



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Tecnologia

Arthur Emanuel de Oliveira Carosia

**Previsão do Mercado de Ações Brasileiro com o uso de
Análise de Sentimentos, Indicadores Técnicos e Valores de
Ações**

Limeira
2022

Arthur Emanuel de Oliveira Carosia

Previsão do Mercado de Ações Brasileiro com o uso de Análise de Sentimentos, Indicadores Técnicos e Valores de Ações

Tese apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Doutor em Tecnologia, na área de Sistemas de Informação e Comunicação.

Orientador: Prof. Dr. Guilherme Palermo Coelho
Coorientadora: Prof. Dra. Ana Estela Antunes da Silva

Este trabalho corresponde à versão final da Tese defendida por Arthur Emanuel de Oliveira Carosia e orientada pelo Prof. Dr. Guilherme Palermo Coelho.

Limeira
2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Felipe de Souza Bueno - CRB 8/8577

C222p Carosia, Arthur Emanuel de Oliveira, 1987-
Previsão do mercado de ações brasileiro com o uso de análise de sentimentos, indicadores técnicos e valores de ações / Arthur Emanuel de Oliveira Carosia. – Limeira, SP : [s.n.], 2022.

Orientador: Guilherme Palermo Coelho.
Coorientador: Ana Estela Antunes da Silva.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Mercado de ações. 2. Aprendizado profundo. 3. Redes neurais (Computação). 4. Análise de sentimentos. I. Coelho, Guilherme Palermo, 1980-. II. Silva, Ana Estela Antunes da, 1965-. III. Universidade Estadual de Campinas. Faculdade de Tecnologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Predicting the Brazilian stock market using sentiment analysis, technical indicators, and stock prices

Palavras-chave em inglês:

Stock market

Deep learning

Neural networks (Computer science)

Sentiment analysis

Área de concentração: Sistemas de Informação e Comunicação

Titulação: Doutor em Tecnologia

Banca examinadora:

Guilherme Palermo Coelho [Orientador]

Thiago Alexandre Salgueiro Pardo

Fabricio Aparecido Breve

Cristiano Torrezan

Ulisses Martins Dias

Data de defesa: 20-06-2022

Programa de Pós-Graduação: Tecnologia

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-6277-0014>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6360630975051205>

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de tese para o Título de Doutor em Tecnologia na área de concentração Sistemas de Informação e Comunicação, a que se submeteu o aluno Arthur Emanuel de Oliveira Carosia, em 20 de junho de 2022 na Faculdade de Tecnologia – FT/UNICAMP, em Limeira/SP.

Prof. Dr. Guilherme Palermo Coelho

Presidente da Comissão Julgadora

Prof. Dr. Thiago Alexandre Salgueiro Pardo

Universidade de São Paulo - USP

Prof. Dr. Fabricio Aparecido Breve

Universidade Estadual Paulista - UNESP

Prof. Dr. Cristiano Torezzan

Universidade Estadual de Campinas - UNICAMP

Prof. Dr. Ulisses Martins Dias

Universidade Estadual de Campinas - UNICAMP

Ata da defesa, assinada pelos membros da Comissão Examinadora, encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós Graduação da Faculdade de Tecnologia.

Agradecimentos

O desenvolvimento desta tese de doutorado não seria possível sem o auxílio de pessoas muito importantes, às quais eu gostaria de agradecer.

Primeiramente, eu gostaria de agradecer à minha noiva Beatriz por todo seu amor, carinho e conselhos, principalmente nos momentos mais difíceis. Aos meus pais, Arthur e Ana, por todo seu apoio e sabedoria.

Eu gostaria de agradecer também a todo auxílio, apoio e conhecimento de meu orientador, Prof. Dr. Guilherme P. Coelho, e de minha coorientadora, Profa. Dra. Ana Estela A. da Silva.

Eu gostaria de agradecer ao Instituto Federal de São Paulo (IFSP), São João da Boa Vista, pelo apoio financeiro e cooperação com esta pesquisa. Além disso, gostaria de agradecer à Intel e à Amazon, pelo apoio tecnológico.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

O desenvolvimento da economia global permite a empresas abrirem o seu capital com o objetivo de obter recursos financeiros para acelerar o seu crescimento. Esse cenário atrai investidores que desejam lucratividade nos mercados financeiros. Assim, a literatura apresenta diversas técnicas voltadas para previsão e análise do mercado financeiro, em especial o mercado de ações. Nesse contexto, destacam-se duas abordagens distintas, denominadas análise técnica e fundamentalista. Enquanto a análise técnica procura prever o mercado tendo como base indicadores calculados considerando o preço passado das ações, a análise fundamentalista considera dados oriundos de notícias, rentabilidade e fatores macroeconômicos. No entanto, a junção de ambas as escolas de pensamento em um único sistema de previsão do mercado de ações ainda é um tópico em aberto na literatura. Por um lado, o uso de Redes Neurais Artificiais (RNAs), principalmente RNAs profundas, tem se mostrado uma abordagem interessante para previsão de valores futuros do mercado de ações, sendo empregadas em muitos problemas voltados à análise técnica. Essas redes têm apresentado melhores resultados de previsão quando comparadas aos modelos estatísticos tradicionais, que utilizam normalmente dados vindos de séries temporais de valores de ações e indicadores técnicos. Por outro lado, mais recentemente, surgiram trabalhos na literatura indicando o uso de Análise de Sentimentos (AS) aplicada a notícias financeiras como um importante indicativo voltado à análise fundamentalista para a previsão da oscilação do mercado acionário. Apesar dos esforços recentes para combinar o uso dessas fontes de dados, a literatura carece de trabalhos em que essas estratégias sejam executadas em uma abordagem totalmente baseada em RNAs profundas, que têm obtido resultados do estado da arte em muitas tarefas de regressão e classificação. Portanto, no sentido de apoiar o processo de tomada de decisão e previsão frente ao grande número de informações disponíveis, este trabalho tem como objetivo principal apresentar uma abordagem de previsão do mercado de ações brasileiro por meio de Aprendizado Profundo, integrando dados voltados tanto à análise técnica quanto à análise fundamentalista, a saber: séries temporais de valores de ações, indicadores técnicos, e de Análise de Sentimentos aplicada a notícias. Os experimentos foram realizados com dados do período de 2010 a 2019, tanto do índice Ibovespa quanto de empresas relevantes do mercado de ações brasileiro: Banco do Brasil, Itaú, Ambev e Gerdau. Os resultados mostram que a combinação de preços de ações, indicadores técnicos e notícias melhora a previsão do mercado de ações considerando tanto o erro de previsão quanto o valor final do investimento.

Palavras-chave: Análise de Sentimentos, Redes Neurais Artificiais, Aprendizado Profundo, Mercado de Ações

Abstract

The global economy has led companies to open their capital to obtain resources to accelerate their growth. This scenario attracts investors who want to profit, stimulating the science community to develop techniques to forecast and analyze the stock market. Thus, two distinct approaches are highlighted, namely technical and fundamental analyses. While technical analysis tries to predict the stock market based on indicators calculated considering the past stock prices, fundamental analysis considers data from news, profitability, and macroeconomic factors. However, the combination of both schools of thought into a single stock market prediction system is still an open challenge in the literature. On the one hand, the use of Artificial Neural Networks (ANNs), mainly deep ANNs, has been shown to be an interesting approach to predicting future stock market values, being employed mainly in technical analysis. These ANNs have shown better prediction results when compared to traditional statistical models, which generally use stock prices and technical indicators as a data source. On the other hand, more recently, studies have indicated that Sentiment Analysis (SA) applied to financial news as an important fundamental analysis indicator to predict the stock market movement. Despite recent efforts to combine the use of these data sources, the literature lacks works in which these strategies are performed in an approach entirely based on deep ANNs, which have obtained state-of-the-art results in many regression and classification tasks. Therefore, in order to support the decision-making and forecasting process considering a large amount of available information, this work aims to present an approach to predict the Brazilian stock market through Deep Learning, integrating data of both technical and fundamental analysis, namely: stock prices, technical indicators, and financial news. The experiments were performed with data from 2010 to 2019, considering the Ibovespa index and the following relevant companies of the Brazilian stock market: Banco do Brasil, Itaú, Ambev, and Gerdau. The results show that the combination of stock prices, technical indicators, and news improves stock market prediction considering both the prediction error and the return on investment.

Keywords: Sentiment Analysis, Artificial Neural Networks, Deep Learning, Stock Market

Lista de Figuras

2.1	Representação de um <i>candlestick</i> de período de baixa (em vermelho) e um de período de alta (em verde). Fonte: próprio autor.	24
2.2	Representação de gráfico de <i>candlestick</i> para o Ibovespa. Período de março a agosto de 2017 (FARIA, 2018).	25
2.3	Representação de indicadores técnicos.	26
3.1	Representação de um conjunto de vetores compondo um <i>Bag of Words</i>	32
3.2	Representação geral funcionamento do algoritmo <i>Word2Vec</i> , modelos <i>CBOW</i> e <i>Skip-gram</i> . Adaptado de (MIKOLOV; CHEN, K. et al., 2013).	34
3.3	Representação do modelo utilizado no algoritmo <i>Word2Vec</i> , modelo <i>Skip-gram</i> . Adaptado de (RONG, 2014).	35
3.4	Representação de uma árvore binária para o modelo softmax hierárquico. (RONG, 2014).	37
3.5	Representação da técnica de amostragem negativa.	38
4.1	Neurônio artificial do tipo Perceptron. Representação matemática. Adaptado de (HAYKIN, 2009).	43
4.2	Gráfico ilustrativo da função tangente hiperbólica. Adaptado de (HAYKIN, 2009).	43
4.3	Gráfico ilustrativo da função logística. Adaptado de (HAYKIN, 2009).	44
4.4	Gráfico ilustrativo da função ReLU. Adaptado de (HAHNLOSER et al., 2000).	44
4.5	Representação de uma RNA do tipo MLP. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2019).	46
4.6	Representação do algoritmo Backpropagation.	47
4.7	Representação de uma RNN de Jordan.	49
4.8	Representação de uma RNN de Elman.	50
4.9	Representação geral de uma RNN. Adaptado de (LIPTON; BERKOWITZ; ELKAN, 2015).	51
4.10	Representação de um neurônio de uma RNN.	52
4.11	Representação de uma célula de memória de uma LSTM. Adaptado de (ZHANG, L.; WANG, S.; LIU, B., 2018).	53
4.12	Representação das camadas de um modelo Bi-LSTM. Adaptado de (ALTHELAYA; EL-ALFY; MOHAMMED, 2018).	56
4.13	Representação de uma rede do tipo CNN. Adaptado de (FARIA, 2018).	57
4.14	Operação de convolução.	58
4.15	Operação de <i>pooling</i> . Aplicação das operações de max e média. Adaptado de (SZE et al., 2017).	59
6.1	Etapas desenvolvidas durante este trabalho.	77

6.2	Metodologia proposta para investigar o relacionamento entre <i>tweets</i> e o mercado de ações brasileiro. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2020).	78
6.3	Metodologia proposta para comparar a influência tanto de postagens da rede social Twitter como de notícias no mercado de ações brasileiro. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2019).	79
6.4	Metodologia proposta para a investigação de estratégias de investimento baseadas em Análise de Sentimento. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2021).	80
6.5	Representação da metodologia proposta.	83
6.6	Representação da Metodologia de Análise de Sentimentos.	85
6.7	Representação dos componentes utilizados para alimentar a RNA.	86
6.8	Representação do Módulo de Predição, que recebe o vetor gerado por meio da Junção de Dados e investiga a arquitetura de RNA mais adequada para a previsão do mercado de ações.	87
7.1	Algoritmo <i>increasing window cross-validation</i> considerando a repartição da série temporal em 3 divisões. Adaptado de (LI, X.; WU, P.; WANG, W., 2020). .	97

Lista de Tabelas

7.1	Índice e ações usados nos experimentos. São apresentados: os valores iniciais e finais de cada ação juntamente com seu retorno máximo possível no período (Ganho Máx.) e número de movimentos de subida (Alta) e descida (Baixa) no período entre janeiro de 2010 a dezembro de 2019.	91
7.2	Resultados da tarefa de AS. Conjunto de dados de testes.	94
7.3	Métricas de Aprendizado de Máquina considerando o baseline SVM aplicado a todas ações.	99
7.4	Métricas de Aprendizado de Máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Ibovespa (BVSP).	100
7.5	Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Banco do Brasil (BBAS3).	100
7.6	Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Itau (ITUB4).	100
7.7	Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Gerdau (GGBR4).	101
7.8	Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Ambev (ABEV3).	101
7.9	Teste estatístico de Mann-Whitney para a métrica MSE considerando a tolerância de 5%, realizado entre a metodologia proposta nesta tese e as demais abordagens, tanto com RNAs profundas (identificado por DL) quanto com o modelo SVM. O marcador X representa que há significância estatística, enquanto que o marcador – representa ausência de significância estatística.	103
7.10	Simulação de investimento considerando a estratégia Buy & Hold.	103
7.11	Simulação de investimento considerando o baseline SVM.	104
7.12	Simulação de investimento usando os modelos de RNAs profundas. Ibovespa (BVSP).	105
7.13	Simulação de investimento usando os modelos de RNAs profundas. Banco do Brasil (BBAS3).	105
7.14	Simulação de investimento usando os modelos de RNAs profundas. Itau (ITUB4).	105
7.15	Simulação de investimento usando os modelos de RNAs profundas. Gerdau (GGBR4).	106
7.16	Simulação de investimento usando os modelos de RNAs profundas. Ambev (ABEV3).	106
7.17	Resultado dos trabalhos correlatos (HÁJEK, 2018) e (PICASSO et al., 2019), utilizando uma RNA do tipo MLP e método léxico.	108
7.18	Resultados do trabalho correlato (LI, X.; WU, P.; WANG, W., 2020), considerando uma RNA do tipo LSTM e método léxico.	109

7.19 Resultados do trabalho correlato (DUARTE; GONZÁLEZ; CRUZ, 2021)
considerando o modelo MLP.

..... 109

Sumário

1	Introdução	14
2	Mercado de Ações	20
2.1	Teorias de Mercado	21
2.2	Análises de Mercado	22
2.3	Considerações Finais	28
3	Análise de Sentimentos	29
3.1	Representação de Palavras	31
3.1.1	Word2Vec	33
3.1.2	FastText	38
3.2	Considerações Finais	40
4	Redes Neurais Artificiais	41
4.1	Aspectos Gerais	42
4.2	Arquiteturas	45
4.2.1	Multilayer Perceptron	45
4.2.2	Redes Neurais Recorrentes	48
4.2.3	Long Short-Term Memory	52
4.2.4	Convolutional Neural Networks	56
4.3	Considerações Finais	60
5	Trabalhos Correlatos	61
5.1	Dados textuais para Previsão do Mercado Financeiro	61
5.2	Valores de Ações e Indicadores Técnicos para Previsão do Mercado Financeiro	67
5.3	Valores de Ações, Indicadores Técnicos e Dados Textuais para Previsão do Mercado Financeiro	71
5.4	Considerações Finais	74
6	Metodologia	76
6.1	Desenvolvimento Geral	76
6.2	Metodologia de Previsão do Mercado de Ações	81
6.2.1	Módulo de Pré-processamento	82
6.2.2	Módulo de Análise de Sentimentos	84
6.2.3	Módulo de Predição	85
6.2.4	Módulo de Avaliação	87

7	Resultados	90
7.1	Fonte de Dados	90
7.2	Análise de Sentimentos em Notícias Financeiras	92
7.3	Módulo de Predição	95
7.3.1	Metodologia Experimental	95
7.3.2	Métricas de Aprendizado de Máquina	98
7.3.3	Métricas de Investimento	102
7.4	Comparação Experimental com Trabalhos Correlatos	107
7.5	Discussão	110
8	Conclusões e Trabalhos Futuros	116
	Referências	118

Capítulo 1

Introdução

A previsão do valor de ações no mercado é um tema que tem atraído muita atenção nas últimas décadas, com muita discussão na comunidade científica sobre o assunto. O trabalho de Malkiel e Fama (1970) apresentou a Hipótese de Mercado Eficiente (HME) afirmando que, em um mercado eficiente, a informação disponível aos investidores já está contida no preço de mercado de um ativo, implicando que o movimento futuro do mercado acionário é imprevisível considerando os preços passados das ações. Por outro lado, críticos dessa hipótese afirmam que pode existir um certo grau previsibilidade no mercado, pelo menos parcialmente (MALKIEL, 2003). Além disso, muitos trabalhos na literatura apresentam resultados satisfatórios na previsão do mercado, como o descrito por Atsalakis e Valavanis (2009), que apresenta uma revisão de mais de 100 artigos publicados na área. De fato, conforme apresentado pela Hipótese de Mercado Adaptativo (HMA) (LO, 2004), que considera aspectos evolutivos e comportamentais para o estudo da atuação dos investidores no mercado, é possível lucrar com uma carteira de ações explorando falhas na eficiência do mercado.

De forma complementar, a Teoria Dow (RHEA, 1993) acrescenta mais informações que permitem melhor compreender o mercado de ações. Segundo essa teoria, os movimentos dos preços no mercado de ações ocorrem de acordo com tendências. Assim, a fim de compreender as tendências do mercado, os investidores desenvolveram duas estratégias principais: Análise Técnica e Análise Fundamentalista. Enquanto os analistas técnicos tentam prever os preços futuros no mercado de ações de acordo com os preços anteriores das ações e indicadores técnicos, os analistas fundamentalistas baseiam suas decisões de negociação nas condições financeiras, notícias e nos indicadores macroeconômicos da

empresa. Como mostrado por Atsalakis e Valavanis (2009), sabe-se que a Análise Técnica é a abordagem mais frequentemente empregada nos trabalhos publicados. A maioria dos investidores, no entanto, naturalmente observa dados de ambas análises.

A análise financeira atual se desenvolveu bastante devido à disponibilidade cada vez maior de recursos computacionais e ao avanço em ciência de dados (PICASSO et al., 2019). Inicialmente, esse desenvolvimento ocorreu por meio da aplicação de métodos estatísticos sobre dados passados com o objetivo de prever valores futuros do mercado, tais como Médias Móveis Autoregressivas (do inglês, *Autoregressive Moving Average* - ARMA) e Médias Móveis Integradas Autorregressivas (do inglês, *Autoregressive Integrated Moving Averages* - ARIMA). No entanto, tais modelos se baseiam na suposição de que as séries temporais financeiras são geradas por um processo linear, o que raramente acontece no mercado financeiro, que é naturalmente complexo (KUMBURE et al., 2022). Desse modo, posteriormente foram propostas técnicas computacionais baseadas em Inteligência Artificial, sendo de grande destaque as técnicas de Aprendizado de Máquina. Por exemplo, dentre essas técnicas, atualmente as Redes Neurais Artificiais são extensivamente utilizadas na literatura (KUMAR, D.; SARANGI; VERMA, 2021; KUMAR, G.; JAIN; SINGH, 2021). Mais recentemente, técnicas de Aprendizado Profundo têm ganhado bastante destaque na literatura financeira (THAKKAR; CHAUDHARI, 2021).

Por outro lado, os avanços nas tecnologias da *Web* também levaram à disponibilização de grandes quantidades de informações que poderiam ser exploradas, atraindo a atenção de investidores para plataformas como sites de notícias, mídia social e *blogs*. Muitos autores estudaram a influência desses dados textuais *online* no mercado de ações (BOLLEN; MAO; ZENG, 2011; LI, X.; WU, P.; WANG, W., 2020). Segundo Ozbayoglu, Gudelek e Sezer (2020) e Kumbure et al. (2022), a mineração de textos financeiros recentemente tem atraído muita atenção na literatura, principalmente com foco em obter o sentimento do investidor e correlacioná-lo com o mercado de ações. Além disso, dados da Análise Fundamentalista podem ser apresentados de forma textual, o que naturalmente aumenta a complexidade de seu tratamento computacional em relação aos dados da Análise Técnica (KHADJEH NASSIRTOUSSI et al., 2014).

Nesse sentido, os trabalhos existentes na literatura, em sua maioria, se concentram em um tipo de informação para prever o mercado de ações. Normalmente, se dividem da seguinte forma: (1) utilizando principalmente dados textuais, como notícias ou postagens em

mídias sociais (SCHUMAKER et al., 2012; KRAUS; FEUERRIEGEL, 2017); ou (2) utilizando principalmente preços históricos de ações ou indicadores técnicos (NELSON; PEREIRA, A. C.; OLIVEIRA, R. A. de, 2017; PANG, X. et al., 2018).

Apenas recentemente os trabalhos têm procurado integrar mais fontes de informação sobre o mercado. Por exemplo, os trabalhos de Picasso et al. (2019), Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020) e Shi et al. (2021) apresentam um sistema para prever o mercado de ações concatenando dados textuais, preços de ações e indicadores técnicos. No entanto, como afirmado por Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), ainda é um desafio de pesquisa em aberto a previsão do mercado de ações combinando valores históricos de ações e dados textuais, ou seja, combinando dados oriundos tanto da Análise Técnica quanto da Análise Fundamentalista. Dentre os trabalhos recentes que combinam ambas fontes de dados, foram identificadas as seguintes lacunas:

1. Trabalhos que utilizam dados textuais por meio de métodos lexicais (PICASSO et al., 2019; LI, X.; WU, P.; WANG, W., 2020). Nesse caso, as abordagens baseadas em métodos lexicais necessitam de conhecimento específico sobre a linguagem e o domínio de aplicação, além de apresentarem resultados inferiores quando comparados às técnicas de Aprendizado de Máquina (TABOADA et al., 2011; AVANÇO; NUNES, 2014; JANUÁRIO et al., 2021).
2. Trabalhos em que há o uso de informações de redes sociais para previsão do mercado (SHI et al., 2021). Em uma investigação anterior, publicada em Carosia, Coelho e A. E. A. Silva (2019), verificamos que notícias financeiras apresentam maior correlação com o mercado de ações do que postagens em redes sociais. Além disso, verificamos em Carosia, Coelho e A. E. A. Silva (2021) que o uso de notícias como parte de uma estratégia de investimento em ações pode levar a lucros no curto e longo prazo.
3. Trabalhos que usam notícias, séries históricas e indicadores técnicos (DUARTE; GONZÁLEZ; CRUZ, 2021), mas que não optam pelo uso de RNAs profundas, que atualmente têm se mostrado muito promissoras em tarefas de previsão aplicadas ao mercado de ações (KUMAR, D.; SARANGI; VERMA, 2021).
4. Trabalhos publicados com pouco foco no mercado de ações brasileiro (OZBAYOGLU; GUDELEK; SEZER, 2020), apesar de sua grande relevância na economia global. Recentemente, Ozbayoglu, Gudelek e Sezer (2020) revisaram 140 artigos, dentre os

quais apenas três utilizaram dados do mercado de ações brasileiro; o trabalho de Bustos e Pomares-Quimbaya (2020) revisou 53 artigos e apenas dois utilizaram dados brasileiros. Além disso, Kumbure et al. (2022) revisaram 138 trabalhos cobrindo os anos de 2000 a 2019 e mostraram que a maioria das publicações são voltadas ao mercado norte-americano e asiático. Apesar dos poucos estudos sobre o mercado brasileiro, a B³ (Brasil, Bolsa, Balcão), a bolsa de valores oficial do Brasil, é uma das maiores bolsas de valores do mundo em valor de mercado. Muitos fatores tornam a bolsa brasileira atrativa, incluindo a alta volatilidade dos preços negociados das ações. Ademais, segundo consta, existem poucos recursos computacionais voltados para Análise de Sentimentos em língua portuguesa (PEREIRA, D. A., 2021).

5. Nem todos os trabalhos publicados na literatura realizam uma simulação de investimento investigando a lucratividade e o risco do modelo proposto (KUMAR, D.; SARANGI; VERMA, 2021; KUMAR, G.; JAIN; SINGH, 2021; KUMBURE et al., 2022).

Assim, para resolver as questões apresentadas, propomos aqui uma abordagem de previsão do mercado de ações brasileiro, incorporando notícias financeiras, indicadores técnicos e os valores históricos das ações em um único sistema. O processo todo é realizado com Redes Neurais Artificiais (RNAs) profundas, tanto para processar o sentimento presente em notícias quanto para prever o mercado. Ainda, realizamos uma simulação de investimento para analisar a eficácia da metodologia proposta neste trabalho. Nossa abordagem foi testada em dados do mercado de ações brasileiro, coletados entre janeiro de 2010 e dezembro de 2019, por meio do índice Ibovespa e de empresas relevantes do índice: Banco do Brasil, Itaú, Ambev e Gerdau.

Como primeiro passo, é utilizada uma abordagem baseada em aprendizado profundo para realizar a Análise de Sentimentos de notícias brasileiras (portanto, escritas em português). Em seguida, o resultado dessa etapa alimenta outra RNA profunda, juntamente com indicadores técnicos e preços históricos de ações. Foram consideradas as seguintes RNAs profundas: *Multilayer Perceptron* (MLP), *Long Short-Term Memory* (LSTM) e *Bidirecional Long Short-Term Memory* (Bi-LSTM). Como *baseline* para os experimentos, foi utilizado o algoritmo de *Support Vector Machines* (SVM), comumente empregado em trabalhos da área. Essas diferentes arquiteturas de RNAs foram comparadas considerando-se que são as arquiteturas mais utilizadas na área financeira (OZBAYOGLU; GUDELEK; SEZER, 2020).

Para avaliar os de modelos de Aprendizado de Máquina no domínio financeiro, nossos experimentos foram divididos em duas etapas (PICASSO et al., 2019): (1) avaliação de

métricas relacionadas à área de Aprendizado de Máquina; (2) avaliação de uma simulação de investimento a partir do valor final do investimento e das medidas de risco *Maximum Drawdown* (MDD) e *Sharpe Ratio* (SR). Picasso et al. (2019) apontam que a primeira etapa comprova a eficácia do treinamento do modelo, enquanto que a segunda etapa comprova a qualidade das previsões do modelo em seu domínio de aplicação.

As principais contribuições do nosso trabalho são:

1. O uso de Aprendizado Profundo em todo o processo de previsão do mercado de ações (tanto aplicado a notícias quanto a dados históricos do mercado de ações).
2. O estudo de caso do mercado de ações brasileiro.
3. A proposta de integração de dados provenientes tanto da Análise Técnica quanto da Análise Fundamentalista em um único sistema de previsão de ações.

Nossos resultados mostram que o uso de notícias melhora a previsão do mercado de ações quando as mesmas são incorporadas a modelos de RNAs profundas, considerando tanto métricas de Aprendizado de Máquina quanto simulação de estratégias de investimento. O restante deste trabalho está estruturado como descrito a seguir.

- O Capítulo 2 apresenta informações sobre o mercado de ações, dando ênfase ao mercado brasileiro.
- O Capítulo 3 apresenta a revisão da literatura sobre técnicas de Análise de Sentimentos necessárias a este trabalho, bem como o conceito de representação de palavras, necessário para alimentação das RNAs utilizadas aqui.
- O Capítulo 4 apresenta a revisão da literatura sobre técnicas de Redes Neurais Artificiais e suas diferentes arquiteturas, a saber: *Multilayer Perceptron* (MLP), *Recurrent Neural Network* (RNN), *Long Short-Term Memory* (LSTM), *Bidirectional Long Short-Term Memory* (Bi-LSTM) e *Convolutional Neural Network* (CNN);
- O Capítulo 5 apresenta os trabalhos correlatos a este trabalho, levando em consideração trabalhos de previsão do mercado de ações que utilizem AS ou séries históricas.
- O Capítulo 6 apresenta a metodologia desenvolvida neste trabalho, dividindo-a em: metodologia da Análise de Sentimentos e metodologia da previsão do mercado brasileiro.

- O Capítulo 7 apresenta os resultados obtidos em nossos experimentos, bem como também a comparação experimental com trabalhos correlatos.
- Por fim, o Capítulo 8 apresenta as conclusões e trabalhos futuros para complemento deste trabalho.

Capítulo 2

Mercado de Ações

De acordo com HALFELD (2007), o mercado acionário é como uma associação de agentes, intermediários e corretores de valores, dentre outros, que realizam a negociação de ativos financeiros. O mercado financeiro é o meio em que recursos financeiros são negociados entre empresas ou pessoas físicas com os seguintes intuitos: (1) obter suporte financeiro para seu crescimento por meio do investimento dos recursos adquiridos na abertura de capital; ou (2) obter lucro em negociações (BM&FBOVESPA, 2018). Essa categoria de investimento é denominada como renda variável, visto que sua rentabilidade é incerta, de difícil previsão e, portanto, considerada de alto risco. Dentro da categoria de renda variável, o mercado de ações é preponderante, sendo que o seu funcionamento permite que ações de empresas de capital aberto sejam negociadas na Bolsa de Valores (HALFELD, 2007). Ou seja, na Bolsa de Valores são negociadas ações, que representam um valor emitido por sociedades anônimas (empresas) correspondente à menor fração do capital social da empresa. Uma empresa pode, por exemplo, abrir o seu capital emitindo ações para a implementação de projetos que a desenvolvam. Portanto, esse processo permite ao investidor tornar-se sócio da empresa ao adquirir uma ação de uma empresa de capital aberto. Além disso, vale destacar que o mercado acionário abrange dois tipos de ações: (i) ordinárias (ON), em que os acionistas possuem direito a voto em assembleias de acionistas mas não possuem preferência na distribuição de dividendos (percentual do lucro da empresa); e (ii) preferenciais (PF), em que os acionistas têm preferência no recebimento de dividendos mas não possuem direito a votos em assembleias.

A bolsa brasileira, conhecida como B³ (Brasil, Bolsa e Balcão), que nasceu da fusão entre BM&F BOVESPA e a Central de Custódia e de Liquidação Financeira de Títulos (CETIP), é a

bolsa de valores oficial do Brasil e tem como principal objetivo a inovação e o constante desenvolvimento do mercado financeiro e de capitais. A bolsa B³ pode ser considerada uma das maiores bolsas de mercado de capitais e financeiro do mundo em valor de mercado, possuindo patrimônio de aproximadamente US\$ 13 bilhões, movimentando cerca de 7 a 8 bilhões de dólares todos os dias (FARIA, 2018). Além disso, vale destacar que, considerando o mercado acionário do Brasil, o índice denominado Ibovespa reflete a média da performance dos preços das ações mais negociadas na bolsa, ponderado por um fator proporcional ao volume negociado das mesmas.

Dessa forma, diversas oportunidades de negociar, por meio da compra e venda de ações no mercado, tornaram-se atividades comuns aos investidores. Para auxiliar o processo de tomada de decisão dos investidores surgiram teorias e técnicas de análise de mercado, que estão apresentadas nas seções a seguir.

2.1 Teorias de Mercado

O valor de uma ação de uma determinada empresa está sujeito às leis de oferta e demanda, fazendo com que essa forma de investimento seja classificada como de alto risco. Além disso, o valor de uma ação no mercado está diretamente ligado à situação da empresa perante o cenário global. Nesse sentido, a escolha de uma operação de compra ou venda por parte do investidor não é uma tarefa trivial, existindo estudos para auxiliar investidores em seu processo de análise do mercado financeiro.

Algumas teorias relevantes na literatura sobre o funcionamento do mercado são: a Hipótese do Mercado Eficiente (MALKIEL; FAMA, 1970), a Teoria *Random Walk* (FAMA, 1995) e a Hipótese de Mercado Adaptativo (LO, 2004). Além disso, as análises técnicas e fundamentalistas são abordagens práticas que foram desenvolvidas a partir das teorias de mercado.

A Hipótese de Mercado Eficiente (do inglês, *Efficient Market Hypothesis* - EMH), apresentada por Malkiel e Fama (1970), afirma que em um mercado eficiente, o preço do mercado atual reflete toda a informação disponível. Dessa forma, entende-se que o passado não contém qualquer informação que já não esteja incorporada no preço atual e os preços variam com a chegada de novas informações. Os autores propõem três formas de eficiência de mercado: (i) *fraca*, que sugere que somente preços passados e informação histórica

estejam incorporados no preço atual; (ii) *semi-forte*, que sugere que os preços incorporam o seu comportamento passado e o restante da informação pública; e (iii) *forte*, que sugere que os preços refletem, além de toda a informação pública, toda a informação que pode ser obtida, seja ela oculta ou privilegiada. Para essa teoria, as únicas formas de se obter grandes retornos no mercado de ações seriam: aleatoriamente ou fazendo investimentos de grande risco.

Já a teoria *Random Walk* (RW), apresentada por Fama (1995), considera a previsão de mercado difícil, visto que os preços se formam aleatoriamente no mercado. Nessa teoria, a informação pública estaria disponível para todos, assim como no modo semi-forte da EMH. Dessa forma, essa teoria também considera que superar o mercado, obtendo elevados retornos em transações, seria impossível.

Por outro lado, a Hipótese de Mercado Adaptativo (HMA - do inglês, *Adaptative Market Hypothesis*), definida por Lo (2004), combina aspectos evolutivos e comportamentais ao mercado financeiro, indicando ser possível obter retornos positivos com uma carteira de ações explorando falhas na eficiência do mercado. As teorias previamente citadas supõem que, no mercado, os atores agem sempre em busca do auto-interesse, tomando decisões ótimas e com base na racionalidade. No entanto, a HMA mostra que muitos elementos presentes no mercado acabam por violar a racionalidade humana, como: aversão à perda, superconfiança e reação exagerada, dentre outros vieses comportamentais. Assim, tais comportamentos diante de incertezas trazem ao investidor resultados indesejados tanto em suas operações de investimento como em suas buscas pelo bem-estar econômico. Dessa forma, argumentam os autores que, dentre outros: (1) investimentos tendem a experimentar ciclos de alta e baixa acima do esperado; (2) a eficiência de mercado não é sempre total ou ausente, mas uma característica que varia continuamente no tempo e entre diferentes mercados.

Como aplicações práticas das teorias de mercado apresentadas estão as análises técnica e fundamentalista, discutidas a seguir.

2.2 Análises de Mercado

Dentre as técnicas para análise de mercado, pode-se mencionar a Análise Fundamentalista e a Análise Técnica (BULKOWSKI, 2012; MATSURA, 2013).

A Análise Fundamentalista procura fatores que afetam a oferta e demanda das ações. Para isso, o investidor deve se atentar aos fundamentos macroeconômicos, industriais e específicos da empresa. Essa análise deve ser conduzida por meio do entendimento da saúde financeira da empresa, bem como também da consideração de fatores que envolvam o setor de atuação da mesma. Portanto, a Análise Fundamentalista deve ainda levar em conta pontos que afetem a relação entre oferta e demanda do ativo no mercado, ponderando variáveis que eventualmente afetem o mercado no presente e no futuro. Para esse tipo de análise, é importante destacar que notícias financeiras possuem grande importância para os investidores por apresentarem informações que podem alterar o preço de um ativo.

Dessa forma, esse é o tipo de análise indicada para investidores que desejam estabelecer um relacionamento de longo prazo com a empresa por meio da técnica *Buy & Hold* (Comprar e Segurar), mantendo em sua posse ações enquanto a empresa atende às suas expectativas.

Por outro lado, a Análise Técnica é uma investigação que leva em conta a observação da mudança do mercado considerando representações gráficas de seu comportamento histórico, obtidas a partir de indicadores de valor de compra, venda, volume negociado, dentre outros (OLIVEIRA, F. A. de; NOBRE; ZÁRATE, 2013). A Análise Técnica é uma evolução da Teoria de Dow (RHEA, 1993), que mostra que os preços de um ativo seguem três tendências de períodos diferentes, sendo a tendência primária de longo prazo (podendo durar um ano ou mais), a secundária de médio prazo (podendo durar de semanas a meses) e a terciária de curto prazo (podendo durar de dias a semanas).

