



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE ENGENHARIA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

GABRIEL SARAIVA ESPESCHIT

**ALGORITMO DE PREVISÃO DO MERCADO DE CAPITAIS UTILIZANDO ANÁLISE
DE SENTIMENTOS EM NOTÍCIAS FINANCEIRAS**

BELO HORIZONTE

2020

GABRIEL SARAIVA ESPESCHIT

**ALGORITMO DE PREVISÃO DO MERCADO DE CAPITAIS UTILIZANDO ANÁLISE
DE SENTIMENTOS EM NOTÍCIAS FINANCEIRAS**

Monografia apresentada durante o Seminário dos Trabalhos de
Conclusão de Curso de Graduação em Engenharia Elétrica da
UFMG como parte dos requisitos necessários para a obtenção
do título de Engenheiro Eletricista.
Orientador: Prof. Frederico Gadelha Guimarães

BELO HORIZONTE

2020

AGRADECIMENTOS

Não existem palavras para agradecer todos aqueles que estiveram comigo até esse ponto da minha jornada. As coisas que aprendi, as minhas conquistas e, é claro, esse trabalho, foram graças a vocês e jamais me esquecerei disso. Olho para trás e vejo o quanto crescemos juntos e quanto aprendi com vocês e, por isso, devo a vocês esse agradecimento.

À minha mãe, por ter sempre me ajudado com as minhas tarefas e ter sido minha amiga até nos momentos difíceis. Ao meu pai, por ter me ajudado a pensar sempre que minha cabeça não estava conseguindo compreender a situação. À minha irmã, por ter sido uma inspiração e me feito um irmão melhor. A toda minha família, agradeço pela educação, carinho, amor e suporte que me deram ao longo de todos esses anos. Sem sua força e sabedoria, as pedras no caminho seriam mais pesadas.

À Marcela, minha melhor metade, por ter sido minha melhor amiga ante de todos os gargalos e dores de cabeça que esse caminho inevitavelmente nos traz. Por ter sido a paciência, a calma, a alegria e a energia durante toda essa jornada. Que essa seja mais uma de muitas conquistas que teremos juntos!

A todos os meus amigos que estiveram do meu lado: meus colegas do Santo Agostinho, da UFMG, da CPE Jr., do Milhagem e do CSEM Brasil, obrigado por terem tornado esse período mais fácil. Pelas conversas e pelos desabafos, só chegamos aqui, apesar das incertezas e da exaustão, graças a vocês. Se pudesse, colocaria uma foto de cada um de vocês aqui.

Agradeço também ao corpo docente e de serviços da UFMG. O ensino público do nosso país está sendo sucateado pelo desgoverno, mas vocês mantêm a chama da UFMG mais forte que nunca. O esforço de todos vocês me motiva a lutar por uma educação melhor para nosso país e a ser um engenheiro que orgulhará minha instituição e o ensino público brasileiro. O papel de todos vocês, servidores públicos, é essencial para nossa sociedade e por isso eu agradeço!

Esse trabalho vai para todos vocês. Muito obrigado!

RESUMO

O mercado de capitais é definido por muitos profissionais do setor como sendo objetivo e regido pela análise. No entanto, é impossível remover a parcialidade humana das atividades do mercado. Sendo assim, propôs-se investigar o impacto no desempenho de algoritmos de aprendizado de máquina na previsão do mercado financeiro quando se introduz a análise sentimental de notícias no modelo. Utilizando um léxico para língua portuguesa, obteve-se a polaridade de textos extraídos de um site popular de notícias. As médias diárias das polaridades das notícias foram calculadas e agrupadas com uma série temporal contendo os preços de fechamento do índice Ibovespa daquele mesmo dia. Esse conjunto de dados multivariados foi utilizado para treinar dois algoritmos de aprendizado de máquina. Os resultados alcançados foram comparados aos obtidos quando se utiliza somente os preços de fechamento para treinar o modelo. Em ambos os casos, observou-se que o uso de uma série temporal multivariada, contendo os preços de fechamento e a polaridade das notícias do dia, teve um desempenho superior ao uso somente de uma série contendo os preços de fechamento. Quando submetidos a uma estratégia de investimento simples, observou-se um desempenho semelhante para os dois tipos de dados nos dois algoritmos. Porém, quando comparados com uma estratégia de *buy and hold*, ambos obtiveram um desempenho vastamente superior. O estudo demonstra que o uso de algoritmos de aprendizado de máquina são uma forma eficiente para realizar a previsão do mercado financeiro e que a introdução de dados exógenos ao modelo pode melhorar ainda mais seu desempenho. Nesse estudo, a utilização da polaridade diária de notícias se provou uma das variáveis exógenas que pode ser utilizada para aprimorar a performance do algoritmo.

Palavras-chaves: Mercado Financeiro. Análise de Sentimentos. Redes Neurais. LSTM. Séries Temporais Nebulosas.

ABSTRACT

Financial markets are defined as being objective and analysis oriented by many professionals in the field. However, it is impossible to remove human subjectivity from market analysis. For that reason, it was proposed to investigate the impact on the performance of machine learning algorithms in the forecast of the financial market when sentimental analysis of news was introduced to the model. Using a lexicon for the Portuguese language, the polarity of texts extracted from a popular news website was obtained. The daily averages of news polarities were calculated and grouped with a time series containing the closing prices of the Ibovespa index for the same day. This set of multivariate data was used to train two machine learning algorithms. The results achieved were compared to those obtained when using only closing prices to train the model. In both cases, it was observed that the use of a multivariate time series, containing the closing prices and the polarity of the daily news, outperformed the use of a series containing only the closing prices. When used in a simple investment strategy, a similar performance was observed for both types of data, in both algorithms. However, when compared to a buy and hold strategy, they achieved a vastly superior performance. The study shows that the use of machine learning algorithms is an efficient way to forecast the financial market and that the introduction of exogenous data to the model can further improve its performance. In this study, the use of daily news polarity proved to be one of the exogenous variables that can be used to improve the algorithm's performance.

Keywords: Financial Markets. Sentiment Analysis. Neural Networks. LSTM. Fuzzy Time Series.

LISTA DE FIGURAS

Figura 1 - Formulação generalizada de aplicações de análise de sentimentos.....	23
Figura 2 - Exemplo de uma série temporal.....	26
Figura 3 - Decomposição de uma série temporal.....	27
Figura 4 - Visualização dos conjuntos nebulosos para idade.....	29
Figura 5 - Exemplos de <i>underfitting</i> e <i>overfitting</i>	32
Figura 6 - Representação de um neurônio.....	33
Figura 7 - Representação de uma rede neural profunda com duas camadas escondidas.....	33
Figura 8 - Exemplo de uma rede neural recorrente.....	35
Figura 9 - Estrutura geral de uma rede LSTM.....	36
Figura 10 - Célula LSTM com a correia transportadora em destaque.....	37
Figura 11 - Representação da camada da porta de esquecimento de uma LSTM.....	37
Figura 12 - Definição dos valores candidatos para atualizar o estado da célula.....	38
Figura 13 - Atualização do estado da célula da rede LSTM.....	38
Figura 14 - Definição da saída de uma célula da rede.....	38
Figura 15 - Diagrama geral para previsão do mercado de ações considerando dados textuais.....	39
Figura 16 - Arquitetura geral do projeto.....	44
Figura 17 - Arquitetura da rede LSTM desenvolvida.....	51
Figura 18 - Variação do Ibovespa e a polaridade das notícias do dia.....	55
Figura 19 - Correlação de Pearson e média móvel da correlação.....	56
Figura 20 - Previsão para a LSTM da seção 4.2.1.....	57
Figura 21 - Previsão para a LSTM da seção 4.2.2.....	58
Figura 22 - Previsões para a LSTM da seção 4.2.3.....	60
Figura 23 - Previsões para a LSTM da seção 4.2.4.....	61
Figura 24 - Previsões FTS univariado com conjunto de dados completos.....	63
Figura 25 - Previsões FTS univariado com conjunto de dados dividido em cinco.....	64
Figura 26 - Previsões FTS univariado com conjunto de dados dividido em sete.....	64
Figura 27 - Previsões FTS multivariado com conjunto de dados completos.....	65
Figura 28 - Previsões FTS multivariado com conjunto de dados dividido em cinco.....	66
Figura 29 - Previsões FTS multivariado com conjunto de dados dividido em sete.....	66
Figura 30 - Previsões FTS multivariado com pesos com conjunto de dados completos.....	67

Figura 31 - Previsões FTS multivariado com pesos com conjunto de dados dividido em cinco.....	68
Figura 32 - Previsões FTS multivariado com pesos com conjunto de dados dividido em sete.....	68
Figura 33 - Exemplo de operações realizadas.....	70

LISTA DE TABELAS

Tabela 1 - Representação do funcionamento de um léxico classificando a sentença “O tempo está bom hoje”.....	24
Tabela 2 - Relação de estudos anteriores.....	41
Tabela 3 - Amostragem dos dados textuais obtidos para esses estudo.....	46
Tabela 4 - Amostragem dos dados do índice Ibovespa utilizados no estudo.....	46
Tabela 5 - Retorno do LeIA para um trecho de uma das notícias analisadas.....	47
Tabela 6 - Amostragem dos dados de entrada.....	48
Tabela 7 - Resultados para rede LSTM da seção 4.2.1.....	58
Tabela 8 - Resultados para rede LSTM da seção 4.2.2.....	59
Tabela 9 - Resultados para rede LSTM da seção 4.2.3.....	59
Tabela 10 - Resultados para rede LSTM da seção 4.2.4.....	62
Tabela 11 - Resultados modelos univariado.....	63
Tabela 12 - Resultados modelos multivariado.....	65
Tabela 13 - Resultados modelos multivariado com pesos.....	67
Tabela 14 - Resultado da aplicação de investimento para alguns dos modelos.....	70

LISTA DE ABREVIATURAS E SIGLAS

AMH	Hipótese do Mercado Adaptativo
API	Interface de Programação de Aplicativos
EMH	Hipótese do Mercado Eficiente
LeIA	Léxico para Inferência Adaptada
LSTM	<i>Long Short Term Memory</i>
SMAPE	Erro Percentual Absoluto Médio Simétrico
MINDS	<i>Machine Intelligence and Data Science Lab</i>
MSE	Erro Quadrático Médio
N.D.	Não disponível
ReLU	Unidade Linear Retificada
RNA	Redes Neurais Artificiais
RNR	Redes Neurais Recorrentes
STN	Series Temporais Nebulosas
SVM	Máquina de Vetores de Suporte
UFMG	Universidade Federal de Minas Gerais
VADER	<i>Valence Aware Dictionary and Sentiment Reasoner</i>

SUMÁRIO

1 INTRODUÇÃO.....	15
1.1 CONTEXTUALIZAÇÃO.....	15
1.2 MOTIVAÇÃO.....	15
1.3 OBJETIVOS.....	16
1.4 LOCAL DE REALIZAÇÃO.....	16
1.5 ESTRUTURA DA MONOGRAFIA.....	17
2 REVISÃO SISTEMÁTICA DA LITERATURA.....	18
2.1 MOTIVAÇÃO.....	18
2.2 CONCEITOS MERCADOLÓGICOS.....	19
2.2.1 Hipótese do Mercado Eficiente.....	19
2.2.2 Economia Comportamental.....	20
2.2.3 Hipótese do Mercado Adaptativo.....	20
2.2.4 Análises Fundamentalistas e Análises Técnicas.....	21
2.3 CLASSIFICAÇÃO TEXTUAL.....	22
2.3.1 Análise de Sentimentos.....	22
2.3.2 Análise Léxica e o LeIA.....	24
2.4 ALGORITMOS DE PREVISÃO.....	25
2.4.1 Séries Temporais.....	26
2.4.2 Conjuntos Nebulosos e Lógica Fuzzy.....	27
2.4.3 Series Temporais Nebulosas.....	29
2.4.4 Redes Neurais Artificiais.....	32
2.4.5 Redes Neurais Recorrentes e LSTM.....	35
2.5 TRABALHOS ANTERIORES.....	39
2.5.1 Diagrama Geral das Soluções.....	39
2.5.2 Resultados Relevantes.....	40
2.6 LACUNAS NA PESQUISA.....	40
3 METODOLOGIA.....	44
3.1 AQUISIÇÃO DE DADOS.....	44
3.1.1 Dados Textuais.....	45
3.1.2 Dados de Mercado.....	46
3.2 PRÉ-PROCESSAMENTO.....	47

3.3 ALGORITMOS DE APRENDIZADO DE MÁQUINA.....	49
3.3.1 Rede LSTM.....	49
3.3.2 Series Temporais Nebulosas.....	51
3.4 ESTRATÉGIAS DE INVESTIMENTOS.....	52
3.5 ANÁLISE DOS RESULTADOS.....	54
4 RESULTADOS.....	55
4.1 RELAÇÃO MERCADO X NOTÍCIAS.....	55
4.2 REDES LSTM.....	56
4.2.1 Conjunto de Dados Multivariado Completo.....	57
4.2.2 Conjunto de Dados Univariado Completo.....	58
4.2.3 Conjunto de Dados Multivariado Dividido em Cinco.....	59
4.2.4 Conjunto de Dados Univariado Dividido em Cinco.....	60
4.2.5 Discussão dos Resultados Obtidos para LSTM.....	62
4.3 SÉRIES TEMPORAIS NEBULOSAS.....	62
4.3.1 Modelo Univariado.....	63
4.3.2 Modelo Multivariado.....	65
4.3.3 Modelo Multivariado com Pesos.....	67
4.3.4 Discussão dos Resultados Obtidos para STN.....	69
4.4 COMPARAÇÃO LSTM E STN.....	69
4.5 ESTRATÉGIA DE INVESTIMENTOS.....	70
5 CONCLUSÃO.....	71
REFERÊNCIAS.....	73

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A subjetividade humana é paradoxal por natureza: guia a criatividade, conceito que define o que é ser humano, e, simultaneamente, atíca o caos, podendo causar especulações errôneas. No mundo moderno, a inteligência artificial consegue contornar majoritariamente essa parcialidade intrínseca ao ser humano, resolvendo problemas de forma ágil e objetiva. No entanto, para as tarefas atreladas à subjetividade, essa característica pode ser importante e sua introdução de maneira mais direta no algoritmo de aprendizado ser extremamente necessária.

