0,47% em abril ante março, com ajuste, afirma Banco Central	https://www.infomoney.com.br/mercados/ibc-br-cai-047-em-abril-ante-marco-com-ajuste-afirma-banco-central/
18/06/2019	97633	Queda	Ação da B3 salta 5% com possível reviravolta no “caso CSLL”; Vale e siderúrgicas avançam com commodities	https://www.infomoney.com.br/mercados/acao-da-b3-salta-5-com-possivel-reviravolta-no-caso-csll-vale-e-siderurgicas-avancam-com-commodities/
26/06/2019	100095	Queda	Vale concluiu 49 acordos individuais de indenização em Brumadinho	http://agenciabrasil.ebc.com.br/geral/noticia/2019-06/vale-concluiu-49-acordos-individuais-de-indenizacao-em-brumadinho
27/06/2019	100689	Alta	Mercosul e União Europeia fecham acordo de livre comércio	http://agenciabrasil.ebc.com.br/internacional/noticia/2019-06/mercosul-e-uniao-europeia-fecham-acordo-de-livre-comercio