Assim, a Análise Técnica se preocupa em identificar tendências e antecipar comportamentos futuros do mercado acionário. Ou seja, essa categoria de análise de mercado assume inicialmente que os investidores apresentam a tendência de repetir comportamentos passados de negociação, além de assumir também que eles apresentam um nível razoável de conhecimento sobre o mercado no instante em que tomam suas decisões. Isto é, a Análise Técnica é focada na compreensão, por parte do investidor, do comportamento histórico dos valores que determinam os preços ou condições atuais do mercado. Dentre as suas estratégias, destaca-se o uso de métodos estatísticos e matemáticos com o intuito de prever os preços e tendências posteriores.

A sequência de preços das ações negociadas pode ser representada por uma série temporal. Tais séries temporais podem ser representadas em gráficos formados por *candlesticks*, que são um dos gráficos mais utilizados para Análise Técnica. O *candlestick*, ilustrado na Figura 2.1,

representa graficamente a variação de preços de uma ação em um determinado período. Cada barra de *candlestick* é um período de tempo composto por quatro preços, a saber: abertura e fechamento (ilustrados no corpo da barra), mínima e máxima (ilustrados, respectivamente, pela sombra inferior e superior). Além disso, duas cores definem o corpo do gráfico. Na Figura 2.1, o *candlestick* vermelho representa um período de baixa (preço de abertura maior que o de fechamento) e o *candlestick* verde representa um período de alta (preço de abertura menor que o de fechamento).

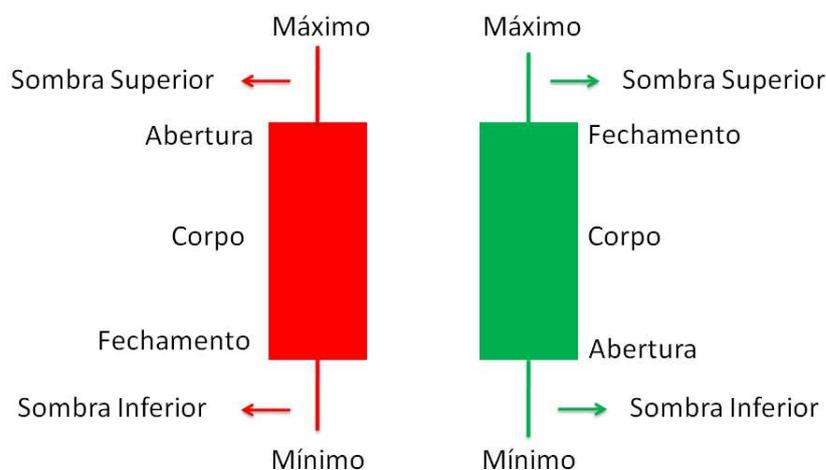


Figura 2.1: Representação de um *candlestick* de período de baixa (em vermelho) e um de período de alta (em verde). Fonte: próprio autor.

O *candlestick* pode representar diversos períodos, sejam dias, meses, horas, etc. Considerando uma análise diária, cada *candlestick* representa um dia de negociações para uma determinada ação, sendo composto de quatro preços: abertura, fechamento, máximo e mínimo. A Figura 2.2 apresenta um gráfico *candlestick* da série temporal das cotações diárias do índice Bovespa (Ibovespa), que sumariza as principais empresas do mercado brasileiro negociadas na B3.

Além disso, a Análise Técnica normalmente é realizada pelo investidor considerando a visualização gráfica dos chamados indicadores técnicos dos gráficos de *candlestick* para os valores históricos de cada ação de interesse. O intuito dos indicadores técnicos é fornecer ao investidor uma possível previsão do movimento futuro dos preços no mercado.

Os indicadores técnicos são medidas que usam cálculos matemáticos a partir do histórico de uma determinada ação a fim de estudar o comportamento do mercado e, também, para a eliminação de ruídos gerados pelo grande volume de negociação de um determinado ativo.



Figura 2.2: Representação de gráfico de *candlestick* para o Ibovespa. Período de março a agosto de 2017 (FARIA, 2018).

Dessa forma, existem diversos tipos de indicadores, cada um utilizando uma fonte de dados diferente e abrangendo uma perspectiva diferente de um mesmo ativo (ACHELIS, 2001).

Dentre as suas classificações, os indicadores podem ser vistos como indicadores de *leading* ou *lagging* (ACHELIS, 2001). Enquanto o primeiro tenta prever uma futura variação no preço do ativo analisado, o segundo fornece uma informação atrasada confirmando tendências de mercado. A Figura 2.3 apresenta alguns dos indicadores técnicos comumente utilizados por investidores, que serão detalhados a seguir.

Um dos indicadores técnicos mais conhecidos são as *Médias Móveis*, normalmente utilizadas pelo investidor para verificar tendências por meio do cruzamento de médias de diferentes períodos, além de atenuar ruídos de variações de preços de um ativo em um determinado prazo. As *Médias Móveis* abrangem diversas variações, sendo as mais comuns: a média móvel simples (do inglês, *Simple Moving Average* - SMA), representada na Figura 2.3 (A), a média móvel exponencial (do inglês *Exponential Moving Average* - EMA) e a média móvel ponderada (do inglês *Weighted Moving Average* - WMA), todas calculadas considerando um período de ordem n (MATSURA, 2013).

Uma adaptação do indicador de *Médias Móveis* é o chamado indicador *Convergência e Divergência de Médias Móveis* (do inglês, *Moving Average Convergence and Divergence* - MACD), representado na Figura 2.3 (B). Esse indicador mostra a relação entre médias móveis exponenciais de preços, com o principal objetivo de identificar tendências. O MACD é gerado



Figura 2.3: Representação de indicadores técnicos comumente utilizados por investidores: (A) *Simple Moving Average* - SMA, (B) *Moving Average Convergence and Divergence* - MACD; (C) *Relative Strength Index* - RSI; (D) *Money Flow Index* - MFI; (E) *Money Flow Index* - BB; e (F) *On Balance Volume* - OBV.

a partir da subtração de duas médias móveis exponenciais, cujos valores tradicionais de período são 26 e 12, respectivamente, seguida de uma comparação com uma média móvel exponencial denominada sinal, de período 9. O resultado desse indicador é um histograma que indica a tendência de alta ou baixa de um ativo (ACHELIS, 2001).

Outro indicador presente na Análise Técnica é o *Índice de Força Relativa* (do inglês, *Relative Strength Index* - RSI), representado na Figura 2.3 (C). Esse indicador mede a aceleração do movimento de preços de uma ação, além de mostrar indicações quando a sua velocidade é reduzida. Esse indicador é exibido ao investidor em um gráfico de linha, mostrando os seguintes sinais de um ativo: (1) sobrecompra (i.e., quando um ativo está

significativamente valorizado e há uma expectativa que seu valor reverta para queda) quando seu valor está, geralmente, acima de 80; ou (2) sobrevenda (i.e., quando um ativo está significativamente desvalorizado e há uma expectativa que seu valor reverta para alta) quando seu valor está, geralmente, abaixo de 20 (MATSURA, 2013).

O indicador *Fluxo de Dinheiro* (do inglês, *Money Flow Index* - MFI), representado na Figura 2.3 (D), mostra a intensidade do fluxo de dinheiro para dentro e para fora de um ativo. Esse indicador, assim como o RSI, é capaz de identificar sinais de sobrecompra ou sobrevenda em um ativo. No entanto, o MFI usa dados de preço acrescido de dados de volume de negociações. Dentre as suas aplicações, um valor de MFI acima de 80 em um determinado ativo é considerado sobrecompra, enquanto que um valor abaixo de 20 é considerado sobrevenda (ACHELIS, 2001).

O indicador *Bandas Bollinger* (do inglês, *Money Flow Index* - BB), representado na Figura 2.3 (E), mostra para o investidor, graficamente, o dobro do desvio padrão acima e abaixo de uma média móvel de 20 períodos do valor do ativo. Este indicador mostra a volatilidade de um ativo, aumentando as bandas quando o mercado está instável e diminuindo-as em um período de estabilidade. Além disso, à medida em que os preços se movem rumo à banda superior, há um indicativo de sobrecompra, enquanto que à medida em que os preços se movem para a banda inferior, há um indicativo de sobrevenda do ativo.

Além desses, o indicador *On Balance Volume* (OBV), representado na Figura 2.3 (F), usa as mudanças de volume para prever as mudanças de preço de um ativo. O objetivo desse indicador é identificar tendências de alta e baixa no mercado, considerando que o volume é uma força importante nos principais movimentos do mercado.

Por fim, vale a pena ressaltar que existem vários indicadores técnicos na literatura além dos apresentados neste trabalho (ACHELIS, 2001; MATSURA, 2013), sendo a sua escolha dependente das preferências do investidor. Conforme observado, cada indicador técnico apresenta uma visão parcial do mercado e dificilmente fomenta uma estratégia de investimento por si só. Assim, os indicadores normalmente são utilizados em combinações pelo investidor, cuja escolha deve ser baseada em quais indicadores são adequados a cada situação ou ativo no mercado.

Neste trabalho, para a escolha dos indicadores técnicos utilizados, procuramos abordar os principais indicadores apresentados nos trabalhos de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020) e Picasso et al. (2019).

2.3 Considerações Finais

Este capítulo introduziu ao leitor os principais conceitos de mercado financeiro, tendo o objetivo de facilitar o entendimento do restante deste trabalho. Desse modo, foram apresentadas teorias de mercado (Hipótese de Mercado Eficiente, *Random Walk* e Hipótese de Mercado Adaptativo) e suas análises mais comumente utilizadas (Análise Fundamentalista e Análise Técnica), bem como também conceitos essenciais como *candlestick* e indicadores técnicos.

Os conceitos aqui discutidos serão aplicados ao longo do desenvolvimento deste trabalho, servindo como base para o desenvolvimento do Módulo de Previsão de mercado, apresentado em detalhes no Capítulo 6.

Capítulo 3

Análise de Sentimentos

A *World Wide Web* (WWW) é um dos meios fundamentais de publicações de conteúdo e opiniões atualmente. Existe uma grande quantidade de informações disponível *online*, seja por meio de redes sociais, *sites* de notícias ou mesmo *blogs*. Dessa forma, recentemente tem surgido uma grande quantidade de trabalhos cujo principal objetivo é investigar esse conhecimento disponível por meio da extração de opiniões e sentimentos, com o uso da técnica de Análise de Sentimentos (MARTINS; PEREIRA, A.; BENEVENUTO, 2015).

A Análise de Sentimentos (AS) pode ser vista como uma técnica de Processamento de Linguagem Natural (PLN) que tem como objetivo analisar opiniões, sentimentos e emoções presentes em dados não-estruturados (PANG, B.; LEE, L. et al., 2008; HUSSEIN, 2018). Assim, uma atividade muito comum em trabalhos que usam AS é o reconhecimento de duas classes de sentimentos dentro de um documento textual: positiva e negativa. Alguns trabalhos consideram também uma terceira classe de sentimento, o neutro. A AS é realizada a partir do desenvolvimento de classificadores, por meio de, principalmente, duas técnicas, a saber: métodos lexicais e modelos de Aprendizado de Máquina (AM).

Métodos lexicais dependem de conhecimento linguístico para serem desenvolvidos, e são robustos quando portados para outros domínios de aplicação (TABOADA et al., 2011). Nessa abordagem, normalmente é utilizada uma lista contendo um conjunto de pares de palavras e a respectiva polaridade de sentimentos associados a elas. Avanço e Nunes (2014) mostraram o uso de um método léxico para a tarefa de AS em língua portuguesa. O dicionário utilizado pelos autores apresenta polaridade para cada uma de suas palavras considerando o respectivo sentimento. Ainda, os autores incrementaram a metodologia de AS adicionando conhecimento linguístico, considerando o contexto ao qual a palavra está incluída na

sentença a ser analisada. Por outro lado, o estudo de Martins, Adriano Pereira e Benevenuto (2015) apresentou a metodologia denominada SentiPipe, que realiza a AS em português com a combinação de um conjunto de métodos léxicos.

Apesar de existirem técnicas para incrementar métodos léxicos mais simples, como apresentado em Taboada et al. (2011), raramente esses métodos conseguem atingir resultados em tarefas de AS que se equivalem aos resultados obtidos por técnicas de AM (AVANÇO; NUNES, 2014). Além disso, em um trabalho anterior (JANUÁRIO et al., 2021), verificamos que, para o domínio do mercado financeiro, métodos de Aprendizado de Máquina superaram uma abordagem lexical, ainda que essa tenha utilizado uma estratégia de AS composta de um dicionário específico do domínio financeiro. Nesse sentido, esta tese foca principalmente em métodos de AM para Análise de Sentimentos.

As técnicas de AM mais utilizadas para Análise de Sentimentos, conforme Sun, Luo e Junyu Chen (2017), são: *Naive Bayes*, *Maximum Entropy* (ME) e *Support Vector Machines* (SVM) (MOCHERLA; DANEHY; IMPEY, 2017). A técnica *Naive Bayes* utiliza o Teorema de Bayes para o cálculo de probabilidade de um documento d se associar a uma determinada classe c (no caso, de um sentimento específico). Do mesmo modo, a técnica *maximum entropy* (ME) faz uma verificação similar, por meio de uma função de probabilidade denominada entropia máxima, para realizar a classificação. Por fim, a técnica SVM busca determinar um hiperplano que divida os documentos em diferentes classes, maximizando a distância entre eles. Todas as três técnicas têm apresentado bons resultados em termos de acurácia ao classificar textos, conforme verificado por Bo Pang, Lillian Lee e Vaithyanathan (2002).

Além disso, ultimamente Redes Neurais Artificiais (RNAs) têm atraído bastante atenção na literatura científica para uso em tarefas de classificação e predição (KRAUS; FEUERRIEGEL, 2017). Segundo Haykin (1994), uma RNA é formada por conexões entre neurônios artificiais, que são unidades de processamento matemático, com o objetivo de simular a ação de um neurônio biológico. As RNAs têm apresentado resultados promissores em tarefas de classificação de documentos textuais, especialmente em se tratando de RNAs profundas (KRAUS; FEUERRIEGEL, 2017).

Nos últimos 10 anos, o desenvolvimento das RNAs profundas fez grandes avanços e produziu resultados referentes ao estado da arte em muitos domínios de aplicação, começando pela visão computacional, depois pelo reconhecimento de fala e PLN (COLLOBERT et al., 2011). O crescimento do uso das RNAs pode ser atribuído a muitos

fatores. Os mais importantes incluem: a disponibilidade de poder de computação devido aos avanços em *hardware* (por exemplo, GPUs) e à disponibilidade de grandes quantidades de dados de treinamento (ZHANG, L.; WANG, S.; LIU, B., 2018). RNAs são abordadas em mais detalhes no Capítulo 4.

Algoritmos de AM que tratam de tarefas de PLN, em especial Análise de Sentimentos, necessitam que exista uma representação para dados que estão no formato textual. Nesse sentido, a Seção 3.1 detalha como preparar dados textuais para um formato adequado a esses algoritmos.

3.1 Representação de Palavras

Tarefas de PLN necessitam que dados textuais e documentos sejam representados de modo que os algoritmos possam processá-los. Por meio do Modelo de Espaço Vetorial (do inglês, *Vector Space Model* – VSM) (SALTON; WONG; YANG, 1975), dados textuais são representados de forma geométrica. Ou seja, no VSM os dados são representados em um espaço t -dimensional, em que cada elemento corresponde a uma palavra do vocabulário. Nesse sentido, pontos próximos entre si no espaço podem ser considerados com semântica similar, sendo possível obter a sua similaridade por meio do cálculo da distância entre eles ou por meio da função *coseno* de similaridade (FARIA, 2018).

Dentre os modos de representação VSM, o vetor *one-hot* é um dos mais simples e utilizados em tarefas de PLN. Esse vetor é esparso, do tamanho do vocabulário do documento a ser considerado, com representação composta de valores 1 se a palavra pertence ao documento, e 0 se a palavra não pertence ao documento.

Vale destacar também que existe uma versão utilizada nos modelos VSM que é ponderada pelo cálculo da frequência do termo (do inglês, *term frequency* - tf), pelo inverso da frequência no documento (do inglês, *inverse document frequency* - idf) ou pela frequência do termo-inverso da frequência do documento (do inglês, *term frequency-inverse document frequency* - tfidf). O cálculo de frequência do termo corresponde à frequência do termo i no documento j . O cálculo da frequência inversa do documento realiza um cálculo para diminuir a relevância do peso dos termos que ocorrem frequentemente nos documentos, ao passo que beneficia termos que aparecem raramente. E, por fim, o cálculo frequência do termo-frequência inversa do

documento é a medida que indica a importância de uma palavra em um documento em relação a uma coleção de documentos.

A união de todos os vetores compreende uma matriz de características denominada de “saco de palavras” (do inglês, *Bag-of-Words* - BOW). Nessa matriz, cada linha representa um documento (d_n) e cada coluna representa uma palavra (t_n), que pode ser 1 ou valores relacionados à frequência do termo. A Figura 3.1 ilustra um conjunto de vetores BOW com 5 termos (t_1, \dots, t_5) e seus respectivos documentos (d_1, \dots, d_n).

	t_1	t_2	t_3	t_4	t_5
d_1	1	0	1	0	1
d_2	0	1	1	0	0
.					
.					
.					
d_n	0	1	1	0	0

Figura 3.1: Representação de um conjunto de vetores compondo um *Bag of Words*.

Uma desvantagem dessa representação é que a ordem das palavras não é levada em consideração, descartando a estrutura presente em frases, sentenças e parágrafos do documento. Além disso, essa representação possui dimensionalidade dependente do número de palavras, que em geral é elevado, além de depender do número de documentos pertencente à coletânea de documentos analisada.

Para contornar essa questão, muitos modelos de Aprendizado de Máquina em PLN fazem uso de uma estratégia de representação de palavras como entrada de dados denominadas *word embeddings* (COLLOBERT et al., 2011). *Word embeddings* são técnicas para modelagem de linguagem e aprendizado de características que transformam palavras pertencentes a um vocabulário em vetores de números reais. Normalmente, essa técnica reduz um vetor esparsa e de alta dimensão a um vetor denso e de baixa dimensionalidade, em que cada dimensão representa uma característica da palavra em questão.

A representação vetorial de palavras por meio de *word embeddings* pode também ser utilizada para representar o significado das palavras, sendo que os vetores que representam palavras similares podem ser vizinhos no espaço de representação. Esse tipo de representação está associada à hipótese da distribuição, que afirma que palavras que estão presentes em contextos similares tendem a ter significados similares (HARRIS, 1954).

Dentre diversos métodos disponíveis, o princípio da hipótese distribuída pode ser aplicado para a criação de *word embeddings* por meio de métodos preditivos. Esses métodos tentam realizar a previsão de uma palavra a partir de suas palavras vizinhas por meio de algoritmos de AM (BENGIO; DUCHARME et al., 2003). O método pioneiro para criação de *word embeddings* é denominado *Word2Vec* (MIKOLOV; CHEN, K. et al., 2013), que consiste em um modelo de RNA que aprende características das palavras de um texto e as codifica em vetores. Essa técnica pode ser implementada de duas formas: *continuous bag-of-words* (CBoW) e *Skip-gram* (SG). Enquanto CBoW prevê uma palavra alvo (por ex.: laranja) por meio de suas palavras de contexto ("suco de _", onde "_" significa a palavra alvo), SG faz o contrário, ou seja, prevê as palavras de contexto dada a palavra alvo.

Uma extensão do modelo *Word2Vec* é o denominado *FastText* (BOJANOWSKI et al., 2017), que é um dos algoritmos mais populares para geração de representações de palavras, permitindo o treinamento rápido de grandes *corpora*, superando o *Word2Vec*, além de ser um dos algoritmos mais modernos para representação de morfologia de palavras. No desenvolvimento deste trabalho, utilizamos o modelo *FastText*.

Os algoritmos *Word2Vec* e *FastText* são apresentados em detalhes nas seções a seguir.

3.1.1 Word2Vec

Mikolov, Kai Chen et al. (2013) propuseram o algoritmo *Word2Vec*, um método preditivo que utiliza a hipótese da distribuição de palavras, conforme discutido na Seção 3.1, responsável por criar uma representação vetorial de palavras (*word embeddings*) por meio de técnicas de RNAs com arquitetura rasa, ou seja, de poucas camadas. Com isso, um dos efeitos desse método é que os vetores capturam, indiretamente, propriedades linguísticas, ao contrário de abordagens como BOW. Os vetores resultantes podem, portanto, ser utilizados em diferentes aplicações de PLN: classificação de textos, agrupamento de documentos, análise de sentimentos, dentre outras, além de poderem ser utilizados como entradas de modelos de RNAs profundas.

O algoritmo *Word2Vec* agrupa palavras com semântica similar no espaço vetorial, isto é, o algoritmo produz uma representação distribuída mantendo relações semânticas entre as palavras e seus respectivos vizinhos no espaço vetorial. Por exemplo, em Mikolov, Sutskever et al. (2013) mostra-se que é possível realizar operações entre vetores de palavras, como em $v(\textit{king}) - v(\textit{man}) + v(\textit{women}) = v(\textit{queen})$, cujo resultado representa a versão feminina de rei $v(\textit{king})$. Nesse caso, $v(\textit{palavra})$ é a representação distribuída em *Word2Vec* de *palavra*. Vale

destacar que não é necessário fornecer nenhum conhecimento sintático ou semântico ao algoritmo, que é capaz de identificar tais relacionamentos por meio da verificação da posição das palavras, bem como também do contexto em que estão inseridas nos documentos. Ou seja, o algoritmo é capaz de aprender representações vetoriais de alta qualidade e baixa dimensionalidade (geralmente, os valores de dimensionalidade variam entre 50 e 200). Ainda, o *Word2Vec* permite identificar automaticamente padrões linguísticos presentes nos dados, sendo eficiente em representações semânticas devido à sua compactação e geometria, capturando, desse modo, informações de relacionamento entre as palavras.

O algoritmo *Word2Vec* pode ser implementado por meio de dois métodos: *Continuous Bag-of-Words* (CBOW) e *Skip-gram* (SG). Ambos apresentam funcionamento similar, porém, enquanto o primeiro prevê a palavra alvo por meio do contexto, o segundo prevê o contexto a partir da palavra alvo. Esse processo é apresentado em alto nível na Figura 3.2, em que $w(t)$ significa a palavra alvo e as demais representam as palavras do contexto, sendo todas representadas em formato de vetor *one-hot*. Assim, o algoritmo detalhado a seguir será a versão SG, sendo que para o funcionamento do algoritmo CBOW basta alterar a entrada e saída da RNA, invertendo-se contexto e alvo.

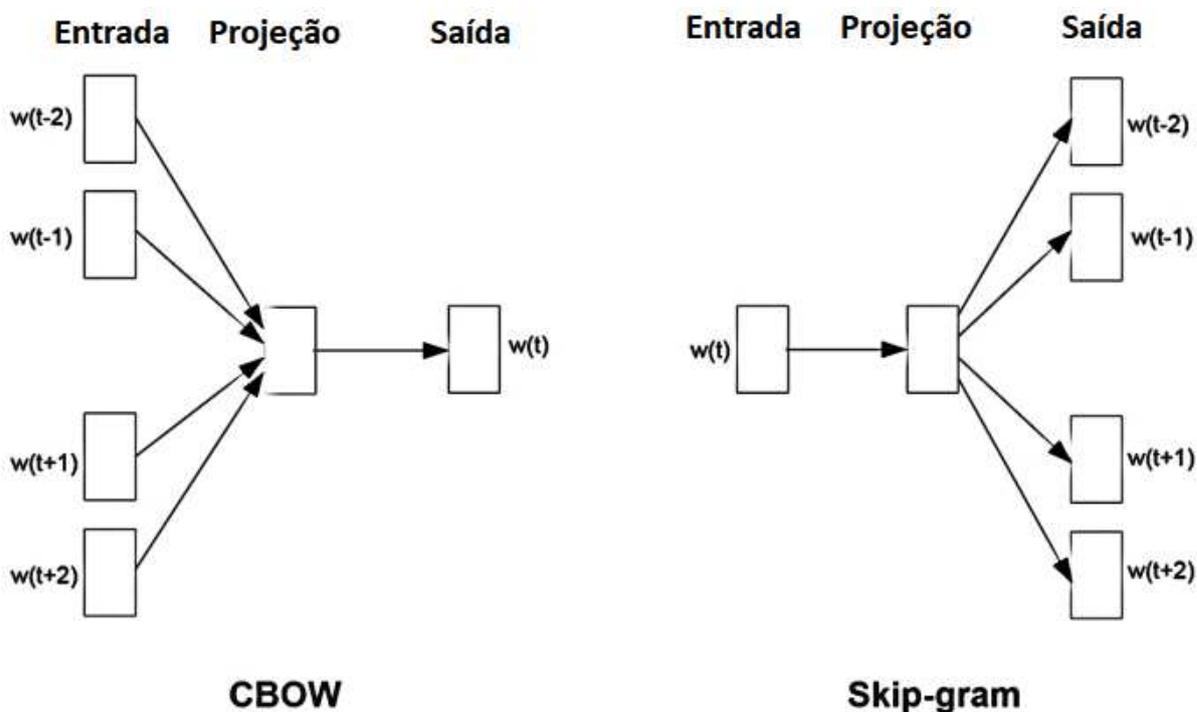


Figura 3.2: Representação geral funcionamento do algoritmo *Word2Vec*, modelos *CBOW* e *Skip-gram*. Adaptado de (MIKOLOV; CHEN, K. et al., 2013).

O modelo *Skip-gram* é um modelo de rede neural com uma camada oculta que é treinado utilizando como entrada a palavra corrente e como saída esperada o contexto. A Figura 3.3 ilustra o funcionamento do modelo SG, sendo: V a dimensão do vocabulário; N o número de neurônios na camada oculta; x representa a entrada e y representa a saída da rede, sendo ambos vetores *one-hot*; $W_{V \times N}$ representa a matriz de pesos sinápticos que está entre a camada de entrada e a camada oculta; $W'_{N \times V}$ referencia os pesos sinápticos entre a camada oculta e a camada de saída; e C representa a dimensão do contexto, ou seja, o número de palavras vizinhas da palavra alvo.

Na Figura 3.3, o vetor de entrada x_k é um vetor *one-hot* que representa a palavra alvo. Nesse vetor, todas as posições possuem valor zero, exceto a posição correspondente à palavra k . A saída do modelo é um conjunto de vetores, em que $y_{C,j}$ corresponde à palavra pertencente ao contexto C da palavra de entrada k . Note que, por se tratar do modelo SG, espera-se que exista apenas uma entrada para a rede neural, correspondente à palavra alvo, e uma ou mais saídas, correspondentes ao contexto da palavra alvo.

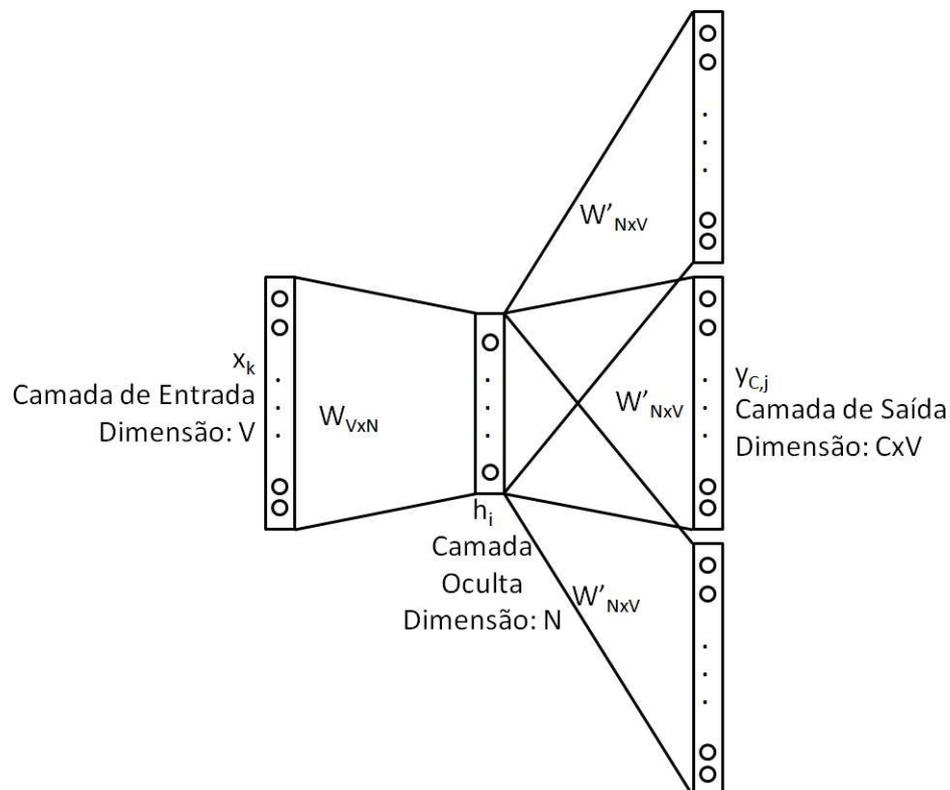


Figura 3.3: Representação do modelo utilizado no algoritmo Word2Vec, modelo *Skip-gram*. Adaptado de (RONG, 2014).

Dessa forma, a função de ativação dessa rede neural é definida pelo modelo *softmax* (Equação 3.1), que representa a probabilidade da predição da palavra w_i dada a palavra w_j ,

considerando: (i) uma sequência de palavras do conjunto de treinamento (w_1, \dots, w_n) fornecida à rede; e (ii) os vetores u_w e v_w , que são aprendidos durante o treinamento da rede, representando, respectivamente, vetores de entrada e saída da palavra w .

$$p(w_i|w_j) = \frac{\exp(u_{w_i}^T v_{w_j})}{\sum_{l=1}^V \exp(u_l^T v_{w_j})} \quad (3.1)$$

Perceba que u_w e v_w são duas representações da palavra w_j . Ou seja, vetor u_w vem das linhas da matriz de $W_{V \times N}$, que é a matriz de pesos sinápticos que conecta as camadas de entrada e oculta da rede neural, enquanto que o vetor v_w vem das colunas que representam a matriz de pesos sinápticos $W'_{N \times V}$, que conecta as camadas oculta e de saída da rede. Portanto, vale a pena destacar que a representação em *Word2Vec* de uma palavra w_j é dada pelos pesos sinápticos representados pelo vetor v_w .

Além disso, a função do erro da rede neural é apresentada pela Equação 3.2, onde C é o tamanho da janela de contexto do treinamento e T é o número de palavras. Dessa forma, o treinamento dessa rede é efetuado pelo algoritmo *Backpropagation*, apresentado no Capítulo 4.

$$J(\theta) = \frac{1}{T} \sum_{t=1}^T \left[\sum_{j=-C, j \neq 0}^C \log(p(w_{t+j}|w_t)) \right] \quad (3.2)$$

O algoritmo SG com função *softmax* é considerado de alto custo computacional por seus autores, pois o seu processamento dependerá da quantidade de palavras pertencentes ao vocabulário, que frequentemente é muito grande (MIKOLOV; CHEN, K. et al., 2013). Como alternativas, os autores propõem duas estratégias: o *softmax* hierárquico e amostragem negativa (do inglês, *negative sampling*).

O *softmax* hierárquico é uma otimização do algoritmo *Word2Vec* realizada por meio de técnicas de aproximação para o cálculo das probabilidades de escolha de cada uma das palavras. Nessa técnica, ocorre uma redução do número de nós de saída que são avaliados de V para $\log_2(V)$, de acordo com uma representação em forma de árvore binária para todas as palavras do vocabulário V . Ou seja, cada palavra em V representa um nó folha da árvore com um caminho único desde a raiz. Esse caminho é, portanto, utilizado para estimar as probabilidades das palavras dos nós folhas, reduzindo o número de cálculo de probabilidades necessário ao algoritmo e, conseqüentemente, o seu custo computacional.

Por exemplo, considere a Figura 3.4, que representa uma árvore binária para o modelo *softmax* hierárquico. Nela, as unidades em branco (nós folha) (w_V) representam as palavras do

vocabulário V e $n(w, j)$ significa o nó j no caminho da raiz à palavra w . Essa figura apresenta um caminho detalhado para a palavra w_2 , que possui tamanho 4, indicando que devem ser computadas e multiplicadas as seguintes probabilidades para w_2 : probabilidade de percorrer o caminho à esquerda do nó 1, probabilidade de percorrer o caminho à esquerda do nó 2 e a probabilidade de percorrer o caminho à direita do nó 3.

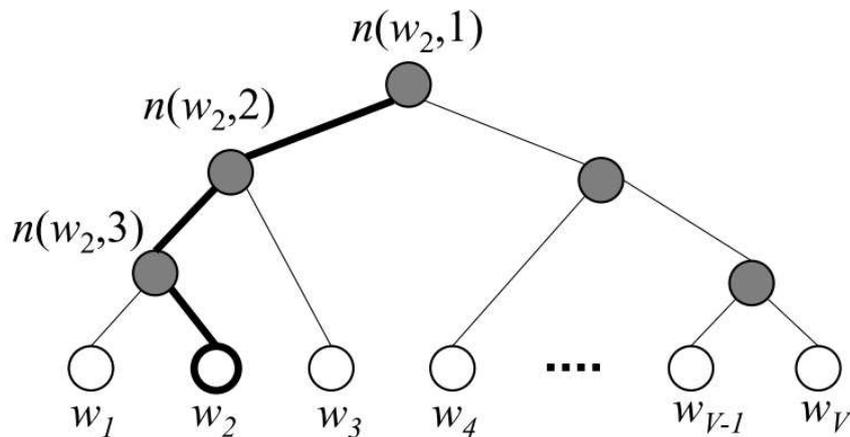


Figura 3.4: Representação de uma árvore binária para o modelo softmax hierárquico. (RONG, 2014).

Por outro lado, a técnica de amostragem negativa (*negative sampling*) (MIKOLOV; SUTSKEVER et al., 2013), para lidar com a dificuldade de ter muitos vetores de saída que precisam ser atualizados, atualiza apenas uma amostra deles. Para isso, cria-se uma amostragem em que a palavra de saída deve ser inicialmente mantida como correta, adicionada de algumas palavras selecionadas como amostras negativas.

Assim, o modelo é treinado para discriminar a palavra correta de N palavras ruidosas de acordo com um determinado contexto, conforme apresentado na Figura 3.5. Por exemplo, considerando que a amostragem correta envolva a palavra “suco” para o contexto “laranja”, as amostragens negativas poderiam incluir: “rei” e “laranja”, “livro” e “laranja”, “cabeça” e “laranja” e assim por diante.

Como método de seleção das N palavras ruidosas da amostragem, (MIKOLOV; SUTSKEVER et al., 2013) indicam que a distribuição Unigrama, $U(w)$, apresentada na Eq. 3.3, foi a distribuição que obteve melhores resultados nos experimentos, onde $f(w)$ representa a frequência da palavra no *corpus* analisado e $3/4$ é um parâmetro encontrado experimentalmente.

Entrada		Saída
Contexto	Palavra	Alvo
Laranja	Suco	1
Laranja	Rei	0
Laranja	Livro	0
Laranja	Cabeça	0



Figura 3.5: Representação da técnica de amostragem negativa.

$$P(w_i) = \frac{f(w_i)^{(3/4)}}{\sum_{j=0}^N f(w_j)^{(3/4)}} \quad (3.3)$$

Por fim, vale destacar que o algoritmo *Word2Vec* foi apresentado nesta seção por ser um dos trabalhos pioneiros da área na literatura, além de ter servido como base para o desenvolvimento de estratégias posteriores, como o *FastText*, apresentado na próxima seção.