Um campo em que a parcialidade muitas vezes é vista de forma a atrapalhar resultados é o financeiro. O mercado de capitais é definido por muitos profissionais do setor como sendo regido pela análise. No entanto, é impossível remover a parcialidade humana das atividades do mercado. Isso pode ser percebido pela reação de investidores às notícias, que na maioria das vezes é sujeita à subjetividade, causando especulações que alteram por completo o panorama das ações. Aumentam-se os investimentos quando se tem um fluxo alto de notícias positivas e diminuem-se quando se tem notícias negativas. Assim, o mercado é extremamente influenciado pela parcialidade dos envolvidos e é um fator que precisa ser levado em conta de forma direta.

1.2 MOTIVAÇÃO

A predição das ações utilizando inteligência artificial sempre foi um tópico bastante estudado, e é considerado muito complexo e desafiador devido ao número de variáveis ligadas ao assunto e à volatilidade do mercado. Em pesquisas desenvolvidas

anteriormente, o uso de inteligência artificial (na forma de redes neurais artificiais, algoritmos genéticos ou sistemas nebulosos), é limitado. Tomamos o exemplo da rede neural: a limitação decorre dos riscos associados ao sobre ajuste ou ao número alto de parâmetros ao se fazer uma análise histórica ou uma análise gráfica. Além disso, não é levado em consideração o importante fator que a mídia empenha sobre o mercado de capitais, algo que influencia diretamente no comportamento dos investidores.

Observando esses aspectos, a motivação desse estudo é expandir o patamar atualmente trilhado nessa área, criando um algoritmo de inteligência artificial que incorpora o sentimento de notícias na previsão do mercado de capitais. A classificação de uma notícia será feita por meio de um léxico. A classificação das notícias alimentará o software desenvolvido com o intuito de melhorar o algoritmo.

1.3 OBJETIVOS

O principal objetivo desse estudo é implementar e realizar a análise do desempenho de dois algoritmos de regressão para a previsão do mercado financeiro. Além disso, pretende-se avaliar se, quando incorporamos a polaridade de notícias no algoritmo, temos um desempenho superior ou não.

1.4 LOCAL DE REALIZAÇÃO

Esse estudo será realizado como Trabalho de Conclusão do Curso de Engenharia Elétrica com foco em Computação na Universidade Federal de Minas Gerais (UFMG). O trabalho será desenvolvido na plataforma Python podendo fazer uso de outras plataformas, conforme for necessário. O trabalho conta com o acesso à infraestrutura do laboratório MINDS da UFMG.

1.5 ESTRUTURA DA MONOGRAFIA

A monografia é dividida em 5 capítulos:

- Capítulo 1 – Introdução: apresenta o contexto e o foco do trabalho, os objetivos e a estrutura geral;
- Capítulo 2 – Revisão Sistemática da Literatura: relata o estado da arte referente ao assunto pesquisado, reunindo todo material utilizado para embasar o estudo;
- Capítulo 3 – Metodologia: detalha o desenvolvimento do algoritmo e sua estrutura, como foi feita a aquisição de dados e como foi realizado a análise dos resultados;
- Capítulo 4 – Resultados: apresenta os resultados discutindo a sua significância para o projeto e para o contexto do estudo;
- Capítulo 5 – Conclusão: determina, por fim, se os objetivos do projeto foram atingidos, apresentando as possibilidades de continuação do estudo e pontos de melhorias.

2 REVISÃO SISTEMÁTICA DA LITERATURA

Nesse capítulo exploraremos os conceitos e pincelaremos o que já foi feito em trabalhos anteriores com contribuições significativas na área de previsão de desempenho de ações no mercado de capitais com base em notícias. Vamos começar denotando alguns conceitos interdisciplinares importantes para compreender e embasar o estudo de forma objetiva. Em seguida descreveremos as metodologias adotadas em trabalhos publicados. Por fim, falaremos sobre a lacuna nas pesquisas que o presente estudo pretende preencher.

2.1 MOTIVAÇÃO

De maneira geral, a previsão do desempenho de mercados é dividido em duas escolas: técnica e fundamentalista (APPLE, 2005). Elas se diferem em seus parâmetros de entrada, aquela dependendo de dados históricos de mercado, como análise gráfica da ação e correlações entre ativos, e essa de dados de caráter informativo, como notícias, relatórios e informações privilegiadas. Pela facilidade de manipulação dos dados, disponibilidade de acesso e seu caráter quantitativo e gráfico, a metodologia de análise técnica para a previsão de mercados usando aprendizado de máquina foi objeto de vários estudos passados.

A análise fundamentalista, por sua vez, apresenta um maior desafio. Apesar de estudos anteriores terem explorado dados estruturados: dados macroeconômicos ou relatórios financeiros de bancos, governos e empresas (CHATRATH et al., 2014), o foco tem mudado. O uso de textos não estruturados tem se tornado objeto de pesquisa, requerendo o uso de técnicas de análise sentimental e inteligência artificial. Esses dados são obtidos em sites de notícias, redes sociais, blogs e fóruns.

Existem alguns conceitos fundamentais para a compreensão do uso de modelos computacionais e inteligência artificial para a identificação da relação entre informações textuais e a economia (KHADJEH NASSIRTOUSSI et al., 2014). Abordaremos esses e

outros conceitos que englobam três áreas de conhecimento diferentes (linguística, aprendizado de máquina e economia) para melhor compreender as metodologias dos trabalhos anteriores.

2.2 CONCEITOS MERCADOLÓGICOS

Como todo trabalho de reconhecimento de padrões e inteligência artificial, uma boa compreensão da área em estudo é necessária para fazer a análise correta dos resultados alcançados. Sendo assim, começaremos a revisão bibliográfica estudando alguns conceitos fundamentais na área de mercados financeiros. Essas noções serão importantes para conseguir correlacionar os resultados com a teoria de mercado atualmente adotada por profissionais da área.

2.2.1 Hipótese do Mercado Eficiente

A EMH (FAMA, 1965), afirma que os mercados não são previsíveis e não é possível conseguir um retorno superior à média do mercado (dado um certo risco). Isso se deve ao fato de que as informações disponíveis (públicas e privilegiadas) estão embutidas no preço do ativo. Sendo assim, seria impossível ter um desempenho acima da média no mercado usando técnicas diversas como *timing* de mercado ou seleção de ações. O retorno acima da média de mercado seria devido ao acaso ou investimentos de alto risco.

No entanto, FAMA (1970) defende que existem 3 classes de mercado e o comportamento relatado só é percebido em mercados denominados como “fortes”. Em mercados denotados “semifortes”, os preços das ações e ativos incorporam todas as informações publicamente disponíveis e análises técnicas e fundamentais não funcionariam nesse tipo de mercado. Por fim, são classificados como mercados “fracos”

aqueles que refletem apenas as informações de negociações (mudanças de preço e volume) dos ativos, impossibilitando análises técnicas.

Nessa hipótese, a eficiência de um mercado está diretamente relacionada com a disponibilidade de informações e em mercados em que o acesso às informações não é aberto e instantâneo para todos os participantes, a teoria enfraquece.

2.2.2 Economia Comportamental

A economia comportamental estuda os efeitos de fatores emocionais, psicológicos, sociais e cognitivos nas decisões econômicas do indivíduo e empresas (MULLAINATHAN; THALER, 2000). Os preços dos ativos são atribuídos aos sentimentos dos investidores como: bias cognitivos (confiança e otimismo em excesso, bias de status quo, aversão de perdas e efeito posse ou dotação e bias de confirmação), heurísticas (fenômeno de ancoragem, disponibilidade de informações, padrões representativos e comportamento de manada) e o efeito de enquadramento (LIN, 2011).

Como diferentes indivíduos possuem interpretações e bias diferentes a respeito da mesma informação, seu comportamento em relação a um ativo pode variar drasticamente. Sendo assim, a economia comportamental determina que o sentimento otimista ou pessimista de investidores em relação a mercados futuros reflete diretamente em seus padrões de investimentos.

2.2.3 Hipótese do Mercado Adaptativo

A AMH propõe conciliar as visões opostas de pensamento (EMH e a economia comportamental), empregando fundamentos evolucionários para o mercado financeiro (LO, 2004, 2005). LO (2004) defende que os preços dos ativos refletem diretamente as informações ditadas por uma combinação de condições ambientais e os grupos de participantes do mercado, ou *espécies*.

Se existirem muitas *espécies* (ou um volume grande de indivíduos da mesma *espécie*) competindo por recursos escassos dentro de um determinado mercado, é provável que esse mercado seja fortemente eficiente. Caso um número pequeno de *espécies* esteja competindo por recursos abundantes, esse mercado seria menos eficiente. Sendo assim, a eficiência do mercado não teria relação somente com a disponibilidade de informações, mas com toda a *ecologia* do mercado, como o número de competidores, as oportunidades de lucro e a adaptabilidade dos participantes do mercado.

Nesse caso, a AMH se diferencia da EMH pois estabelece que existem oportunidades para arbitragem e que estratégias de investimentos (fundamentalistas, técnicas e outras) poderão ter um desempenho bom em alguns *ecossistemas* e um desempenho ruim em outros. Lo conclui que o principal objetivo é a *sobrevivência* e para obter um nível de retorno consistente é necessário adaptar às condições de um mercado mutável. A chave para a sobrevivência e para se adaptar às mudanças na relação de risco/ganho é a inovação (CLOWES, 2005).

2.2.4 Análises Fundamentalistas e Análises Técnicas

Para fazer previsões de mercados, investidores costumam a adotar duas ferramentas diferentes de análise: técnicas e fundamentalistas. A primeira utiliza de dados históricos de mercado como preços históricos e volume. Analistas técnicos acreditam que a história se repete e todos os fundamentos que influenciam os mercados estão imediatamente incorporados no preço, por isso, utilizam reconhecimentos de padrões de movimentações passadas, análises gráficas e outras ferramentas analíticas para avaliar as forças ou fraquezas de um ativo (HAYES, [s.d.]).

A análise fundamentalista, por sua vez, define que para um determinado ativo, é importante estabelecer os principais atributos de uma empresa antes de investir. Entre esses atributos estão: o tamanho da empresa, capitalização, lucros, dividendos, entre outros aspectos financeiros (CAVALCANTE et al., 2016). Além disso, analistas

fundamentais acreditam que o preço de um ativo reflete diretamente aspectos políticos e econômicos internos e externos à empresa, instituição ou país.

É importante ressaltar que os investidores utilizam as duas análises. Para horizontes de tempo curtos, a análise técnica tem se provado mais eficiente. Esse tipo de análise também tem sido melhor para prever tendências de mercado e pontos de reviravolta. Isso pode estar associado a algo que já foi dito antes nesse capítulo: análise técnica é um tópico de estudo computacional mais simples e frequente que a fundamentalista. O grande desafio que se tem hoje com análises fundamentalistas acopladas ao uso computacional está relacionado a estruturação e quantidade de dados.

2.3 CLASSIFICAÇÃO TEXTUAL

Para extrair os sentimentos do mercado para textos de notícias, é importante que o computador saiba interpretá-los. Existem diversas formas de fazer isso e abordaremos as maneiras adotadas por alguns autores na seção cinco desse capítulo. Nesse estudo optou-se por fazer a análise de sentimentos por meio de léxicos. Para tal foi utilizada o módulo para linguagem *Python*, LeIA (Léxico para Inferência Adaptada)¹.

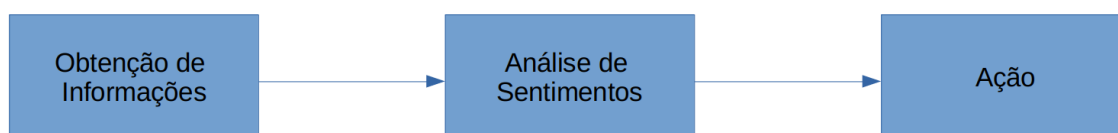
2.3.1 Análise de Sentimentos

Análise sentimental lida com a detecção de emoção ou sentimentos preservada em textos por meio de análises semânticas especializadas. Normalmente são atribuídas três classes de emoções diferentes para um texto, negativa, neutra ou positiva. Esse tipo de análise é utilizado em diversas aplicações, especialmente em redes sociais, para determinar a percepção popular a respeito de um certo produto ou de uma certa marca. Estudos na área de mineração de opinião do mercado financeiro vêm crescendo e existem indicações que o sentimento de investidores e a forma como se expressam em

1 Disponível em <<https://github.com/rafjaa/LeIA>>. Acesso em 13 set. 2020.

redes sociais e fóruns influencia diretamente os preços de ativos (DAS et al., 2001). A Figura 1 abaixo demonstra como funcionam as aplicações de análise de sentimentos de forma geral.

Figura 1 - Formulação generalizada de aplicações de análise de sentimentos



Fonte: Elaborada pelo autor, adaptado de AlphaBold²

Estudos nessa área ainda são muito recentes e utilizam estratégias diversas e, por isso, obtiveram resultados abrangentes. SCHUMAKER et al. (2012) propôs avaliar o sentimento embutido em notícias financeiras utilizando um sistema de avaliação dos substantivos próprios nesses artigos. Eles obtiveram resultados aquém das expectativas, obtendo um acerto de direção do preço de apenas 59% em comparação à taxa de 50% de acerto que se tem com o acaso.

Em um exemplo mais bem-sucedido, YU et al. (2013) utilizou um sistema de entropia contextual para escalonar a emoção por trás de palavras “sementes” e palavras candidatas, semelhantes em contexto e ocorrência dentro de textos de notícias financeiras. A partir disso, conseguiu classificar o sentimento dos artigos. Os resultados do estudo mostraram que o uso de palavras com emoções expandidas (escalonadas) melhorou a eficiência da classificação e a incorporação de intensidades ao algoritmo permitiu que os resultados tivessem uma precisão de 91,5%.

Outros estudos exploram conceitos mais amplos de emoções, como a multidimensionalidade e o multinível de emoções, atuam com noções de ontologia e exploram causas de emoções e não a emoção embutida no texto em si (KHADJEH NASSIRTOUSSI et al., 2014).

2 Disponível em <<https://alphabold.com/sentiment-analysis-the-lexicon-based-approach/>>. Acesso em 13 set. 2020.

2.3.2 Análise Léxica e o LeIA

A análise léxica é uma das formas usadas para realizar a análise de sentimentos. A análise léxica utiliza de lista atributos léxicos (como, por exemplo, palavras) para realizar a classificação de uma frase como positiva, negativa ou neutra (SONI, 2019). Tomemos como exemplo a frase “O tempo está bom hoje.” Um analisador léxico bem treinado entenderia que essa frase representa um sentimento positivo analisando as palavras, expressões, pontuação entre outros atributos. Para cada atributo o léxico daria uma pontuação e o sentimento da frase seria uma operação linear em cima dessas pontuações. Na Tabela 1 abaixo temos uma possível forma de realizar uma classificação léxica em que, -1 representa para o sentimento negativo, 0 neutro e +1 positivo.

Tabela 1 - Representação do funcionamento de um léxico classificando a sentença “O tempo está bom hoje”

Atributos	O	tempo	está	bom	hoje	.	SOMA
Pontuação	0	0	0	1	0	0	1

Fonte: Elaborada pelo autor

Para esse estudo, optou-se pelo uso do módulo LeIA para fazer a classificação léxica de textos:

LeIA (Léxico para Inferência Adaptada) é um fork do léxico e ferramenta para análise de sentimentos VADER (*Valence Aware Dictionary and sEntiment Reasoner*) adaptado para textos em português, com suporte para emojis e foco na análise de sentimentos de textos expressos em mídias sociais – mas funcional para textos de outros domínios. (ALMEIDA, 2018)

Optou-se pelo uso dessa biblioteca, pois não há necessidade de fazer pré-processamentos dos textos a serem analisados e os resultados foram consistentes quando submetidos a testes pelo autor.

O módulo funciona por meio da classificação por valência. Isto é, atributos podem assumir um espectro de valores que varia de -4 (extremamente negativa) a +4

(extremamente positiva) (BURCHELL, 2017). Dessa forma, as frases “O tempo está bom hoje.” e “O tempo está EXCELENTE hoje!!” são classificadas de maneiras diferentes, pois a segunda é objetivamente mais positiva que a primeira. A classificação feita pelo LeIA também considera aspectos como pontuação, se a letra está em caixa-alta ou não e *emojis*.

A classificação nos retorna um dicionário com os seguintes campos (ALMEIDA, 2018):

- *pos*: porcentagem positiva do texto;
- *neg*: porcentagem negativa do texto;
- *neu*: porcentagem neutra do texto;
- *compound*: valor de sentimento geral normalizado variando de -1 (extremamente negativo) a +1 (extremamente positivo).

Tomemos como exemplo as duas frases “O tempo está bom hoje.” e “O tempo está EXCELENTE hoje!!”. Quando submetidas ao analisador de sentimentos, retornam respectivamente: *neg: 0.0; neu: 0.517; pos: 0.483; compound: 0.4215* e *neg: 0.0; neu: 0.356; pos: 0.644; compound: 0.7519*.

2.4 ALGORITMOS DE PREVISÃO

Para fazer a previsão do mercado financeiro é importante entender o que são e como funcionam séries temporais e conjuntos nebulosos. Após isso, será introduzido ao leitor o conceito de redes neurais artificiais (RNA) e, em seguida vamos abordar as duas técnicas utilizadas nesse estudo para fazer a previsão do mercado: Redes Neurais *Long Short Term Memory* (LSTM) e Series Temporais Nebulosas (STN).

2.4.1 Séries Temporais

De acordo com SILVA (2018), séries temporais são conjuntos de dados que representam o comportamento de uma (ou mais) variável aleatória ao longo do tempo e sua principal característica é que os registros sucessivos dessa variável não são independentes entre si e sua análise deve levar em conta a ordem em que foram coletados. Em outras palavras, no estudo de séries temporais, podemos olhar o comportamento passado da série para tentar prever como ela irá se comportar no futuro. A Figura 2 trás um exemplo de uma série temporal, o histórico da cotação do índice Ibovespa futuros.

Figura 2 - Exemplo de uma série temporal



Fonte: Investing.com ³

Uma série temporal depende de 3 principais fatores: estacionariedade sazonalidade e autocorrelação (PEIXEIRO, 2019). Além disso, podemos introduzir na análise de séries temporais um quarto fator que é a aleatoriedade.

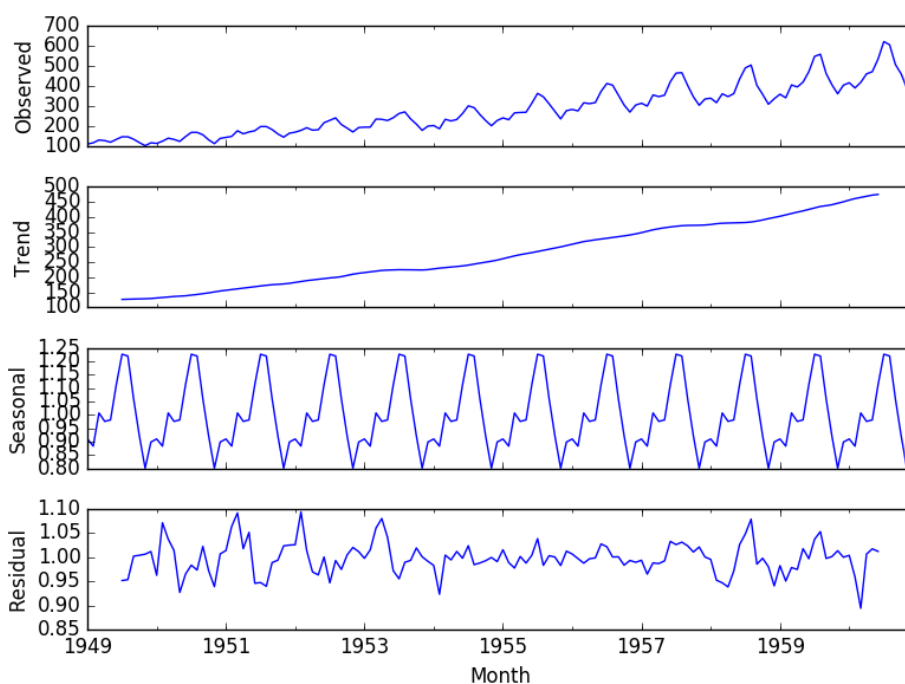
A estacionariedade se refere a manutenção de uma média e uma variância constante ao longo do período analisado. A média ou variância também podem ter um aumento ou queda constante ao longo do tempo, o que caracteriza uma tendência. A sazonalidade, por sua vez, reflete o caráter cíclico de uma série temporal. Por exemplo, a venda de brinquedos segue um caráter sazonal, pois toda época de natal, aumentam. A

3 Disponível em <<https://www.investing.com/indices/ibovespa-futures-chart>>. Acesso em 16 set. 2020.

autocorrelação reflete a semelhança entre os pontos da nossa série como uma função do tempo passado entre elas. Por fim, temos a aleatoriedade, que tem um papel introduzir incertezas na série analisada.

Na Figura 3, temos um exemplo de uma série temporal destrinchada em 3 partes diferentes: tendência, sazonalidade e resíduo. O último pode ser interpretado como a aleatoriedade do sistema, mas também pode ser visto como autocorrelações, pois podemos observar altas variâncias no começo e no final de todo ano mostrando que esses dados são correlacionados com o período em que se encontram (BROWNLEE, 2017).

Figura 3 - Decomposição de uma série temporal



Fonte: BROWNLEE (2017) ⁴

2.4.2 Conjuntos Nebulosos e Lógica Fuzzy

Para a compreensão das técnicas de previsão do mercado usando séries temporais nebulosas, é necessário entender os conjuntos nebulosos e a lógica fuzzy. Os

4 Disponível em <<https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>>. Acesso em 25 set. 2020.

conjuntos nebulosos, diferente de conjuntos matemáticos clássicos, são usados em problemas em que não se tem uma opção clara a respeito de qual classe determinado objeto pertence. Tomemos como exemplo uma classificação de idade: sabendo-se que uma pessoa pode ser criança, jovem, adulta ou idosa, podemos ter idades que se encaixam em duas dessas classes (SILVA, 2018a).

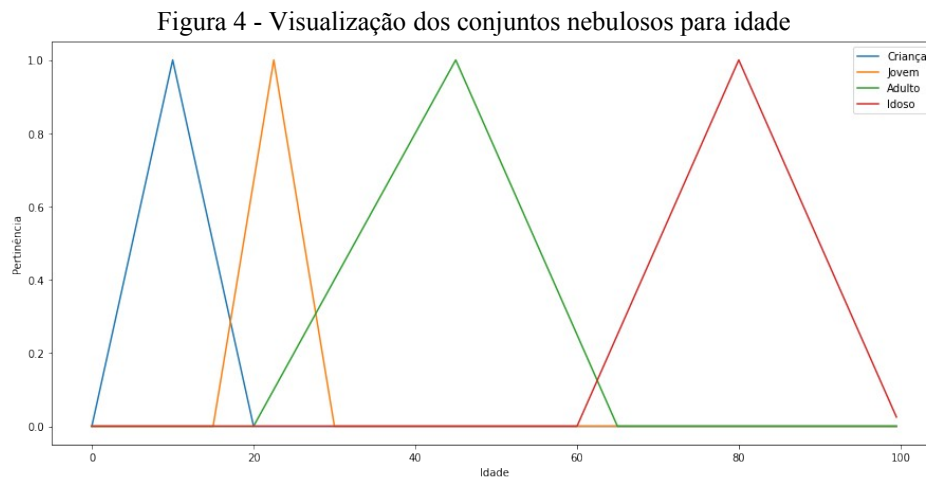
A lógica fuzzy nos diz que o pertencimento de um dado elemento a um conjunto não é binário, mas sim contínuo. Sendo assim, um conjunto fuzzy atribui um valor, tipicamente um número real que esteja no intervalo $[0, 1]$, de tal modo que esse número represente a pertinência do elemento ao conjunto (CINTULA; FERMÜLLER; NOGUERA, 2017).

Um espaço nebuloso pode ser composto por vários conjuntos nebulosos definidos nos intervalos $[x^i, y^i]$ de tal modo que, o começo ou final de um dos conjuntos pode sobrepor o começo ou final de outro, ao menos que esse conjunto seja o primeiro ou último do nosso espaço. A partir disso, podemos definir funções para calcular a pertinência de um determinado elemento ao conjunto delimitado por aquele intervalo.

Uma função comumente usada é a função triangular, em que são definidos um par de pontos (a, b) , como sendo os intervalos de limitação dos conjuntos, e um ponto (c) como o ponto de máxima pertinência. A partir disso, traça-se um triângulo passando pelas coordenadas $(a, 0)$; $(c, 1)$ e $(b, 0)$. Sendo assim, qualquer ponto x pertencente ao intervalo $[a, b]$ tem sua pertinência ao conjunto obtida por meio da função:

$$P(x) = \max \left(\min \left(\frac{(x-a)}{(b-a)}, \frac{(c-x)}{(c-b)} \right), 0 \right)$$

Caso o ponto x não se encontre no intervalo $[a, b]$, ele está fora do conjunto. Podemos usar esse conceito para criar um espaço para nossos conjuntos nebulosos do exemplo de idade que discutimos anteriormente. O exemplo de espaço criado pode ser visto na Figura 4.