3.1.2 FastText

Uma grande desvantagem das técnicas de *word embeddings*, como *Word2Vec*, é a sua incapacidade de lidar com palavras fora do *corpus* para o qual o modelo foi treinado. Nesse sentido, o modelo gerador de *word embeddings* denominado *FastText* foi desenvolvido para o aprendizado eficiente de representações de palavras e classificação de sentenças, estendendo o algoritmo *Word2Vec* (BOJANOWSKI et al., 2017; JOULIN; GRAVE; BOJANOWSKI; MIKOLOV, 2017; JOULIN; GRAVE; BOJANOWSKI; DOUZE et al., 2016).

O objetivo principal dos *word embeddings* gerado pelo *FastText* é levar em consideração a estrutura interna das palavras, uma abordagem extremamente útil para línguas morfológicamente ricas, de modo que as representações para diferentes formas morfológicas das palavras sejam treinadas no modelo de forma independente.

Em detalhes, enquanto que os *word embeddings* gerados pelo algoritmo *Word2Vec* levam em conta cada palavra como a menor unidade cuja representação vetorial é gerada, o algoritmo *FastText* pressupõe que uma palavra é formada por unidades menores dentro da mesma palavra, denominadas *n*-gramas, em que *n* pode variar de 1 ao comprimento da palavra. Por exemplo, a palavra *termo* é composta dos seguintes 3-gramas: $\langle te, ter, erm, rmo, mo \rangle$, sendo que os elementos \langle e \rangle indicam início e fim do *token* para o modelo, respectivamente.

Essa nova representação de palavras feita pelo *FastText* apresenta as seguintes vantagens (BOJANOWSKI et al., 2017):

- Maior facilidade para encontrar a representação vetorial de palavras raras. Como as palavras raras ainda podem ser divididas em caracteres n -gramas, elas podem compartilhar esses n -gramas com palavras comuns usadas para treinar o modelo.
- Pode fornecer as representações vetoriais para as palavras não presentes no dicionário, considerando que também podem ser divididas em n -gramas de caracteres. O algoritmo *Word2Vec* não fornece nenhuma representação vetorial para palavras que não estejam no dicionário.

Para ilustrar, considere que a palavra *domínio* esteja no conjunto de dados de teste de uma aplicação de PLN com *FastText*, mas não tenha nenhuma representação no conjunto de treinamento do modelo. Supondo que o conjunto de treinamento possua a representação vetorial de todos os seus n -gramas, é possível, por exemplo, calcular a média da representação vetorial de todos os n -gramas e obter o *word embedding* dessa palavra.

Assim, dois parâmetros que são fornecidos ao algoritmo são os valores mínimo e máximo de n -gramas. Embora valores baixos de n -gramas permitam uma maior representação de porções menores de palavras, facilitando uma futura representação de palavras que não foram utilizadas no modelo, há também um maior gasto de memória quando se considera o desempenho do algoritmo.

O modelo *FastText* funciona do seguinte modo. Há uma janela que segue percorrendo o texto de entrada aplicando a mesma forma de treinamento do algoritmo *Word2Vec*, isto é, os algoritmos *Skip-gram* ou CBOW (SANTOS, I.; NEDJAH; MACEDO MOURELLE, 2017). O aprendizado do modelo (conforme apresentado na seção anterior), portanto, pode ser visto como uma série de atualizações sobre uma RNA com três camadas de neurônios. As duas camadas externas possuem cada uma um neurônio para cada palavra no vocabulário, enquanto que a camada do meio apresenta o número de neurônios de acordo com as dimensões definidas pelo tamanho dos *word embeddings*.

A abordagem desse algoritmo é muito semelhante ao *Word2Vec* (exceto pelo uso de subpartes de palavras, que traz as vantagens discutidas anteriormente). Vale a pena destacar que inclusive ambos algoritmos foram desenvolvidos tendo em comum um mesmo autor, Tomas Mikolov (BOJANOWSKI et al., 2017; MIKOLOV; CHEN, K. et al., 2013).

Como discutido, o algoritmo *FastText* apresenta vantagens que aprimoram o aprendizado do vocabulário, principalmente em idiomas que apresentam muitas flexões na linguagem, como o português. Nesse sentido, esse é o modelo de representação de linguagem escolhido para o desenvolvimento do restante deste trabalho.

3.2 Considerações Finais

Este capítulo apresentou os conceitos relacionados a AS e representação de palavras, essenciais para o desenvolvimento do restante deste trabalho. Além disso, apresentou o conceito de *word embeddings*, que são necessários para alimentar Redes Neurais Artificiais que desempenham tarefas de Processamento de Linguagem Natural, por meio dos algoritmos *Word2Vec* e *FastText*. No restante deste trabalho, o conteúdo apresentado servirá como base para a extração de características de notícias da área financeira para serem utilizadas como dados voltados à previsão do mercado acionário.

Capítulo 4

Redes Neurais Artificiais

As Redes Neurais Artificiais (RNAs) (do inglês *Artificial Neural Networks*) começaram a ser exploradas com maior ênfase a partir dos anos 80, apesar de suas primeiras pesquisas terem surgido há mais de 50 anos (SILVA, I. da; SPATTI; FLAUZINO, 2010). Em meados dos anos 80, os modelos de RNAs denominados MLP (do inglês *Multilayer Perceptron*) foram popularizados, apresentando significativa capacidade de generalização a partir de dados fornecidos em sua etapa de treinamento. Já nas últimas décadas, as arquiteturas de RNAs têm chamado a atenção por meio de modelos de aprendizado profundo (do inglês *Deep Learning Neural Networks* - DLNN), que são caracterizados por uma representação em várias camadas hierárquicas. Nesses modelos, características de mais alto nível são obtidas por meio da composição de características de níveis inferiores. Modelos baseados em RNAs profundas têm apresentado resultados que são o estado da arte em muitas aplicações de Inteligência Artificial, tais como: reconhecimento de imagem, áudio, na área de Processamento de Linguagem Natural, dentre outros (JOHNSON; ZHANG, T., 2015; LECUN; BENGIO; HINTON, G., 2015; KENTON; TOUTANOVA, 2019; BUSTOS; POMARES-QUIMBAYA, 2020; KUMBURE et al., 2022).

Resumidamente, pode-se dizer que as RNAs são modelos matemáticos baseados em características do cérebro humano, cuja funcionalidade é adquirida por meio de aprendizado e generalização. As RNAs reconhecem e classificam padrões para posteriormente generalizar o conhecimento adquirido (HAYKIN, 2009).

Neste capítulo são discutidos os seguintes conteúdos. A Seção 4.1 apresenta pontos gerais sobre RNAs, enquanto as seções 4.2.1 a 4.2.4 apresentam, respectivamente, as seguintes arquiteturas de RNAs: *Multilayer Perceptron*, *Recurrent Neural Network*, *Long Short-Term*

Memory, *Bidirectional Long Short-Term Memory* e *Convolutional Neural Network*. Por fim, a Seção 4.3 apresenta as considerações finais do capítulo.

4.1 Aspectos Gerais

Uma RNA é composta por conexões entre diversos neurônios artificiais, cujo principal objetivo é simular um neurônio biológico. Em 1943, McCulloch e Pitts (1943) descreveram pela primeira vez como seria um neurônio artificial, cujo funcionamento é realizado por meio de uma unidade que recebe entradas e produz saídas, representando, respectivamente, os dendritos e axônios em um neurônio biológico. O neurônio artificial dispara quando suas entradas, moduladas pelas respectivas sinapses, ultrapassam um determinado limiar. Esse neurônio pode representar apenas dois estados: ativado ou desativado. Como principal limitação, ele não possuía uma técnica de aprendizado, mantendo fixos seus pesos sinápticos.

Em 1958 foi introduzido o *Perceptron* (ROSENBLATT, 1958), que é um modelo de neurônio para reconhecimento de padrões. Esse trabalho introduziu o algoritmo de treinamento denominado algoritmo de convergência do perceptron, que permite que seus pesos sinápticos aprendam a saída correta de acordo com atualizações iterativas.

A Figura 4.1 ilustra o funcionamento do *Perceptron*. Primeiramente, há a ponderação dos valores de entrada X_n por seus respectivos pesos sinápticos W_n , e que em seguida passam por um somatório (representado pelo símbolo \sum) juntamente com um valor de viés (do inglês, *bias*). Por fim, o resultado obtido produz um potencial de ativação r , cujo uso posterior é por meio de uma função de ativação $f(r)$. Finalmente, o resultado obtido por meio da aplicação da função de ativação determina o valor de saída do neurônio (HAYKIN, 2009).

Existem diversas funções de ativação disponíveis na literatura. Primeiramente, pode-se destacar a *função tangente hiperbólica*, apresentada na Equação 4.1 e ilustrada na Figura 4.2, cujo resultado é compreendido entre os valores -1 e 1 (HAYKIN, 2009).

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (4.1)$$

Outra função de ativação que pode ser utilizada, dependendo do contexto do problema a ser tratado, é a *função logística*, que limita a saída de um neurônio entre os valores 0 e 1, apresentada na Equação 4.2 e ilustrada na Figura 4.3 (SILVA, I. da; SPATTI; FLAUZINO, 2010;

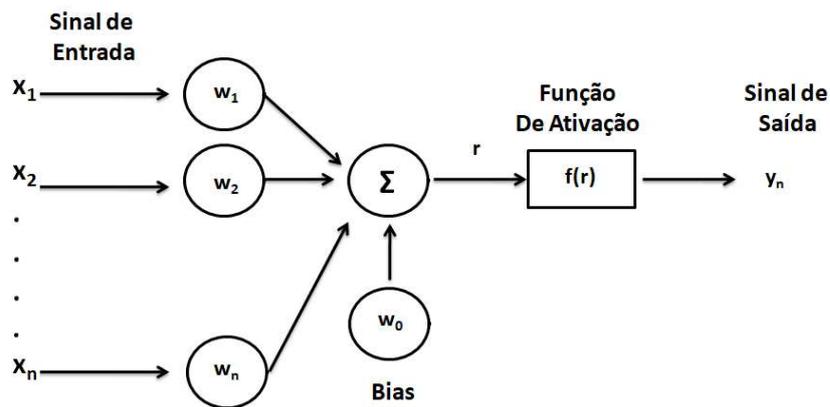


Figura 4.1: Neurônio artificial do tipo Perceptron. Representação matemática. Adaptado de (HAYKIN, 2009).

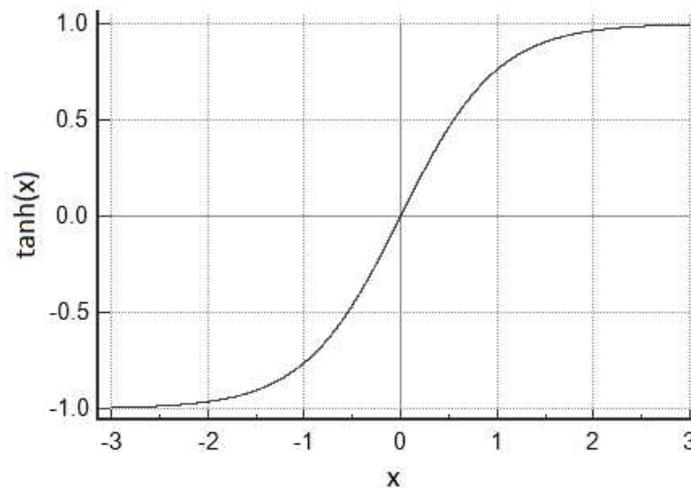


Figura 4.2: Gráfico ilustrativo da função tangente hiperbólica. Adaptado de (HAYKIN, 2009).

HAYKIN, 2009). Essa função é muito utilizada por modelos que devem prever probabilidades como saída (FARIA, 2018).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.2)$$

Além disso, a função *ReLU* (do inglês *REctified Linear Unit*) (HAHNLOSER et al., 2000), apresentada na Equação 4.3 e ilustrada na Figura 4.4, tem sido uma das funções mais utilizadas no treinamento de RNAs profundas na última década. Seu valor de saída é o valor máximo entre 0 e x , assumindo zero quando x é menor que zero e x quando esse valor é maior do que zero. Foi demonstrado que essa função permite um melhor treinamento de Redes Neurais Profundas, gerando menor erro durante o treinamento quando comparada com as funções de ativação tangente hiperbólica e logística. De acordo com LeCun, Bengio e Geoffrey Hinton

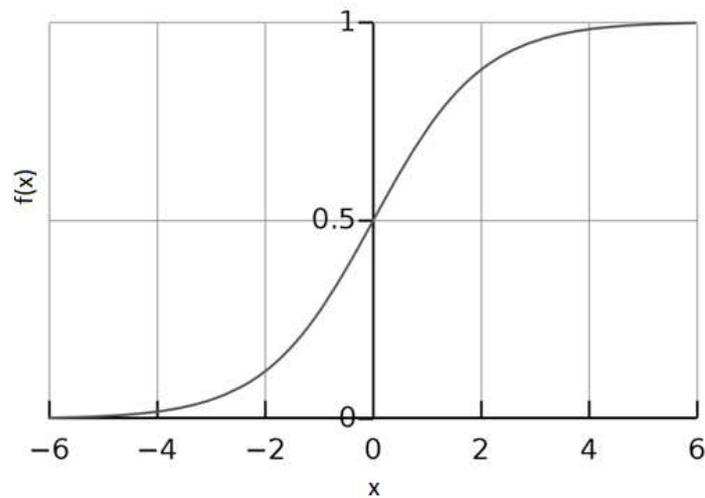


Figura 4.3: Gráfico ilustrativo da função logística. Adaptado de (HAYKIN, 2009).

(2015), essa é atualmente uma das funções de ativação mais utilizadas para treinamento de redes neurais profundas devido à sua simplicidade, cujo principal benefício é permitir acelerar o treinamento das RNAs.

$$f(x) = \max(0, x) \quad (4.3)$$

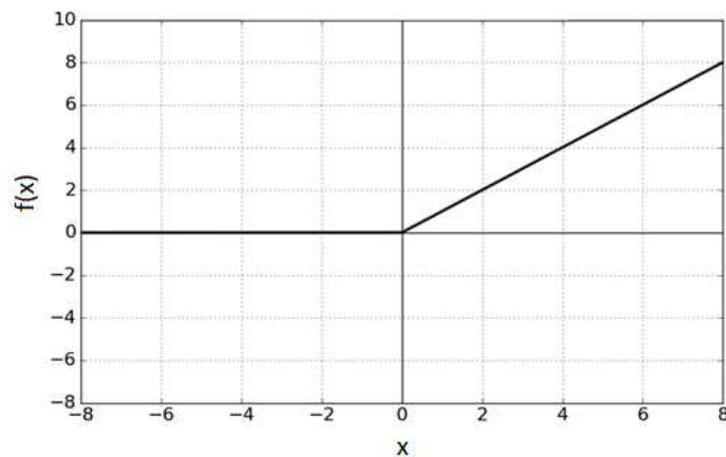


Figura 4.4: Gráfico ilustrativo da função ReLU. Adaptado de (HAHNLOSER et al., 2000).

Por fim, vale a pena complementar que a saída de um neurônio do tipo *Perceptron* pode ser usada como entrada para um neurônio seguinte, criando RNAs divididas em diferentes arranjos denominados *camadas*. Existem diversos modelos de arquiteturas de RNAs, que serão discutidos na seção a seguir.

4.2 Arquiteturas

Apesar de as primeiras tentativas de se simular uma Rede Neural Artificial de múltiplas camadas datarem a década de 60, em trabalhos como o de Ivakhnenko e Lapa (1967), apenas ao final da década de 90 e início dos anos 2000 que ocorreram avanços relevantes no uso de modelos de arquiteturas profundas. Esse avanço deve-se aos seguintes fatores. Primeiramente, o desenvolvimento das Unidades de Processamento Gráfico (do inglês, *Graphic Processing Units* - GPUs) permitiu aumentar a velocidade do processamento dos dados. Além disso, vale a pena enfatizar que existe, atualmente, uma grande quantidade de dados disponíveis, principalmente *online*, seja na forma de imagens, áudios, mapas e textos, dentre outros, que são necessários para a etapa de treinamento de uma Rede Neural Artificial.

Ainda, vale destacar que métodos de aprendizado profundo, ou seja, métodos que contenham muitas camadas de processamento ou que sejam hierárquicos, conforme apresentado por Bengio et al. (2009), são mais robustos em lidar com problemas complexos do que os métodos que utilizam modelos cuja arquitetura é rasa. Vale a pena destacar aqui que, a partir do momento que uma RNA possui mais de uma camada oculta, ela pode ser denominada como uma Rede Neural Artificial Profunda (*Deep Learning Neural Network* - DLNN) (LECUN; BENGIO; HINTON, G., 2015).

Nesse sentido, esta seção apresenta diversas arquiteturas de RNAs profundas: MLP (*Multilayer Perceptron*), RNN (*Recurrent Neural Network*), LSTM (*Long Short-Term Memory*), Bi-LSTM (*Bidirectional Long Short-Term Memory*) e CNN (*Convolutional Neural Network*).

Como observação para a leitura desta seção, vale a pena ressaltar que, durante a descrição das Redes Neurais Artificiais, algumas notações podem ser diferentes entre as diversas redes apresentadas. Isso se deve ao fato de optarmos por manter as notações das fontes em que as mesmas foram baseadas.

4.2.1 Multilayer Perceptron

As RNAs do tipo *Multi-Layer Perceptron* (MLP) são uma generalização do *Perceptron* com apenas uma camada, proposto por Rosenblatt (1958). Nessas redes, a saída de um *Perceptron* está conectada à entrada de um *Perceptron* seguinte, criando uma rede cujos neurônios estão ligados em camadas. Portanto, cria-se uma arquitetura de Rede Neural Artificial organizada em camadas, contendo pelo menos uma camada oculta entre a camada de entrada e saída.

Alguns autores também podem denominar essas estruturas como redes *feedforward* com múltiplas camadas, uma vez que a camada anterior é utilizada para fornecer dados a uma camada posterior (SILVA, I. da; SPATTI; FLAUZINO, 2010). As RNAs do tipo MLP podem ser consideradas como aproximadoras universais de funções (HAYKIN, 2009), sendo uma das estruturas de redes neurais mais versáteis (SILVA, I. da; SPATTI; FLAUZINO, 2010).

A Figura 4.5 ilustra uma MLP com a seguinte configuração: uma camada de entrada, duas camadas ocultas e uma camada de saída. A primeira camada (entrada) provê os dados à rede, enquanto que as duas camadas seguintes (ocultas) usam seus respectivos pesos sinápticos para ponderá-los, além da aplicação das respectivas funções de ativação. Finalmente, os valores resultantes são transferidos à última camada (saída), que também aplica pesos e função de ativação, resultando na saída da rede.

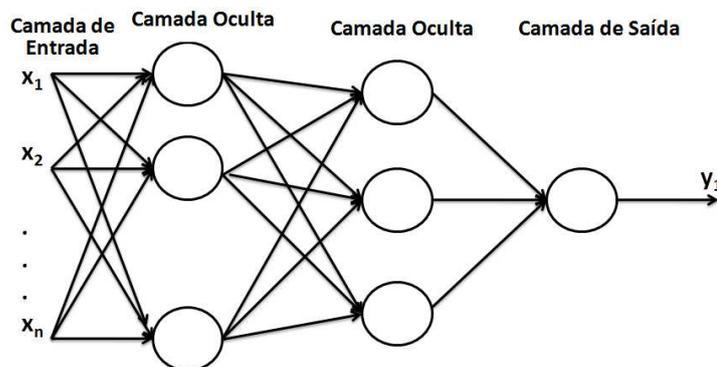


Figura 4.5: Representação de uma RNA do tipo MLP. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2019).

Existem diversos algoritmos para treinamento de redes MLP, sendo o algoritmo *Backpropagation*, introduzido por Rumelhart, Geoffrey E Hinton, Williams et al. (1986), um dos mais populares entre eles. Esse algoritmo foi um dos grandes responsáveis por popularizar o treinamento das redes MLP e pelo avanço do uso das RNAs em diferentes áreas. O funcionamento do algoritmo *Backpropagation* consiste em duas fases. Na primeira, denominada propagação adiante (*Forward*), uma amostra dos dados é fornecida à RNA, que as transfere a seus neurônios e produz uma saída. Assim, os sinais da amostra fornecida são comparados às saídas desejadas para produção do sinal de erro, fornecendo um valor representativo do erro gerado pela rede neural. Na segunda etapa, denominada retropropagação (*Backward*), os pesos dos neurônios são ajustados por meio de uma regra de correção de erro calculada por meio do gradiente da função de erro. Ou seja, essa etapa consiste, portanto, na retropropagação do erro pela rede neural por meio do seu uso como

parâmetro para ajuste dos pesos sinápticos (RUMELHART; HINTON, G. E.; WILLIAMS et al., 1986; SILVA, I. da; SPATTI; FLAUZINO, 2010). Nesse processo também é utilizado um parâmetro denominado *taxa de aprendizado* (do inglês, *learning rate*), que pode ser ajustado entre 0 e 1, cujo objetivo é garantir que a mudança dos pesos sinápticos da RNA seja proporcional à taxa fornecida. O processo geral do treinamento de uma RNA está ilustrado na Figura 4.6.

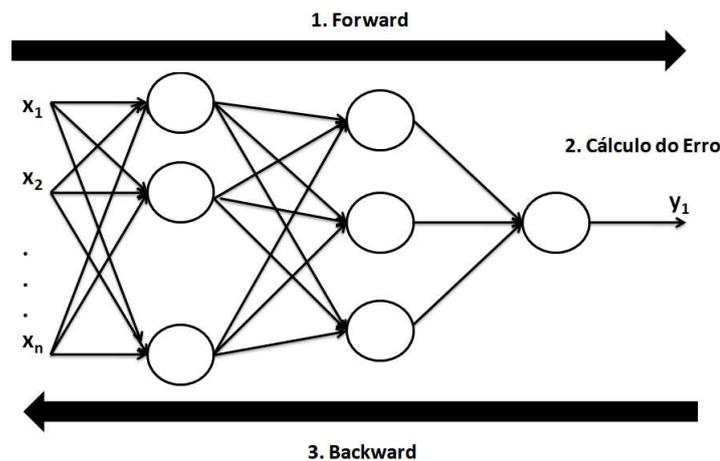


Figura 4.6: Representação do algoritmo Backpropagation.

O processo de treinamento via algoritmo *Backpropagation* se repete por um número determinado de vezes (*épocas*) ou até que não haja redução significativa do erro calculado. Nesse processo, a função de erro mais adequada depende do tipo de problema a ser tratado.

Dentre as diversas funções de erro disponíveis para problemas de regressão, pode-se utilizar o erro quadrático médio (do inglês *Mean Squared Error - MSE*), apresentado na Equação 6.3, em que Y é o valor estimado, S é o valor esperado e n é o número de amostras. Para problemas de classificação, pode-se utilizar a função de erro de entropia Cruzada (do inglês *Cross-Entropy - CE*), apresentada na Equação 4.5, em que y representa a probabilidade estimada para a classe i , y'_i representa a probabilidade de fato de um objeto pertencer à classe i e C representa o número de classes a serem analisadas.

$$MSE = \sum_{i=1}^n \frac{(Y_i - S_i)^2}{n} \quad (4.4)$$

$$CE = - \sum_i^C y'_i \log(y_i) \quad (4.5)$$

Para ocorrer o treinamento supervisionado de RNAs, ou seja, com o fornecimento de amostras que contenham as saídas desejadas, é recomendado particionar o conjunto de dados em três partes: treinamento, validação e testes (HAYKIN, 2009). Essa divisão é realizada como uma das formas de se evitar que o modelo se ajuste em demasia a um dado conjunto de amostras e não apresente bons resultados (do inglês, *overfitting*) quando aplicado a outros conjuntos de dados não conhecidos, isto é, a rede não apresenta boa capacidade de generalização. Assim, o conjunto de treinamento é utilizado para o ajuste da RNA durante o treinamento, o conjunto de validação é utilizado para avaliar, a cada época do treinamento, a capacidade da rede em generalizar os seus resultados para um conjunto de dados desconhecidos e o conjunto de testes é utilizado para analisar se a solução obtida, após todo o treinamento da rede, está de acordo para um conjunto de dados não conhecido.

Outra técnica comumente utilizada em RNAs profundas para reduzir o *overfitting* é o *dropout*, conforme apresentado em Srivastava et al. (2014). Essa técnica consiste em definir, com um dado valor de probabilidade p , pesos sinápticos nulos às saídas de alguns neurônios de uma determinada camada da RNA profunda. O restante dos neurônios são treinados normalmente pelo algoritmo *Backpropagation*. Os autores argumentam que a técnica de *dropout* faz com que os neurônios aprendam características mais robustas dos dados, melhorando a sua generalização ao conjunto de testes.

4.2.2 Redes Neurais Recorrentes

As RNAs também podem ser organizadas em arquiteturas denominadas Recorrentes (do inglês *Recurrent Neural Network* - RNN) ou redes *feedback*. Esse tipo de RNA é semelhante a uma MLP, mas a saída de alguns neurônios, que podem ser tanto de camadas ocultas como da camada de saída, pode ser utilizada para realimentar a rede neural (SILVA, I. da; SPATTI; FLAUZINO, 2010). Esse tipo de arquitetura de RNA permite o uso da noção de tempo no modelo, podendo ser aplicado a tarefas como: predição de séries temporais, processamento de linguagem natural, legenda automática de imagens, dentre outros (WELCH; RUFFING; VENAYAGAMOORTHY, 2009; LIPTON; BERKOWITZ; ELKAN, 2015).

Uma arquitetura de RNN foi introduzida por Jordan (1986). Tal arquitetura é uma RNA *feedforward* em que a realimentação da rede ocorre a partir da camada de saída para a camada oculta. Nessa rede, foi utilizada uma camada de contexto, cujas entradas não possuem pesos associados e seus neurônios não possuem função de ativação. Por outro lado,

as ligações que se relacionam às conexões da camada de contexto direcionadas para a camada oculta apresentam ponderações. Intuitivamente, essas arestas permitem o envio de informações através de múltiplos passos de tempo. Essa arquitetura é apresentada na Figura 4.7.

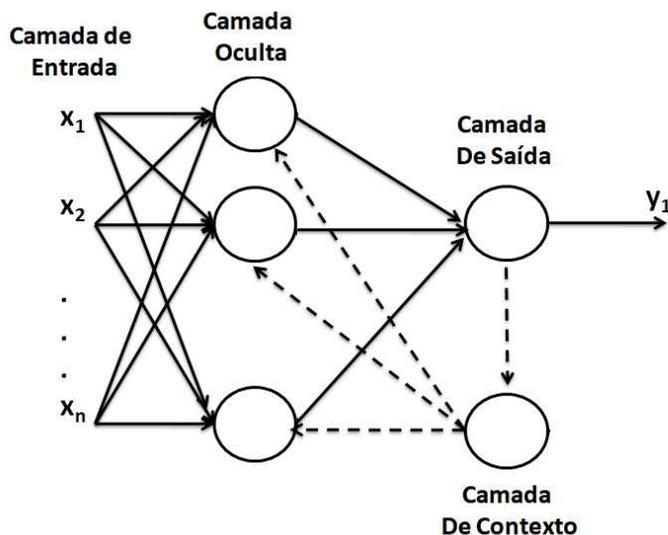


Figura 4.7: Representação de uma RNN de Jordan.

Outra arquitetura de RNN é a rede neural de *Elman*, apresentada na Figura 4.8 e introduzida por Elman (1990). Nessa arquitetura, existe uma unidade de contexto associada a cada unidade da camada oculta. Cada unidade de contexto recebe como entrada o estado do nó correspondente da camada oculta no tempo $t - 1$ e então o utiliza como entrada no mesmo nó durante o tempo t .

A Figura 4.9 ilustra uma RNN genérica ao longo de três passos de tempo. Nessa figura, x_i indica a entrada da rede no tempo i , y_i significa a saída da rede no tempo i e a_i representa a informação presente na rede neural ao longo dos passos do tempo. Ou seja, essa figura representa uma RNN que recebe, a cada passo de tempo ($i - 1$, i e $i + 1$), uma entrada x_i em conjunto com o seu estado interno anterior a_{i-1} , resultando em uma saída y_i . Nesse caso, é importante notar que, a cada passo de tempo, além da entrada, a rede recebe o seu estado interno associado ao tempo anterior, fazendo com que a rede considere não apenas a entrada no tempo atual, mas também uma informação que represente o histórico de entradas à qual foi submetida ao longo de seu treinamento.

A Figura 4.10 ilustra o funcionamento interno de um neurônio em uma RNN utilizada atualmente na literatura. Cada uma das etapas indicadas na figura pelos números de (1) a (4)

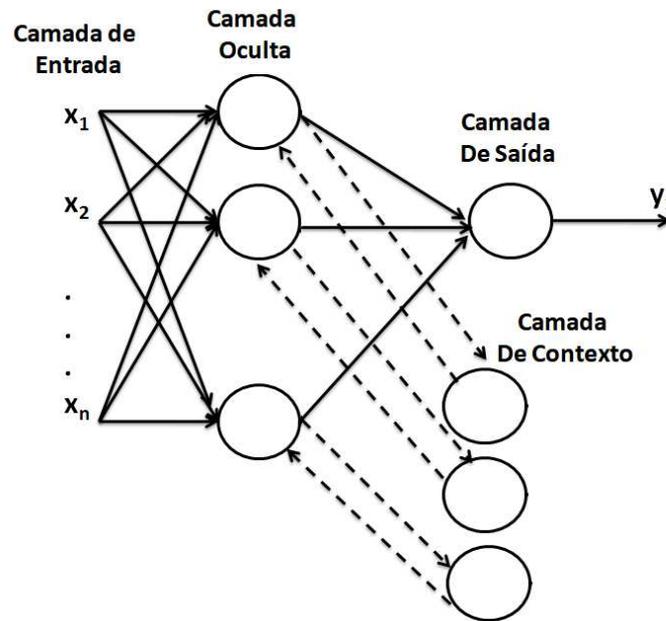


Figura 4.8: Representação de uma RNN de Elman.

podem ser descritas, respectivamente, pelas Equações 4.6 a 4.9, de acordo com Goodfellow, Bengio e Courville (2016).

A etapa (1), correspondente à Equação 4.6, recebe como entradas x^t e a^{t-1} que são, respectivamente, a entrada da RNN no tempo t e o valor do estado anterior da RNN no tempo $t - 1$. Esses valores são ponderados por seus respectivos pesos sinápticos, W_{xh} e W_{ah} , adicionados ao valor correspondente ao bias do neurônio, b_h . Essas etapas resultam no valor denominado h^t , que corresponde ao valor da operação resultante entre o estado antigo e a nova entrada do neurônio. A etapa (2), correspondente à Equação 4.7, é definida pela aplicação da função de ativação \tanh sobre o resultado obtido da etapa (1), resultando no valor indicado por a^t . Em seguida, a etapa (3), correspondente à Equação 4.8, é responsável por receber como entrada o valor interno ao neurônio, a^t , e ponderá-lo com os pesos sinápticos de saída do neurônio e seu respectivo bias, indicados por, W_{ao} e b_o , e obter o valor de saída do neurônio, indicado por o^t . Por fim, a etapa (4), correspondente à Equação 4.9, aplica a função de ativação do neurônio, indicada por softmax (HAYKIN, 2009), ao resultado obtido na etapa anterior. Esse processo gera o valor de saída y^t do neurônio. Além disso, esse processo é responsável por transferir o valor interno do neurônio, a^t , para o próximo intervalo de tempo de processamento da RNN, realizando a propagação de seu estado interno pelo tempo.

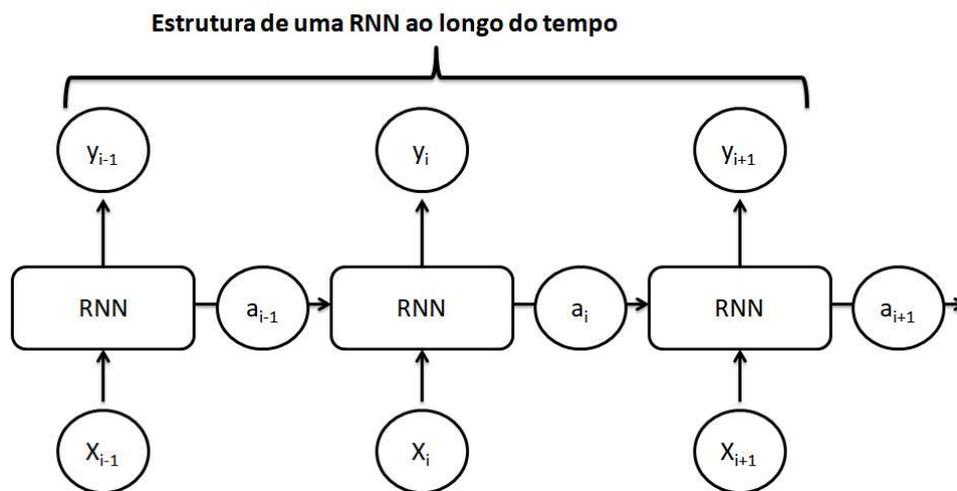


Figura 4.9: Representação geral de uma RNN. Adaptado de (LIPTON; BERKOWITZ; ELKAN, 2015).

$$h^t = (W_{xh} \cdot x^t) + (W_{ah} \cdot a^{t-1}) + b_h \quad (4.6)$$

$$a^t = \tanh(h^t) \quad (4.7)$$

$$o^t = (W_{ao} \cdot a^t) + b_o \quad (4.8)$$

$$y^t = \text{softmax}(o^t) \quad (4.9)$$

As Redes Neurais Recorrentes são comumente treinadas por meio do algoritmo *backpropagation through time* (BPTT), uma adaptação do algoritmo *Backpropagation* para lidar com a noção de tempo, e que foi introduzido por Werbos et al. (1990). Um problema comumente enfrentado pela RNN é a dificuldade em aprender, durante o treinamento, as dependências de longo prazo (BENGIO; SIMARD; FRASCONI et al., 1994; HOCHREITER; BENGIO et al., 2001), além dos problemas do desaparecimento (do inglês, *vanishing*) e explosão (do inglês, *exploding*) do gradiente. Esses problemas ocorrem quando os valores calculados para o gradiente, propagados pelo algoritmo BPTT por muitos intervalos de tempo, têm como resultados valores muito pequenos (ou muito grandes). Dessa forma,

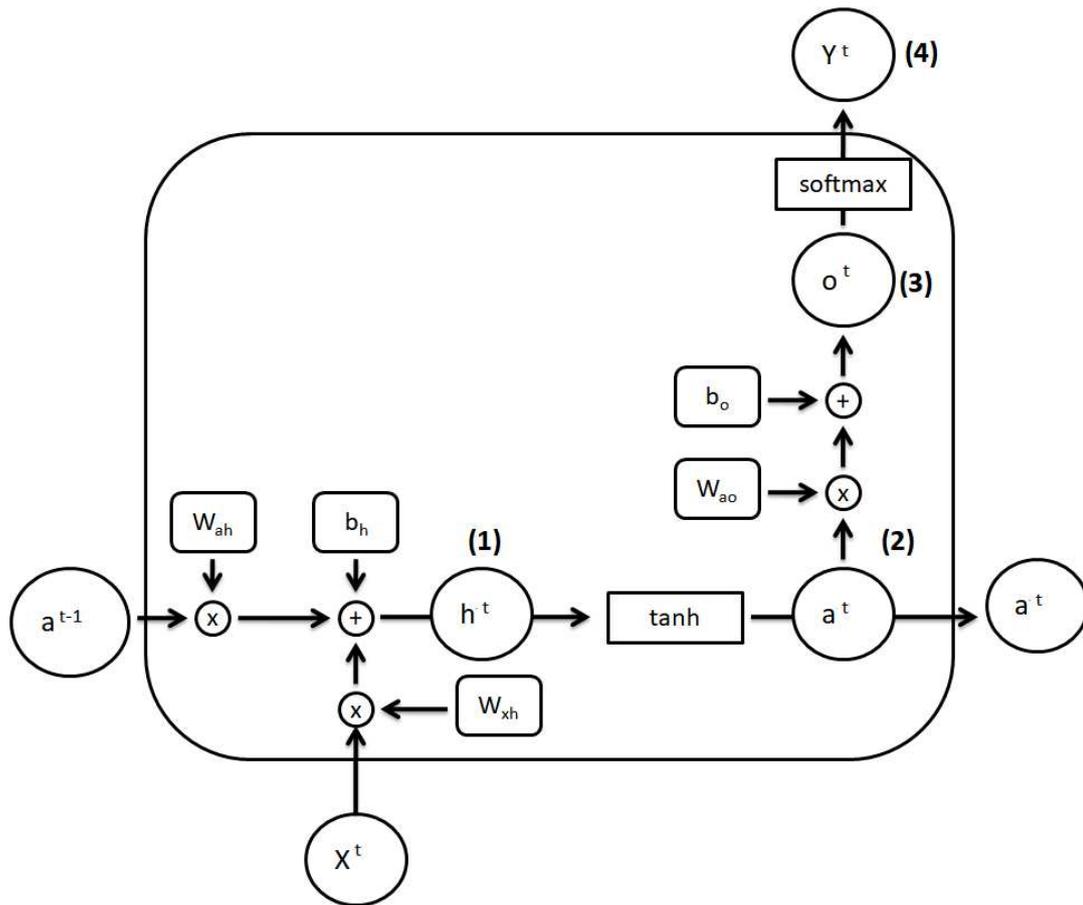


Figura 4.10: Representação de um neurônio de uma RNN.

quanto menor (ou maior) o valor do gradiente fica, mais difícil é para a RNN atualizar seus pesos sinápticos, além de dificultar cada vez mais que ela chegue ao seu resultado final.