Fonte: Elaborada pelo autor, adaptado de SILVA (2018)

Nesse caso, uma pessoa com 23 anos teria um pertencimento de 93% ao conjunto de jovens e 12% ao conjunto de adultos.

2.4.3 Series Temporais Nebulosas

Podemos unir os dois conceitos vistos nas seções 2.4.1 e 2.4.2 para a criação de séries temporais nebulosas (STN). Os princípios de STN foram inicialmente desenvolvidos para resolver o problema de prever a matrícula de novos alunos, representada como uma série temporal (SONG; CHISSOM, 1993). CHEN (1996) expandiu o modelo por meio da divisão de séries temporais em conjuntos nebulosos e traçando a relação entre os conjuntos, criando regras para como os conjuntos se ordenam. O desenvolvimento de modelos de STN pode ser dividido em sete etapas (PRIANDA, 2019; SILVA, 2018a):

- i. Definição do universo do problema;
- ii. Determinação do número de conjuntos e o intervalo de cada conjunto;
- iii. Criação do espaço dos conjuntos nebulosos;
- iv. Fuzzificação de dados históricos;
- v. Determinação das relações temporais;
- vi. Determinação das regras de sequência;
- vii. Defuzzificação.

Na primeira etapa, definimos o universo U para o nosso problema. Esse universo é delimitado pelos limites da série temporal analisada extrapolada por alguma porcentagem a mais (normalmente 20%). Em seguida, preenchemos esse universo com conjuntos nebulosos (etapa ii e iii). Sendo assim, partimos o nosso universo U criado em n intervalos sobrepostos e, em cada intervalo, vamos definir uma função para representar os dados utilizando a lógica fuzzy.

A forma com que particionamos o universo, o número de conjuntos que criamos e a função que escolhemos para representar os dados são parâmetros muito importantes a serem considerados na hora projetar um modelo de STN (SILVA, 2018a).

A próxima etapa consiste na transformação de valores numéricos em uma representação do conjunto nebuloso, ao qual o valor tem a maior probabilidade de pertencer. Para facilitar a interpretação dessa transformada, utilizaremos o espaço nebuloso da Figura 4.

Suponhamos uma série temporal que representa a média das idades do consumidor de uma determinada loja, por mês do ano: {15,2; 19,2; 22,3; 30,2; 44,3; 62,1; 72,8; 54,6; 36,8; 22; 12,7; 10,4}. Podemos representar esses valores como uma sequência de conjuntos ao qual esses valores possuem maior probabilidade de pertencerem: {C, C, J, A, A, A, I, A, A, J, C, C}, em que C representa criança, J representa jovem, A representa adulto e I representa idoso.

Em seguida, determinamos as relações temporais, isto é, como dois conjuntos, o precedente e o conseqüente, relacionam-se. Para nosso exemplo, teríamos as seguintes relações: $C \rightarrow C$, $C \rightarrow J$, $J \rightarrow A$, $A \rightarrow A$, $A \rightarrow A$, $A \rightarrow I$, $I \rightarrow A$, $A \rightarrow A$, $A \rightarrow J$, $J \rightarrow C$, $C \rightarrow C$.

Na etapa seis, criamos o conjunto de regras para o nosso modelo. Esse é definido como sendo o agrupamento de conjuntos conseqüentes que compartilham o mesmo precedente. Dessa forma, teríamos as seguintes regras:

$$\begin{aligned} C &\rightarrow C, J \\ J &\rightarrow A, C \\ A &\rightarrow A, I \\ I &\rightarrow A \end{aligned}$$

Essas regras definem o nosso modelo de STN e serão usadas para prever os próximos padrões da série, dado que ela seja estacionária. Além disso, há vantagens no

uso de STN quando aplicadas a séries com grandes volumes de dados e/ou que são constantemente atualizadas (SILVA, 2018a).

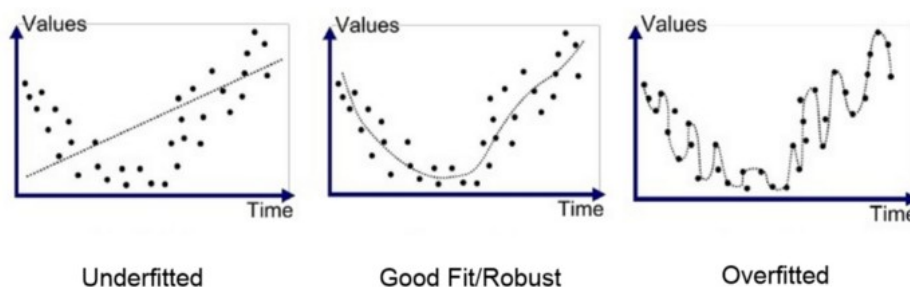
A etapa sete é utilizada para fazer as previsões para o instante seguinte ao dado lido pelo modelo no instante t . Para tal, o modelo recebe o dado de entrada, fuzzifica esse dado (encontra o conjunto ao qual ele pertence), encontra a regra que prevê o próximo conjunto da sequência e desfuzzifica o dado. Para fazer isso existem diversas técnicas. A adotada nesse estudo foi o método do centros de massas:

$$\frac{1}{n} \sum_{i=0}^i P_i, \text{ em que } P \text{ é o centro dos } n \text{ conjuntos consequentes localizados pela regra.}$$

Logo, para o nosso exemplo, caso queiramos prever qual seria a média de idade dos consumidores do próximo mês, dado que a desse mês foi de 78,4, seguimos os seguintes passos: a fuzzificação de 78,4 para I; a localização da(s) regra(s) que prevê o próximo conjunto, que, nesse caso, seria o conjunto A; e, por fim, a defuzzificação dos dados usando a técnica dos centros de massa. Como nossa regra só tem um conjunto com valor central de 45, nosso modelo teria a saída de 45 como previsão de média da idade para o próximo mês.

Outras técnicas já foram desenvolvidas para melhorar o desempenho de STN. Dentre elas, vamos discutir duas que utilizaremos nesse estudo: uso de pesos e uso de dados multivariados. O uso de pesos pondera as regras do modelo que são mais influentes, melhorando a sua acurácia. Já o uso de dados multivariados é o uso de outras séries temporais para auxiliar a predição de uma série temporal original (SILVA, 2018b).

Ambas as técnicas ajudam a separar os dados de ruídos e a compreender o comportamento generalizado da série estudada. Isso ajuda a reduzir o viés (ajuste) e a variância (generalização) do modelo. Um modelo com viés alto é dito um modelo *underfitted*. Um modelo com variância alta é um modelo *overfitted* (BHANDE, 2018). Na Figura 5 temos um exemplo de modelos com bias alto, bias e variância adequados e variância alta.

Figura 5 - Exemplos de *underfitting* e *overfitting*

Fonte: BHANDE (2018) ⁵

2.4.4 Redes Neurais Artificiais

Uma das formas de resolver o problema de previsão do mercado financeiro é utilizando Redes Neurais Artificiais (RNA). RNA se inspiram no cérebro humano e buscam encontrar uma forma de relacionar matematicamente dois conjuntos de dados, um chamado de conjunto de dados de entrada e o outro de conjunto de dados de saída. Essas relações matemáticas são encontradas pelo que chamamos de neurônios.

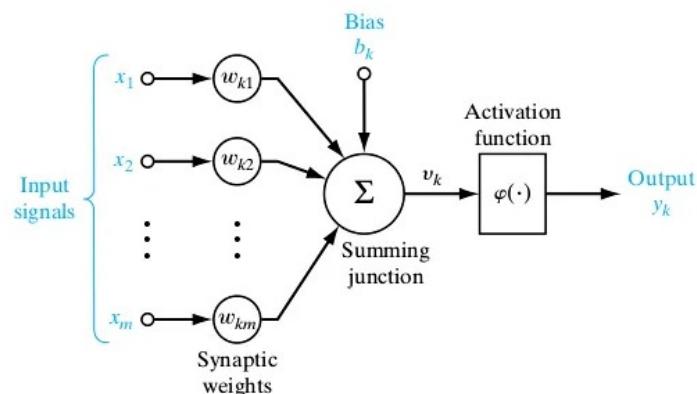
As RNA são compostas por uma série de neurônios. Para cada neurônio, dado um vetor de entrada X de tamanho n , um conjunto de pesos W , também de tamanho n e um bias b , faz-se o somatório das entradas multiplicadas pelos pesos, soma-se o bias b e, usando uma função $g(v)$, comumente chamada de função de ativação, mapeia-se esse somatório a uma saída Y (MANNARSWAMY, 2017). Sendo assim, temos que:

$$Y = g(X * W^T + b) = g\left(\sum_{i=0}^n x_i * w_i + b\right)$$

Algumas funções de ativação comuns são: a sigmoideal, a ReLU e a tangente hiperbólica. O neurônio simples está representado na Figura 6.

5 Disponível em <<https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>>. Acesso em 26 set. 2020.

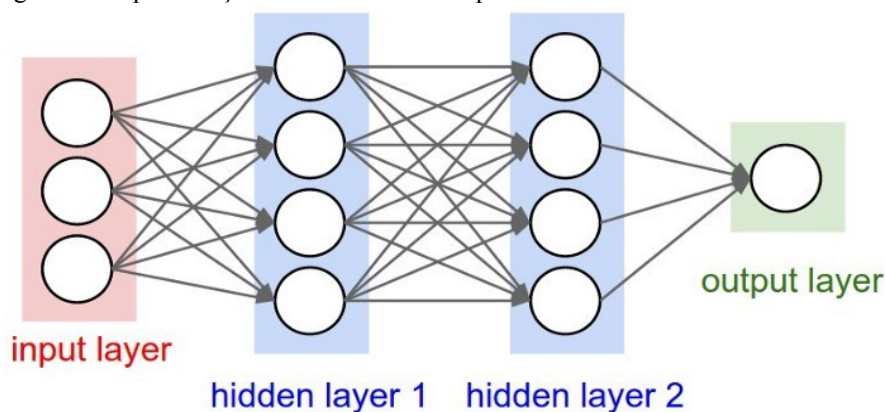
Figura 6 - Representação de um neurônio



Fonte: MANNARSWAMY (2017) ⁶

Toda rede neural tem no mínimo 3 camadas: a camada de entrada, onde se tem os dados ou atributos a serem correlacionados com uma saída; uma ou mais camada(s) escondida(s); e a camada de saída aonde se tem a saída prevista pela rede. Uma camada escondida é composta por um número arbitrário de neurônios, definido pelo projetista. Uma rede neural com mais de uma camada escondida é considerada uma rede neural profunda. A Figura 7 representa uma rede neural profunda com duas camadas escondidas.

Figura 7 - Representação de uma rede neural profunda com duas camadas escondidas



Fonte: GUPTA (2017) ⁷

6 Disponível em <<https://www.opensourceforu.com/2017/03/neural-networks-in-detail/>>. Acesso em 16 set. 2020.

7 Disponível em <<https://medium.com/@rajatgupta310198/getting-started-with-neural-network-for-regression-and-tensorflow-58ad3bd75223>>. Acesso em 25 set. 2020.

As RNA utilizam de uma técnica chamada propagação reversa para fazer a atualização dos pesos, de modo a reduzir o erro entre a saída prevista por ela e a saída real. Esse método se baseia no princípio de gradiente descendente e o hiperparâmetro *alpha* define a taxa de aprendizado, isto é, o quão rápido a rede chega no ponto ótimo (em que o erro é zero). O erro pode ser calculado de diversas formas, no caso de regressão a forma mais comum é o erro quadrático médio (MSE) (BROWNLIE, 2019).

Quando trabalhamos com redes neurais, normalmente temos duas etapas na construção de um modelo: a etapa de treinamento e a de validação. Na etapa de treinamento, entregamos os dados de entrada e os dados de saída para a rede. Inicialmente a rede produzirá uma saída aleatória pois os pesos foram inicializados de maneira randômica. A rede calcula o erro entre a saída esperada (ou saída real) e a saída calculada pela rede (ou saída prevista). Por meio da propagação reversa mencionada anteriormente, a rede neural calcula novos pesos de modo a reduzir o erro entre a saída prevista e a saída real para aquela iteração. Isso se repete por diversas iterações até que o erro esteja abaixo de um determinado patamar arbitrário ou até um número, também arbitrário, de iterações ser atingido.

Em vez de iterações pelo conjunto de treino completo, o projetista pode optar por treinar a rede utilizando épocas e *mini-batches*. Um *mini-batch* é um conjunto de tamanho arbitrário dos dados de treinamento selecionados de forma aleatória. Uma época é uma passada pelos dos dados do *mini-batch* em que se faz o processo de treinamento da rede discutido anteriormente.

Na etapa de validação se passa apenas dados de entrada. Esses dados preferencialmente são diferentes dos dados utilizados para treino. Nessa etapa o objetivo é ver o quão bem a rede consegue prever a saída com dados de entrada nunca vistos antes. As saídas previstas pela rede são então comparadas com as saídas reais e o projetista poderá fazer alterações na arquitetura da rede ou nos hiperparâmetros para melhorar seu desempenho.