Dentre as alternativas para contornar o problema do gradiente em RNNs, Hochreiter e Schmidhuber (1997) introduziram uma RNN denominada *Long Short-Term Memory* (LSTM), que usa neurônios desenvolvidos especialmente com o objetivo de lidar com dependências de longo prazo. Essa arquitetura tem se mostrado uma das mais eficientes em executar tarefas que exigem dados com dependência temporal (FISCHER; KRAUSS, 2018). Essa arquitetura é apresentada em mais detalhes na Seção 4.2.3.

4.2.3 Long Short-Term Memory

A arquitetura de RNN *Long Short-Term Memory*, desenvolvida para aprendizado de grandes sequências de dados, foi apresentada por Hochreiter e Schmidhuber (1997). Sua principal contribuição é introduzir uma célula de memória (do inglês, *memory cell*), que é uma unidade de computação que substitui os nós tradicionais das camadas ocultas de uma RNA. Essa

unidade permite que as redes superem as dificuldades do treinamento encontradas em RNN anteriores, como o já mencionado problema do desaparecimento do gradiente, que dificulta a convergência do treinamento da rede. Na prática, a LSTM tem mostrado uma capacidade superior de aprender dependências de longo prazo em comparação com RNNs simples.

O termo *Long Short-Term Memory* vem da seguinte intuição. RNN simples possuem memória de longo prazo na forma de pesos sinápticos que são transferidos ao longo do tempo e que mudam lentamente durante o treinamento, armazenando conhecimento geral sobre os dados. Além disso, elas também possuem memória de curto prazo, mantida na forma da transferência de dados entre curto espaço de tempo. A arquitetura LSTM introduz, nesse conceito, um tipo intermediário de armazenamento por meio da célula de memória. A célula de memória é uma unidade construída a partir de nós simples em um padrão de conexão específico, conforme ilustrado na Figura 4.11. Todos os elementos de uma célula LSTM são discutidos a seguir.

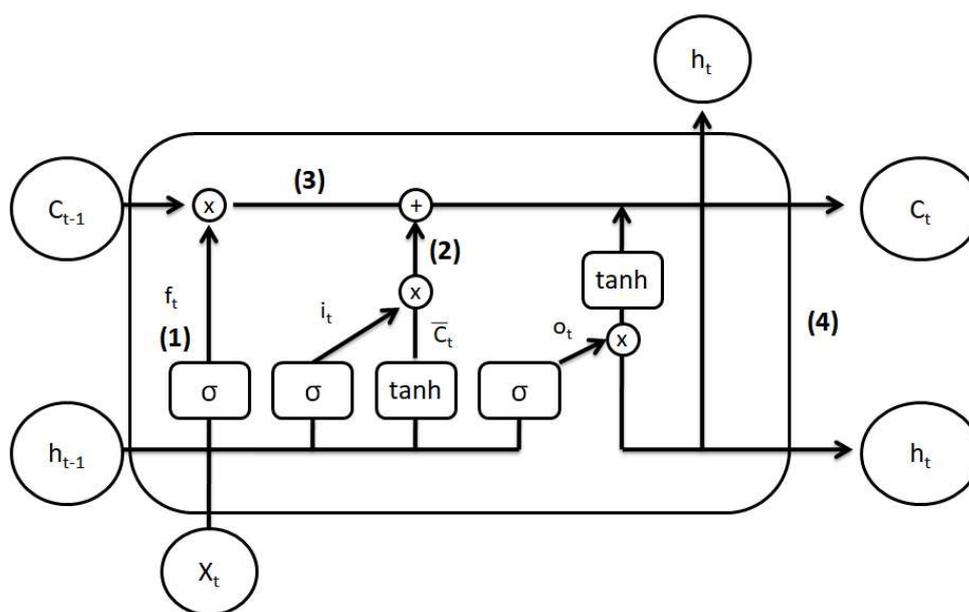


Figura 4.11: Representação de uma célula de memória de uma LSTM. Adaptado de (ZHANG, L.; WANG, S.; LIU, B., 2018).

O ponto essencial de uma célula de memória LSTM é o estado interno da célula (do inglês, *cell state*). Esse estado corresponde a um condutor de informações, denominado C_t , responsável por ligar passos de tempo adjacentes, passando informações sobre o estado interno da célula da LSTM no tempo atual t à célula seguinte, no tempo $t + 1$. Além disso, uma célula LSTM possui portas (do inglês, *gates*), que são responsáveis por definir se informações serão adicionadas ou removidas do estado interno da célula. Uma porta é uma unidade em que há uma função de

ativação que é seguida de uma operação de multiplicação entre elementos. Em conjunto, essas inovações são responsáveis por evitar o problema do desaparecimento do gradiente.

O primeiro passo em uma célula LSTM, ilustrado na Figura 4.11 (1), é relacionado à chamada porta de esquecimento (do inglês, *Forget Gate*), denominada f_t . Essas portas foram introduzidas posteriormente por Gers, Schmidhuber e Cummins (2000) à arquitetura da LSTM e não constam em seu *design* original. As portas de esquecimento são responsáveis por prover um método no qual a rede pode esquecer informações de seu estado interno, C_t , quando não são mais necessárias. Nessa etapa, ocorre a aplicação de uma função de ativação (σ) aos valores h_{t-1} e x_t , respectivamente representando a saída da célula no tempo $t - 1$ e a entrada de dados no tempo t . Seu resultado corresponde a um número entre 0, significando que a informação deve ser totalmente esquecida, e 1, significando que a informação deve ser integralmente mantida. Note que a ilustração de aplicação dos pesos sinápticos (W_f) foi omitida na figura para facilitar a visualização. A equação referente a esta etapa é apresentada a seguir.

$$f_t = \sigma(W_f \cdot h_{t-1} + W_f \cdot x_t + b_f) \quad (4.10)$$

Em seguida, é necessário calcular qual informação será adicionada ao estado interno da célula, ilustrado na Figura 4.11 (2). Em primeiro lugar, a camada denominada porta de entrada (do inglês, *Input Gate*) recebe o dado de entrada x_t juntamente com a saída da célula do instante de tempo anterior, h_{t-1} , e aplica a função de ativação. Em seguida, aplica-se uma função de ativação, \tanh , para a criação de valores candidatos \bar{C}_t que podem ser adicionados ao estado interno. As seguintes equações representam esse processamento:

$$i_t = \sigma(W_i \cdot h_{t-1} + W_i \cdot x_t + b_i) \quad (4.11)$$

$$\bar{C}_t = \tanh(W_c \cdot h_{t-1} + W_c \cdot x_t + b_c) \quad (4.12)$$

Para atualizar o estado interno da célula, C_{t-1} para o seu novo valor, C_t , ilustrado na Figura 4.11 (3), são considerados: o resultado da porta do esquecimento para decidir o que deve ou não ser retirado do estado, f_t ; o valor anterior do estado interno da célula, i_t ; e o valor de candidatos calculados no passo anterior, \bar{C}_t . A seguinte equação sintetiza essa etapa:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \bar{C}_t \quad (4.13)$$

Por fim, é calculado o valor de saída da célula de memória, h_t , ilustrado na Figura 4.11 (4), que também irá alimentar a célula de memória do passo de tempo posterior na rede. Para isso, primeiro é considerado a porta de saída (do inglês, *output gate*), o_t , que opera sobre a entrada atual, x_t , e a saída do passo de tempo anterior, h_{t-1} , aplicando-os a uma função de ativação. Em seguida, a saída, h_t , é calculada por meio da aplicação da função de ativação *tanh* ao estado interno da célula multiplicado pelo resultado do portal de saída. As equações a seguir mostram esse processo.

$$o_t = \sigma(W_o \cdot h_{t-1} + W_o \cdot x_t) + b_o \quad (4.14)$$

$$h_t = o_t * \tanh(C_t) \quad (4.15)$$

Além disso, vale destacar que existem algumas variações da RNN do tipo LSTM conforme apresentado nesta seção. Uma variação que tem apresentado bastante destaque é a denominada *Gated Recurrent Unit* (GRU), apresentada por Cho et al. (2014). Nesse modelo, as portas de esquecimento e entrada são compostas por uma única porta de atualização (do inglês, *update gate*), além de outras mudanças menores. No entanto, neste trabalho será utilizada a implementação convencional da LSTM, conforme apresentada nesta seção, adicionada da porta de esquecimento, visto que também se trata de uma das arquiteturas de RNN que compõem o estado da arte atualmente, de acordo com Kraus e Feuerriegel (2017).

Outro modelo de rede neural do tipo recorrente é a *Bidirectional Long Short-Term Memory* (Bi-LSTM) (SCHUSTER; PALIWAL, 1997), que realiza uma pequena extensão no neurônio tradicional da rede LSTM, adicionando uma característica que considera as sequências de entrada passadas e futuras, enquanto a rede LSTM original lida apenas com os dados do passado. Dessa forma, o reconhecimento de padrões em grandes sequências é apurado.

A rede neural Bi-LSTM é baseada na conexão de duas camadas para processar os dados de entrada no instante do treinamento da rede do seguinte modo: enquanto a primeira camada opera considerando as sequências de entrada anteriores (ou seja, $t - 1, t, t + 1, \dots$), a segunda camada opera considerando a direção oposta do intervalo de tempo (ou seja, $t + 1, t, t - 1, \dots$). Finalmente, os resultados de ambas as camadas são mesclados, conforme apresentado na Figura 4.12. Existem, na literatura, trabalhos que mostram a eficácia do Bi-LSTM sobre a rede LSTM tradicional (ALTHELAYA; EL-ALFY; MOHAMMED, 2018; GRAVES; SCHMIDHUBER, 2005). Desse modo, a rede neural do tipo Bi-LSTM também será considerada neste trabalho.

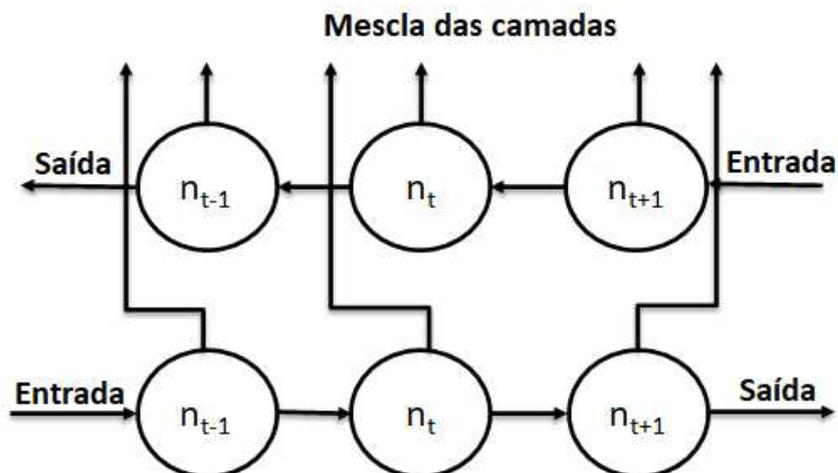


Figura 4.12: Representação das camadas de um modelo Bi-LSTM. Adaptado de (ALTHELAYA; EL-ALFY; MOHAMMED, 2018).

4.2.4 Convolutional Neural Networks

Redes Neurais Convolucionais (do inglês, *Convolutional Neural Networks* - CNNs) foram desenvolvidas com base no trabalho de inspiração neurobiológica proposto por Hubel e Wiesel (1962) e Hubel e Wiesel (1977), em que os autores mostraram que o córtex visual de gatos se organiza de modo hierárquico, de acordo com células sensíveis a pequenas sub-regiões chamadas campos receptivos. Assim, existe uma estrutura de arranjo entre células simples e complexas, sendo as primeiras responsáveis por reconhecer padrões simples e as últimas ativadas por padrões complexos. Em conjunto, ambas produzem a percepção visual. Dessa forma, a combinação entre células simples e compostas inspiraram o desenvolvimento dos trabalhos com RNAs do tipo CNN.

As CNNs foram primeiramente desenvolvidas por Fukushima (1975), na época denominadas como *Neocognitron*. Assim como as CNNs atuais, eram feitas sobre uma base hierárquica possuindo várias camadas, que permitem ao computador ser treinado por meio de padrões visuais. Além disso, também possuíam várias camadas de convolução e camadas de subamostragem. Por outro lado, diferentemente das versões de CNNs atuais, que comumente utilizam o algoritmo *Backpropagation*, seu treinamento era realizado por meio de técnicas de aprendizado não supervisionado (FUKUSHIMA, 2013). Adiante, LeCun, Boser et al. (1989) desenvolveram uma rede neural com várias camadas, conforme proposta por Fukushima (1975), incrementada com o algoritmo de treinamento *Backpropagation*. Já em LeCun, Bottou et al. (1998), houve uma revisão dos métodos apresentados no trabalho anterior aplicado ao reconhecimento de caracteres manuscritos por meio de redes CNNs com

várias camadas. Os resultados desse trabalho foram promissores, ao passo que é considerado um dos trabalhos mais significativos da área, apesar das limitações da disponibilidade de recursos de computação na época.

Devido às suas características, CNNs são empregadas com sucesso atualmente no campo da visão computacional, em que é necessário o uso de dados com características espaciais (LECUN; BENGIO; HINTON, G., 2015). Além disso, existem também, na literatura, trabalhos que empregam CNNs no campo do Processamento de Linguagem Natural, como em Kim (2014) e Johnson e Tong Zhang (2015), em que foram utilizadas para classificação de sentenças.

Resumidamente, uma rede neural do tipo CNN pode ser vista como uma rede *feedforward* que recebe entradas de dados que são passadas por uma série de camadas convolucionais, de subamostragem (do inglês, *pooling*) e funções de ativação, que por fim alimentam camadas totalmente conectadas, similares a uma MLP, para então produzir uma saída. A saída pode ser uma classe ou uma probabilidade de pertencimento a uma determinada classe. A Figura 4.13 apresenta uma arquitetura genérica de uma rede CNN com apenas uma camada de convolução e *pooling*, sendo que estas, em muitos casos, podem aparecer em maior número. A camada de convolução gera mapas de características, que são utilizados para alimentar uma camada de *pooling*, que por sua vez alimenta uma camada totalmente conectada.

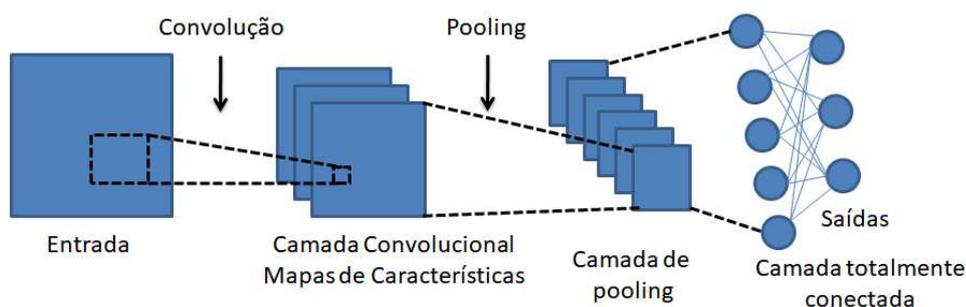


Figura 4.13: Representação de uma rede do tipo CNN. Adaptado de (FARIA, 2018).

A seguir são detalhados cada um dos principais componentes de uma rede do tipo CNN: camadas convolucionais, camadas de *pooling* e a camada totalmente conectada. Além disso, serão abordados hiperparâmetros comuns de redes CNNs, como *stride* e *padding*.

Em primeiro lugar, a camada convolutiva (do inglês, *convolution layer*), de um modo geral, serve para extração de características de seus dados de entrada por meio da aplicação de um filtro de aprendizado. Ou seja, um mapa de características é obtido por meio da operação de convolução da matriz de entrada por um filtro linear. Esse processo é apresentado na Figura 4.14, em que uma matriz de entrada é passada por um filtro de detecção de bordas em imagens

(PARKER, 2010). Na figura, a operação de convolução é realizada por toda a matriz de entrada, com o deslocamento de um pixel por vez. O tamanho do deslocamento é um hiperparâmetro de redes CNN, denominado passo (do inglês, *striding*). Em aplicações de computação visual, normalmente seu valor é igual a 1, ou seja, o filtro é deslocado pela imagem um pixel por vez. Assim, para cada janela correspondente ao filtro que desliza um determinado passo sobre a matriz, ocorre o produto entre cada elemento do filtro e o elemento da entrada correspondente. Por fim, os resultados dessa operação são somados para obter a saída daquele local. Esse processamento resulta no mapa de características (FARIA, 2018).

Vale destacar, no entanto, que em vez de definir os filtros que serão aplicados a cada camada, precisamos apenas treinar os filtros nas CNNs. Dessa forma, as CNNs são independentes do conhecimento prévio e da interferência humana na extração de características (LIU, W. et al., 2017).

Outro hiperparâmetro necessário nessa operação é o *stride*, que define como o filtro deve ser aplicado a elementos da borda da matriz. Uma primeira abordagem nesse caso é o *padding*, que preenche com zeros as bordas da matriz de entrada, resultando em um processo denominado *wide convolution*, permitindo aplicar o filtro a todos os elementos da matriz de entrada. Quando não há uso de *padding*, e, portanto, o filtro não pode ser aplicado a todos elementos da matriz de entrada, o processo é denominado *narrow convolution*.

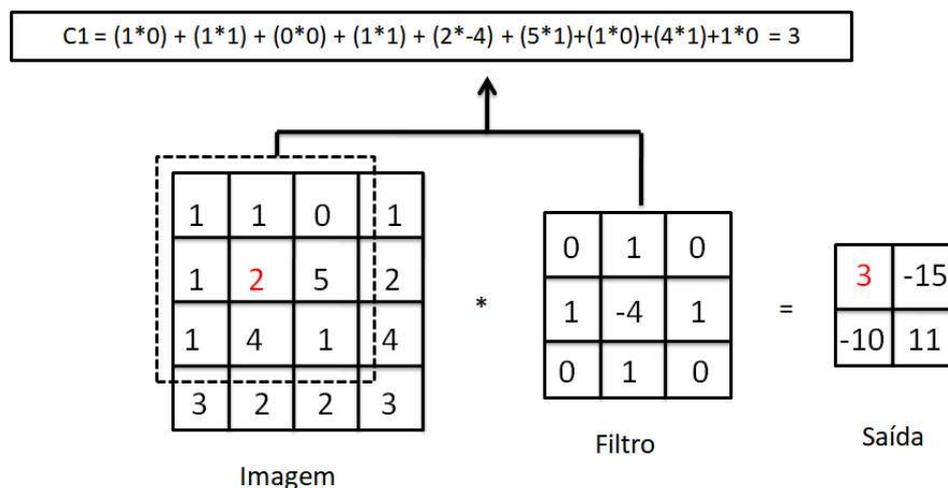


Figura 4.14: Operação de convolução.

Assim que o processo de convolução é encerrado, seu resultado é utilizado como entrada para uma função de ativação não-linear, que pode ser *sigmoid*, *tanh* ou mesmo *ReLU*.

De acordo com Goodfellow, Bengio e Courville (2016), devido às características de suas camadas convolucionais, as CNNs normalmente armazenam menos parâmetros quando comparadas com RNAs tradicionais, reduzindo requisitos de memória e processamento.

Além disso, as CNNs são compostas por um tipo de camada denominada *pooling*, aplicadas imediatamente após às camadas convolucionais, conforme apresentado na Figura 4.15. As camadas de *pooling* têm como objetivo encontrar a informação mais importante obtida nas camadas convolucionais por meio da redução de dimensionalidade do mapa de características. Existem diferentes operações de *pooling*, destacando-se como *max* aquela que determina o valor máximo em uma dada região e como a *média* o cálculo da média dos valores de uma determinada região (SZE et al., 2017). Assim, conforme apresentado na Figura 4.15, é possível verificar que essa operação reduz a dimensionalidade da matriz (no caso, de 4x4 para 2x2), fazendo com que apenas as características mais significativas fiquem armazenadas. Como consequência, há também redução de parâmetros para as camadas posteriores da rede.

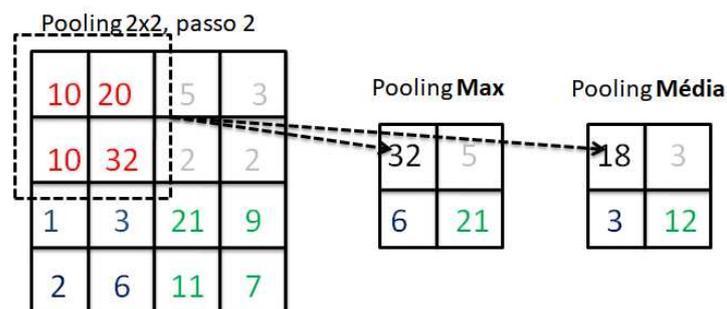


Figura 4.15: Operação de *pooling*. Aplicação das operações de max e média. Adaptado de (SZE et al., 2017).

Por fim, uma CNN apresenta uma ou mais camadas totalmente conectadas (do inglês, *fully connected*), ilustrada à direita na Figura 4.13, com o objetivo de usar o mapa de características de mais alto nível na hierarquia da rede, gerado pelas camadas anteriores de convolução e *pooling*, para classificar os dados de entrada com base em um conjunto de treinamento. Como entrada nessa camada, deve ocorrer a operação de *flatten*, em que os mapas de características são transpostos da forma matricial para um formato vetorial de apenas uma dimensão (por exemplo, mapas de características de dimensão 10x4x4 seriam mapeados para serem usados como entradas em 160 neurônios). Nessa camada, todos os neurônios da camada anterior são conectados a todos neurônios da camada posterior até a camada de saída, em que cada neurônio de saída corresponde a uma classe do modelo. Vale destacar que essa camada é muito similar a uma rede MLP.

A função usada como saída da camada totalmente conectada é denominada *softmax*, em que Y corresponde ao vetor de saída do classificador. Essa função (definida pela Equação 4.16), recebe como entrada z , relacionada aos pesos das conexões sinápticas da camada anterior à camada de saída, e retorna um vetor de probabilidades que indicam a chance do dado de entrada pertencer a uma classe i .

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{j \in Y} e^{z_j}} \quad (4.16)$$

O treinamento de redes CNN é realizado por meio do algoritmo *Backpropagation*, conforme apresentado na Seção 4.2.1, cujo objetivo é minimizar a diferença entre a saída da rede e a saída esperada até que um determinado critério de parada seja satisfeito.

4.3 Considerações Finais

Neste capítulo foram introduzidos os conceitos principais sobre RNAs, bem como algumas de suas diversas arquiteturas, treinamento e principais aplicações. Vale a pena destacar que as RNAs são apresentadas na literatura como uma das mais promissoras técnicas para modelar o mercado de ações, sendo facilmente adaptável às mudanças do mercado (GURESEN; KAYAKUTLU; DAIM, 2011). A literatura mostra que o uso de RNA, como método de predição de séries temporais, apresenta resultados interessantes, sendo uma alternativa aos métodos tradicionais de previsão (ZHANG, G.; PATUWO; HU, 1998).

Conforme apresentado por Ozbayoglu, Gudelek e Sezer (2020), Deepak Kumar, Sarangi e Verma (2021) e Kumbure et al. (2022), há uma predominância de RNNs, principalmente LSTMs, em trabalhos relacionados à previsão do mercado de ações. Essas redes apresentam os resultados mais notáveis em publicações da área. Na verdade, a natureza dos dados de ações é ideal para as RNNs, pois apresentam dependência temporal. No entanto, outros modelos também são considerados nos trabalhos desta área, entre eles, a MLP. Por fim, a CNN somente recentemente tem recebido mais atenção na área financeira, sendo utilizada tanto como modelo para classificação quanto para regressão. No entanto, sua desvantagem é que o modelo requer uma transformação de dados para uma visualização similar a uma imagem e, portanto, ainda é menos utilizada na literatura da área.

Capítulo 5

Trabalhos Correlatos

Este capítulo apresenta os trabalhos correlatos a esta tese. Visto que o objetivo geral deste trabalho é integrar Análise de Sentimentos aplicada a notícias, séries históricas de ações e indicadores técnicos na previsão do mercado financeiro com o uso de Redes Neurais Artificiais, este capítulo está dividido em três seções. A Seção 5.1 apresenta os trabalhos que utilizam principalmente dados textuais como entrada para preditores e analisa seus respectivos impactos no mercado acionário. A Seção 5.2 apresenta os trabalhos cujo enfoque é voltado para o uso de séries históricas e indicadores técnicos como entradas principais para modelos de previsão de séries temporais relacionadas ao mercado financeiro. Por fim, a Seção 5.3 apresenta os trabalhos com todas as três fontes de dados citadas anteriormente, a saber: dados textuais, indicadores técnicos e série histórica de ações, como entrada para modelos de previsão.

5.1 Dados textuais para Previsão do Mercado Financeiro

É possível notar que existe um grande interesse na literatura na previsão do mercado financeiro por meio de AS, seja como método principal de previsão ou como método auxiliar em um Sistema de Suporte à Decisão. Essa revisão de diversos artigos da área mostra como, recentemente, tem havido um interesse crescente em analisar o sentimento dos investidores em relação ao mercado de ações (MA et al., 2021; KUMBURE et al., 2022). Ela ainda aponta que os principais métodos dos trabalhos da área têm se concentrado em: métodos baseados em léxicos e métodos baseados em Aprendizado de Máquina.

Um dos trabalhos pioneiros da área foi desenvolvido por Bollen, Mao e Zeng (2011), cujo foco foi calcular o sentimento presente em publicações no Twitter e usá-lo para prever o mercado de ações. Assim, duas ferramentas foram utilizadas com o intuito de analisar os sentimentos: *OpinionFinder*, que classifica o sentimento de um determinado documento como positivo e negativo; e *Google-Profile of Mood States* (GPOMS) que determina o sentimento em seis categorias. Foram utilizados como fontes de dados *Tweets* considerando o período de 28 de fevereiro a 19 de dezembro de 2008. Os resultados apresentaram precisão de 86,7% na previsão do movimento do mercado para o índice Dow Jones Industrial Average (DJIA).

O trabalho de Schumaker et al. (2012) apresentou um estudo da relação presente entre notícias e os movimentos do mercado de ações. Assim, os autores propuseram uma combinação entre o sistema *Arizona Financial Text System* e a biblioteca *OpinionFinder*. Como fontes de dados, os autores utilizaram notícias de empresas listadas no índice S&P 500 considerando o período entre 26 de outubro de 2005 e 28 de novembro de 2005. Como resultados, o trabalho mostrou que notícias podem levar a um lucro de 3,3% ao mesmo tempo em que são capazes de prever o movimento do mercado em até 59% das vezes.

O trabalho descrito por Arias, Arratia e Xuriguera (2014) investigou se um indicador de sentimento público extraído de mensagens diárias do Twitter pode melhorar a previsão de indicadores sociais, econômicos ou comerciais. Como estrutura, utilizou uma *summary-tree*, baseada em árvores de decisão. Os dados utilizados foram obtidos do Twitter, de março de 2011 a 2013, e pertencem a dois domínios distintos: mercado de ações e receita de bilheteria. Os resultados mostraram que modelos não-lineares aproveitam os dados do Twitter ao prever tendências financeiras, enquanto os lineares falham sistematicamente ao prever qualquer tipo de série temporal financeira. Em relação à previsão de tendências de receitas de bilheteria, o trabalho mostrou que *Support Vector Machines* (SVM) é a técnica que apresentou os melhores resultados.

Makrehchi, Sameena Shah e Liao (2013) apresentaram um trabalho cujo objetivo é realizar previsões sobre o próximo movimento do mercado de ações considerando o sentimento agregado de *Tweets* diários. Como técnica de AS, os autores utilizaram a abordagem lexical considerando palavras ou frases específicas. O índice S&P 500 foi utilizado como principal fonte de dados, considerando o período entre março e junho de 2012. Como resultados, o estudo mostrou que o sentimento predominante diário se correlaciona com o

movimento do mercado. Além disso, os autores desenvolveram uma estratégia de investimento no mercado de ações que foi capaz de obter até 20% de lucro.

Yoshihara et al. (2014) mostraram uma estratégia de previsão do mercado de ações com o uso de Análise de Sentimentos feito com as RNAs profundas recorrentes. Como fontes de dados, foram selecionados artigos de jornais relacionadas a dez empresas do índice Nikkei, considerando o período que compreende de 1999 a 2008. Os resultados mostraram que a metodologia do trabalho apresenta redução da taxa de erro de previsão para 40,05% em média, quando comparada com as abordagens *baselines SVM* e *Deep Belief Networks*.

Qing Li et al. (2014) desenvolveram uma estratégia de negociação quantitativa por meio do impacto da mídia nos mercados de ações. Como metodologia, o trabalho extraiu substantivos com um etiquetador *Part of Speech* e palavras que indicam sentimento para representar uma notícia como um vetor de termos. Além disso, para realizar a AS, foram utilizadas palavras previamente rotuladas como positivas ou negativas, incorporando informações do mercado de ações para enriquecer as associações de palavras. Foram utilizados 72 *websites* financeiros da China, sites de discussão e de dados financeiros. O trabalho mostrou que a informação principal dos artigos de notícias sobre uma determinada empresa pode enriquecer o conhecimento dos investidores e afetar suas atividades de negociação. Além disso, verificou-se que os sentimentos públicos causam flutuações emocionais nos investidores e intervêm em sua tomada de decisão. Por fim, o trabalho indicou que o impacto da mídia nas empresas varia de acordo com características da empresa e conteúdo da notícia.

Nofer e Hinz (2015) mediram o sentimento em publicações no Twitter para correlacioná-lo com a previsão do mercado financeiro. Para isso, usaram a versão alemã do *Profile of Mood States* (POMS), que consiste em 19 adjetivos que pertencem a 5 dimensões do humor: tristeza, desesperança, cansaço, raiva e humor positivo. A fonte de dados consistiu de *Tweets* da Alemanha que abrangem o período de 01 de janeiro de 2011 e 17 de março de 2012. O trabalho mostrou que a correlação entre *tweets* e mercado financeiro existe ao se considerar o número de seguidores do Twitter em cada publicação. Com a abordagem proposta, a carteira de ações obteve até 36% de retorno dentro de um período de seis meses no mercado financeiro alemão.

Dois trabalhos mostraram o uso de RNAs profundas e representação de palavras (ie., *word embeddings*). O trabalho proposto por Ding et al. (2015), fez uso dessas técnicas na previsão

do retorno diário do índice S&P 500 com resultados de 55,21% de acurácia. Do mesmo modo, o trabalho proposto por Peng e Jiang (2016) utilizou ambas as técnicas para previsão do movimento do preço de ações. Os experimentos foram realizados considerando uma janela de 1 a 5 dias. Dentre os resultados reportados pelos autores, os de melhor desempenho obtiveram acurácia de 52,44% para ações não mencionadas nas notícias e previstas através de um gráfico de correlações.

Zhao et al. (2016) apresentam uma técnica de previsão de movimento do mercado financeiro por meio de AS da rede social Sina Weibo. A técnica empregada para AS foi análise lexical por meio de seleção de características e palavras em conjunto com SVM. Além dos dados textuais, o trabalho utilizou também dados de séries históricas para previsão do valor de mercado do Índice Composto de Shanghai, no período de setembro a dezembro de 2015. A AS apresentou, na média, de 62% a 68% de precisão, enquanto que a previsão de mercado apresentou de 53% a 60% de precisão na média.

Feuerriegel e Prendinger (2016) desenvolveram estratégias de negociação que utilizam notícias textuais para tomada de decisão com base em novas informações que entram no mercado. Para isso, foram utilizados aprendizado supervisionado e aprendizado por reforço. Para a análise das notícias, foi proposta uma abordagem baseada em regras e dicionário. Como fonte de dados, foram utilizados anúncios regulamentados de empresas em língua inglesa no período de janeiro de 2004 até o final de junho de 2011, além de dados do índice financeiro CDAX. Como resultado, o trabalho apresentou um sistema de suporte à decisão que demonstra como os dados baseados em notícias podem ser incorporados a um sistema de investimento.

Em Yan et al. (2016), o trabalho procurou estabelecer uma relação entre o mercado de ações chinês e a rede social chinesa Microblog. O trabalho utilizou duas técnicas de AS voltadas à língua chinesa: *ROST Content Mining* e *C-POMS*. Além disso, o trabalho utilizou *Probabilistic Neural Network* para previsão do mercado financeiro. Os dados textuais utilizados foram provenientes do microblog chinês no período de março a junho de 2014. Como resultados, a técnica de SVM se mostrou melhor na previsão do mercado financeiro chinês quando comparada com a técnica *Probabilistic Neural Network*.

Kraus e Feuerriegel (2017) compararam o uso de Redes Neurais Artificiais profundas do tipo LSTM e RNN, aplicadas a divulgações financeiras, com abordagens tradicionais de aprendizado de máquina que utilizam *Bag-of-Words*, além de explorar o uso de aprendizado

por transferência, no qual uma rede neural foi pré-treinada em um corpus de contexto diferente. Como fonte de dados, o trabalho usou 13.135 anúncios regulamentados de empresas alemãs. Os resultados revelam uma maior precisão na previsão do movimento do mercado de ações utilizando técnicas de Aprendizado Profundo em comparação com o aprendizado de máquina tradicional com *Bag-of-Words*. Como destaque, o trabalho mostrou que uma Rede Neural Artificial do tipo LSTM produz uma melhoria de 6,8% nos resultados, enquanto que uma Rede Neural Artificial do tipo LSTM com transferência de aprendizado produz uma melhoria de 7,1%.

Vargas, B. S. Lima e Evsukoff (2017) usaram Aprendizado Profundo para previsão de índice S&P 500 com o uso de títulos de notícias e indicadores técnicos. A tarefa de Análise de Sentimentos foi realizada com uma composição de RNAs do tipo CNN e RNN. Como fontes de dados, foram utilizadas notícias do site *Reuters*, considerando o período que compreende de outubro de 2006 a novembro de 2013. Os resultados mostraram que a RNA do tipo CNN supera os resultados da RNA do tipo RNN em relação à semântica dos textos; e que a RNA do tipo RNN superou a concorrente CNN com relação às características temporais da previsão de mercado.

O trabalho de Huynh, Dang e Duong (2017) foi desenvolvido para previsão do índice S&P 500. Nele, os autores introduziram um modelo de previsão chamado *Bidirectional Gated Recurrent Unit* (BGRU). Para analisar o impacto de notícias financeiras sobre o preço das ações, os autores consideraram diferentes intervalos de tempo (1, 2, 5, 7 e 10 dias). Os resultados do modelo proposto atingiram quase 60% de acurácia na previsão do índice S&P 500 e 65% na previsão de ações das empresas Google, Wal-Mart e Boeing.

O trabalho descrito por Hájek (2018) apresentou o uso da técnica de Análise de Sentimentos aplicada a relatórios de empresas para prever retornos anormais de ações, isto é, retornos superiores a retornos obtidos em uma determinada carteira. Como técnicas utilizadas, a etapa de AS fez uso de dicionários voltados ao domínio financeiro e a previsão do mercado foi realizada com uma RNA *Multilayer Perceptron*. As fontes de dados utilizadas consideraram notícias de empresas norte-americanas listadas na Bolsa de Valores de Nova York ou na Nasdaq. Como resultados, os autores mostraram que a abordagem proposta obteve resultados similares ao *baseline Naive Bayes*, mas que os demais métodos de Aprendizado de Máquina testados como *baselines* foram superados.

Além disso, George Guan-Ru Wu, Hou e Lin (2019) analisaram os impactos das notícias na previsão do mercado financeiro de Taiwan. A AS do trabalho foi desenvolvida por meio de mineração de dados em notícias de economia, convertendo-as em indicadores numéricos e unindo-as a indicadores de macroeconomia para auxiliar na previsão do mercado por meio de modelos de regressão. A conversão das notícias para indicadores numéricos foi realizada por meio de uma contagem de frequência de palavras a partir de um conjunto de palavras-chave, rotuladas de acordo com sua polaridade (positivo, negativo e neutro). O trabalho utilizou dados do mercado financeiro de Taiwan, de janeiro de 2008 a dezembro de 2014. Os resultados do trabalho mostraram que o modelo contendo notícias e indicadores financeiros apresentou melhores resultados do que o modelo que continha apenas indicadores macroeconômicos.

O trabalho apresentado por Eck et al. (2021) apresenta uma metodologia para prever flutuações de preços no mercado de ações alemão com dados de informações de artigos financeiros. Os artigos são convertidos em uma matriz com base nas ocorrências de palavras nas notícias, que serve como entrada para algoritmos de classificação com o objetivo de prever a direção provável do impacto no mercado. Os resultados mostraram que o algoritmo SVM alcançou alta precisão na previsão do mercado de ações, com acurácia de até 70%.

O trabalho de Valle-Cruz et al. (2022) mostra uma análise de dados do Twitter e índices financeiros mundiais para analisar como a polaridade gerada pelas postagens do Twitter influenciou o comportamento dos índices financeiros durante as pandemias de COVID-19 e H1N1. Para isso, os autores utilizaram uma abordagem baseada em léxico em *tweets* de contas financeiras do Twitter e calcularam as correlações entre as polaridades dos indicadores do mercado financeiro e as postagens na rede social. Os resultados identificaram correlações entre dias próximos a postagens do Twitter e as reações do mercado. O método que levou aos melhores resultados para detectar as correlações foi o léxico SenticNet.

É importante destacar que existem poucos trabalhos na literatura que consideram Análise de Sentimentos em português e sua relação com o mercado de ações do Brasil. Dentre eles, o trabalho de H. S. Santos, Laender e A. C. Pereira (2015) estuda a presença de relação entre postagens em redes sociais e o mercado de ações brasileiro. Verificou-se que as notícias geram uma alta em publicações no Twitter e, em seguida, há uma certa influência das mesmas no próximo dia de negociações no mercado. Entretanto, esse estudo não apresentou foco em Análise de Sentimentos e sua relação com o mercado. O trabalho de Milson L Lima et al. (2016)

mostra o uso da técnica SVM para realizar Análise de Sentimentos em *tweets* sobre a empresa Petrobras. No entanto, o estudo não investiga a relação entre o sentimento predominante calculado sobre a empresa e o seu movimento no mercado de ações. Ainda, em Araújo Júnior e Marinho (2018), há o emprego de técnicas de AS em notícias dos principais jornais *online* brasileiros para identificação de tendências no mercado brasileiro. No entanto, os autores empregaram um método automático de tradução para o inglês e não demonstraram a criação de um classificador de sentimentos, se restringindo a utilizar classificadores já implementados.

Em geral, verificamos que existem poucos métodos para AS para o idioma português com o uso de técnicas de Aprendizado de Máquina, como é o caso de Redes Neurais Artificiais (PEREIRA, D. A., 2021). Ainda, os métodos existentes para o idioma inglês produzem resultados inferiores quando aplicados a conteúdos na língua portuguesa devido a características da língua portuguesa tais como: verbos irregulares, expressões típicas do idioma, duplos sentidos, metáforas, dentre outros. Além disso, o processo de tradução automática de um idioma para outro é uma atividade sujeita a erros (MARTINS; PEREIRA, A.; BENEVENUTO, 2015). Nesse sentido, este trabalho pretende utilizar a abordagem de Aprendizado de Máquina, em especial com o uso de diferentes arquiteturas de RNA, para contornar as dificuldades apresentadas na AS da linguagem portuguesa.

5.2 Valores de Ações e Indicadores Técnicos para Previsão do Mercado Financeiro

A aplicação de técnicas de RNAs em séries temporais ou indicadores do mercado acionário é notável. A seguir, é apresentada uma lista de trabalhos que fazem uso dessas técnicas aplicadas à previsão do mercado financeiro.

O trabalho apresentado por An-Sing Chen, Leung e Daouk (2003) teve como objetivo prever a direção do índice de mercado *Taiwan Stock Index*. Nesse trabalho, os autores fizeram uso da rede *Probabilistic Neural Network* (PNN) para prever a direção de movimento do índice analisado. Como fontes de dados, foram usados dados históricos do *Taiwan Stock Index*, de janeiro de 1982 a agosto de 1992. Como resultados, os autores apontaram que o uso de PNN mostra retorno superior aos métodos de aprendizado com filtro de Kalman e o modelo *Random Walk*.

Por outro lado, Guresen, Kayakutlu e Daim (2011) avaliaram a eficácia dos seguintes modelos de redes neurais nas previsões do mercado de ações: MLP, *Dynamic Artificial Neural Network* (DAN2), além de redes neurais híbridas. Como fontes de dados, foi analisado o índice NASDAQ, com dados de 7 de outubro de 2008 até 26 de junho de 2009. Os resultados apontaram que a arquitetura DAN2 se comporta como um método estatístico em vez de uma Rede Neural Artificial. Ainda, os resultados gerais mostram que o modelo clássico de RNA MLP traz resultados mais confiáveis na previsão de séries temporais quando comparados aos demais. Ainda, os métodos híbridos não conseguiram melhorar os resultados da previsão.

Kara, Boyacioglu e Baykan (2011) desenvolveram um trabalho para prever a direção do movimento na Bolsa de Valores de Istambul (ISE) com o uso de RNAs e *Support Vector Machines* (SVM). A arquitetura de Redes Neurais Artificiais escolhida foi do tipo MLP profunda com 3 camadas. A configuração apresentada pelos autores foi a seguinte: para a camada de entrada, foram utilizados 10 nós e 1 nó para a camada de saída, que prevê o sentido do movimento do mercado financeiro. Os dados do experimento compreenderam séries históricas do *Istanbul Stock Exchange* (ISE) National 100 Index de 1997 a 2007. Por fim, os resultados experimentais mostraram que o desempenho médio do modelo de RNA (75,74%) foi melhor que o do modelo SVM (71,52%).

O trabalho de Fagner A. de Oliveira, Nobre e Zárate (2013) buscou compreender as informações disponíveis no mercado financeiro e identificar as variáveis chave que impulsionam os preços das ações. Além disso, o trabalho realizou um estudo de caso com a aplicação de RNA para previsão das ações da empresa brasileira Petrobrás. A arquitetura de Rede Neural Artificial escolhida foi uma MLP com 3 camadas. Além disso, a fonte de dados conteve observações de séries temporais de janeiro de 2000 a dezembro de 2011 da Bovespa para ações da Petrobrás. Como resultados, o trabalho mostrou que a taxa de previsão de direção do mercado foi de 93,62%.

Oriani e Coelho (2016) analisaram o impacto de diferentes indicadores técnicos na previsão dos preços de fechamento de ações de cinco empresas listadas no Ibovespa. Para isso, os autores utilizaram RNAs do tipo MLP. Os resultados mostraram que a combinação de diferentes indicadores técnicos pode melhorar o desempenho da previsão do mercado de ações e que os indicadores Média Móvel Exponencial e Média Móvel Ponderada melhoram a precisão da previsão quando comparadas às previsões feitas apenas com a série histórica de preços.

Nelson, Adriano CM Pereira e Renato A de Oliveira (2017) utilizaram RNAs do tipo LSTM para prever a direção dos movimentos do mercado financeiro. Como fontes de dados, foram utilizadas séries temporais de dados históricos de preços de diferentes ações de bolsa de valores brasileira (Bovespa) de 2008 a 2015. Os resultados mostraram melhorias quando comparados a outros métodos de Aprendizado de Máquina e estratégias de investimento, chegando a uma média de 55,9% de precisão ao prever se o preço de uma determinada ação vai subir ou não no futuro próximo.

O trabalho de Fischer e Krauss (2018) mostrou o uso de RNAs do tipo LSTM para prever a direção dos movimentos do mercado financeiro. Como fontes de dados, foram utilizadas séries temporais do Índice S&P 500 de 1992 até 2015. Os resultados mostraram que as RNAs do tipo LSTM superaram métodos de classificação, tais como *Random Forest* e classificador de regressão logística. Além disso, o trabalho formalizou uma estratégia de investimento em ações de curto prazo que garantiu retorno de 23%.

Xiongwen Pang et al. (2018) realizaram previsões do mercado financeiro por meio de RNAs profundas do tipo LSTM. Como fonte de dados, foram utilizados dados históricos do Índice Composto de Shanghai no período entre de 1 de janeiro de 2006 e 19 de outubro de 2016. Os resultados mostraram que a LSTM apresentou previsão correta em 57,2% dos casos em seu melhor resultado.

Ainda, Hiransha et al. (2018) apresentaram quatro arquiteturas de RNAs profundas para previsão do mercado financeiro: MLP, RNN, LSTM e CNN, aplicadas a dois mercados: *National Stock Exchange* (NSE) e *New York Stock Exchange* (NYSE). Os dados utilizados da NSE correspondiam ao período de janeiro de 1996 a junho de 2015, enquanto que os dados da NYSE compreendiam de janeiro de 2011 a dezembro de 2016. O trabalho observou que a CNN superou os demais modelos no processo, capturando melhor as mudanças abruptas em valores de ações no mercado, além de superar a técnica de previsão linear ARIMA.

Sang e Di Pierro (2019) mostraram o uso de RNAs profundas do tipo LSTM para melhorar algoritmos de negociação que utilizam Análise Técnica. Os dados utilizados no trabalho são do índice S&P 500. Os resultados mostraram que a integração entre LSTM e Análise Técnica obtiveram melhores resultados do que apenas a Análise Técnica. Ainda, Kusuma et al. (2019) utilizaram RNAs profundas (CNNs) e *candlesticks* para composição de um *framework* de apoio ao investidor na previsão do mercado financeiro. Os dados utilizados no trabalho são de 50 empresas do mercado de Taiwan e 10 empresas do mercado da Indonésia, compreendendo

os anos de 2000 a 2018. O trabalho apresentou expressivos 92% de retorno nos mercados de Taiwan e Indonésia.

O trabalho de Ming-Che Lee et al. (2021) apresentou um estudo com RNAs profundas conjuntamente com indicadores técnicos para prever movimentos de preços de ações de curto prazo. O objeto do estudo foi a bolsa de valores de Taiwan, e os experimentos utilizaram dados de 2017 a 2019. Para isso, foi utilizado um modelo de RNA do tipo LSTM juntamente com indicadores técnicos, como RSI e MACD, combinados com o preço de abertura, fechamento, alta e baixa diária. Os resultados do trabalho mostraram que a combinação proposta pode atingir até 83,6% de precisão na previsão do movimento do mercado.

Ainda, recentemente Ezen-Can (2020) propuseram um sistema de suporte à decisão baseados em RNAs profundas do tipo LSTM e indicadores técnicos. Os dados utilizados foram do mercado indiano e os resultados mostraram que, considerando as métricas erro quadrático médio, erro médio absoluto e erro percentual médio absoluto, as melhoras em comparação com as técnicas *baselines* foram, respectivamente, 4,13%, 3,24% e 1,21%.

Por fim, vale destacar a revisão da literatura apresentada em Thakkar e Chaudhari (2021), que discute o uso de RNAs profundas no mercado financeiro, principalmente nos anos de 2017-2020, mostrando que o número de publicações da área tem aumentado nos últimos anos significativamente. O trabalho de Chhajer, Manan Shah e Kshirsagar (2022) abordou uma revisão do uso de RNAs convencionais, RNAs profundas do tipo LSTM e do algoritmo SVM na literatura como métodos de previsão do mercado de ações, mostrando que as redes LSTMs alcançam, de modo geral, melhores resultados na previsão do mercado dentre os trabalhos analisados. Além disso, o trabalho também apresenta que 82% dos líderes de companhias usam técnicas de Inteligência Artificial e Aprendizado de Máquina para determinar caminhos de investimento e tomar decisões de investimento adequadamente.

Com base nos trabalhos apresentados, é possível notar que existe uma mudança de tendência de trabalhos que utilizavam RNAs do tipo MLP para trabalhos que utilizam RNAs profundas, principalmente LSTMs, visto que a característica principal desta é lidar com dados que possuem ordem temporal. Desse modo, é possível perceber a capacidade promissora das RNAs em prever o mercado acionário, seja por meio do seu movimento ou valor da ação.

5.3 Valores de Ações, Indicadores Técnicos e Dados Textuais para Previsão do Mercado Financeiro

Recentemente, conforme apresentado pelas revisões da literatura em Ozbayoglu, Gudelek e Sezer (2020) e Bustos e Pomares-Quimbaya (2020), as publicações da área de predição do mercado financeiro têm se voltado para o uso de dados textuais por meio de tarefas de mineração de textos. Dentre as abordagens, o uso de sentimentos em notícias e redes sociais tem apresentado grande crescimento em publicações recentes relacionadas ao mercado financeiro, incorporando-os aos comumente usados dados de séries históricas financeiras e indicadores técnicos.

O trabalho de Maqsood et al. (2020) apresentou uma abordagem de Aprendizado Profundo com Análise de Sentimentos para investigar eventos significativos nos mercados de ações usando empresas de diferentes países. Os dados consistem em *tweets* de 2012 a 2016. A tarefa de AS foi realizada por meio de uma lista de palavras, que posteriormente é usada como entrada para uma CNN. Os experimentos mostraram que os sentimentos melhoram as previsões do mercado de ações.

O trabalho de Nti, Adekoya e Weyori (2020) mostrou um modelo baseado em AS para prever o movimento do mercado de ações de Gana usando uma rede neural MLP. O modelo proposto foi testado com várias janelas de tempo para prever os preços das ações. Os resultados mostram que as informações disponíveis em notícias financeiras, *tweets* e fóruns *online* são úteis para os investidores, apresentando acurácias que variam de 49% a 55%.

O trabalho de Qian, Zirao Li e Yuan (2020) apresentou o uso de uma rede neural CNN para classificar tendências a partir de comentários *online*, além da investigação das inclinações dos usuários sobre a volatilidade e retornos do mercado. O trabalho também apresenta modelos de previsão de retornos aplicados aos índices de Xangai e Shenzhen. Os resultados mostraram que as tendências negativas dos usuários *online* se correlacionam com a volatilidade e os retornos do mercado.

Além disso, o trabalho de Picasso et al. (2019) apresentou a combinação de notícias e preços históricos das ações para prever o mercado de ações. Para realizar a Análise de Sentimento das notícias, utilizou um dicionário financeiro e, para prever o mercado, utilizou uma rede neural MLP. Além disso, experimentos mostraram que o uso de notícias melhora a previsão do

mercado de ações, considerando tanto métricas de aprendizado de máquina quanto métricas financeiras.

O trabalho de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020) apresentou uma estratégia de previsão do mercado de ações que integra notícias e indicadores técnicos, usados como entrada para uma rede neural LSTM aplicada à Bolsa de Valores de Hong Kong. O trabalho realiza Análise de Sentimentos em notícias por meio de um dicionário financeiro. Os experimentos mostram que o modelo proposto supera os modelos *baseline* em relação às métricas de treinamento dos modelos e que a combinação entre notícias e indicadores técnicos supera as estratégias com apenas sentimentos de notícias ou apenas indicadores técnicos.

Duarte, González e Cruz (2021) apresentaram o uso de notícias e preços de ações para prever quedas de preços, a fim de evitar quedas no mercado de ações brasileiro. O conjunto de dados compreende notícias de agosto de 2016 a maio de 2018 e o trabalho usa como técnicas de previsão 11 modelos clássicos de Aprendizado de Máquina. Os resultados mostram que as notícias se correlacionam com a variação dos preços das ações no mercado de ações brasileiro. Além disso, a abordagem proposta supera os *baselines Buy & Hold* e cruzamento de médias móveis ao considerar uma simulação de investimento.

O trabalho de Shi et al. (2021) mostrou um sistema de análise de sentimento com RNAs profundas aplicado a comentários sobre ações em redes sociais financeiras da China. O modelo de AS utiliza RNAs do tipo RNN e CNN juntamente com *Word2Vec*. Os resultados mostraram que o método de AS proposto apresentou melhora de 9% em relação ao algoritmo de regressão logística. Além disso, os autores propuseram um método de previsão do mercado, que usa sentimentos e o método de Aprendizado de Máquina SVM, e que apresentou 1,25% de melhoria entre 150 ações chinesas no conjunto de dados de teste. Para ações americanas, no entanto, os autores apontam que as informações de sentimento reduziram os resultados da previsão.

É possível notar que são poucos os trabalhos que combinam notícias e preços de ações em um sistema de previsão, e esse ainda é um tópico de pesquisa em aberto e recente na literatura. Assim, vale destacar algumas diferenças entre nosso trabalho e as abordagens mais semelhantes encontradas nessa revisão.

Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020) e Maqsood et al. (2020) não desenvolveram simulações de investimentos para comprovar a eficácia das previsões no domínio financeiro, restringindo-se às próprias métricas de Aprendizado de Máquina

relativas ao treinamento de seus modelos preditores. Conforme notado anteriormente, isso pode ser um problema na avaliação do modelo, uma vez que melhores medidas de desempenho nas métricas de Aprendizado de Máquina podem não implicar em melhor desempenho nas simulações de investimentos.

Nesse sentido, os autores do trabalho Picasso et al. (2019) realizaram uma simulação de investimento além da avaliação das métricas de Aprendizado de Máquina, de maneira similar à metodologia proposta aqui. No entanto, o trabalho de Picasso et al. (2019) usa dados de aproximadamente 1 ano, restringindo o período analisado do mercado de ações. Além disso, vale ressaltar que nosso trabalho, de acordo com o nosso conhecimento e revisão da literatura, se destaca no sentido de abordar a questão da inclusão de notícias em um sistema de previsão sem uma lista de palavras ou dicionário para realizar a Análise de Sentimento, como fazem os trabalhos Picasso et al. (2019) e Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020).

Já Duarte, González e Cruz (2021) tentam prever o mercado de ações brasileiro com o uso de Aprendizado de Máquina, mas difere desta tese em relação às abordagens consideradas. Enquanto o trabalho de Duarte, González e Cruz (2021) usa principalmente modelos tradicionais de Aprendizado de Máquina, este trabalho de doutorado se concentra em RNAs profundas. Outro fator que difere nosso trabalho é que usamos o algoritmo *FastText* para extrair recursos das notícias, enquanto Duarte, González e Cruz (2021) usam as frequências de termos presentes nas sentenças para alimentar modelos utilizados.

Por fim, o trabalho de Shi et al. (2021) realiza uma etapa de AS similar à realizada aqui, com o uso de Aprendizado Profundo e redes CNN para tal tarefa. No entanto, ainda utiliza *Word2Vec* como estratégia para extração de características de palavras, além de usar como fontes de dados informações de redes sociais, que, conforme já observamos em outro trabalho com resultados iniciais desta tese (CAROSIA; COELHO; SILVA, A. E. A., 2019), apresentam menor correlação ao mercado brasileiro em comparação às notícias. Por fim, o trabalho de Shi et al. (2021) utiliza o algoritmo SVM para prever o mercado, evitando o uso de Aprendizado Profundo nessa etapa. Conforme será possível observar no Capítulo 7 desta tese, o algoritmo SVM normalmente apresenta resultados inferiores quando comparado com RNAs.

5.4 Considerações Finais

Por meio dos trabalhos apresentados neste capítulo, pode-se verificar que, em relação ao tema análise de sentimentos e mercado financeiro, a maioria busca identificar correlação entre sentimentos de mídias sociais ou notícias e movimentos do mercado financeiro. Assim, duas técnicas se destacaram: (i) análise léxica por meio de palavras que indicam sentimentos a serem rastreados; e (ii) Aprendizado Profundo por meio do treinamento de Redes Neurais Artificiais Profundas.

Por outro lado, é possível verificar que os trabalhos que realizam previsão do mercado financeiro com o uso das técnicas de Redes Neurais Artificiais utilizam, principalmente, as seguintes arquiteturas: *LSTM* e *MLP*. Os trabalhos demonstraram que existem melhoras em seus resultados quando a técnica de previsão utiliza Redes Neurais Artificiais, comparados com demais técnicas de Aprendizado de Máquina.

A análise da literatura sobre previsão de mercado financeiro também permite identificar limitações da literatura científica. Dentre os trabalhos analisados, poucos trataram de dados do mercado de ações brasileiro. Ademais, nenhum trabalho identificado nessa busca tratou de análise de sentimentos especificamente em português, correlacionando-a com o mercado financeiro. Além disso, a maior parte dos trabalhos foca somente em análise de sentimentos ou previsão de séries temporais, sendo muito recente a integração dessas fontes de dados em um único sistema de previsão. Dessa forma, existe uma oportunidade de desenvolvimento de trabalhos na área, seja por meio da união desses dois tópicos, abordando tanto análise de sentimentos em português quanto o uso de dados históricos de ações do mercado financeiro brasileiro.

Nesse sentido, é possível verificar que o uso de RNAs compreende o estado da arte em diversas áreas relacionadas à previsão do mercado financeiro.

Considerando os trabalhos discutidos aqui, as principais contribuições do presente trabalho são:

1. Todo o processo de previsão do mercado de ações utiliza técnicas de RNAs profundas, tanto aplicadas à AS de notícias quanto aos preços de ações. Isso traz uma vantagem, principalmente no processo de Análise de Sentimentos. Conforme apresentado, a maioria dos trabalhos relacionados usa léxicos para realizar essa tarefa. Ainda, contribuimos também em relação à tarefa de Análise de Sentimentos em língua

portuguesa, que, conforme a revisão de Denilson Alves Pereira (2021), ainda possui poucos trabalhos publicados.

2. O mercado de ações brasileiro como estudo de caso, que não é muito explorado na literatura da área financeira. As revisões da literatura mostram que existe uma quantidade muito baixa de artigos publicados que consideram o mercado de ações brasileiro (OZBAYOGLU; GUDELEK; SEZER, 2020; BUSTOS; POMARES-QUIMBAYA, 2020; KUMBURE et al., 2022), concentrando-se, em sua maioria, nos mercados norte-americano e asiático.
3. A apresentação de uma proposta que utiliza dados provenientes tanto da Análise Técnica quanto da Análise Fundamentalista. Conforme discutido anteriormente, normalmente os investidores se concentram em uma forma de análise. Desse modo, este trabalho foca na junção de ambas em um sistema de previsão para uma análise mais ampla do mercado.

Além dos pontos já levantados, vale destacar também que, neste trabalho, realizamos a comparação das arquiteturas de RNAs mais utilizadas nesta área de aplicação, i.e., o mercado de ações (OZBAYOGLU; GUDELEK; SEZER, 2020). Por fim, em nossos experimentos, também damos ênfase à importância de avaliar os modelos propostos considerando tanto métricas de Aprendizado de Máquina quanto métricas do domínio financeiro, conforme também realizado por Picasso et al. (2019).

Capítulo 6

Metodologia

O desenvolvimento do presente trabalho de doutorado contou com investigações que levaram a publicações cujos resultados foram utilizados como base para as decisões tomadas nesta tese (CAROSIA; COELHO; SILVA, A. E. A., 2021, 2020, 2019), além de uma contribuição no trabalho de Januário et al. (2021). Nesse sentido, para explicitar as tomadas de decisões ao longo desta pesquisa, a Seção 6.1 apresenta em detalhes a metodologia de cada um dos trabalhos publicados, bem como seus resultados e o ponto de convergência entre eles e a presente tese. Por fim, a Seção 6.2 apresenta a metodologia proposta nesta tese de doutorado.

6.1 Desenvolvimento Geral

Ao longo do desenvolvimento deste trabalho, pelo menos quatro investigações foram realizadas buscando atingir o objetivo final de proposição de uma metodologia de predição do mercado de ações brasileiro integrando elementos da Análise Fundamentalista (notícias) e da Análise Técnica (indicadores técnicos e histórico de ações), conforme mostrado nos passos de 1 a 4 na Figura 6.1. Primeiramente, nos voltamos à investigação da relação entre postagens da rede social Twitter e o mercado de ações (CAROSIA; COELHO; SILVA, A. E. A., 2020) (Figura 6.1(1)). Uma vez analisado tal relacionamento, investigamos se a forma mais adequada de previsão dos movimentos do mercado de ações seria com o uso de notícias ou *tweets* (CAROSIA; COELHO; SILVA, A. E. A., 2019) (Figura 6.1(2)). Dessa análise, verificamos que notícias de jornais brasileiros *online* são uma fonte de dados mais adequada para operar no mercado de ações do que *tweets*. Assim, realizamos uma terceira análise para investigar se seria possível utilizar ou desenvolver estratégias de investimento baseadas em sentimento de

notícias para operar no mercado acionário (CAROSIA; COELHO; SILVA, A. E. A., 2021) (Figura 6.1(3)). Uma vez observada a relevância do sentimento presente em notícias para a realização de investimentos, por fim, a quarta investigação deste trabalho de doutorado (Figura 6.1(4)) utiliza resultados das três investigações anteriores para estudar se a junção do sentimento presente em notícias com o histórico de valores de ações e indicadores técnicos é capaz de melhorar a previsão do mercado de ações. Essa última etapa da investigação corresponde à presente tese. A seguir, são detalhados cada um dos desenvolvimentos publicados na literatura ao longo deste trabalho de doutorado, e que serviram como base para a presente tese.

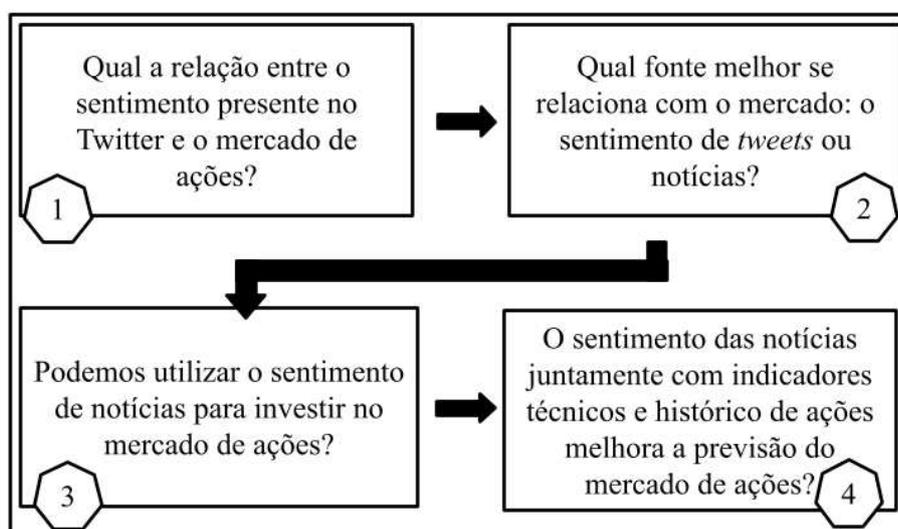


Figura 6.1: Etapas desenvolvidas durante este trabalho.

No início deste trabalho, avaliamos o impacto do sentimento presente em *tweets* no mercado de ações brasileiro (CAROSIA; COELHO; SILVA, A. E. A., 2020). Para isso, utilizamos, conforme apresentado na Figura 6.2, a técnica de *Bag-of-Words* para extração de características de *tweets* e comparamos os seguintes modelos como classificadores de sentimentos: *Naive Bayes*, SVM e MLP. Para a fase de estudo da relação entre os *tweets* publicados e o movimento do mercado de ações, utilizamos um total de 16.352 *tweets* não rotulados. Os nossos resultados mostraram que a rede MLP é o melhor modelo para AS em português dentre os modelos testados. Além disso, observamos que o sentimento médio nas mídias sociais se relaciona com o mercado de ações, aumentando a acurácia presente entre sentimento e movimento de mercado à medida em que o tamanho das janelas (número de dias) entre a postagem e o movimento do mercado aumenta.

Em Carosia, Coelho e A. E. A. Silva (2020), verificamos que há um certo atraso entre a publicação dos *tweets* analisados e sua influência no mercado, o que pode dificultar sua utilização em negociações no mesmo dia (ou em outros intervalos curtos) dentro do mercado. Além disso, dois desafios identificados a partir desse estudo foram: (1) os *tweets* são normalmente escritos em linguagem informal, o que pode dificultar a extração de características textuais; e (2) muitas contas com pouca relevância na rede social podem fazer postagens, o que introduz um certo ruído na análise do relacionamento entre postagens e movimento do mercado. Ainda, por se tratar de um estudo inicial, no artigo não mostramos como utilizar essas informações como estratégia de investimento.

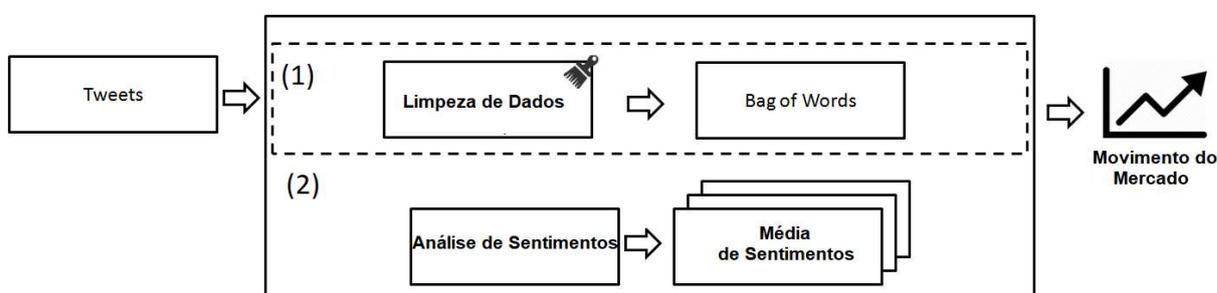


Figura 6.2: Metodologia proposta para investigar o relacionamento entre *tweets* e o mercado de ações brasileiro. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2020).

Nesse sentido, verificamos a necessidade de se estudar uma outra fonte de dados para a investigação da possível existência de relações entre sentimentos e o mercado de ações: notícias. Assim, comparamos o impacto de notícias e *tweets* com o movimento do mercado acionário brasileiro (CAROSIA; COELHO; SILVA, A. E. A., 2019). Para esse trabalho, a arquitetura proposta, apresentada na Figura 6.3, considerou uma RNA do tipo MLP para analisar os sentimentos de *tweets* e notícias de jornais *online* brasileiros. Como incremento em relação ao trabalho anterior, desenvolvemos uma métrica que determina o sentimento predominante diário tanto de notícias quanto de *tweets*, que foi utilizada para o estudo de sua relação com o movimento do mercado de ações brasileiro. Os resultados dessa publicação mostraram que o sentimento presente em notícias permite prever o movimento do mercado de ações brasileiro com maior acurácia do que postagens da rede social Twitter. Esse relacionamento foi investigado tanto no que diz respeito aos preços de abertura quanto em relação ao volume de transações. Portanto, dado o resultado obtido em Carosia, Coelho e A. E. A. Silva (2019), decidimos focar em notícias como fonte de dados nas etapas subsequentes da pesquisa.

O trabalho Carosia, Coelho e A. E. A. Silva (2019), no entanto, ainda não apresentava o aprofundamento necessário no sentido de investigar a correlação entre notícias e mercado de ações, a fim de permitir o desenvolvimento de estratégias de investimentos mais robustas com o uso de notícias, além de incrementar a tarefa de AS com modelos mais avançados e melhores técnicas de extração de características de texto, o que foi tratado em Carosia, Coelho e A. E. A. Silva (2021).

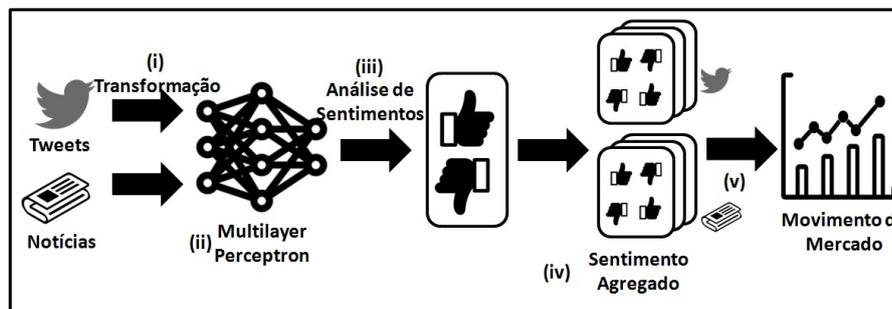


Figura 6.3: Metodologia proposta para comparar a influência tanto de postagens da rede social Twitter como de notícias no mercado de ações brasileiro. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2019).

Em Carosia, Coelho e A. E. A. Silva (2021), investigamos o uso de estratégias de investimento baseadas exclusivamente em Análise de Sentimentos aplicadas ao mercado de ações brasileiro. Para isso, as seguintes etapas foram desempenhadas (Figura 6.4): (1) desenvolvimento de uma arquitetura baseada em Aprendizado Profundo, alimentada com *FastText* para AS de notícias financeiras; (2) um estudo de causalidade, com o teste de causalidade de Granger (HOOVER, 2008), para investigar a correlação entre o sentimento predominante em notícias financeiras e o movimento do mercado; (3) uso do resultado do teste de causalidade para propor estratégias de investimento com base em AS; e (4) aplicação das estratégias em uma simulação de investimento no mercado brasileiro. Os resultados obtidos no trabalho nos mostraram que a arquitetura de RNA do tipo CNN é a mais adequada à tarefa de AS; que existe uma causalidade de Granger significativa entre o sentimento predominante em notícias e o movimento do mercado de ações; e que o uso de estratégias de investimento baseadas em AS podem trazer lucros significativos no curto e longo prazo, superando os *baselines Buy & Hold* e *Random Walk*.