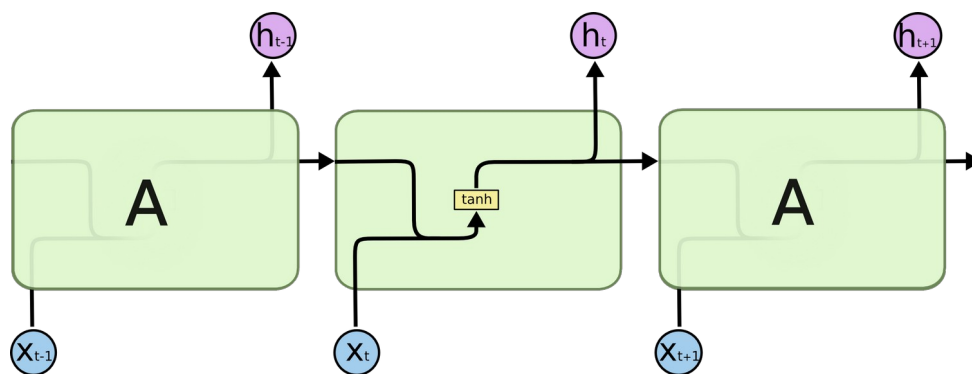
O projetista de RNA deverá ter o cuidado para que sua rede represente bem a relação entre os conjuntos de entrada e saída, mas também consiga generalizar essa representação o suficiente para que os dados de validação possam ser trabalhados. Essa ideia está relacionada diretamente ao conceito de *overfitting* e *underfitting* que discutimos na seção 2.4.3. Além disso, o projetista deve se atentar aos hiperparâmetros

e outros quesitos da rede, como taxa de aprendizado, número de épocas e qual otimizador utilizar.

2.4.5 Redes Neurais Recorrentes e LSTM

Redes neurais recorrentes (RNR) são aplicações de RNA para séries temporais ou sequenciais, em que a saída prevista pela rede pode ser realimentada na entrada e influenciar na previsão da próxima saída. Sendo assim, para um problema de séries temporais, RNR objetivam construir um sequenciamento observando os padrões dos dados de entrada e os dados previstos pela rede anteriormente (AMIDI; AMIDI, [s.d.]). A Figura 8 representa uma RNR. Em cada estado escondido dessa RNR temos uma cadeia de funções de ativação (no caso abaixo, a tangente hiperbólica).

Figura 8 - Exemplo de uma rede neural recorrente



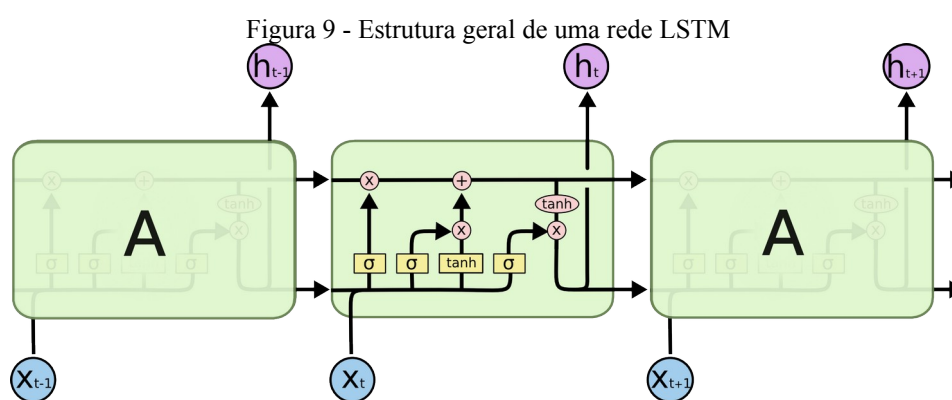
Fonte: (OLAH, 2015) ⁸

As RNR tradicionais, entretanto, possuem uma desvantagem. Quando estamos tratando de um sequenciamento longo, as RNR tem um comportamento de gradiente explosivo ou diminutivo, podendo ser muito instáveis ou acabar não reter a informação ao longo da cadeia como um todo. Para contornar esse problema, podemos introduzir os conceitos de portas lógicas e estados de célula que permitem ou não a ativação da memória, independente da distância entre os elementos da sequência (VENKATACHALAM, 2019).

⁸ Disponível em <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em 26 set. 2020.

As redes Long Short Term Memory (LSTM) são uma adaptação de RNR capazes de reter informações de sequenciamentos longos. Foram propostas por HOCHREITER e SCHMIDHUBER (1997) e são vastamente usadas em aplicações que vão desde o mercado financeiro até processamento de linguagem natural.

Assim como as RNN, redes LSTM também são compostas por um módulo que se repete ao longo da cadeia da rede, no entanto a sua estrutura é muito diferente. A Figura 9 mostra como é a estrutura de uma rede LSTM. Dividiremos a estrutura interna de uma célula da rede LSTM em quatro partes.



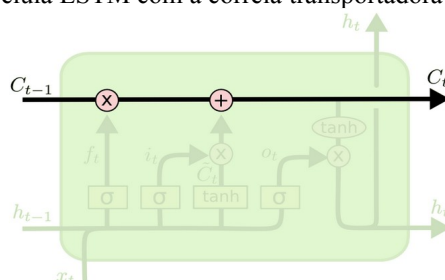
Fonte: OLAH (2015) ⁹

Antes de entender o funcionamento interno das redes LSTM é importante entender o conceito chave por trás delas. No topo da célula, temos uma espécie de correia transportadora que percorre todo o comprimento da cadeia com pouquíssimas alterações, chamada de estado da célula.

Portas lógicas controlam o fluxo de informações para essa correia. Essas portas são compostas por uma camada de uma RNA com função de ativação sigmoideal e um produto escalar. A função sigmoideal pode assumir qualquer valor entre 0 e 1, deixando poucos ou muitos dados passarem, respectivamente. A Figura 10 destaca a correia transportadora que percorre toda extensão da cadeia de uma rede LSTM.

⁹ Disponível em <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em 26 set. 2020.

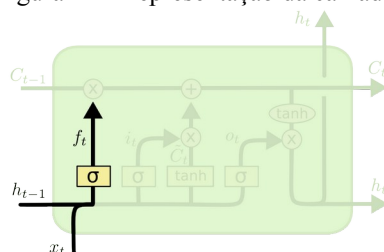
Figura 10 - Célula LSTM com a correia transportadora em destaque



Fonte: OLAH (2015) ¹⁰

O primeiro passo de uma LSTM é decidir quais dados armazenados no estado da célula vamos manter. Essa decisão é tomada pela rede sigmoideal discutida anteriormente, chamada de “camada da porta de esquecimento”. Essa camada observa a entrada x_t e a saída da célula anterior h_{t-1} e decide, para cada dado em C_{t-1} , quais vamos “manter totalmente” (sigmoideal igual a 1) e quais iremos “descartar totalmete” (sigmoideal igual a 0). Isso nos dará um vetor que contem valores de 0 a 1, que chamaremos de vetor de esquecimento (Figura 11).

Figura 11 - Representação da camada da porta de esquecimento de uma LSTM



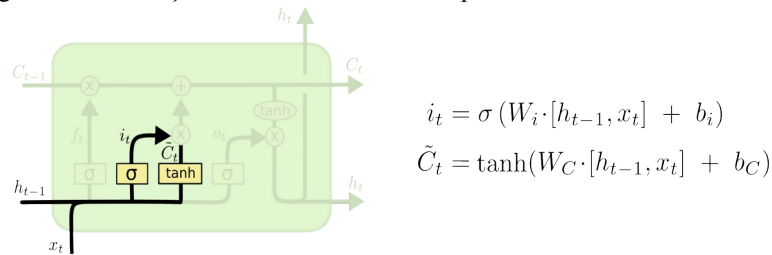
$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Fonte: OLAH (2015) ¹⁰

Em seguida, decidimos quais dados armazenaremos no estado da célula. Isso é feito em duas partes: primeiro a “camada da porta de entrada” decide quais dados da correia atualizaremos, nos dando o vetor selecionador dos dados candidatos. Em seguida, uma camada que usa da função de ativação tangente hiperbólica transforma os dados de entrada, nos dando o vetor contendo os valores candidatos a entrar no estado da célula (Figura 12).

¹⁰ Disponível em <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em 26 set. 2020.

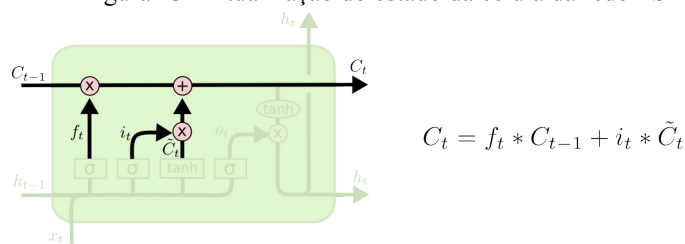
Figura 12 - Definição dos valores candidatos para atualizar o estado da célula



Fonte: OLAH (2015) ¹¹

Dado nosso vetor de esquecimento, o vetor selecionador dos dados candidatos e o vetor dos valores candidatos, podemos atualizar o estado da célula. Primeiramente multiplicamos o vetor C_{t-1} pelo vetor de esquecimento. Em seguida, somamos ao vetor resultante o produto escalar entre o vetor selecionador dos dados candidatos e o vetor contendo os dados candidatos e, assim, temos o nosso novo valor C_t (Figura 13).

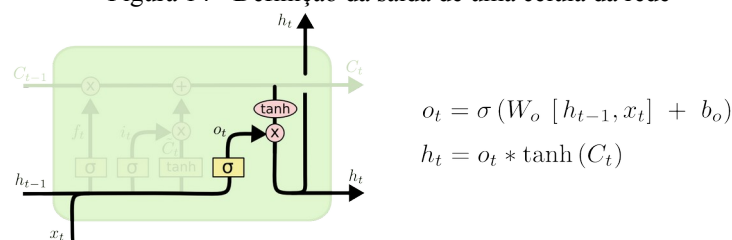
Figura 13 - Atualização do estado da célula da rede LSTM



Fonte: OLAH (2015) ¹¹

Por fim, teremos a nossa saída. Essa saída será baseada no estado da célula, mas somente nos dados que a rede decidir passar. Dessa maneira, fazemos uma seleção dos dados por meio de uma camada com uma função sigmoide. Em seguida, passamos os dados do estado da célula para uma camada com a função tangente hiperbólica. Multiplicamos o resultado de uma camada com o resultado da outra, nos dando a saída h_t daquela célula.

Figura 14 - Definição da saída de uma célula da rede



Fonte: OLAH (2015) ¹¹

¹¹ Disponível em <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em 26 set. 2020.

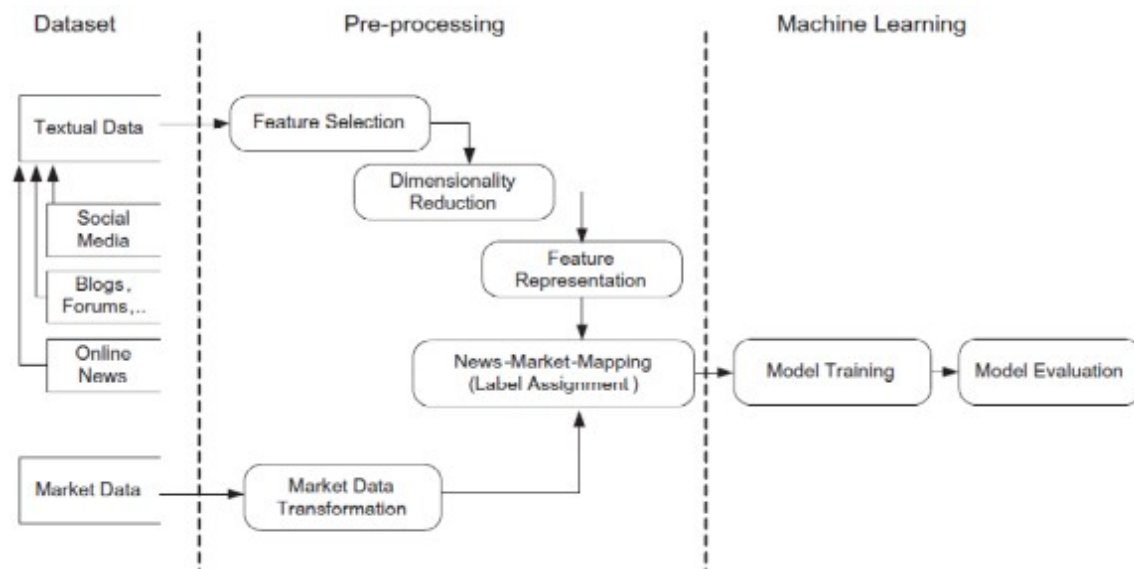
2.5 TRABALHOS ANTERIORES

Nessa seção, faremos um compilado simples dos trabalhos notáveis já feitos na área de estudo vislumbrada nessa monografia.

2.5.1 Diagrama Geral das Soluções

De maneira geral os modelos estudados seguem um diagrama muito semelhante ao demonstrado na Figura 15 (KHADJEH NASSIRTOUSSI et al., 2014). Os dados de entrada são dados textuais, advindos de sites de notícias, redes sociais e/ou blogs e fóruns. Na saída se tem os valores de mercado previstos.

Figura 15 - Diagrama geral para previsão do mercado de ações considerando dados textuais



Fonte: KHADJEH NASSIRTOUSSI et al. (2014, p. 4)

2.5.2 Resultados Relevantes

KHADJEH NASSIRTOUSSI et al. (2014) levanta que a métrica mais utilizada para apresentar resultados dos trabalhos da área é a precisão. Na maior parte dos casos avaliados, a precisão está na faixa de 50-70%, apelando que resultados acima do acaso (50%) são bons. Qualquer resultado acima de 55% é considerado digno de reportagem. No entanto, é levantado que muitos estudos não mencionam se os dados utilizados são balanceados ou não. De modo geral, os pesquisadores utilizam parâmetros de avaliação que diferem, tornando uma comparação de resultados objetiva pouco efetiva.

2.6 LACUNAS NA PESQUISA

Nesse estudo, deseja-se verificar o desempenho de um algoritmo de previsão do mercado de capitais por meio de notícias financeiras usando análise sentimental. Em uma aplicação na bolsa brasileira, estudos desse tipo são escassos e a metodologia desenvolvida nesse trabalho é única. Além disso, o estudo pretende analisar o desempenho do algoritmo em duas métricas diferentes: precisão e taxa de retorno.

A Tabela 2 abaixo relaciona os trabalhos anteriores com o trabalho proposto nesse estudo.

Tabela 2 - Relação de estudos anteriores

Referência	Índice de Mercado	Fonte de dados	Análise sentimental	Algoritmo	Tipo de previsão	Resultados
Wuthrich et al. (1998)	DJIA, Nikkei 225, FT100, Hang Seng e Singapore Straits	The Wall Street Journal, Financial Times, Reuters, Dow Jones, Bloomberg	Bag of words	Experimento multi-algoritmico	Categórico: aumenta, mantém, desce	Acurácia média de 44,4% considerando 5 mercados diferentes.
Peramunetilleke e Wong (2002)	Taxas de câmbio (USD-DEM, USD-JYP)	HFDF93 via www.olsen.ch	Bag of words	Regras de Decisões ou Árvores	Categórico: aumenta, mantém, desce	Melhor que o acaso (50%).
Pui Cheong Fung et al. (2003)	33 ações do Hang Seng	Reuters Market 3000 Extra	Bag of words	SVM	Categórico: sobe, desce	O retorno monitorando séries temporais múltiplas é o dobro daquele que monitora séries únicas.
Werner e Myrray (2004)	DJIA e Dow Jones Internet	Yahoo! Finance, Raging Bull, Wall Street Journal	Bag of words	Experimento multi-algoritmico	Categórico: comprar, vender e segurar	Mensagens ajudam a prever volatilidade dos mercados.
Mittermayer (2004)	Preços de ações	N.D.	Bag of words	SMV	Categórico: notícia boa, ruim ou indiferente	Lucro médio de 11% comparado com lucro de um trader randômico de 0%.
Das e Chen (2007)	24 setores de tecnologia do Morgan Stanley High Tech	Páginas de mensagens	Bag of words, Triplets	Algoritmos Combinatórios	Índice de sentimento agregado	Regressão tem baixo poder explicativo.
Soni et al. (2007)	11 empresas de gás e petróleo	FT Intelligence	Visualização	SVM	Categórico: positivo, negativo	Taxa de acerto do classificador: 56,2%.
Zhai et al. (2007)	Ações da BHP Billiton Ltd.	Australian Financial Review	Bag of words	SVM	Categórico: sobe, desce	Acurácia de até 70,1% usando dados de preço e notícias e retorno de 5,1%.
Rachlin et al. (2007)	5 ações da NASDAQ	Forbes.com, today.reuters.com	Bag of words, valores financeiros comuns	Regras de Decisões ou Árvores	Categórico: sobe, sobe levemente, mantém, desce levemente, desce	Não se pode melhorar a predição usando dados textuais comparando com dados numéricos.
Tetlock et al. (2008)	Firmas individuais do S&P 500	Wall Street Journal, Serviço de Notícias Dow Jones	Bag of words para palavras negativas	Algoritmos de Regressão	Regressão de medidas	Entre outras conclusões, a previsibilidade é maior para as histórias que se concentram nos fundamentos

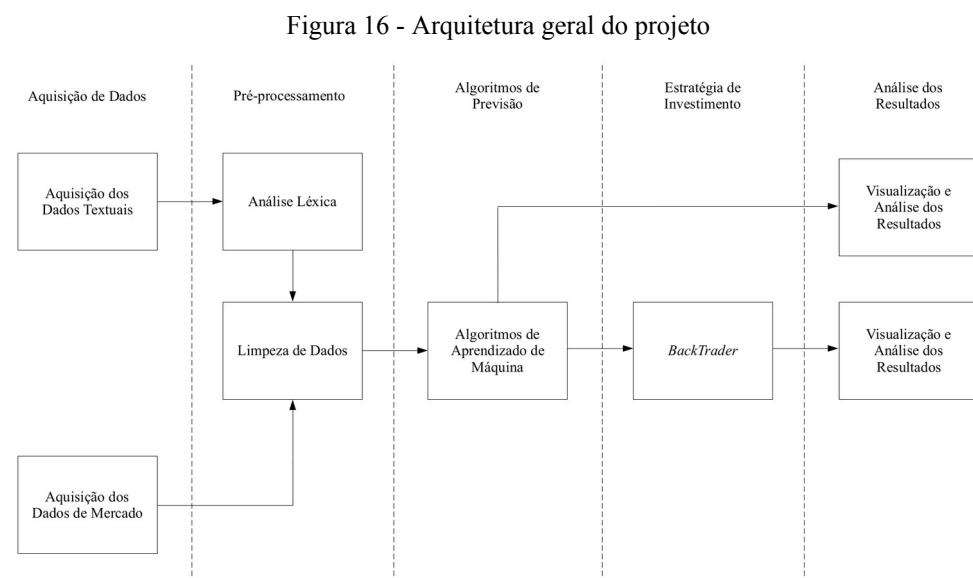
Mahajan et al. (2008)	Sansex	N.D.	Latent Dirichlet Allocation	Algoritmos Combinatórios	Categórico	Acurácia de 60%
Butler e Kešelj (2009)	Deriva de 1 ano do mercado	Sites das empresas	Character n-Grams	Algoritmos Combinatórios	Categórico: desempenho acima ou abaixo do S&P500	Acurácia de 62,81% e precisão de 67,8%
Schumaker e Chen (2009)	Ações do S&P 500	Yahoo! Finance	Bag of words, substantivos próprios e entidades	SVM	Categórico: numérico discreto	Acurácia direcional de 57,1%, retorno de 2,06% e proximidade de 0,042.
Li (2010)	Diversos	Site da SEC Edgar	Bag of words, tom e conteúdo	Naive Bayes	Categórico: positivo, negativo, neutro, incerto	Acurácia de 67% para tom e 63% para conteúdo
Huang, Liao, Yang, Chang e Luo (2010)	Bolsa de Valores de Taiwan	Principais jornais eletrônicos de Taiwan	Termos simultâneos e pares ordenados	Regras de Decisões ou Árvores	Atribuição do grau de significância	Precisão de previsão e taxa de recall de até 85,2689% e 75,3782% em média, respectivamente
Huang, Chuang et al. (2010)	Bolsa de Valores de Taiwan	Principais jornais eletrônicos de Taiwan	Termos simultâneos e pares ordenados	Regras de Decisões ou Árvores	Atribuição do grau de significância	Precisão de previsão e taxa de recall de até 85,2689% e 75,3782% em média, respectivamente
Groth E Muntermann (2011)	TRTH	Comunicados corporativos	Bag of words	Experimento multi-algoritmico	Categórico: positivo, negativo	Precisão (ligeiramente) acima do valor de referência equivalente a 75%
Schumaker et al. (2012)	S&P 500	Yahoo! Finance	Tom e polaridade geral de acordo com OpinionFinder	Algoritmos de Regressão	Regressão	Melhor desempenho com 59% de acurácia direcional e 3,3% de retorno
Yu, Duan, et al. (2013)	Retornos anormais de 824 empresas	Blogs, fóruns, notícias e microblogs (e.g. Twitter)	Bag of words	Naive Bayes	Categórico: positivo, negativo	Polaridade com acurácia de 79% e F1 score de 0,86
Hagenau et al. (2013)	Específico por empresa	DGAP, EuroAdhoc	Bag of words, substantivos próprios, combinações de palavras e n-grams	Algoritmos de Regressão	Categórico: positivo, negativo	Acurácias de até 76%
Jin et al. (2013)	Taxas de câmbio	Bloomberg	Latent Dirichlet Allocation (LDA)	Algoritmos de Regressão	Regressão	Precisão em torno de 0,28

Chatrath et al. (2014)	Taxas de câmbio	Bloomberg	Dados estruturados	Algoritmos de Regressão	Categórico: pulos positivos ou negativos	Entre outras conclusões, os preços respondem rapidamente após o comunicado de imprensa
Bollen E Huina (2011)	DJIA	Twitter	OpinionFinder	Algoritmos Combinatórios	Análise de regressão	O sentimento “calmo” tem a maior relação de causalidade com o DJIA
Vu et al. (2012)	(NASDAQ: AAPL, GOOG, MSFT e AMZN)	Twitter	Twitter Sentiment Tool	Regras de Decisões ou Árvores	Categórico: sobe, desce	Acurácia média de 78,51% para 4 empresas
Esse estudo	Ibovespa	Folha de São Paulo	Análise Léxica usando LeIA	LSTM e STN	Regressão	SMAPE abaixo de 5% em média. Retornos 2 vezes maior que <i>buy and hold</i> .

Fonte: Elaborada pelo autor

3 METODOLOGIA

Como foi dito na seção 1.3, o objetivo principal desse estudo é analisar se a análise de sentimentos aplicado a notícias poderá melhorar o desempenho de algoritmos de inteligência artificial na previsão dos mercados financeiros. Com base nos conceitos discutidos no capítulo 2, nessa seção discutiremos as etapas necessárias para a construção desse estudo. Na Figura 16 temos uma representação da arquitetura geral do projeto.



Fonte: Elaborada pelo autor

3.1 AQUISIÇÃO DE DADOS

Como todo em problema de inteligência artificial, dados são uma necessidade. O primeiro passo para esse estudo foi a aquisição de dados. Existem dois tipos de dados de extrema importância para a realização dessa pesquisa: os dados textuais e os dados do mercado de ativos. Nessa sessão discutiremos as possibilidades levantadas para a aquisição de dados e explicaremos a motivação por trás do caminho adotado nesse estudo.

3.1.1 Dados Textuais

Para aquisição de dados textuais, foram levantadas duas opções: *scrapping* de sites de notícias e redes sociais ou o uso de bancos de dados já elaborados por outros pesquisadores. Nesse estudo, optou-se pela segunda opção, visto que muitos sites de notícias possuem métodos para impedir *web-crawling* ou dificultam o acesso às notícias de dias anteriores. Optou-se por usar um intervalo diário para os dados da bolsa e por isso, é necessário que tenhamos um banco de notícias que contemple muitos dias. O site do “Folha de São Paulo”, por exemplo, tem sua página de “Últimas Notícias” limitada à visualização de no máximo 130 itens (aproximadamente 2 dias de notícias).

Estudou-se também o uso de API (*Application Programming Interface*) oficiais para recolher dados em redes sociais, visto que muitas delas têm perfis que compartilham notícias relevantes ao mercado financeiro (como Twitter e Reddit). No entanto, essas aplicações também restringem a quantidade de buscas que podem ser feitas e o número de dias que se pode explorar no passado. O API do Twitter por exemplo, permite a exploração somente dos 7 dias anteriores e a funcionalidade para explorar Tweets mais antigos é paga. Existem módulos e bibliotecas que conseguem contornar esse problema, mas seu funcionamento é altamente instável.

Dadas essas limitações, apesar de a técnica de *scrapping* poder fornecer mais flexibilidade em relação aos dados textuais, optou-se por utilizar um banco de dados já criado. O banco escolhido foi o *News of the Brazilian Newspaper*¹² que reúne mais de 167 mil notícias recolhidas do site do jornal “Folha de São Paulo” entre janeiro de 2015 e setembro de 2017. O banco contém 6 atributos: data de publicação, título das notícias, texto das notícias, categoria, subcategoria e o link para acessar as notícias (MARLESSON, 2019). A Tabela 3 apresenta uma pequena amostragem dos dados textuais utilizados nesse estudo. Os textos de cada notícia foram omitidos para facilitar a visualização da tabela.