A partir dos resultados publicados em Carosia, Coelho e A. E. A. Silva (2021), pudemos delimitar alguns pontos importantes para o restante do desenvolvimento desta tese. Primeiramente, foi possível verificar a causalidade existente entre notícias e mercado,

confirmando a importância do uso dessa fonte de dados em um sistema de previsão. Em segundo lugar, nesse trabalho verificamos a eficiência do uso da rede neural CNN como modelo para AS. Por fim, também foi possível superar as limitações dos trabalhos anteriores, uma vez que foram desenvolvidas estratégias que demonstraram que o uso de AS para investimento no mercado de ações pode levar a lucros.

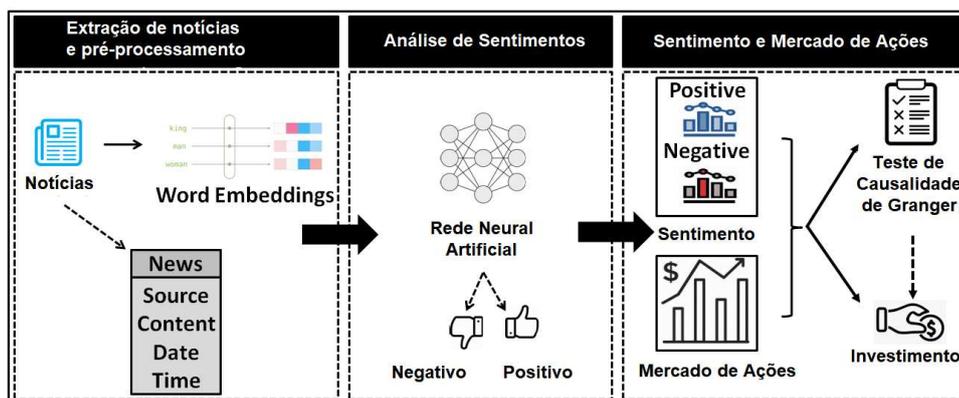


Figura 6.4: Metodologia proposta para a investigação de estratégias de investimento baseadas em Análise de Sentimento. Adaptado de (CAROSIA; COELHO; SILVA, A. E. A., 2021).

Além dos trabalhos discutidos aqui, esta pesquisa de doutorado também contribuiu com o trabalho de Januário et al. (2021), que verificou que algoritmos de Aprendizado de Máquina superam métodos léxicos especificamente voltados à área financeira em tarefas de AS em português.

Considerando o exposto, as etapas anteriormente estudadas e publicadas levaram à presente tese de doutorado. Ou seja, partimos de resultados de trabalhos anteriores em que identificamos notícias como a fonte de dados textual mais adequada ao mercado de ações brasileiro (CAROSIA; COELHO; SILVA, A. E. A., 2020, 2019), e selecionamos uma RNA do tipo CNN para desempenhar a tarefa de AS (CAROSIA; COELHO; SILVA, A. E. A., 2021).

Assim, a presente tese apresenta o passo final da investigação apresentada na Figura 6.1(4), isto é, a proposição de uma estratégia de previsão do mercado de ações brasileiro considerando as seguintes fontes de dados: sentimentos presentes em notícias, série histórica de valores de ações e indicadores técnicos. Para isso, o processo foi desenvolvido com RNAs profundas, que apresentam resultados superiores quando comparados a métodos tradicionais de previsão de mercado de ações (KUMBURE et al., 2022; HOUSSEIN et al., 2021; KUMAR, D.; SARANGI; VERMA, 2021). Dessa forma, são combinados aqui dados pertencentes tanto à Análise Fundamentalista (notícias) quanto à Análise Técnica (indicadores técnicos e histórico

de ações) em um único sistema de predição. A seção a seguir apresenta, portanto, a metodologia proposta para o último passo deste trabalho de doutorado, conforme ilustrado na Figura 6.1(4).

6.2 Metodologia de Previsão do Mercado de Ações

Como discutido na seção anterior, a metodologia proposta aqui consiste em uma abordagem de previsão do mercado de ações que combina o sentimento presente em notícias, indicadores técnicos e preços de ações. Além disso, tal metodologia usa o mercado de ações brasileiro como estudo de caso. A Figura 6.5 ilustra a abordagem proposta, que é composta pelos módulos e etapas descritos a seguir.

- **Módulo de Pré-processamento:** primeiramente, as notícias coletadas são pré-processadas para geração de *word embeddings* que servirão de entrada aos modelos subsequentes. Além disso, nesse módulo ocorre também a obtenção dos valores de ações e os cálculos de seus respectivos indicadores técnicos. A etapa de pré-processamento como um todo é descrita em detalhes na Seção 6.2.1.
- **Módulo de Análise de Sentimentos:** essa etapa consiste em usar os dados resultantes do pré-processamento (ou seja, os *word embeddings* de notícias) para alimentar uma RNA profunda com o objetivo de realizar a tarefa de AS. Essa etapa está detalhada na Seção 6.2.2.
- **Junção dos Dados:** essa etapa intermediária consiste na concatenação dos dados necessários para alimentar o Módulo de Predição, por meio da geração de um vetor composto de sentimentos diários presente em notícias, indicadores técnicos e preços de ações. Essa etapa está descrita na Seção 6.2.3.
- **Módulo de Predição:** o vetor gerado na etapa anterior é usado como entrada para uma segunda RNA profunda, que realiza uma tarefa de regressão para prever o retorno do investimento do dia seguinte no mercado de ações. Essa etapa é descrita em detalhes na Seção 6.2.3.
- **Módulo de Avaliação:** a última etapa avalia os resultados da previsão considerando duas abordagens: métricas de aprendizado de máquina e métricas financeiras relativas

a uma tarefa de simulação de investimento. Esta etapa é descrita em detalhes na Seção 6.2.4.

As seções a seguir detalham cada um dos módulos propostos em nosso sistema.

6.2.1 Módulo de Pré-processamento

Os dados usados neste trabalho são de três tipos diferentes: preços de ações, indicadores técnicos e notícias. Cada um destes tipos de dados exige um pré-processamento diferente, conforme descrito nas subseções a seguir.

Preços de Ações e Indicadores Técnicos

A série temporal dos preços das ações consiste nos seguintes dados: preço de abertura, alta, baixa, preço de fechamento e volume (do inglês, *Open, High, Low, Close, Volume* - OHLCV). Depois de baixados, tais dados são usados tanto para alimentar as RNAs quanto para cálculo dos Indicadores Técnicos, que também alimentam as RNAs.

Os indicadores técnicos são indicadores matemáticos calculados pelos investidores para estudar o comportamento do mercado. Neste trabalho, juntamente com notícias e preços de ações, também usamos vários indicadores técnicos como entrada das RNAs. Tais indicadores podem fornecer sinais de tendências de mercado e foram escolhidos considerando o trabalho de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), juntamente com indicadores que indicam volatilidade e volume (ACHELIS, 2001). Assim, os seguintes indicadores técnicos são considerados neste processo:

- **Indicadores de tendência:** MACD e Médias Móveis. Enquanto o MACD mostra a relação entre duas médias móveis de preços para identificar tendências, as Médias Móveis mostram o valor médio de um ativo ao longo do tempo.
- **Indicadores de Momentum:** RSI, MFI e OBV. Enquanto o RSI mostra sinais de sobrecompra (geralmente acima de 80) ou sobrevenda (geralmente abaixo de 20) de um ativo, o MFI mostra a intensidade do fluxo de dinheiro que entra e sai de um ativo. Além disso, o indicador OBV relaciona as mudanças de volume às mudanças de preço de um ativo.

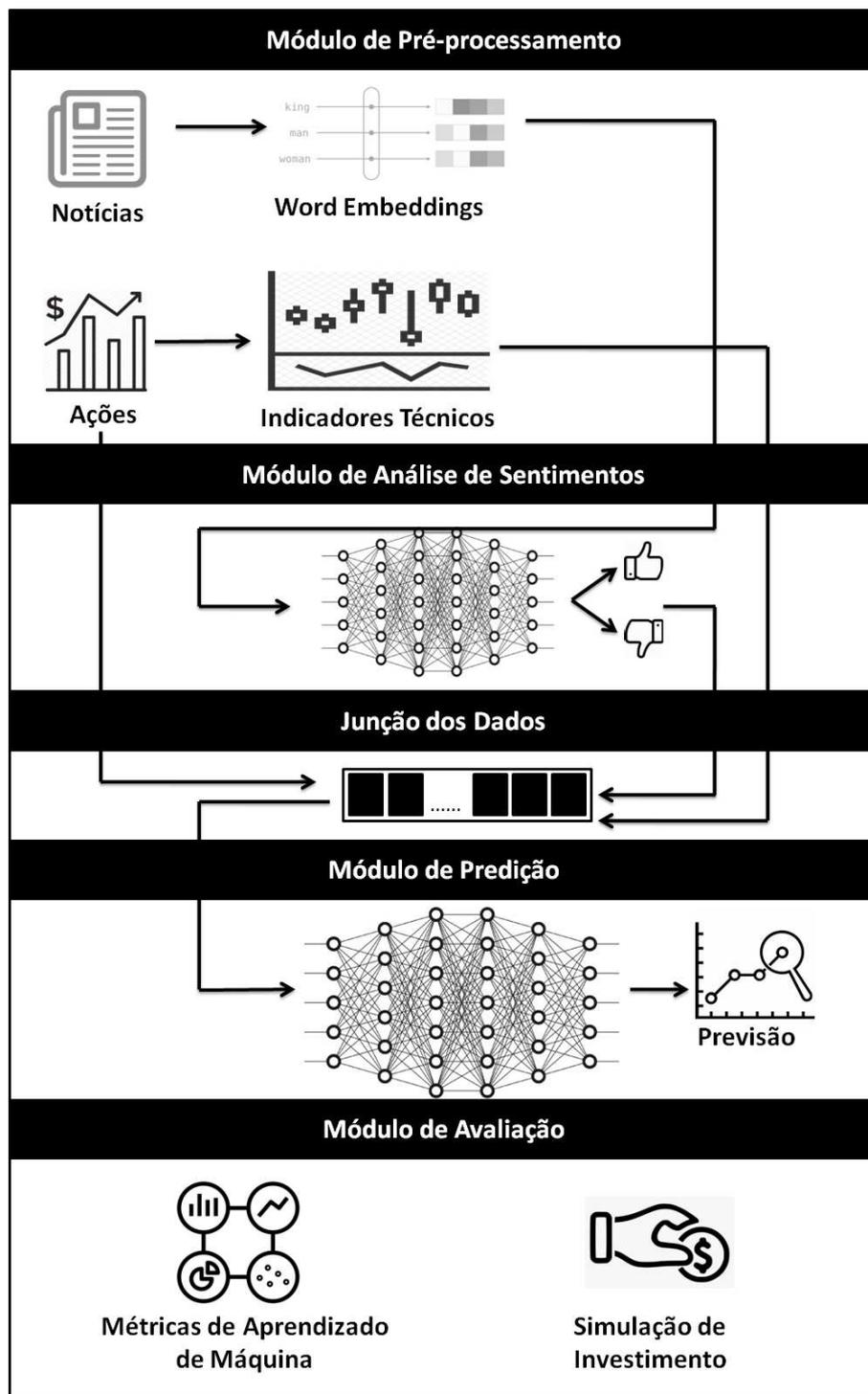


Figura 6.5: Representação da metodologia proposta. A metodologia consiste, resumidamente, nas seguintes etapas: processar os dados de entrada, ou seja, notícias, indicadores técnicos e preços de ações, identificar o sentimento de cada notícia e juntá-los em um vetor de características; prever o retorno do investimento do dia seguinte na bolsa de valores; e avaliar os resultados considerando métricas da área de Aprendizagem de Máquina em uma simulação de investimento.

- **Indicador de volatilidade:** *Bollinger Bands*, que representa um nível de desvio padrão da média móvel de uma ação. Este indicador aumenta quando o mercado está instável e diminui em um período mais estável.

Notícias

Primeiramente, vale a pena destacar que, neste trabalho, utilizamos apenas os títulos das notícias para prever o mercado de ações, conforme recomendado por Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020). Assim, a etapa de pré-processamento dos títulos de notícias adapta o documento de seu formato textual original para um formato que permita sua classificação como positivo ou negativo por uma RNA. Isto envolve duas etapas.

1. **Limpeza de dados:** que consiste na eliminação de símbolos, números, *emoticons* e *links* indesejados ao conteúdo do documento.
2. **Representação de palavras:** que consiste na criação de *word embeddings* para cada documento utilizando *FastText* (BOJANOWSKI et al., 2017), que é um dos algoritmos mais populares da área para representação morfológica de palavras. O resultado dessa etapa é um vetor de 128 dimensões representando cada documento, contendo a união de todos os seus *word embeddings*.

Além disso como passo final é necessário levar em conta os sentimentos das notícias diárias, ou seja, criar um vetor com duas dimensões em que a primeira é o número de notícias positivas calculadas em um dia e a segunda é o número de notícias negativas do mesmo dia. Essas saídas são utilizadas como entradas no Módulo de Predição.

6.2.2 Módulo de Análise de Sentimentos

Conforme discutido no Capítulo 5 (e corroborado pelos trabalhos descritos na Seção 6.1), existem estudos recentes mostrando resultados consistentes na previsão do mercado de ações utilizando, dentre outras, técnicas de RNAs e Análise de Sentimentos. Assim, o presente passo desta metodologia tem como objetivo desenvolver o módulo de Análise de Sentimentos para posterior aplicação ao mercado de ações brasileiro. Especificamente, a metodologia desse processo é destacada na Figura 6.6 e resumida nas seguintes etapas.

1. **Pré-processamento**, em que as notícias coletadas são pré-processadas, conforme discutido na Seção 6.2.1, com o objetivo de treinar o modelo de RNA para a execução da tarefa de AS.
2. **Modelo de Análise de Sentimentos**, em que as RNAs são usadas para classificação de sentimentos e determinar cada notícia como positiva ou negativa. Como resultado

desse módulo, é possível criar uma série temporal diária com a quantidade de notícias positivas e negativas.

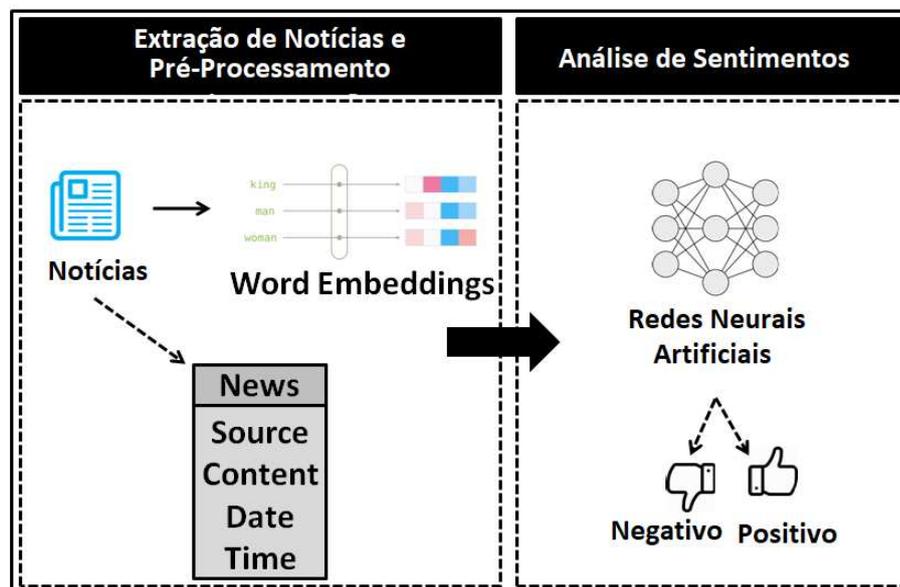


Figura 6.6: Representação da Metodologia de Análise de Sentimentos. As notícias são pré-processadas, gerando *word embeddings* de palavras que são usadas como entrada para uma RNA que realiza a AS. Adaptada de (CAROSIA; COELHO; SILVA, A. E. A., 2021).

Com o objetivo de realizar a tarefa de Análise de Sentimentos com RNAs, as seguintes arquiteturas foram consideradas: MLP, LSTM e CNN. A metodologia experimental, bem como a escolha da RNA mais adequada para o desenvolvimento da tarefa de AS estão apresentadas na Seção 7.2.

6.2.3 Módulo de Predição

O Módulo de Predição proposto neste trabalho consiste também em uma RNA profunda, que recebe como entrada um vetor com as seguintes características combinadas: (1) informações sobre o preços das ações (ou seja, abertura, baixa, alta, fechamento) e volume, totalizando 5 dimensões (2) indicadores técnicos (conforme apresentado na Seção 6.2.1), totalizando 10 dimensões; e (3) os sentimentos das notícias diárias (ou seja, o número de notícias positivas e o número de notícias negativas), totalizando 2 dimensões. Assim, a dimensão do vetor final concatenado com todas informações é 17, conforme apresentado na Figura 6.7.

Para que a concatenação de dados de diferentes fontes aconteça, é necessário um estágio intermediário entre as etapas de Pré-Processamento, Análise de Sentimentos e a etapa de

Predição. Tal etapa foi denominada Junção dos Dados, conforme apresentado na metodologia da Figura 6.5.



Figura 6.7: Representação dos componentes utilizados para alimentar a RNA: (1) Valores Históricos: valor de abertura, fechamento, baixa e alta diária, além do volume diário de negociações; (2) Indicadores Técnicos: MA (*Moving Averages*), RSI (*Relative Strenght Index*), MFI (*Money Flow Index*), MACD (*Moving Average Convergence Divergence*), OBV (*On Balance Volume*) e *Bollinger Bands*; e (3) Sentimentos de Notícias: sentimentos positivos e sentimentos negativos.

Após a Junção de Dados, os vetores gerados passam por um processo de normalização antes de serem fornecidos para a RNA. Como no trabalho de Picasso et al. (2019), usamos uma técnica de normalização Min-Max (Eq. 6.1) para limitar os dados ao intervalo $[0,1]$, onde x_{norm} representa o valor normalizado do atributo x , x_{min} é o valor mínimo do atributo e x_{max} é o valor máximo do atributo sendo normalizado:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (6.1)$$

Assim, para a implementação do Módulo de Previsão foi feita uma investigação para a escolha da RNA mais adequada para previsão do mercado, considerando as seguintes arquiteturas com diferentes complexidades: MLP, LSTM e BI-LSTM. A RNA escolhida deve realizar a tarefa de regressão com o objetivo de prever o retorno do dia seguinte na bolsa de valores. A metodologia experimental e os resultados para esse módulo são apresentados na Seção 7.3.

A saída do módulo de previsão é o retorno do preço de fechamento do dia seguinte na bolsa, ou seja, o valor do preço de fechamento do dia $d + 1$, subtraindo o preço de fechamento do dia atual, d , como dado na Eq. 6.2.

$$\text{Retorno} = \text{Fechamento}_{d+1} - \text{Fechamento}_d \quad (6.2)$$

A Figura 6.8 ilustra o funcionamento geral do Módulo de Previsão.

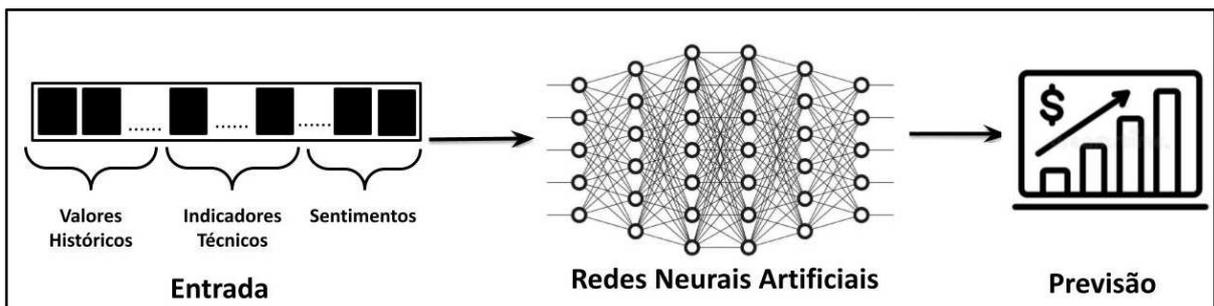


Figura 6.8: Representação do Módulo de Predição, que recebe o vetor gerado por meio da Junção de Dados e investiga a arquitetura de RNA mais adequada para a previsão do mercado de ações.

6.2.4 Módulo de Avaliação

Conforme discutido anteriormente, os resultados de nosso trabalho foram avaliados considerando uma etapa de testes com dados referentes ao período de janeiro de 2017 a dezembro de 2019. Devido às dificuldades para avaliar um algoritmo de Aprendizado de Máquina na área financeira, dividimos o processo de avaliação em duas etapas, considerando a abordagem proposta por Picasso et al. (2019).

Métricas de Aprendizado de Máquina

Nessa etapa, avaliamos os resultados de predição considerando as seguintes métricas comuns em tarefas de regressão: *Erro Quadrático Médio* (do inglês, *Mean Squared Error* - MSE), dado pela Eq. 6.3, e *Erro Médio Absoluto* (do inglês, *Mean Absolute Error* - MAE), dado pela Eq. 6.4, onde y_i é a saída real, y'_i representa o valor previsto e n é o número de amostras no conjunto de dados.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (6.3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (6.4)$$

Métricas de Investimento

Nessa etapa, realizamos uma simulação de investimento para medir a qualidade das previsões obtidas pelo modelo. Para avaliar os resultados, utilizamos as seguintes métricas: valor final do investimento, *Sharpe Ratio* (SR) and *Maximum Drawdown* (MDD).

O valor final do investimento representa o valor resultante de uma série de operações de compra e venda no período considerado. A métrica SR é uma medida de risco calculada a partir do retorno médio obtido (R) menos a taxa livre de risco (r) dividido pela volatilidade ou risco total, ou seja, o desvio padrão (σ_R) do valor ganho. No Brasil, o valor da taxa livre de risco é geralmente o *Certificado de Depósito Interbancário* (conhecido como CDI), principal referência para a rentabilidade das aplicações de renda fixa. O SR é definido conforme apresentado a seguir.

$$SR = \frac{R - r}{\sigma_R} \quad (6.5)$$

Finalmente, a métrica MDD é um indicador de risco que mede a queda do valor da ação ao longo de um período de tempo especificado. Esse indicador considera a perda máxima de um pico (*Max*) a um mínimo (*Min*) de uma série temporal de preços de ações, conforme apresentado a seguir.

$$MDD = \frac{Max - Min}{Max} \quad (6.6)$$

Para realizar a simulação de investimento, foi desenvolvida uma estratégia de investimento simples a ser aplicada considerando o resultado do Módulo de Predição, conforme apresentado no Algoritmo 1. O algoritmo considera um fator de segurança (*MedidaSeguro*) indicado pelo usuário. A partir do retorno do dia atual e da previsão do dia seguinte (linhas 1 e 2), o algoritmo inicia com a posição *Vendido* para o investidor e calcula o percentual do retorno previsto pelo modelo de RNA (linha 4). O restante do código (linhas 5 a 12) é responsável pela avaliação da situação do investimento. Se a previsão de retorno for maior que fator de segurança definido pelo usuário, o algoritmo utiliza todo o valor disponível na carteira para comprar a ação (ou retém se a ação já foi comprada), ficando com seu estado como *comprado* (linhas 5 a 8); se a

previsão for menor que o segurança, o algoritmo vende todas as ações em sua posse (ou não faz nada se o usuário ainda não comprou a ação), ficando com o estado *vendido* (linhas 9 a 12). Finalmente, o algoritmo retorna a posição atual do usuário (linha 13), após a sua operação de investimento.

Algorithm 1: Algoritmo da Estratégia de Investimento

```
input : MedidaSeguro
output: Operação de Investimento
1 hoje ← retornoHoje()
2 pred ← preveRetornoAmanha()
3 posicao ← Vendido
4 percentual ← obterPorcentagemRetorno(hoje, pred)
5 if percentual ≥ MedidaSeguro and posicao = Vendido then
6 |   posicao ← Comprado
7 else if percentual ≥ MedidaSeguro and posicao = Comprado then
8 |   posicao ← Comprado
9 else if percentual < MedidaSeguro and posicao = Vendido then
10 |  posicao ← Vendido
11 else if percentual < MedidaSeguro and posicao = Comprado then
12 |  posicao ← Vendido
13 return (posicao)
```

Capítulo 7

Resultados

Este capítulo apresenta a metodologia experimental e os resultados obtidos neste trabalho, considerando a metodologia proposta no capítulo anterior. Primeiramente, os dados usados em nossos experimentos são apresentados na Seção 7.1. A Seção 7.2 mostra os resultados relacionados ao módulo de Análise de Sentimentos, abordando a escolha da melhor arquitetura de RNA profunda para a tarefa de AS. A Seção 7.3 mostra os resultados de predição do mercado de ações brasileiro considerando tanto métricas de Aprendizado de Máquina quanto de investimento. Além disso, a Seção 7.4 apresenta os resultados comparativos de experimentos com trabalhos relacionados. Por fim, a Seção 7.5 apresenta a discussão dos resultados obtidos.

7.1 Fonte de Dados

Para o desenvolvimento desta tese, são necessárias bases de dados para dois módulos: o Módulo de Análise de Sentimentos e o Módulo de Predição. Assim, as seções seguintes detalham cada uma das bases de dados utilizadas e seus respectivos módulos.

Conjunto de Dados do Módulo de Análise de Sentimentos

Para a tarefa de treinamento de cada modelo que irá desempenhar a AS, foi utilizada uma base de dados composta por 5.507 notícias rotuladas. A primeira parte dessa base foi apresentada no trabalho de Martins, Adriano Pereira e Benevenuto (2015), e é composta por 2.132 notícias financeiras. A essa base de dados foram adicionadas outras 3.373 notícias coletadas e rotuladas em 2019, durante o desenvolvimento deste trabalho e publicada parcialmente em Januário et

al. (2021). A base de dados final possui 45% de documentos com sentimentos positivos e 55% de documentos com sentimentos negativos. Esse conjunto de dados foi criado considerando notícias de empresas que integram o índice Ibovespa.

Todos os documentos estão em língua portuguesa e foram rotulados manualmente de acordo com a sua polaridade. Ou seja, os títulos de notícias que influenciam positivamente o mercado financeiro foram rotulados como positivos e os títulos de notícias que influenciam negativamente o mercado financeiro foram rotulados como negativos.

Vale a pena destacar que, quanto à escolha das classes de notícias, este trabalho de doutorado considera a investigação prévia realizada por Faria (2018), que mostra que o uso das classes positivo e negativo traz resultados superiores em relação ao uso das classes positivo, negativo e neutro em trabalhos voltados ao mercado financeiro.

Bases de Dados do Módulo de Previsão

Os dados utilizados no Módulo de Previsão deste trabalho foram coletados por meio de *scripts* considerando o período de janeiro de 2010 a dezembro de 2019. Nos experimentos, foram utilizados dados do principal índice da bolsa brasileira, o Ibovespa, e das seguintes empresas representativas do mercado acionário brasileiro: Banco do Brasil e Itaú, do setor bancário, Gerdau, do setor de siderurgia e metalurgia, e Ambev, do setor de bebidas. A Tabela 7.1 apresenta cada uma das ações citadas com o respectivo identificador (*ticker*), valorização no período, quantidade de movimentos de alta e de baixa no mercado acionário, além do máximo retorno possível no período.

Tabela 7.1: Índice e ações usados nos experimentos. São apresentados: os valores iniciais e finais de cada ação juntamente com seu retorno máximo possível no período (Ganho Máx.) e número de movimentos de subida (Alta) e descida (Baixa) no período entre janeiro de 2010 a dezembro de 2019.

Ação	Ticker	Valor Inicial	Valor Final	Ganho	Ganho Máx.	Altas	Baixas
Ibovespa	BVSP	R\$ 59.589,00	R\$ 115.964,00	94,61%	523%	394	336
Banco do Brasil	BBAS3	R\$ 27,54	R\$ 52,82	91,79%	884%	384	340
Itaú	ITUB4	R\$ 22,35	R\$ 37,10	66,00%	676%	373	355
Gerdau	GGBR4	R\$ 10,84	R\$ 20,00	84,50%	897%	354	370
Ambev	ABEV3	R\$ 16,31	R\$ 18,67	14,47%	426%	361	355

As notícias foram baixadas dos três jornais *online* brasileiros mais relevantes (ANJ, 2019): G1, Estadão e Folha de São Paulo. Para isso, foram desenvolvidos *scripts* de acesso aos buscadores de cada um desses sites, os quais trazem notícias que contenham o nome de cada

uma das empresas. Esse processo resultou em: 79.943 notícias sobre o Ibovespa, 111.708 notícias sobre o Banco do Brasil, 53.589 notícias sobre o Itaú, 1.934 notícias sobre a Gerdau e 8.817 notícias sobre a Ambev no período considerado.

As séries temporais dos preços das ações foram coletadas no site Yahoo Finance¹. Como os indicadores técnicos são calculados sobre os preços das ações, eles não precisam ser baixados.

7.2 Análise de Sentimentos em Notícias Financeiras

A primeira fase de experimentos consiste na determinação da arquitetura de RNA mais adequada para a tarefa de Análise de Sentimentos de notícias em Português, conforme discutido na Seção 6.2.2. Para isso, foram consideradas as seguintes RNAs: MLP, LSTM e CNN. Como modelos *baselines* para esse experimento foram consideradas as técnicas de Aprendizado de Máquina mais utilizadas para Análise de Sentimentos (SUN; LUO; CHEN, J., 2017): *Naive Bayes*, *Regressão Logística* e *Support Vector Machines* (SVM).

A etapa de Pré-Processamento desse experimento foi desenvolvida em linguagem *Python* (versão 3.10) com auxílio da biblioteca *Natural Language Toolkit*² (versão 3). Para a implementação do algoritmo *FastText*, foi utilizada a biblioteca *gensim*³ (versão 4.1.2), enquanto que, para a modelagem das RNAs, foi escolhida a biblioteca *Keras*⁴ (versão 2.7.0). Por fim, para a implementação dos demais algoritmos de Aprendizado de Máquina, foi utilizada a biblioteca *scikit-learn*⁵ (versão 1.0.0).

Além disso, o conjunto de dados rotulado foi dividido aleatoriamente de acordo com a proporção de 80% dos dados como conjunto de treinamento e 20% dos dados como o conjunto de testes. Para verificar quais seriam os parâmetros mais adequados para a tarefa de AS, foi aplicado o método *k-Fold Cross Validation* ao conjunto de treinamento, com valor de $k = 10$. Desse modo, assim que fossem determinados os parâmetros, cada um dos modelos selecionados seriam avaliados no conjunto de dados de testes.

Para as redes neurais do tipo MLP e LSTM, foi realizada uma busca por parâmetros para verificar qual a arquitetura mais adequada ao problema, considerando os seguintes valores: (i) 1, 3, 5, 10 e 20 camadas; e (ii) 1, 3, 5, 10, 25, 50 e 100 neurônios por camada oculta. Para a

¹<https://finance.yahoo.com/>

²<https://www.nltk.org/>

³<https://pypi.org/project/gensim/>

⁴<https://keras.io/>

⁵<https://scikit-learn.org/>

investigação desses parâmetros, consideramos o trabalho que publicamos em Carosia, Coelho e A. E. A. Silva (2021).

Os parâmetros da rede CNN, que é frequentemente usada na área de visão computacional, foram definidos como apresentado em Ye Zhang e Wallace (2017), também utilizados em Igor Santos, Nedjah e Macedo Mourelle (2017) e avaliados em nossa publicação prévia Carosia, Coelho e A. E. A. Silva (2021). Esses trabalhos adaptaram a arquitetura da CNN para lidar com a tarefa de classificação de sentenças. Em tais trabalhos, a entrada da rede CNN é definida pela concatenação dos *word embeddings* que representam cada uma das palavras dos títulos das notícias financeiras, de modo a serem dispostos um sobre o outro, de forma similar a uma matriz (a dimensão resultante é 128 no eixo x e n no eixo y , sendo n o número de palavras contido no título de uma notícia).

Ademais, os parâmetros da configuração da rede CNN são: três regiões de convolução, com dois filtros cada (de tamanhos 2, 3 e 4), que irão gerar mapas de características. Os mapas gerados alimentam camadas de *pooling* com a função *max*. Por fim, há a criação de um vetor que, na camada de saída, terá aplicado a função de ativação *softmax* para classificar a sentença. As RNAs foram treinadas considerando os seguintes parâmetros:

- Taxa de aprendizado: 0,001;
- Épocas: 100;
- *Word embeddings*: gerados pelo algoritmo *FastText* com 128 dimensões, valor definido experimentalmente;
- Função de ativação de camadas ocultas: ReLu, cujo cálculo é simples e permite convergência rápida durante o treinamento (ZHANG, L.; WANG, S.; LIU, B., 2018);
- Função de ativação da camada de saída: *sigmoid* (i.e., função logística), que retorna um valor no intervalo entre 0 e 1, representando o pertencimento da saída a uma determinada classe.

Para avaliar os resultados, foram utilizadas as seguintes métricas:

- *F1-Score*: média harmônica entre as métricas precisão e revocação. Enquanto a precisão é dada pela relação de amostras classificadas corretamente em uma determinada classe e todas as amostras classificadas como pertencentes àquela classe, a revocação é dada

pela a relação entre as amostras classificadas como uma determinada classe e o total de amostras que fazem parte dessa classe.

- *Acurácia*: número de instâncias classificadas corretamente dividido pelo número total de instâncias.

Os resultados obtidos para o conjunto de testes, após o treinamento e seleção dos modelos por meio da aplicação do método *K-Fold Cross Validation* ($k = 10$), são apresentados na Tabela 7.2.

Tabela 7.2: Resultados da tarefa de AS. Conjunto de dados de testes.

Algoritmo	F1-Score	Acurácia
Naive Bayes	67,2%	70,6%
Logistic Regression	69,8%	71,7%
SVM	69,3%	72,0%
MLP	80,9%	81,2%
LSTM	82,9%	83,9%
CNN	86,3%	86,5%

Os resultados mostraram que, dentre os modelos comparados, a arquitetura mais adequada para a tarefa de Análise de Sentimentos em português é a CNN, inicialmente apresentada em Ye Zhang e Wallace (2017). O modelo apresentou 86,3% de F1-Score e 86,5% de precisão, os maiores dentre todas as arquiteturas testadas. Dessa forma, esse modelo foi o escolhido para o restante dos experimentos deste trabalho quando envolverem AS. Considerando as demais arquiteturas, os parâmetros que obtiveram os melhores resultados para AS foram: MLP com 3 camadas e 5 neurônios por camada; LSTM com 1 camada com 5 neurônios por camada.

Dessa forma, os resultados dessa etapa do trabalho reiteram algumas observações apontadas por Vargas, B. S. Lima e Evsukoff (2017) de que, normalmente, redes do tipo CNN obtêm melhores resultados do que RNAs do tipo recorrentes na análise da semântica de texto. Além disso, os autores também apontam que RNAs do tipo recorrentes, por outro lado, se mostram melhores considerando características de dependências temporais, como previsão do mercado de ações. Ainda, assim como observado também por Qian, Zirao Li e Yuan (2020) e Igor Santos, Nedjah e Macedo Mourelle (2017), é possível verificar que os modelos de RNAs do tipo CNN têm se apresentado como uma solução promissora na classificação de textos.

7.3 Módulo de Predição

Esta seção apresenta os resultados de nossos experimentos considerando a previsão do mercado financeiro brasileiro e está organizada da seguinte forma: (i) metodologia experimental, descrevendo os detalhes e parâmetros utilizados em cada experimento, apresentada na Seção 7.3.1; (ii) avaliação do modelo proposto segundo as métricas de Aprendizado de Máquina, apresentadas na Seção 7.3.2; (iii) avaliação dos modelos selecionados de acordo com as métricas financeiras e a simulação de investimento, apresentadas na Seção 7.3.3.

7.3.1 Metodologia Experimental

A metodologia desse experimento envolve a definição de parâmetros para: indicadores técnicos, treinamento das RNAs e simulação de investimento, conforme apresentado nas subseções a seguir.

Indicadores Técnicos

Os indicadores técnicos utilizados nos experimentos foram configurados por parâmetros diferentes, para refletir tendências de mercado ou flutuações considerando diferentes aspectos. Os parâmetros foram escolhidos considerando o trabalho de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), juntamente com indicadores que indicam volatilidade e volume (ACHELIS, 2001) para melhor representação dos diferentes aspectos do mercado. Assim, os seguintes parâmetros foram escolhidos para os indicadores técnicos considerados neste trabalho:

- **MACD**: com período de 12 e 26 dias e sinal de 9 dias.
- **Médias Móveis**, com período de 10, 20 e 30 dias.
- **RSI**: com os parâmetros: 6, 12 e 24 dias.
- **MFI**: com período de tempo 14.
- **OBV e Bollinger Bands**: sem a necessidade de parâmetros adicionais.

Treinamento de RNAs

Para desenvolver o processo de treinamento, utilizamos o MSE como função de erro. MSE é uma das métricas comumente usadas para tarefas de regressão (MAQSOOD et al., 2020) e é dado na Eq. 6.3.

As RNAs foram treinadas com o algoritmo *Root Mean Square Prop* (RMSProp), que geralmente é uma boa escolha para redes neurais conforme apresentado em Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020). Os seguintes parâmetros foram fixados: a taxa de aprendizado inicial foi definida como 0,0001; o tamanho do *batch* fornecido à rede foi definido em 32; o número de épocas foi definido como 500 e o treinamento foi realizado com um mecanismo de parada precoce (do inglês, *early stopping*), que interrompe o treinamento se a função de erro parar de reduzir por 10 épocas. A função de ativação foi definida como ReLu, pois é fácil de computar e acelera a convergência do treinamento de acordo com Lei Zhang, Shuai Wang e Bing Liu (2018).

Neste módulo, também verificamos as seguintes RNAs, com complexidade crescente, para encontrar a arquitetura que melhor executa a tarefa de regressão para o nosso problema: MLP, LSTM e Bi-LSTM. Além disso, para encontrar os hiperparâmetros mais adequados para as RNAs testadas, usamos a busca em grade (do inglês, *grid search*) juntamente com um método de validação cruzada de janela crescente para encontrar os valores ideais dentre os seguintes parâmetros:

- Número de camadas: 3 e 5.
- Número de neurônios por camada: 5, 10, 20, 50 e 100.
- *Dropout* (mecanismo de regularização para evitar *overfitting* em RNAs (LI, X.; WU, P.; WANG, W., 2020)): 0, 0,125, 0,375, 0,25 e 0,5.

Uma vez que as RNAs são projetadas para prever séries temporais de preços de ações, a etapa de treinamento segue o método *increasing window cross-validation*, conforme apresentado em Ratto et al. (2018) e também em Picasso et al. (2019) e Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020). Neste trabalho, usamos a validação cruzada dividida em três passos (do inglês, *step*) no tempo para treinar e validar cada modelo proposto, conforme apresentado na Figura 7.1. Nessa abordagem, a série temporal é dividida em k partes para treinar e a divisão de $k + 1$ para validar a rede. Em seguida, selecionamos a

configuração do modelo que apresenta melhor desempenho considerando a média de todos os conjuntos de validação para medir seus resultados no conjunto de teste, conforme apresentado na próxima seção.



Figura 7.1: Algoritmo *increasing window cross-validation* considerando a repartição da série temporal em 3 divisões. Adaptado de (LI, X.; WU, P.; WANG, W., 2020).

O particionamento do conjunto de dados em conjunto de treinamento e testes para treinamento das RNAs desse módulo é o seguinte. A porção que abrange o período de janeiro de 2010 a dezembro de 2016, é considerado conjunto de treinamento; a porção de janeiro de 2017 a dezembro de 2019, é considerado o conjunto de teste, ou seja, são dados fora da amostra de treinamento para medir o desempenho de cada modelo.

Como *baseline* para nossos experimentos, usamos *Support Vector Machines* (SVM), assim como em Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), Picasso et al. (2019) e Maqsood et al. (2020). O SVM é um dos métodos mais populares de Aprendizado de Máquina aplicados ao campo financeiro. No caso desta tese, utilizamos o algoritmo voltado para a tarefa de regressão, denominado também como *Support Vector Regression*. Para encontrar a configuração mais adequada ao algoritmo, foi utilizado o Kernel Gaussiano (assim como em Picasso et al. (2019)) e foi realizada a busca dentro dos seguintes valores do parâmetro C : 10^{-3} , $10^{-2,5}$, 10^{-2} , ..., $10^{2,5}$, 10^3 .

Simulação de Investimento

Em relação à etapa de simulação de investimento, consideramos um investimento inicial de R\$ 100.000,00 no início do período dos dados de teste (i.e., janeiro de 2017). Como taxa de negociação, foi levado em conta o valor de R\$ 2,50 por cada operação de compra ou venda. O baseline considerado é o *Buy & Hold*, comumente considerado em trabalhos da área

(KUMBURE et al., 2022), e que consiste em comprar um ativo e mantê-lo sob posse do investidor durante todo o período considerado no experimento.

Em nossos experimentos, testamos como valores de segurança para o Algoritmo 1 o seguinte intervalo: 0%, 1%..., 5%. Nesse caso, consideramos dois tipos de investidores: os *conservadores* (ou seja, com os valores da faixa de segurança maiores, que consideram investir em um ambiente menos arriscado, mesmo que tenham retornos menores) e os *arrojados* (ou seja, com a faixa de segurança menores, consideram investir em um ambiente de risco, mesmo que tenham algumas perdas no processo). Essa abordagem foi considerada levando-se em conta resultados de um trabalho prévio em que notamos que diferentes tipos de perfis de investimento podem trazer vantagens quanto ao lucro obtido Carosia, Coelho e A. E. A. Silva (2021).

Ainda, nos experimentos dessa fase, consideramos três cenários para investigar se as notícias de fato melhoram os resultados em um sistema de previsão do mercado de ações.

- O primeiro cenário usa como entrada para uma RNA apenas os preços históricos das ações, sem notícias e indicadores técnicos.
- O segundo cenário usa como entrada os preços históricos das ações e os indicadores técnicos.
- O terceiro cenário usa como entrada todas as fontes disponíveis: notícias, indicadores técnicos e históricos de preços de ações.

7.3.2 Métricas de Aprendizado de Máquina

Para executar a metodologia proposta neste trabalho, utilizamos uma rede neural CNN para realizar a Análise de Sentimentos, conforme publicado anteriormente em Carosia, Coelho e A. E. A. Silva (2021) e discutido nos resultados da seção anterior.

Assim, a primeira etapa de nossos experimentos consiste em avaliar as métricas da área de Aprendizado de Máquina. Os resultados na Tabela 7.3 foram obtidos a partir do *baseline*, *Support Vector Machines* (SVM). As tabelas 7.4 a 7.8 apresentam os resultados obtidos com as RNAs profundas e os respectivos parâmetros da arquitetura das RNAs que tiveram melhor desempenho nos experimentos. As métricas consideradas neste caso são MAE e MSE.

Para realizar esses experimentos, consideramos três cenários de dados de entrada para os preditores: (1) apenas os preços históricos das ações (chamado de *OHLCV*, representando

nos resultados o preço de abertura, alta, baixa, fechamento e volume, do inglês, *Open, High, Low, Close, Volume*); (2) os preços históricos das ações e os indicadores técnicos (chamado de *Tech. Ind.* nos resultados); e (3) notícias, indicadores técnicos e históricos de preços das ações (chamado de *Sentiment* nos resultados).

Tabela 7.3: Métricas de Aprendizado de Máquina considerando o baseline SVM aplicado a todas ações.

Ação	Entrada	MAE	MSE
BVSP	OHLCV	0,8939	1,3987
	Tech. Ind.	0,8977	1,4097
	Sentiment	1,1096	2,0054
BBAS3	OHLCV	0,9061	1,5436
	Tech. Ind.	0,9037	1,5383
	Sentiment	0,9042	1,5388
ITUB4	OHLCV	1,0383	1,8818
	Tech. Ind.	0,9745	1,7051
	Sentiment	0,9715	1,698
GGBR4	OHLCV	0,7199	0,8844
	Tech. Ind.	0,7167	0,8837
	Sentiment	0,7365	0,9347
ABEV3	OHLCV	0,8259	1,2955
	Tech. Ind.	0,8366	1,3276
	Sentiment	0,8411	1,3325

Os resultados mostram que, nos experimentos, o uso de preços de ações ao lado de Indicadores Técnicos e sentimentos de notícias melhora a capacidade de predição das RNAs profundas, superando o *baseline* SVM. Na verdade, tanto MAE quanto MSE são mais baixos quando sentimentos de notícias também são usados (nos conjuntos de dados BVSP, BBAS3, ITUB4 e ABEV3), exceto para a empresa GGBR4, em que o uso de apenas o preço das ações levou a um melhor desempenho com uma arquitetura Bi-LSTM. Vale ressaltar que a arquitetura que apresentou melhor desempenho na maioria dos casos foi Bi-LSTM, em 4 conjuntos de dados; e a rede LSTM teve melhor desempenho em apenas 1 conjunto de dados. Esse resultado é esperado, uma vez que ambas as arquiteturas são projetadas para lidar com dados de séries temporais e Bi-LSTMs são uma evolução em relação às redes LSTMs.

Por fim, a Tabela 7.9 apresenta o Teste U não paramétrico de Mann-Whitney (CORDER; FOREMAN, 2014), utilizado para verificar a significância estatística entre os modelos desenvolvidos, considerando a métrica MSE e uma tolerância de 5%. O teste foi aplicado

Tabela 7.4: Métricas de Aprendizado de Máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Ibovespa (BVSP).

Ação	Entrada	Modelo	MAE	RMSE	Camadas	Neurônios	Dropout
BVSP	OHLCV	MLP	0,8836	1,1695	3	100	0,125
		LSTM	0,8834	1,1692	5	5	0
		BILSTM	0,8837	1,1696	5	5	0,25
	Tech. Ind.	MLP	0,8828	1,1687	3	100	0,125
		LSTM	0,8825	1,1684	5	5	0
		BILSTM	0,8833	1,1695	5	5	0,5
	Sentiment	MLP	0,8814	1,1668	3	100	0,125
		LSTM	0,8810	1,1665	3	5	0,125
		BILSTM	0,8809	1,1664	3	5	0

Tabela 7.5: Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Banco do Brasil (BBAS3).

Ação	Entrada	Modelo	MAE	RMSE	Camadas	Neurônios	Dropout
BBAS3	OHLCV	MLP	0,9054	1,2436	3	100	0,125
		LSTM	0,9041	1,2421	5	5	0,5
		BILSTM	0,9044	1,2424	5	5	0,5
	Tech. Ind.	MLP	0,9048	1,2434	3	100	0,125
		LSTM	0,9040	1,2423	5	5	0
		BILSTM	0,9031	1,2412	3	10	0,5
	Sentiment	MLP	0,9033	1,2412	3	100	0,125
		LSTM	0,9029	1,2408	5	5	0,5
		BILSTM	0,9018	1,2394	5	5	0,5

Tabela 7.6: Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Itau (ITUB4).

Ação	Entrada	Modelo	MAE	RMSE	Camadas	Neurônios	Dropout
ITUB4	OHLCV	MLP	0,9755	1,3089	3	100	0,125
		LSTM	0,9741	1,3071	3	10	0,25
		BILSTM	0,9743	1,3073	5	5	0
	Tech. Ind.	MLP	0,9889	1,3253	3	100	0,125
		LSTM	0,9887	1,3251	5	5	0
		BILSTM	0,9900	1,3259	5	5	0,125
	Sentiment	MLP	0,9710	1,3028	3	100	0,125
		LSTM	0,9708	1,3026	5	5	0
		BILSTM	0,9724	1,3043	5	5	0,5

Tabela 7.7: Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Gerdau (GGBR4).

Ação	Entrada	Modelo	MAE	RMSE	Camadas	Neurônios	Dropout
GGBR4	OHLCV	MLP	0,7106	0,9296	3	100	0,125
		LSTM	0,7105	0,9293	3	50	0,125
		BILSTM	0,7100	0,9282	3	100	0,125
	Tech. Ind.	MLP	0,7123	0,9317	3	100	0,125
		LSTM	0,7123	0,9317	5	5	0
		BILSTM	0,7118	0,9305	3	5	0,125
	Sentiment	MLP	0,7133	0,9324	3	100	0,125
		LSTM	0,7132	0,9323	3	20	0
		BILSTM	0,7135	0,9324	3	5	0

Tabela 7.8: Métricas de aprendizado de máquina de RNAs profundas: MLP, LSTM e Bi-LSTM. Ambev (ABEV3).

Ação	Entrada	Modelo	MAE	RMSE	Camadas	Neurônios	Dropout
ABEV3	OHLCV	MLP	0,8314	1,1448	5	5	0,375
		LSTM	0,8311	1,1449	5	5	0,5
		BILSTM	0,8328	1,1458	5	5	0,5
	Tech. Ind.	MLP	0,8775	1,3014	5	5	0,5
		LSTM	0,8918	1,3668	5	5	0,5
		BILSTM	0,9032	1,3983	5	5	0,5
	Sentiment	MLP	0,8247	1,1406	3	100	0,125
		LSTM	0,8234	1,1395	5	5	0
		BILSTM	0,8229	1,1379	3	20	0,125

comparando a metodologia proposta neste trabalho (no experimento anterior, apresentada como *Sentiment*, juntamente com RNAs profundas) e cada uma das demais abordagens (no experimento anterior, apresentadas como: RNAs profundas com OHLCV e Tech. Ind.; e SVM com OHLCV, Tech. Ind. e *Sentiment*). Para as abordagens baseadas em RNAs profundas, são consideradas as RNAs que obtiveram menores medidas de erros no experimento anterior.

Os resultados mostram que, considerando a maior parte dos casos, a métrica resultante obtida pela metodologia proposta neste trabalho é estatisticamente diferente dos demais classificadores. No entanto, o classificador baseado em RNAs profundas, valores históricos dos preços e indicadores técnicos (identificado como (DL) Tech. Ind.) é aquele que em mais situações (i.e., para as ações BVSP, BBAS3 e GGBR4) não difere estatisticamente da metodologia proposta nesta tese. Por outro lado, é importante ressaltar que, para esse caso, observamos na etapa de simulação de investimento, detalhada na Seção 7.3.3, que a metodologia proposta nesta tese supera significativamente o modelo mencionado, o que justifica o uso da metodologia aqui proposta na prática. Finalmente, é possível verificar que existe um maior número de ações em que há diferenças estatísticas significativas entre a metodologia proposta e os demais modelos comparados, sejam baseados no modelo SVM ou mesmo RNAs profundas.

7.3.3 Métricas de Investimento

Uma vez que os modelos adequados foram selecionados na seção anterior, projetamos uma simulação de investimento de R\$ 100.000,00, no período de janeiro de 2017 a dezembro de 2019 no mercado de ações brasileiro (ou seja, o conjunto de teste), conforme explicado na Seção 6.2.4. Ainda, foi considerado como custo de operação o valor de R\$ 2,50 (taxas e corretagem). O primeiro *baseline* do experimento é a estratégia de investimento *Buy & Hold*, cujos resultados são apresentados na Tabela 7.10.

O resultado do modelo *baseline* deste experimento, SVM, é apresentado na Tabela 7.11, considerando as métricas valor final de investimento, MDD e SR. Finalmente, os resultados das RNAs profundas são apresentados nas Tabelas 7.12 a 7.16.

Os resultados obtidos nessa etapa mostram que a combinação de preços de ações, indicadores técnicos e o sentimento de notícias também melhora os resultados dos investimentos no mercado de ações brasileiro, superando os *baselines Buy & Hold* e SVM considerando valor final de investimento, MDD e SR. Na verdade, o valor final de

Tabela 7.9: Teste estatístico de Mann-Whitney para a métrica MSE considerando a tolerância de 5%, realizado entre a metodologia proposta nesta tese e as demais abordagens, tanto com RNAs profundas (identificado por DL) quanto com o modelo SVM. O marcador X representa que há significância estatística, enquanto que o marcador – representa ausência de significância estatística.

BVSP				
(SVM) OHLCV	(SVM) Tech. Ind.	(SVM) Sentiment	(DL) OHLCV	(DL) Tech. Ind.
X	X	X	X	–
BBAS3				
(SVM) OHLCV	(SVM) Tech. Ind.	(SVM) Sentiment	(DL) OHLCV	(DL) Tech. Ind.
X	X	X	X	–
ITUB4				
(SVM) OHLCV	(SVM) Tech. Ind.	(SVM) Sentiment	(DL) OHLCV	(DL) Tech. Ind.
X	X	X	–	X
GGBR4				
(SVM) OHLCV	(SVM) Tech. Ind.	(SVM) Sentiment	(DL) OHLCV	(DL) Tech. Ind.
X	X	–	–	–
ABEV3				
(SVM) OHLCV	(SVM) Tech. Ind.	(SVM) Sentiment	(DL) OHLCV	(DL) Tech. Ind.
X	X	X	X	X

Tabela 7.10: Simulação de investimento considerando a estratégia Buy & Hold.

Ação	Valor Final	Maximum Drawdown	Sharpe Ratio
BVSP	R\$ 194.606,39	0.04	2.39
BBAS3	R\$ 191.793,75	0.141	0.89
ITUB4	R\$ 165.971,01	0.101	0.63
GGBR4	R\$ 184.501,85	0.201	1.94
ABEV3	R\$ 114.469,66	0.21	-0.01

Tabela 7.11: Simulação de investimento considerando o baseline SVM.

Ação	Entrada	Valor Final	Maximum Drawdown	Sharpe Ratio
BVSP	OLHCV	R\$ 99.005,97	0,029	-1,83
	Tech. Ind.	R\$ 128.835,80	0,058	0,13
	Sentiment	R\$ 101.983,84	0,07	0,00
BBAS3	OLHCV	R\$ 124.937,40	0,015	0,32
	Tech. Ind.	R\$ 110.916,08	0,029	-0,36
	Sentiment	R\$ 104.654,59	0,050	0,00
ITUB4	OLHCV	R\$ 165.971,01	0,101	0,63
	Tech. Ind.	R\$ 176.008,45	0,090	0,96
	Sentiment	R\$ 175.908,37	0,090	0,96
GGBR4	OLHCV	R\$ 181.831,35	0,140	0,59
	Tech. Ind.	R\$ 133.374,81	0,194	0,20
	Sentiment	R\$ 86.822,29	0,199	-1,26
ABEV3	OLHCV	R\$ 98.152,00	0,180	-0,23
	Tech. Ind.	R\$ 134.242,23	0,129	0,19
	Sentiment	R\$ 140.401,48	0,145	0,23

investimento aumenta quando são usados sentimentos de notícias (nos conjuntos de dados BVSP, BBAS3, ITUB4 e GGBR4), exceto para a empresa ABEV3, em que a entrada considerando apenas o preço das ações teve melhor desempenho com uma arquitetura MLP. Nesse sentido, é possível verificar que o uso de sentimentos aplicados a notícias acrescenta informações importantes para o modelo no sentido de prever a direção do mercado acionário.

As arquiteturas com melhor desempenho foram: MLP e Bi-LSTM. Vale a pena observar que, embora uma arquitetura tenha um desempenho melhor considerando as métricas de Aprendizado de Máquina, isso não significa que a mesma arquitetura levará a melhores resultados considerando as métricas financeiras, ou seja, quando aplicado ao domínio da aplicação. Esse é um ponto importante a se observar, visto que muitos trabalhos na literatura analisam apenas métricas de Aprendizado de Máquina, e não realizam a etapa de simulação de investimento.

Além disso, os resultados mostram que existe uma compensação entre o valor final de investimento e outras métricas financeiras, como SR e MDD. Por exemplo, ITUB4 apresenta melhor valor final de investimento com sentimentos de notícias, mas melhores medidas de risco SR e MDD com Indicadores Técnicos. Esses valores demonstram que o investidor precisa

Tabela 7.12: Simulação de investimento usando os modelos de RNAs profundas. Ibovespa (BVSP).

Ação	Entrada	Modelo	Valor Final	Maximum Drawdown	Sharpe Ratio
BVSP	OHLCV	MLP	R\$ 199.591,19	0,050	2,32
		LSTM	R\$ 194.606,39	0,04	2,39
		BILSTM	R\$ 194.606,39	0,04	2,39
	Tech. Ind.	MLP	R\$ 196.167,47	0,063	3,70
		LSTM	R\$ 194.606,39	0,04	2,39
		BILSTM	R\$ 194.606,39	0,04	2,39
	Sentiment	MLP	R\$ 220.303,43	0,04	1,16
		LSTM	R\$ 239.513,37	0,059	1,14
		BILSTM	R\$ 274.583,62	0,04	5,61

Tabela 7.13: Simulação de investimento usando os modelos de RNAs profundas. Banco do Brasil (BBAS3).

Ação	Entrada	Modelo	Valor Final	Maximum Drawdown	Sharpe Ratio
BBAS3	OLHCV	MLP	R\$ 241.724,86	0,068	1,981
		LSTM	R\$ 223.654,98	0,068	1,690
		BILSTM	R\$ 191.793,75	0,141	0,886
	Tech. Ind.	MLP	R\$ 252.152,85	0,056	0,653
		LSTM	R\$ 242.307,09	0,094	0,881
		BILSTM	R\$ 248.908,77	0,097	0,650
	Sentiment	MLP	R\$ 263.786,82	0,179	0,717
		LSTM	R\$ 206.585,53	0,127	1,628
		BILSTM	R\$ 262.968,51	0,126	1,265

Tabela 7.14: Simulação de investimento usando os modelos de RNAs profundas. Itau (ITUB4).

Ação	Entrada	Modelo	Valor Final	Maximum Drawdown	Sharpe Ratio
ITUB4	OLHCV	MLP	R\$ 183.344,62	0,101	0,733
		LSTM	R\$ 173.266,43	0,090	0,839
		BILSTM	R\$ 165.971,01	0,101	0,633
	Tech. Ind.	MLP	R\$ 202.182,70	0,034	1,133
		LSTM	R\$ 186.396,09	0,035	0,854
		BILSTM	R\$ 184.037,13	0,101	0,812
	Sentiment	MLP	R\$ 203.882,99	0,085	0,753
		LSTM	R\$ 195.289,24	0,048	0,869
		BILSTM	R\$ 180.524,40	0,030	0,769

Tabela 7.15: Simulação de investimento usando os modelos de RNAs profundas. Gerdau (GGBR4).

Ação	Entrada	Modelo	Valor Final	Maximum Drawdown	Sharpe Ratio
GGBR4	OLHCV	MLP	R\$ 217.859,74	0,194	0,814
		LSTM	R\$ 184.501,85	0,201	1,944
		BILSTM	R\$ 184.749,99	0,144	0,793
	Tech. Ind.	MLP	R\$ 212.697,50	0,144	0,651
		LSTM	R\$ 184.501,85	0,201	1,944
		BILSTM	R\$ 184.501,85	0,201	1,944
	Sentiment	MLP	R\$ 230.759,73	0,117	1,151
		LSTM	R\$ 184.501,85	0,201	1,944
		BILSTM	R\$ 184.501,85	0,201	1,944

Tabela 7.16: Simulação de investimento usando os modelos de RNAs profundas. Ambev (ABEV3).

Ação	Entrada	Modelo	Valor Final	Maximum Drawdown	Sharpe Ratio
ABEV3	OLHCV	MLP	R\$ 174.370,72	0,057	0,412
		LSTM	R\$ 121.516,78	0,059	-0,124
		BILSTM	R\$ 119.204,14	0,202	0,029
	Tech. Ind.	MLP	R\$ 155.966,86	0,079	0,369
		LSTM	R\$ 124.058,05	0,194	0,071
		BILSTM	R\$ 138.095,74	0,058	0,226
	Sentiment	MLP	R\$ 165.749,39	0,055	0,514
		LSTM	R\$ 128.233,59	0,105	0,147
		BILSTM	R\$ 122.883,51	0,195	0,060

analisar cada conjunto de dados separadamente e escolher qual arquitetura se encaixa melhor em seu perfil de investimento. Se o investidor estiver disposto a aceitar riscos maiores (i.e., investimentos cujos valores de MDD são maiores e SR são menores), ele poderá escolher a arquitetura MLP com sentimentos de notícias, que trará melhor valor final de investimento com possibilidade de perdas maiores no período; no caso de uma estratégia de investimento conservadora (i.e., investimentos cujos valores de MDD são menores e SR são maiores), pode-se optar por uma arquitetura MLP com Indicadores Técnicos, que podem trazer valor final de investimento inferior, mas com menos risco do que as outras opções.

A Seção 7.5 amplia a discussão considerando os resultados obtidos nos experimentos.

7.4 Comparação Experimental com Trabalhos Correlatos

Esta seção apresenta a comparação experimental dos resultados obtidos neste trabalho com os resultados de trabalhos relacionados que mais se aproximam da abordagem desenvolvida nesta tese.

Foram selecionados cinco trabalhos correlatos, apresentados no Capítulo 5, a saber: Hájek (2018), Picasso et al. (2019), Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), Duarte, González e Cruz (2021) e Shi et al. (2021).

Para os experimentos, algumas etapas dos trabalhos relacionados foram adaptadas para a língua portuguesa, visto que são trabalhos que utilizam, principalmente, notícias em inglês. Nesse sentido, quando necessário utilizamos estratégias baseadas em dicionários em português, conforme proposto em uma publicação prévia nossa (JANUÁRIO et al., 2021), em que observamos que uma estratégia combinada dos seguintes dicionários de palavras obtém melhor resultado no domínio financeiro em tarefas de AS: SentiLex-PT/BR (SILVA, M. J. et al., 2010) e a tradução do dicionário específico do domínio financeiro de Loughran e McDonald (2011).

Além disso, destaca-se que a abordagem utilizada para essa comparação foi a seguinte. Primeiramente, selecionamos o modelo de entrada de dados utilizado pelo trabalho correlato. Em segundo lugar, verificamos qual o método de extração de características de notícias para entrada no modelo (no caso, observamos que os trabalhos em sua maioria utilizam dicionários léxicos). Assim, treinamos o modelo proposto no trabalho correlato e o testamos com os mesmos dados (e o mesmo particionamento de dados) utilizados em nossos

Tabela 7.17: Resultado dos trabalhos correlatos (HÁJEK, 2018) e (PICASSO et al., 2019), utilizando uma RNA do tipo MLP e método léxico.

Ação	Valor Final	MDD	SR	MAE	RMSE
BVSP	R\$ 166.734,96	0,04	1,437	0,940	1,231
BBAS3	R\$ 184.326,85	0,028	0,329	1,124	1,489
ITUB4	R\$ 153.041,86	0,09	2,596	1,202	1,599
GGBR4	R\$ 217.278,89	0,149	3,760	1,108	1,492
ABEV3	R\$ 158.848,48	0,045	0,376	1,108	1,490

experimentos, ou seja, dados de notícias e ações compreendendo de janeiro de 2010 a dezembro de 2019. Por fim, apresentamos aqui os resultados de simulação de investimento de cada um dos trabalhos correlatos considerados levando em conta o conjunto de dados de teste. Para todos os trabalhos correlatos também realizamos uma busca por parâmetros do modelo testado, quando necessário.

O primeiro comparativo foi feito com os trabalhos Hájek (2018) e Picasso et al. (2019), em que são utilizadas RNAs do tipo MLP para previsão do mercado financeiro. Ainda, esses trabalhos empregam dicionários financeiros para o desenvolvimento da etapa de AS e, como entrada, são inseridos na rede neural informações extraídas sobre notícias e séries históricas de ações. Os resultados da execução desse experimento (simulação de investimento), são apresentados na Tabela 7.17. Vale destacar que os resultados para as ações GGBR4 e ABEV3 se mostraram superiores aos resultados do *baseline Buy & Hold* (Tabela 7.10), apesar de ainda serem inferiores aos melhores resultados obtidos pela metodologia proposta nesta tese (Tabelas 7.15 e 7.16).

O segundo experimento comparativo foi desenvolvido com base no trabalho de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020), em que há uma RNA do tipo LSTM sendo alimentada com o sentimento de notícias extraídos com dicionários, indicadores técnicos e série de preços. No entanto, vale destacar que esse trabalho correlato não realiza simulação de investimentos, se limitando a apenas analisar as métricas de Aprendizado de Máquina. Os resultados da simulação de investimento desse trabalho são apresentados na Tabela 7.18. É possível verificar que os resultados do trabalho de Xiaodong Li, Pangjing Wu e Wenpeng Wang (2020) são inferiores aos resultados desta tese (Tabelas 7.12 a 7.16), mas superam o *baseline Buy & Hold* (Tabela 7.10) para a ação ABEV3.

Ainda, avaliamos aqui também a metodologia proposta por Duarte, González e Cruz (2021), que é o trabalho que mais se aproxima ao nosso e que, além disso, abrange o mercado

Tabela 7.18: Resultados do trabalho correlato (LI, X.; WU, P.; WANG, W., 2020), considerando uma RNA do tipo LSTM e método léxico.

Ação	Valor Final	MDD	SR	MAE	RMSE
BVSP	R\$ 188.721,65	0,04	1,872	0,905	1,197
BBAS3	R\$ 157.456,70	0,147	0,25	1,101	1,734
ITUB4	R\$ 165.971,01	0,101	0,633	1,243	1,890
GGBR4	R\$ 114.189,75	0,188	0,028	1,166	1,758
ABEV3	R\$ 139.607,74	0,135	0,219	1,116	1,672

Tabela 7.19: Resultados do trabalho correlato (DUARTE; GONZÁLEZ; CRUZ, 2021) considerando o modelo MLP.

Ação	Valor Final	MDD	SR	MAE	RMSE
BVSP	R\$ 131.327,86	0,1778	0,011	1,443	1,922
BBAS3	R\$ 86.824,32	0,14	0,310	1,068	1,431
ITUB4	R\$ 111.144,43	0,45	0,189	1,880	2,308
GGBR4	R\$ 108.075,03	0,32	0,261	0,776	1,010
ABEV3	R\$ 106.101,02	0,152	0,136	0,870	1,179

brasileiro. Nesse sentido, selecionamos um dos principais modelos utilizados por Duarte, González e Cruz (2021), a rede MLP, e utilizamos como dados de entrada a frequência dos termos juntamente com indicadores técnicos e a série histórica de ações. Esse trabalho, diferentemente dos anteriores, não utiliza dicionários, se limitando à contagem de termos presentes nas notícias. Os resultados desse experimento, apresentados na Tabela 7.19, indicam que, embora seja possível obter lucro nas ações consideradas (exceto BBAS3), os resultados são inferiores tanto aos resultados deste trabalho de doutorado (Tabelas 7.12 a 7.16) quanto aos resultados do *baseline Buy & Hold* (Tabela 7.10).

Ainda, vale a pena considerar que o trabalho correlato de Shi et al. (2021) foi implicitamente comparado nos resultados da seção anterior nos experimentos com o modelo SVM, visto que o trabalho utiliza rede CNN para AS e o modelo SVM para predição do mercado. Nesse caso, também observamos que a metodologia proposta aqui supera os resultados obtidos no trabalho correlato. Os seus resultados estão apresentados nas Tabelas 7.3 e 7.11 e mostram que o *baseline Buy & Hold* (Tabela 7.10) pode ser superado nas ações ITUB4 e ABEV3. No entanto, ainda são inferiores aos resultados obtidos aqui nesta tese.

Como é possível observar, em sua maioria, os trabalhos correlatos apresentaram resultados inferiores aos obtidos pela metodologia proposta aqui. Alguns pontos podem ser considerados.

Primeiramente, é preciso observar que, para certas ações, esses trabalhos apresentam lucro significativo, além de superarem o *baseline Buy & Hold*. Isso mostra que suas abordagens são adequadas ao domínio da aplicação à qual estão voltados, apesar de serem superados pelos resultados obtidos pela metodologia proposta aqui.

Como a maioria desses trabalhos não utiliza Aprendizado Profundo para realizar a tarefa de AS, é provável que esse seja o principal fator de seus resultados inferiores. Além disso, a linguagem que a maioria destes trabalhos emprega é o inglês, que possui mais recursos de PLN disponíveis do que a língua portuguesa, objeto do presente estudo. É justo afirmar que a portabilidade desses trabalhos de língua inglesa ao português pode ter sido prejudicial a alguns de seus resultados.

Além disso, esses trabalhos utilizam em sua maioria dicionários léxicos, que conforme estudo anterior (JANUÁRIO et al., 2021), tendem a apresentar resultados significativamente inferiores aos métodos baseados em Aprendizado de Máquina em tarefas de AS voltadas ao domínio financeiro, algo observado em Avanço e Nunes (2014).

Além disso, no presente trabalho realizamos um comparativo dentre diferentes tipos de arquiteturas e modelos de RNAs. Nos trabalhos correlatos, normalmente há a escolha de um modelo fixo considerando pouca variação de parâmetros, o que pode impactar negativamente os resultados obtidos.

Por fim, o trabalho de Duarte, González e Cruz (2021), que não utiliza dicionários, faz uso de termos de frequência para alimentar os modelos testados. Embora consideramos que essa seja uma estratégia interessante, pois os dicionários em língua portuguesa ainda apresentam resultados inferiores aos métodos de Aprendizado de Máquina, acreditamos que existam outras abordagens que sejam capazes de incrementar tal extração de características, como *Word2Vec* e *FastText*.

A seção a seguir amplia as discussões deste trabalho.

7.5 Discussão

O trabalho de Ozbayoglu, Gudelek e Sezer (2020) mostrou que a mineração de textos financeiros tem recebido cada vez mais atenção na literatura, sendo uma das áreas mais promissoras nos últimos anos. No entanto, há ainda uma questão em aberto na literatura: como combinar as notícias em uma previsão do mercado de ações, ao lado de dados

financeiros? Neste trabalho, propusemos uma abordagem que incorpora sentimentos de notícias, indicadores técnicos, volume e preços de ações em uma metodologia completa. Todo o processo foi realizado com redes neurais profundas, ou seja, tanto o cálculo dos sentimentos das notícias quanto a previsão do mercado de ações. Por fim, propusemos uma estratégia de investimento simples que considera um fator de risco apresentado pelos usuários.

Com relação à combinação de notícias e valores das ações, existem trabalhos recentes que começaram a abordar tal combinação de fontes de dados (LI, X.; WU, P.; WANG, W., 2020; PICASSO et al., 2019; MAQSOOD et al., 2020; DUARTE; GONZÁLEZ; CRUZ, 2021; SHI et al., 2021). Eles também encontraram resultados semelhantes aos nossos no que diz respeito à importância das notícias como parte de um sistema de previsão. De fato, conforme apresentado em Maqsood et al. (2020), a Hipótese de Mercado Adaptativo afirma que existe uma correlação entre o comportamento do mercado e o sentimento público. Ou seja, a probabilidade de maiores investimentos no mercado é maior quando o ânimo do público é positivo. Como tal, a tarefa principal dos trabalhos da área deve ser medir o sentimento do público para prever mudanças no comportamento do mercado.

Além disso, verificamos em um trabalho anterior (CAROSIA; COELHO; SILVA, A. E. A., 2021) que, à medida que notícias são divulgadas em jornais *on-line*, há influência nas operações dos investidores na bolsa de valores, o que, conseqüentemente, afeta a cotação de fechamento do mesmo dia, além de influenciar também no sentimento predominante das notícias do dia seguinte. Verificamos ainda que o sentimento predominante nas notícias em um determinado dia d afeta o valor de abertura do dia $d + 1$, mas apresenta impacto menos significativo no valor de fechamento do dia seguinte. Isso provavelmente se deve ao fato de ocorrerem eventos no dia $d + 1$ que podem mudar a direção do mercado considerando o próximo valor de fechamento. Assim, foi possível verificar que existe uma relação mútua entre notícias e mercado financeiro. Isto é, por um lado o sentimento contido em notícias influencia o sentido do mercado de ações; por outro lado, a direção do mercado de ações influencia o sentimento das notícias.