12 Disponível em <<https://www.kaggle.com/marlesson/news-of-the-site-folhauol>>. Acesso em 28 de set. 2020.

Tabela 3 - Amostragem dos dados textuais obtidos para esses estudo

data	título	texto	categoria	subcategoria	link
2017-06-25	Financiamento público de campanha resolve os problemas dos partidos	Quem considerou, com provável razão, que o fim das doações eleitorais de empresas (...)	colunas	janiodefraitas	http://www1.folha.uol.com.br/colunas/janiodefraitas/2017/06/1895780-financiamento-publico-de-campanha-resolve-os-problemas-dos-partidos.shtml
2017-06-25	'Trabalhar para sobreviver não será mais necessário', diz empreendedor	O empresário Peter Diamandis crê que o avanço tecnológico criará (...)	mercado		http://www1.folha.uol.com.br/mercado/2017/06/1895832-trabalhar-para-sobreviver-nao-sera-mais-necessario-diz-cofundador-da-singularity-university.shtml
2017-06-25	Fábula do choque dos preços administrados é tese frágil	Digamos que eu me proponha a explicar as causas da diversidade de (...)	mercado		http://www1.folha.uol.com.br/mercado/2017/06/1895829-fabula-do-choque-dos-precos-administrados-e-tese-fragil.shtml
2017-06-25	Via de mão dupla	Não tenho sido conhecido exatamente pelo meu otimismo. (...)	colunas	marcos-lisboa	http://www1.folha.uol.com.br/colunas/marcos-lisboa/2017/06/1895785-via-de-mao-dupla.shtml
2017-06-25	Almir Sater e outros músicos sobem serra da Cantareira atrás de sossego	PERI PANE COLABORAÇÃO PARA A FOLHA A cerca de 40 minutos do (...)	sobretudo	morar	http://www1.folha.uol.com.br/sobretudo/morar/2017/06/1895659-almir-sater-e-outros-musicos-sobem-serra-da-cantareira-atras-de-sossego.shtml
2017-06-25	Joaquim Barbosa e o dilema da candidatura à Presidência	O ex-ministro Joaquim Barbosa disse que não é candidato a presidente da República. (...)	colunas	eliogaspai	http://www1.folha.uol.com.br/colunas/eliogaspai/2017/06/1895773-joaquim-barbosa-e-o-dilema-da-candidatura-a-presidencia.shtml

Fonte: Elaborada pelo autor

3.1.2 Dados de Mercado

Para a aquisição dos dados da bolsa de valores, foi utilizado uma biblioteca chamada *pandas_datareader* que faz interface com o API do “Yahoo! Finance”, uma fonte internacional e confiável de dados históricos da bolsa de valores de diversos países.

Por meio desse módulo, obteve-se um histórico dos seguintes dados do índice Ibovespa: máxima diária, mínima diária, preço de abertura, preço de fechamento, volume de negociações e preço de fechamento ajustado. Obteve-se os dados de todos os pregões que ocorreram no período em que temos notícias, de 02 de janeiro de 2015 a 29 de setembro de 2017. A Tabela 4 contém uma amostragem dos dados de mercados obtidos e usados nesse estudo.

Tabela 4 - Amostragem dos dados do índice Ibovespa utilizados no estudo

data	High	Low	Open	Close	Volume	Adj Close
2016-03-28	51149.0	49687.0	49687.0	50838.0	3515800.0	50838.0
2016-03-29	51765.0	50387.0	50839.0	51155.0	4320900.0	51155.0
2016-03-30	52262.0	50900.0	51155.0	51249.0	4730600.0	51249.0
2016-03-31	51248.0	49642.0	51248.0	50055.0	4251900.0	50055.0
2016-04-01	50768.0	49361.0	50054.0	50562.0	3890600.0	50562.0

Fonte: Elaborada pelo autor

3.2 PRÉ-PROCESSAMENTO

Após a aquisição de dados, é necessário prepará-los para que possam ser usados em um algoritmo de aprendizado de máquina. O primeiro passo do pré-processamento dos dados foi a extração da polaridade dos textos utilizando o LeIA. Para tal, foram selecionados do nosso banco de dados de notícias apenas os textos que pertencem à categoria “mercado”, pois se tratam de notícias econômicas mais aplicáveis ao nosso estudo.

Em seguida, os textos de cada notícia foram passados como entrada para o analisador léxico, que retornou a porcentagem positiva, negativa e neutra do texto e o valor composto, como discutimos na sessão 2.3.2. Como o LeIA não necessita de nenhum tipo de pré-processamento ou *tokenização*, pode-se passar os textos como eles foram extraídos do site. Tomemos como exemplo o trecho de um texto de uma das notícias do nosso banco:

“Porém, no Relatório Trimestral de Inflação divulgado ontem pelo BC, a previsão é de inflação de 5,3%. Na manhã de caos de ontem, economistas de bancos e consultorias praticamente faziam campanha para o BC elevar juros, até como um meio de controlar o pânico pontual. Havia mais preocupações. “A piora em cima de piora, a previsão de três anos de déficit somado ao caos dos últimos dias, está fazendo com que apareçam problemas ainda maiores”, diz um executivo.” (FOLHA DE SÃO PAULO, 2015) ¹³

Trata-se de um trecho extremamente negativo e, por consequência, o léxico nos retorna os seguintes parâmetros:

Tabela 5 - Retorno do LeIA para um trecho de uma das notícias analisadas

Porcentagem Negativa	0,298
Porcentagem Neutra	0,702
Porcentagem Positiva	0
Valor Composto	-0,9769

Fonte: Elaborada pelo autor

¹³ Disponível em <<https://www1.folha.uol.com.br/mercado/2015/09/1686300-para-mercado-nem-pacote-fiscal-evitara-degradacao-financeira.shtml>>. Acesso em 28 de set. 2020

Optou-se pelo uso somente do valor composto retornado pelo algoritmo, por ser um ótimo indicador do sentimento do texto. Tendo em mãos os valores compostos de todos os textos, tirou-se a média desses valores para notícias divulgadas no mesmo dia, obtendo-se uma série temporal de polaridade das notícias. Esses valores foram unidos em uma única tabela com os valores de fechamento do mercado do mesmo dia, obtidos conforme explicado na seção 3.1.2. A polaridade de notícias para dias em que não tivemos pregão foram desprezadas, assim como os valores de fechamento para dias em que não tivemos notícias. Dessa forma, o impacto dos finais de semana foram desprezados nesse estudo, o que não é ideal, porém, foi feito para simplificação e validação em um primeiro momento.

A partir desse conjunto de dados, extraiu-se também as variações diárias para fazer uma análise de correlação entre as polaridades e a variação do mercado. Essa variação foi obtida por meio da seguinte fórmula:

$$Variação(t) = \frac{Fechamento(t) - Fechamento(t-1)}{Fechamento(t)}$$

Em que fechamento é o preço de fechamento do dia t . A partir disso, obtivemos um conjunto de dados composto por 670 fileiras e 4 colunas, que foi utilizado para alimentar os algoritmos de aprendizado de máquina, como veremos a seguir, e obter os resultados que serão discutidos no Capítulo 4. Uma amostragem desse conjunto de dados está disposto na Tabela 6 abaixo.

Tabela 6 - Amostragem dos dados de entrada

Data	Fechamento	Polaridade	Variação
2016-05-11	52764.0	-0.5112999999999999	-0.578470351039174
2016-05-12	53241.0	-0.0607848484848485	0.9040254719126678
2016-05-13	51804.0	0.10922380952380953	-2.699047726376289
2016-05-16	51803.0	0.06222666666666667	-0.0019303528685043626

Fonte: Elaborada pelo autor

Deve-se ressaltar que, para o treinamento e teste das redes LSTM, os dados de entrada passaram por uma normalização do tipo *MinMax*, pois as LSTM são muito sensíveis à escala dos dados. A normalização *MinMax* é dada por:

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

Em que, $x(t)$ é o valor a ser normalizado, $\min(x)$ é o valor mínimo da série de dados e $\max(x)$ é o valor máximo.

3.3 ALGORITMOS DE APRENDIZADO DE MÁQUINA

Nesse estudo, optou-se pelo uso de dois algoritmos de aprendizado de máquina: redes LSTM e séries temporais nebulosas. Nessa seção exploraremos a arquitetura por trás de cada um dos algoritmos utilizados, bem como os parâmetros estudados para tentar melhorar o desempenho do algoritmo.

3.3.1 Rede LSTM

Como o problema é de alta complexidade, desenvolveu-se uma rede LSTM empilhada. O objetivo é tornar o modelo mais profundo. Esse tipo de estratégia foi introduzida em 2013 para aplicações de reconhecimento de fala (GRAVES; MOHAMED; HINTON, 2013). Nessa situação, observou-se que redes mais profundas eram mais importantes que mais células de memória. Nesse estudo, para criação das redes neurais, fez-se uso do módulo *Keras* (CHOLLET, 2015)¹⁴.

Para evitar problemas de *overfitting*, utilizou-se de camadas de *dropout*. Essas camadas desconsideram algumas das saídas de camadas anteriores durante o treinamento. Além disso, usou-se uma camada densa, que funciona conforme discutimos na seção 2.4.4 e uma camada de ativação com função de ativação linear.

A arquitetura da rede está disposta na Figura 17. A rede foi treinada com 4 conjuntos de entrada diferentes. Primeiramente, foi feito o treinamento considerando a polaridade e os preços de fechamento, com 70% dos dados totais sendo usados para treino e 30% para teste. Em seguida, dividiu-se os dados em 5 conjuntos de tamanhos

¹⁴ Disponível em <<https://keras.io/>>. Acesso em 30 de set. 2020.

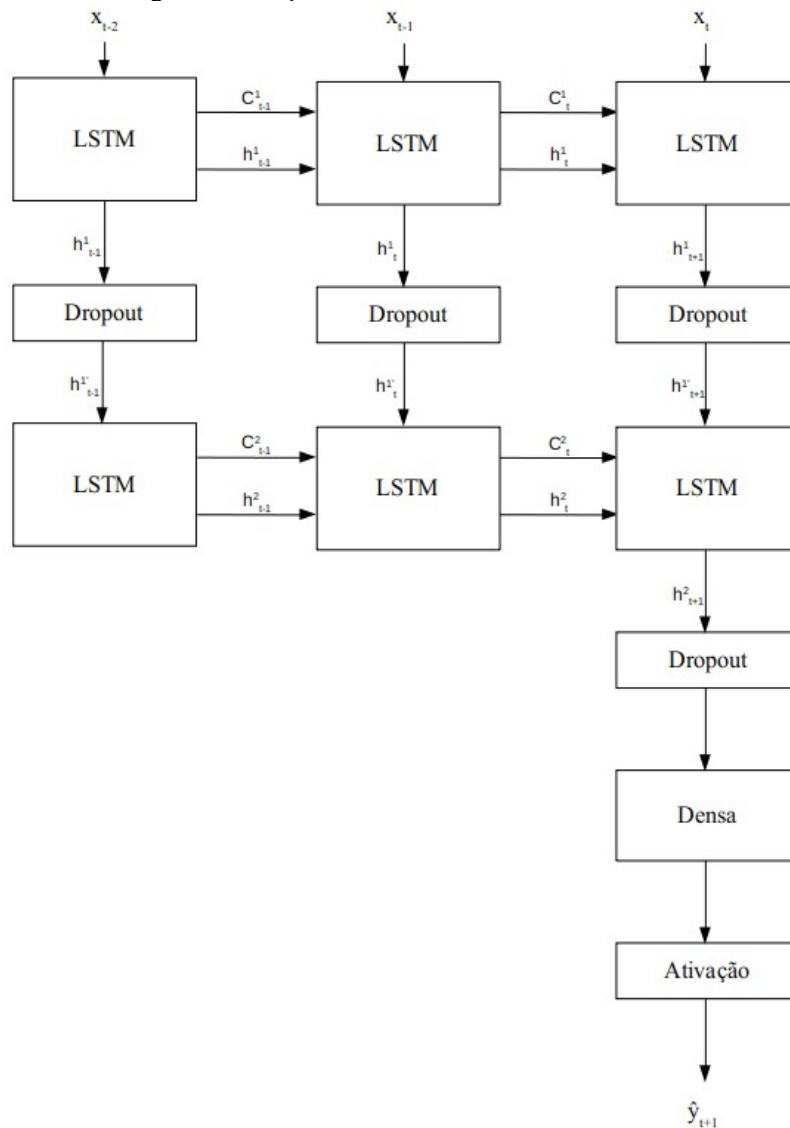
iguais e para cada conjunto se treinou a rede de forma iterativa, com 70% dos dados de cada conjunto sendo passado para treino e 30% sendo usado para teste.

Em seguida fez-se o mesmo processo, mas dessa vez se passou somente os dados de preço de fechamento. Para todas as situações, a rede recebia como entrada os parâmetros referentes ao dia t e aos dois dias anteriores e tinha como saída objetivada o preço de fechamento do dia seguinte ao dia t . A rede foi treinada por 100 épocas, utilizando o otimizador *RMSprop* (BUSHAEV, 2018). A função de perda foi o erro quadrático médio (MSE), dada por:

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

Em que y_i é o valor real, \hat{y}_i é o valor previsto e n é o número de amostras. A escolha por esses parâmetros foi feita de modo empírico, pois eles resultaram no melhor desempenho da rede.