Em um dos poucos trabalhos que consideram o mercado de ações brasileiro no que se refere a notícias e bolsa de valores, Duarte, González e Cruz (2021) enfatizam o alto fator de risco do mercado brasileiro, que também pode estar atrelado a maiores lucros/perdas nas operações. Considerando os dois trabalhos (o presente trabalho de doutorado e o deles), é

possível perceber a maior lucratividade do mercado emergente brasileiro, com retornos superiores a 200% em alguns casos.

Com relação à AS e ao mercado brasileiro, o trabalho de H. S. Santos, Laender e A. C. Pereira (2015) mostrou um estudo de correlações entre redes sociais e o mercado de ações, mas, no entanto, não investigou o uso de AS aplicado às postagens em redes sociais. Por outro lado, Milson L Lima et al. (2016) usaram o modelo SVM em tarefa de Análise de Sentimentos sobre *tweets* da empresa brasileira Petrobras, mas a correlação entre o sentimento predominante e o mercado de ações não foi investigada. Além disso, conforme indicado por Denilson Alves Pereira (2021), a língua portuguesa carece de recursos de PLN. Verificamos que poucos trabalhos realizam Análise de Sentimentos em português do Brasil utilizando técnicas de Aprendizado de Máquina (PEREIRA, D. A., 2021). Martins, Adriano Pereira e Benevenuto (2015) mostraram que as abordagens existentes para o inglês produzem resultados inferiores quando usadas diretamente para textos em português. Assim, o presente trabalho focou na construção de uma metodologia com RNAs profundas para realizar a AS em português diretamente, aplicado ao domínio do mercado de ações. Além disso, vale a pena destacar que, como contribuição indireta, a metodologia proposta no presente trabalho pode ser portada para diferentes idiomas, bastando apenas a coleta de notícias e a criação de um conjunto de dados de treinamento para o módulo de Análise de Sentimentos.

Por fim, nosso trabalho enfatiza a importância de uso de notícias em um sistema de previsão do mercado de ações, tanto de acordo com a avaliação baseada em métricas de Aprendizado de Máquina quanto em métricas do domínio financeiro. De fato, nosso trabalho reitera os resultados de Bustos e Pomares-Quimbaya (2020), que indicam que o uso exclusivo de dados textuais pode não levar a melhores desempenhos do que o uso de indicadores técnicos, mas mostra que as métricas de predição podem ser melhoradas se ambas fontes de dados forem combinadas. No entanto, observamos também que bons resultados em métricas de Aprendizado de Máquina não implicam necessariamente em bons resultados com relação a métricas financeiras. Assim, contribuímos com a literatura existente principalmente com a nossa proposta de um sistema de previsão do mercado de ações com uma abordagem baseada em RNAs profundas combinando notícias, indicadores técnicos e preços de ações, ou seja, contribuímos para a integração de dados que possuem tanto influência na Análise Técnica de mercado quanto na Análise Fundamentalista. Além disso, nosso estudo de caso do mercado

de ações brasileiro contribui para mitigar a escassez de trabalhos que analisam esse mercado em particular. Ainda, cabe destacar que nosso estudo considerou um período significativo para o Brasil (2010-2019), que compreende tantos momentos conturbados quanto positivos para o país.

Nossos experimentos sugerem que há informações importantes presentes nas notícias diárias publicadas, além de mostrar que o sentimento da notícia influencia o mercado de ações e que vale a pena considerar dados oriundos de notícias antes de fazer operações de investimento. Além disso, antes de escolher um dos modelos propostos para investir, acreditamos que o investidor deva ponderar entre o retorno de cada modelo bem como o risco apresentado por cada um deles, observado métricas como SR e MDD.

Limitações e Direções Futuras

Este estudo tem, no entanto, algumas limitações. O primeiro ponto a se considerar é que diferentes períodos ou mercados podem apresentar resultados variados com relação à previsão do mercado de ações por meio de modelos de Aprendizado de Máquina. Acreditamos que o período considerado de 2017 ao final de 2019 é longo o suficiente para captar o comportamento geral dos resultados dos modelos propostos e suas relações com o mercado de ações ao longo do tempo em termos de lucratividade. No entanto, períodos em que ocorrem situações extremas, especialmente considerando uma crise econômica ou instabilidade, podem afetar o desempenho dos modelos treinados. Para mitigar esse efeito, é recomendado que as arquiteturas de RNAs sejam retreinadas ou atualizadas frequentemente para conseguirem apreender os novos padrões de movimentos do mercado de ações.

Adicionalmente, outra limitação da metodologia proposta nesta tese é o tempo necessário para treinamento de cada uma das RNAs utilizadas. Primeiramente, há a necessidade de treinamento de uma RNA dedicada à tarefa de AS. Essa etapa pode ser realizada, de modo geral, em poucas horas considerando todos os modelos e os dados utilizados neste trabalho. Em segundo lugar, há a necessidade de treinamento da RNA voltada à previsão do mercado de ações, que tende a exigir mais tempo computacional. Essa etapa demorou, em geral, alguns dias para cada ação, dada a quantidade de RNAs e seus parâmetros considerados, além do número de notícias que precisam ser convertidas para *word embeddings* e classificadas quanto aos seus sentimentos. Uma vez treinados os modelos, a previsão do valor de uma ação do dia seguinte é realizada em poucos segundos, dadas as informações (notícias, valores e

indicadores) do dia atual. Dessa forma, os momentos de treinamento dos modelos necessários para o funcionamento da metodologia proposta nessa tese devem ser muito bem planejados. Como sugestão, os dias em que não ocorrem negociação durante a semana podem ser utilizados para atualização ou treinamento dos modelos. Outra possibilidade é, uma vez estabelecidos por experimentos anteriores, reduzir a quantidade de parâmetros e modelos de RNAs avaliados, o que pode poupar muito tempo de treinamento ao utilizar cada um dos modelos na prática. Todos experimentos foram desenvolvidos sob a plataforma *Amazon Web Services* (AWS) por meio de uma máquina contendo um disco rígido SSD de 120 *Gigabytes*, 32 *Gigabytes* de memória RAM e processador Intel Xeon.

Finalmente, uma alternativa à nossa abordagem está relacionada aos modelos baseados em *Transformers* e mecanismos de atenção (do inglês, *attention*), como é o caso do modelo BERT (do inglês, *Bidirectional Encoder Representations from Transformers*) (KENTON; TOUTANOVA, 2019), que atualmente é o estado da arte na geração de *word embeddings* e que também pode ser utilizado como uma ferramenta para Análise de Sentimentos. Neste trabalho, treinamos todo um modelo para um determinado domínio, ou seja, textos escritos na língua portuguesa do Brasil e relacionados ao mercado financeiro. No início de desenvolvimento, ainda não havíamos encontrado a disponibilidade do modelo BERT para o Português que realizasse a contento a tarefa de AS desempenhada neste trabalho. Ou seja, no período de início de desenvolvimento deste trabalho, havia apenas a versão BERT *Multilingual* e os resultados para a língua portuguesa ainda não eram satisfatórios. Além disso, conforme apresentado por Ezen-Can (2020), o modelo BERT pode apresentar resultados inferiores aos das redes LSTM quando aplicado a pequenos conjuntos de dados, além de levar mais tempo durante a fase de treinamento. Ademais, o BERT é treinado com um conjunto de dados voltado a um propósito geral e, em nosso trabalho, optamos em um primeiro momento por treinar um modelo inteiramente voltado ao domínio específico da área financeira em português. Por isso, neste trabalho o foco se voltou inicialmente para o algoritmo *FastText*.

No entanto, recentemente, a versão apresentada por Souza, Nogueira e Lotufo (2020), denominada BERTimbau, adaptou o modelo BERT à língua portuguesa do Brasil com sucesso. Dessa forma, o uso do BERT neste trabalho por meio de sua versão BERTimbau pode ser considerado como um possível trabalho futuro, estendendo a nossa proposta. Até o momento, trabalhos na área financeira que utilizam o BERT estão em desenvolvimento e,

acreditamos, portanto, que abordagens utilizando o BERT possam ser melhor exploradas em trabalhos futuros. Nesse sentido, vislumbramos que o uso do modelo BERT por meio de seu *token* [CLS] para a classificação de sentimentos em títulos de notícias possa ser uma abordagem interessante para incrementar a etapa de Análise de Sentimentos. Outra abordagem a ser investigada é o uso do token [CLS] sem a explícita passagem por uma etapa de Análise de Sentimentos, sendo utilizado como entrada diretamente em uma única RNA para desempenhar uma tarefa de regressão e previsão do mercado de ações.

Capítulo 8

Conclusões e Trabalhos Futuros

Este trabalho propôs uma abordagem de previsão do mercado de ações com base em notícias, indicadores técnicos e preços de ações que servem como dados de entrada para modelos de previsão baseados em RNAs profundas. Os experimentos foram realizados com dados do mercado financeiro brasileiro considerando duas etapas.

A primeira etapa considerou o treinamento de uma RNA capaz de desempenhar a tarefa de Análise de Sentimentos voltada ao domínio do mercado financeiro brasileiro. A principal contribuição dessa etapa é a comparação entre diferentes Redes Neurais Artificiais profundas para desempenhar Análise de Sentimentos em Português, algo ainda pouco explorado na literatura.

A segunda etapa deste trabalho considerou a previsão do mercado de ações brasileiro em relação ao período compreendido entre janeiro de 2010 e dezembro de 2019. Esse estudo analisou um período de uma década que compreende diversas crises e avanços no mercado acionário brasileiro, permitindo a exploração de diferentes períodos da economia. Os resultados mostraram a importância de combinar as notícias financeiras aos indicadores da Análise Técnica e históricos de preços das ações para apoiar a decisão dos investidores no mercado de ações. As principais contribuições desse estudo são:

- Uma abordagem inovadora para prever o mercado de ações, incorporando notícias, indicadores técnicos e preços das ações em um processo totalmente realizado com RNAs profundas.
- O uso do mercado brasileiro como estudo de caso, que possui poucos trabalhos publicados até o presente momento.

- Uma proposta de combinação de dados da Análise Técnica e Fundamentalista em um único sistema de previsão do mercado de ações.

Ainda, vale a pena destacar que este trabalho apresentou uma comparação entre as arquiteturas de RNAs mais comumente utilizadas na área financeira para prever o mercado de ações brasileiro. Além disso, os experimentos realizados mostraram que as notícias melhoram a previsão no mercado de ações, tanto considerando métricas de Aprendizado de Máquina quanto métricas do domínio financeiro.

Por fim, foi possível verificar que a AS em notícias financeiras pode ser uma ferramenta útil para os investidores no processo de tomada de decisão no mercado financeiro. Conforme apresentado neste estudo e em estudos relacionados, os investimentos apoiados em técnicas de AS podem trazer lucros quando utilizados de forma adequada.

Como frutos diretos deste trabalho, foram produzidos e publicados os seguintes artigos: (CAROSIA; COELHO; SILVA, A. E. A., 2021), (CAROSIA; COELHO; SILVA, A. E. A., 2020) e (CAROSIA; COELHO; SILVA, A. E. A., 2019). Já os resultados finais da presente tese ainda estão em preparação para publicação.

Trabalhos futuros podem abordar as seguintes tarefas: ampliação dos experimentos, considerando mais empresas e índices; considerar a utilização de uma única RNA, alimentada diretamente com os *word embeddings* de notícias, eliminando a etapa de AS; utilizar o modelo BERT para geração dos *word embeddings*; e utilização de dados de intervalos inferiores a um dia, voltados para *day trade*, como, por exemplo, previsões de mercado em minutos ou horas a partir da publicação de notícias.

Referências

- ACHELIS, S. B. Technical Analysis from A to Z. **McGraw Hill New York**, 2001.
- ALTHELAYA, K. A.; EL-ALFY, E.-S. M.; MOHAMMED, S. Evaluation of bidirectional LSTM for short-and long-term stock market prediction. **International Conference on Information and Communication Systems**, p. 151–156, 2018.
- ANJ. Associação Nacional de Jornais, 2019. Accessed: 2019-10-15.
- ARAÚJO JÚNIOR, J. G. de; MARINHO, L. B. Using Online Economic News to Predict Trends in Brazilian Stock Market Sectors. **Proceedings of the 24th Brazilian Symposium on Multimedia and the Web**, p. 37–44, 2018.
- ARIAS, M.; ARRATIA, A.; XURIGUERA, R. Forecasting with twitter data. **ACM Transactions on Intelligent Systems and Technology**, ACM, v. 5, n. 1, p. 1–24, 2014.
- ATSALAKIS, G. S.; VALAVANIS, K. P. Surveying stock market forecasting techniques–Part II: Soft computing methods. **Expert Systems with Applications**, Elsevier, v. 36, n. 3, p. 5932–5941, 2009.
- AVANÇO, L. V.; NUNES, M. d. G. V. Lexicon-based sentiment analysis for reviews of products in Brazilian Portuguese. **Brazilian Conference on Intelligent Systems**, IEEE, p. 277–281, 2014.
- BENGIO, Y. et al. Learning deep architectures for AI. **Foundations and trends® in Machine Learning**, Now Publishers, Inc., 2009.
- BENGIO, Y.; DUCHARME, R.; VINCENT, P.; JAUVIN, C. A neural probabilistic language model. **Journal of machine learning research**, v. 13, p. 1137–1155, 2003.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. et al. Learning long-term dependencies with gradient descent is difficult. **IEEE transactions on neural networks**, v. 5, n. 2, p. 157–166, 1994.

- BM&FBOVESPA. Acesso em: 2018-11-20, 2018. Disponível em: <<http://www.bmfbovespa.com.br/>>.
- BOJANOWSKI, P.; GRAVE, E.; JOULIN, A.; MIKOLOV, T. Enriching word vectors with subword information. **Transactions of the Association for Computational Linguistics**, MIT Press, v. 5, p. 135–146, 2017.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of computational science**, Elsevier, v. 2, n. 1, p. 1–8, 2011.
- BULKOWSKI, T. N. Fundamental Analysis and Position Trading: Evolution of a Trader. **John Wiley & Sons**, John Wiley & Sons, v. 605, 2012.
- BUSTOS, O.; POMARES-QUIMBAYA, A. Stock market movement forecast: A Systematic review. **Expert Systems with Applications**, Elsevier, v. 156, p. 113464, 2020.
- CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. Analyzing the Brazilian Financial Market through Portuguese Sentiment Analysis in Social Media. **Applied Artificial Intelligence**, Taylor & Francis, v. 34, n. 1, p. 1–19, 2020.
- CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. Investment Strategies Applied to the Brazilian Stock Market: A Methodology Based on Sentiment Analysis with Deep Learning. **Expert Systems with Applications**, Elsevier, v. 184, p. 115470, 2021.
- CAROSIA, A. E. O.; COELHO, G. P.; SILVA, A. E. A. The Influence of Tweets and News on the Brazilian Stock Market Through Sentiment Analysis. **Proceedings of the 25th Brazilian Symposium on Multimedia and the Web**, p. 385–392, 2019.
- CHEN, A.-S.; LEUNG, M. T.; DAOUK, H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. **Computers & Operations Research**, Elsevier, v. 30, n. 6, p. 901–923, 2003.
- CHHAJER, P.; SHAH, M.; KSHIRSAGAR, A. The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction. **Decision Analytics Journal**, Elsevier, v. 2, p. 100015, 2022.
- CHO, K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. **Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing**, Association for Computational Linguistics, 2014.

- COLLOBERT, R. et al. Natural language processing (almost) from scratch. **Journal of machine learning research**, v. 12, Aug, p. 2493–2537, 2011.
- CORDER, G. W.; FOREMAN, D. I. Nonparametric statistics: A step-by-step approach. John Wiley & Sons, 2014.
- DING, X.; ZHANG, Y.; LIU, T.; DUAN, J. Deep learning for event-driven stock prediction. **Proceedings of the Twenty-fourth international joint conference on artificial intelligence**, 2015.
- DUARTE, J. J.; GONZÁLEZ, S. M.; CRUZ, J. C. Predicting Stock Price Falls Using News Data: Evidence from the Brazilian Market. **Computational Economics**, Springer, v. 57, n. 1, p. 311–340, 2021.
- ECK, M. et al. Prediction of Stock Market Performance Based on Financial News Articles and Their Classification. **Data Management, Analytics and Innovation**, Springer, p. 35–44, 2021.
- ELMAN, J. L. Finding structure in time. **Cognitive science**, Wiley Online Library, v. 14, n. 2, p. 179–211, 1990.
- EZEN-CAN, A. A Comparison of LSTM and BERT for Small Corpus. **ArXiv preprint. ArXiv:2009.05451**, 2020.
- FAMA, E. F. Random walks in stock market prices. **Financial analysts journal**, Taylor & Francis, v. 51, n. 1, p. 75–80, 1995.
- FARIA, E. L. de. Redes neurais convolucionais e máquinas de aprendizado extremo aplicadas ao mercado financeiro brasileiro. **Tese de Doutorado. Universidade Federal do Rio de Janeiro**, 2018.
- FEUERRIEGEL, S.; PRENDINGER, H. News-based trading strategies. **Decision Support Systems**, Elsevier, v. 90, p. 65–74, 2016.
- FISCHER, T.; KRAUSS, C. Deep learning with long short-term memory networks for financial market predictions. **European Journal of Operational Research**, Elsevier, v. 270, n. 2, p. 654–669, 2018.
- FUKUSHIMA, K. Cognitron: A self-organizing multilayered neural network. **The Transactions of Electronics and communication Engineers D**, v. 20, n. 3, p. 121–136, 1975.

- FUKUSHIMA, K. Training multi-layered neural network neocognitron. **Neural Networks**, Elsevier, v. 40, p. 18–31, 2013.
- GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with LSTM. **Neural computation**, v. 12, n. 10, p. 2451–2471, 2000.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. **MIT Press**, 2016. Disponível em: <<http://www.deeplearningbook.org>>.
- GRAVES, A.; SCHMIDHUBER, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. **Neural Networks**, v. 18, n. 5, p. 602–610, 2005.
- GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, Elsevier, v. 38, n. 8, p. 10389–10397, 2011.
- HAHNLOSER, R. H. et al. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. **Nature**, Nature Publishing Group, v. 405, n. 6789, p. 947–951, 2000.
- HÁJEK, P. Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. **Neural Computing and Applications**, Springer, v. 29, n. 7, p. 343–358, 2018.
- HALFELD, M. Investimentos: como administrar melhor seu dinheiro. **São Paulo: Fundamento**, 2007.
- HARRIS, Z. S. Distributional structure. **Word**, Taylor & Francis, v. 10, n. 2-3, p. 146–162, 1954.
- HAYKIN, S. Neural networks and learning machines. **Pearson Upper Saddle River**, 2009.
- HAYKIN, S. Neural networks: a comprehensive foundation. **Prentice Hall PTR**, 1994.
- HIRANSHA, M.; GOPALAKRISHNAN, E. A.; MENON, V. K.; SOMAN, K. NSE stock market prediction using deep-learning models. **Procedia computer science**, Elsevier, v. 132, p. 1351–1362, 2018.
- HOCHREITER, S.; BENGIO, Y.; FRASCONI, P.; SCHMIDHUBER, J. et al. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. **A field guide to dynamical recurrent neural networks**, IEEE Press, 2001.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997.

- HOOVER, K. D. Causality in economics and econometrics. **The new Palgrave dictionary of economics**, v. 2, 2008.
- HOUSSEIN, E. H.; DIRAR, M.; HUSSAIN, K.; MOHAMED, W. M. Artificial neural networks for stock market prediction: a comprehensive review. **Metaheuristics in machine learning: theory and applications**, Springer, p. 409–444, 2021.
- HUBEL, D. H.; WIESEL, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **The Journal of physiology**, Wiley Online Library, v. 160, n. 1, p. 106–154, 1962.
- HUBEL, D. H.; WIESEL, T. N. Ferrier lecture-Functional architecture of macaque monkey visual cortex. **Proceedings of the Royal Society of London. Series B. Biological Sciences**, The Royal Society London, v. 198, n. 1130, p. 1–59, 1977.
- HUSSEIN, D. M. E.-D. M. A survey on sentiment analysis challenges. **Journal of King Saud University-Engineering Sciences**, Elsevier, v. 30, n. 4, p. 330–338, 2018.
- HUYNH, H. D.; DANG, L. M.; DUONG, D. A new model for stock price movements prediction using deep neural network. **Proceedings of the Eighth International Symposium on Information and Communication Technology**, ACM, p. 57–62, 2017.
- IVAKHNENKO, A.; LAPA, V. Cybernetics and Forecasting Techniques Modern Analytic and Computational Method in Science and Mathematics. **New York: American Elsevier Publishing Company, Inc**, 1967.
- JANUÁRIO, B.; CAROSIA, A. E. O.; SILVA, A. E. A. da; COELHO, G. P. Sentiment Analysis Applied to News from the Brazilian Stock Market. **IEEE Latin America Transactions**, v. 20, n. 3, p. 512–518, 2021.
- JOHNSON, R.; ZHANG, T. Effective use of word order for text categorization with convolutional neural networks. **Annual Conference of the North American Chapter of the ACL**, Association for Computational Linguistics, p. 103–112, 2015.
- JORDAN, M. I. Serial order: A parallel distributed processing approach. **Technical Report 8604, Institute for Cognitive Science, University of California, San Diego**, 1986.
- JOULIN, A.; GRAVE, E.; BOJANOWSKI, P.; DOUZE, M. et al. FastText.zip: Compressing text classification models. **ArXiv preprint. ArXiv:1612.03651**, 2016.

- JOULIN, A.; GRAVE, E.; BOJANOWSKI, P.; MIKOLOV, T. Bag of Tricks for Efficient Text Classification. **Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers**, Association for Computational Linguistics, 2017.
- 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**, Elsevier, v. 38, n. 5, p. 5311–5319, 2011.
- KENTON, J. D. M.-W. C.; TOUTANOVA, L. K. Bert: Pre-training of deep bidirectional transformers for language understanding. **Proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics**, Association for Computational Linguistics, p. 4171–4186, 2019.
- KHADJEH NASSIRTOUSSI, A.; AGHABOZORGI, S.; YING WAH, T.; NGO, D. C. L. Text mining for market prediction: A systematic review. **Expert Systems with Applications**, v. 41, n. 16, p. 7653–7670, 2014.
- KIM, Y. Convolutional neural network for sentence classification. **Conference on Empirical Methods in Natural Language Processing**, 2014.
- KRAUS, M.; FEUERRIEGEL, S. Decision support from financial disclosures with deep neural networks and transfer learning. **Decision Support Systems**, v. 104, p. 38–48, 2017.
- KUMAR, D.; SARANGI, P. K.; VERMA, R. A systematic review of stock market prediction using machine learning and statistical techniques. **Proceedings of the Materials Today**, Elsevier, 2021.
- KUMAR, G.; JAIN, S.; SINGH, U. P. Stock market forecasting using computational intelligence: A survey. **Archives of Computational Methods in Engineering**, Springer Netherlands, v. 28, n. 3, p. 1069–1101, 2021.
- KUMBURE, M. M.; LOHRMANN, C.; LUUKKA, P.; PORRAS, J. Machine learning techniques and data for stock market forecasting: A literature review. **Expert Systems with Applications**, Elsevier, p. 116659, 2022.
- KUSUMA, R. M. I. et al. Using Deep Learning Neural Networks and Candlestick Chart Representation to Predict Stock Market. **ArXiv preprint. ArXiv:1903.12258**, 2019.

- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 13, n. 1, p. 35, 2015.
- LECUN, Y.; BOSER, B. et al. Backpropagation applied to handwritten zip code recognition. **Neural computation**, MIT Press, v. 1, n. 4, p. 541–551, 1989.
- LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Taipei, Taiwan, v. 86, n. 11, p. 2278–2324, 1998.
- LEE, M.-C.; CHANG, J.-W.; HUNG, J. C.; CHEN, B.-L. Exploring the effectiveness of deep neural networks with technical analysis applied to stock market prediction. **Computer Science and Information Systems**, v. 18, n. 2, p. 401–418, 2021.
- LI, Q. et al. The effect of news and public mood on stock movements. **Information Sciences**, Elsevier, v. 278, p. 826–840, 2014.
- LI, X.; WU, P.; WANG, W. Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. **Information Processing & Management**, v. 57, n. 5, p. 102212, 2020.
- LIMA, M. L. et al. Using sentiment analysis for stock exchange prediction. **International Journal of Artificial Intelligence & Applications (IJAI)**, v. 7, n. 1, p. 59–67, 2016.
- LIPTON, Z. C.; BERKOWITZ, J.; ELKAN, C. A critical review of recurrent neural networks for sequence learning. **ArXiv preprint. ArXiv:1506.00019**, 2015.
- LIU, W. et al. A survey of deep neural network architectures and their applications. **Neurocomputing**, Elsevier, v. 234, p. 11–26, 2017.
- LO, A. W. The adaptive markets hypothesis. **The Journal of Portfolio Management**, Institutional Investor Journals Umbrella, v. 30, n. 5, p. 15–29, 2004.
- LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. **The Journal of finance**, Wiley Online Library, v. 66, n. 1, p. 35–65, 2011.
- MA, H. et al. A Comprehensive Review of Investor Sentiment Analysis in Stock Price Forecasting. **2021 IEEE/ACIS 20th International Fall Conference on Computer and Information Science (ICIS Fall)**, IEEE, p. 264–268, 2021.

- MAKREHCHI, M.; SHAH, S.; LIAO, W. Stock Prediction Using Event-Based Sentiment Analysis. **IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)**, IEEE, p. 337–342, 2013.
- MALKIEL, B. G. The efficient market hypothesis and its critics. **Journal of economic perspectives**, v. 17, n. 1, p. 59–82, 2003.
- MALKIEL, B. G.; FAMA, E. F. Efficient capital markets: A review of theory and empirical work. **The journal of Finance**, Wiley Online Library, v. 25, n. 2, p. 383–417, 1970.
- MAQSOOD, H. et al. A local and global event sentiment based efficient stock exchange forecasting using deep learning. **International Journal of Information Management**, Elsevier, v. 50, p. 432–451, 2020.
- MARTINS, R. F.; PEREIRA, A.; BENEVENUTO, F. An approach to sentiment analysis of web applications in portuguese. **Proceedings of the 21st Brazilian Symposium on Multimedia and the Web**, ACM, p. 105–112, 2015.
- MATSURA, E. Comprar ou Vender? Como investir na bolsa utilizando Análise Técnica. **Editora Saraiva**, 2013.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, Springer, v. 5, n. 4, p. 115–133, 1943.
- MIKOLOV, T.; CHEN, K.; CORRADO, G.; DEAN, J. Efficient estimation of word representations in vector space. **International Conference on Learning Representations**, 2013.
- MIKOLOV, T.; SUTSKEVER, I. et al. Distributed representations of words and phrases and their compositionality. **Advances in neural information processing systems**, p. 3111–3119, 2013.
- MOCHERLA, S.; DANEHY, A.; IMPEY, C. Evaluation of Naive Bayes and Support Vector Machines for Wikipedia. **Applied Artificial Intelligence**, Taylor & Francis, v. 31, n. 9-10, p. 733–744, 2017.
- NELSON, D. M.; PEREIRA, A. C.; OLIVEIRA, R. A. de. Stock market's price movement prediction with LSTM neural networks. **International Joint Conference on Neural Networks (IJCNN)**, p. 1419–1426, 2017.
- NOFER, M.; HINZ, O. Using Twitter to Predict the Stock Market. **Business & Information Systems Engineering**, Springer Fachmedien Wiesbaden, v. 57, n. 4, p. 229–242, 2015.

- NTI, I. K.; ADEKOYA, A. F.; WEYORI, B. A. Predicting Stock Market Price Movement Using Sentiment Analysis: Evidence From Ghana. **Applied Computer Systems**, Sciendo, v. 25, n. 1, p. 33–42, 2020.
- OLIVEIRA, F. A. de; NOBRE, C. N.; ZÁRATE, L. E. Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index – Case study of PETR4, Petrobras, Brazil. **Expert Systems with Applications**, p. 7596–7606, 2013.
- ORIANI, F. B.; COELHO, G. P. Evaluating the impact of technical indicators on stock forecasting. **IEEE Symposium Series on Computational Intelligence (SSCI)**, p. 1–8, 2016.
- OZBAYOGLU, A. M.; GUDELEK, M. U.; SEZER, O. B. Deep learning for financial applications: A survey. **Applied Soft Computing**, Elsevier, v. 93, p. 106384, 2020.
- PANG, B.; LEE, L. et al. Opinion mining and sentiment analysis. **Foundations and Trends® in Information Retrieval**, Now Publishers, Inc., v. 2, n. 1–2, p. 1–135, 2008.
- PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs Up?: Sentiment Classification Using Machine Learning Techniques. **Proceedings of the Conference on Empirical Methods in Natural Language Processing**, Association for Computational Linguistics, p. 79–86, 2002.
- PANG, X. et al. An innovative neural network approach for stock market prediction. **The Journal of Supercomputing**, Springer, v. 76, n. 3, p. 1–21, 2018.
- PARKER, J. R. Algorithms for image processing and computer vision. **John Wiley & Sons**, 2010.
- PENG, Y.; JIANG, H. Leverage financial news to predict stock price movements using word embeddings and deep neural networks. **Proceedings of the Conference of the North American Human Language Technologies**, Association for Computational Linguistics, 2016.
- PEREIRA, D. A. A survey of sentiment analysis in the Portuguese language. **Artificial Intelligence Review**, v. 54, n. 2, p. 1087–1115, 2021.
- PICASSO, A. et al. Technical analysis and sentiment embeddings for market trend prediction. **Expert Systems with Applications**, Elsevier, v. 135, p. 60–70, 2019.

- QIAN, Y.; LI, Z.; YUAN, H. On exploring the impact of users' bullish-bearish tendencies in online community on the stock market. **Information Processing & Management**, Elsevier, v. 57, n. 5, p. 102209, 2020.
- RATTO, A. P. et al. Ensemble of technical analysis and machine learning for market trend prediction. **Symposium Series on Computational Intelligence (SSCI)**, p. 2090–2096, 2018.
- RHEA, R. The Dow theory: An explanation of its development and an attempt to define its usefulness as an aid in speculation. **Fraser Publishing Company**, 1993.
- RONG, X. Word2vec parameter learning explained. **ArXiv preprint. ArXiv:1411.2738**, 2014.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. et al. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533–536, 1986.
- SALTON, G.; WONG, A.; YANG, C.-S. A vector space model for automatic indexing. **Communications of the ACM**, ACM, v. 18, n. 11, p. 613–620, 1975.
- SANG, C.; DI PIERRO, M. Improving trading technical analysis with TensorFlow Long Short-Term Memory (LSTM) Neural Network. **The Journal of Finance and Data Science**, Elsevier, v. 5, n. 1, p. 1–11, 2019.
- SANTOS, H. S.; LAENDER, A. H.; PEREIRA, A. C. A twitter view of the Brazilian stock exchange market. **Lecture Notes in Business Information Processing**, p. 112–123, 2015.
- SANTOS, I.; NEDJAH, N.; MACEDO MOURELLE, L. de. Sentiment analysis using convolutional neural network with fastText embeddings. **IEEE Latin American conference on computational intelligence (LA-CCI)**, p. 1–5, 2017.
- SCHUMAKER, R. P.; ZHANG, Y.; HUANG, C.-N.; CHEN, H. Evaluating sentiment in financial news articles. **Decision Support Systems**, v. 53, n. 3, p. 458–464, 2012.
- SCHUSTER, M.; PALIWAL, K. K. Bidirectional recurrent neural networks. **IEEE Transactions on Signal Processing**, IEEE, v. 45, n. 11, p. 2673–2681, 1997.
- SHI, Y.; ZHENG, Y.; GUO, K.; REN, X. Stock movement prediction with sentiment analysis based on deep learning networks. **Concurrency and Computation: Practice and Experience**, Wiley Online Library, v. 33, n. 6, e6076, 2021.

- SILVA, I. da; SPATTI, D.; FLAUZINO, R. Redes neurais artificiais para engenharia e ciências aplicadas—curso prático. **São Paulo: Artliber**, 2010.
- SILVA, M. J.; CARVALHO, P.; COSTA, C.; SARMENTO, L. Automatic expansion of a social judgment lexicon for sentiment analysis, 2010.
- SOUZA, F.; NOGUEIRA, R.; LOTUFO, R. BERTimbau: pretrained BERT models for Brazilian Portuguese. **Brazilian Conference on Intelligent Systems**, Springer, p. 403–417, 2020.
- SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **The Journal of Machine Learning Research**, v. 15, n. 1, p. 1929–1958, 2014.
- SUN, S.; LUO, C.; CHEN, J. A review of natural language processing techniques for opinion mining systems. **Information Fusion**, v. 36, p. 10–25, 2017.
- SZE, V.; CHEN, Y.-H.; YANG, T.-J.; EMER, J. S. Efficient processing of deep neural networks: A tutorial and survey. **Proceedings of the IEEE**, IEEE, v. 105, n. 12, p. 2295–2329, 2017.
- TABOADA, M. et al. Lexicon-Based Methods for Sentiment Analysis. **Computational Linguistics**, v. 37, n. 2, p. 267–307, 2011.
- THAKKAR, A.; CHAUDHARI, K. A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. **Expert Systems with Applications**, Elsevier, v. 177, p. 114800, 2021.
- VALLE-CRUZ, D.; FERNANDEZ-CORTEZ, V.; LÓPEZ-CHAU, A.; SANDOVAL-ALMAZÁN, R. Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. **Cognitive computation**, Springer, v. 14, n. 1, p. 372–387, 2022.
- VARGAS, M. R.; LIMA, B. S.; EVSUKOFF, A. G. Deep learning for stock market prediction from financial news articles. **IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications**, IEEE, p. 60–65, 2017.
- WELCH, R. L.; RUFFING, S. M.; VENAYAGAMOORTHY, G. K. Comparison of feedforward and feedback neural network architectures for short term wind speed prediction. **International Joint Conference on Neural Networks**, p. 3335–3340, 2009.
- WERBOS, P. J. et al. Backpropagation through time: what it does and how to do it. **Proceedings of the IEEE**, v. 78, n. 10, p. 1550–1560, 1990.

- WU, G. G.-R.; HOU, T. C.-T.; LIN, J.-L. Can economic news predict Taiwan stock market returns? **Asia Pacific Management Review**, Elsevier, v. 24, n. 1, p. 54–59, 2019.
- YAN, D. et al. Predicting stock using microblog moods. **China Communications**, p. 244–257, 2016.
- YOSHIHARA, A.; FUJIKAWA, K.; SEKI, K.; UEHARA, K. Predicting Stock Market Trends by Recurrent Deep Neural Networks. **Pacific rim international conference on artificial intelligence**, Springer, p. 759–769, 2014.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: The state of the art. **International journal of forecasting**, Elsevier, v. 14, n. 1, p. 35–62, 1998.
- ZHANG, L.; WANG, S.; LIU, B. Deep learning for sentiment analysis: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, e1253, 2018.
- ZHANG, Y.; WALLACE, B. C. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. **Proceedings of the The 8th International Joint Conference on Natural Language Processing**, p. 253–263, 2017.
- ZHAO, B.; HE, Y.; YUAN, C.; HUANG, Y. Stock market prediction exploiting microblog sentiment analysis. **International Joint Conference on Neural Networks**, IEEE, p. 4482–4488, 2016.