Figura 17 - Arquitetura da rede LSTM desenvolvida



Fonte: Elaborada pelo autor

3.3.2 Series Temporais Nebulosas

Conforme discutimos na seção 2.4.3, existem algumas variações que podemos implementar em uma STN que podem melhorar seu desempenho. Nesse estudo, foi feito o uso de STN com dados multivariados com e sem pesos, além da metodologia

tradicional apresentada anteriormente. Para criação das séries temporais foi utilizado a biblioteca *pyFTS* (SILVA et al., 2019) ¹⁵.

Assim como para as redes LSTM, foi feito a comparação entre a utilização somente dos dados de fechamento e a utilização dos dados de fechamento e as polaridades das notícias. Para o primeiro caso, usamos o modelo de CHEN (1996), para o segundo usamos um modelo para STN multivariadas. Ambos os modelos já estão implementados pela biblioteca *pyFTS* e basta passar os dados corretamente para as funções do módulo. Além disso, implementou-se também um modelo multivariado com pesos.

Nesse estudo, adotamos o particionamento em grade (*grid partitioning*), em que todas as partições têm o mesmo tamanho, e a função triangular explicada na seção 2.4.2. O único parâmetro que sofreu uma alteração para efeitos comparativos foi o número de partições. Além disso, para todos os modelos criados, passou-se o conjunto de dados completo e o conjunto dividido em cinco e em sete partes. Em todos os casos, 70% dos dados foram usados para treino e 30% para teste.

Essas divisões foram feitas pois os métodos de STN fazem uma partição do intervalo de variação da série, de acordo com os dados de treino, e definem os conjuntos fuzzy nesse intervalo. Séries financeiras possuem muita volatilidade e podem apresentar uma variação grande entre os dados de treinamento e teste, principalmente quando janelas muito grandes são consideradas. Quando isso acontece, o modelo corta nos limites inferiores ou superiores.

3.4 ESTRATÉGIAS DE INVESTIMENTOS

Como estamos tratando de um problema do mercado financeiro, optou-se por desenvolver uma estratégia de investimento simples como uma das formas de averiguar o desempenho dos modelos criados. A estratégia funciona da seguinte forma: se o modelo prever que no dia seguinte o valor do índice subirá e o investidor não tiver posicionado, faça uma operação de compra. Mantenha essa compra até o modelo prever

¹⁵ Disponível em <<https://github.com/PYFTS/pyFTS>>. Acesso em 30 set. 2020

que no dia seguinte a ação cairá. Quando isso acontecer, venda a ação. Se o modelo prever que o preço da ação descera e o investidor não tiver posicionado, faça uma operação de venda. Mantenha essa venda até o modelo prever que no dia seguinte a ação subirá. Quando isso acontecer, compre a ação. Essa estratégia também pode ser descrita pelo Algoritmo 1.

Algoritmo 1: Algoritmo Estratégia de Investimentos

```

1. posicionado = 'nulo';
2. operação = Lista(num_dias_analisados);
3. PARA dia EM dias_analisados:
4.   SE previsão(dia+1) > previsão(dia) E posicionado == 'nulo' ENTÃO:
5.     operação[dia] = 'compra';
6.     posicionado = 'comprado';
7.   FIM SE;
8.   SE previsão(dia+1) > previsão(dia) E posicionado == 'comprado' ENTÃO:
9.     operação[dia] = 'nulo';
10.  FIM SE;
11.  SE previsão(dia+1) <= previsão(dia) E posicionado == 'comprado' ENTÃO:
12.    operação[dia] = 'venda';
13.    posicionado = 'nulo';
14.  FIM SE;
15.  SE previsão(dia+1) <= previsão(dia) E posicionado == 'nulo' ENTÃO:
16.    operação[dia] = 'venda';
17.    posicionado = 'vendido';
18.  FIM SE;
19.  SE previsão(dia+1) <= previsão(dia) E posicionado == 'vendido' ENTÃO:
20.    operação[dia] = 'nulo';
21.  FIM SE;
22.  SE previsão(dia+1) > previsão(dia) E posicionado == 'vendido' ENTÃO:
23.    operação[dia] = 'compra';
24.    posicionado = 'nulo';
25. FIM PARA;
```

A partir dessa estratégia, foram gerados sinais de compra e venda de ações, que foram passadas para uma plataforma de *Backtesting* chamada *BackTrader* (DRO, 2020)¹⁶. Plataformas de *Backtesting* são usadas para testar estratégias de investimento caso elas tivessem sido aplicadas no passado. Cabe ressaltar que o objetivo desse estudo não é explorar as melhores estratégias de investimento para os modelos desenvolvidos, por isso, essa estratégia simples atendeu bem o nosso objetivo.

¹⁶ Disponível em < <https://www.backtrader.com/>>. Acesso em 30 set. 2020.

3.5 ANÁLISE DOS RESULTADOS

Analisou-se tanto os resultados obtidos pelo algoritmo quanto os resultados obtidos pela estratégia de investimento. Para os resultados dos algoritmos de aprendizado de máquina, além da análise visual, optamos pelo uso de duas métricas: o MSE e o SMAPE (erro percentual absoluto médio simétrico). Quanto menor os valores de ambas as métricas, melhor a performance do modelo. A fórmula para calcular o MSE está disposta na seção 3.3.1. O SMAPE é calculado por:

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n 2 \frac{|\hat{y}_t - y_t|}{|y_t| + |\hat{y}_t|}$$

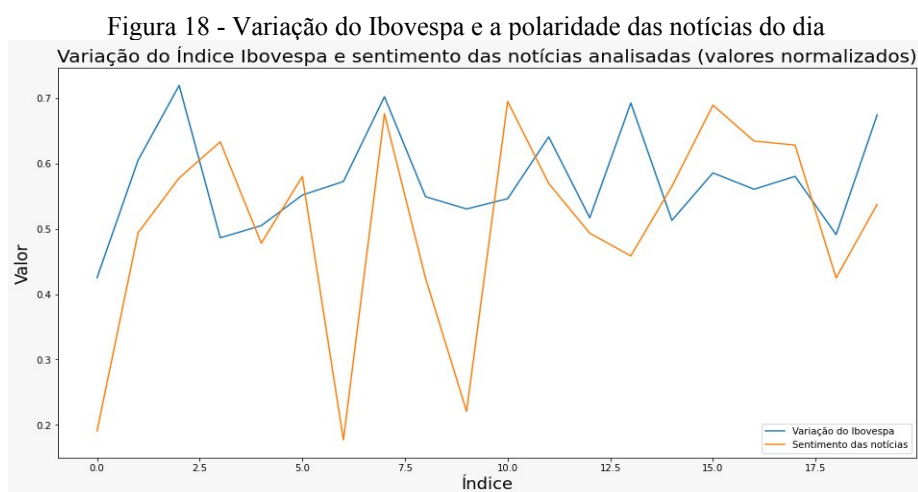
Em que y_t é o valor real, \hat{y}_t é o valor previsto e n é o número de amostras. Para as estratégias de investimentos, comparamos o retorno obtido com o retorno que se obteria fazendo um *buy and hold*, quando se compra uma ação e esta é vendida apenas ao final do período analisado.

4 RESULTADOS

Esse capítulo é dividido em 5 partes. Na primeira seção, apresentamos algumas correlações entre a variação do mercado de capitais e a polaridade das notícias. Em seguida, apresentamos os resultados obtidos com o uso de redes LSTM quando se considera a polaridade das notícias junto aos preços de fechamento e quando se considera somente os preços de fechamento. Após isso, apresentamos os resultados para STN sob as mesmas condições. Depois fazemos uma comparação dos resultados obtidos para STN e LSTM. Por fim, apresentamos e comparamos os resultados obtidos com a estratégia de investimentos para determinados modelo.

4.1 RELAÇÃO MERCADO X NOTÍCIAS

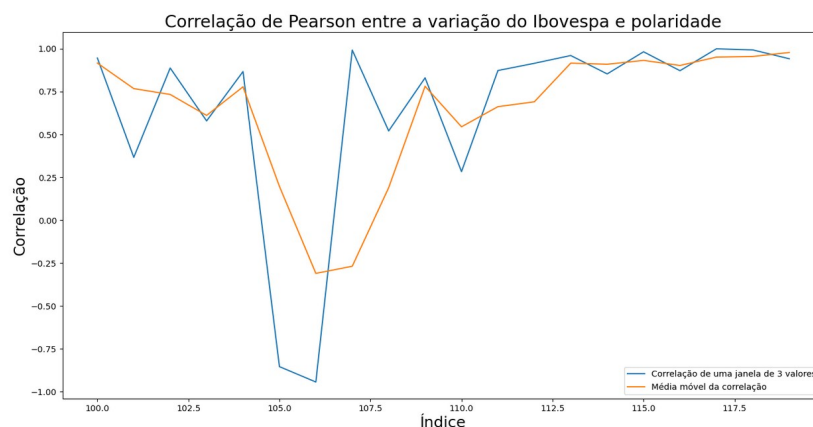
Uma das relações que foi possível observar nesse estudo foi a correlação entre os dados de variação de preço do mercado e os dados da polaridade das notícias. A Figura 18 mostra a variação do preço do índice Ibovespa e a polaridade diária das notícias em um mesmo intervalo de 20 dias escolhido randomicamente. Os valores de ambas as variáveis foram normalizados.



Fonte: Elaborada pelo autor

Sendo assim, foi possível traçar a correlação de Pearson entre as duas variáveis considerando uma janela de 3 valores. Na Figura 19 podemos observar a correlação de Pearson e sua média móvel de 3 períodos.

Figura 19 - Correlação de Pearson e média móvel da correlação



Fonte: Elaborada pelo autor

Como podemos observar, pela maior parte do período analisado, a correlação é positiva e próxima de 1, indicando que quando uma variável sobe ou desce a outra tende fortemente a seguir esse movimento. Esse resultado é importante para o nosso estudo, pois ter uma correlação entre as duas variáveis significa que podemos usar uma para tentar obter a outra. A correlação é uma ferramenta bastante usada no mercado financeiro e investidores costumam usar as correlações entre dois ativos para antecipar seus movimentos (HAYES, 2020).

4.2 REDES LSTM

Nessa sessão, apresentaremos os resultados obtidos para a previsão da série temporal utilizando redes LSTM. Conforme explicado na seção 3.3.1, testamos quatro conjuntos de dados diferentes: um conjunto contendo os dados de fechamento do mercado e polaridade de notícias completos; um conjunto contendo somente os dados de fechamento do mercado completo; um conjunto dividido em 5 contendo os dados de

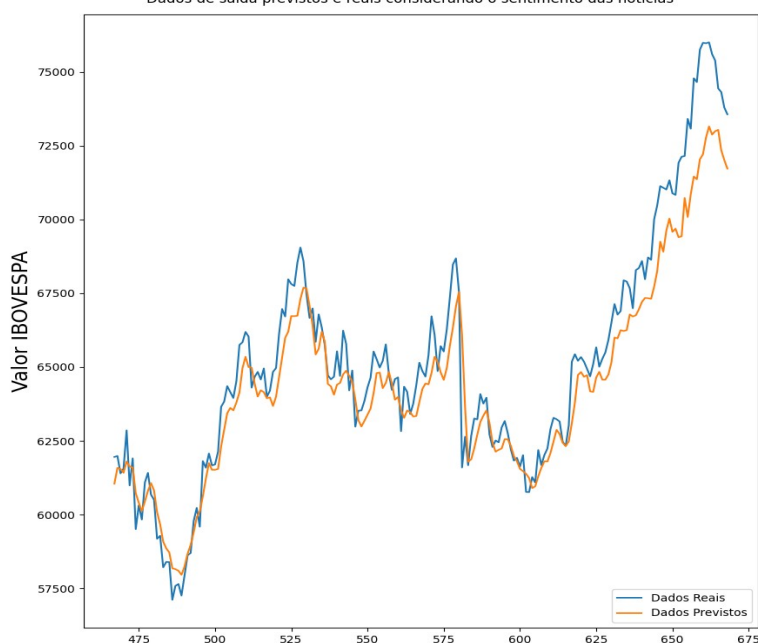
fechamento do mercado e polaridade de notícias; e um conjunto dividido em 5 contendo somente os dados de fechamento do mercado.

4.2.1 Conjunto de Dados Multivariado Completo

Para a criação desse modelo, utilizou-se as 670 amostras que se obteve de dados de fechamento e polaridade, dividindo em 70% para treino e 30% para teste. Na Figura 20, podemos observar como a previsão dos dados de testes se comportou em relação aos dados reais. A rede foi capaz de acompanhar os movimentos gerais do índice Ibovespa muito bem, no entanto, não conseguiu prever os valores máximos e mínimos dos dados precisamente. Além disso, podemos ver que quanto mais longe do ponto de início dos valores de teste, mais distante a previsão fica do valor real.

Figura 20 - Previsão para a LSTM da seção 4.2.1

Dados de saída previstos e reais considerando o sentimento das notícias



Fonte: Elaborada pelo autor

A Tabela 7 dispõe as métricas SMAPE e MSE obtidas para esse modelo. Para as redes LSTM, o MSE foi calculado com os valores normalizados.

Tabela 7 - Resultados para rede LSTM da seção 4.2.1

Resultados LSTM 4.2.1	
MSE	0,0015
SMAPE	4,09%

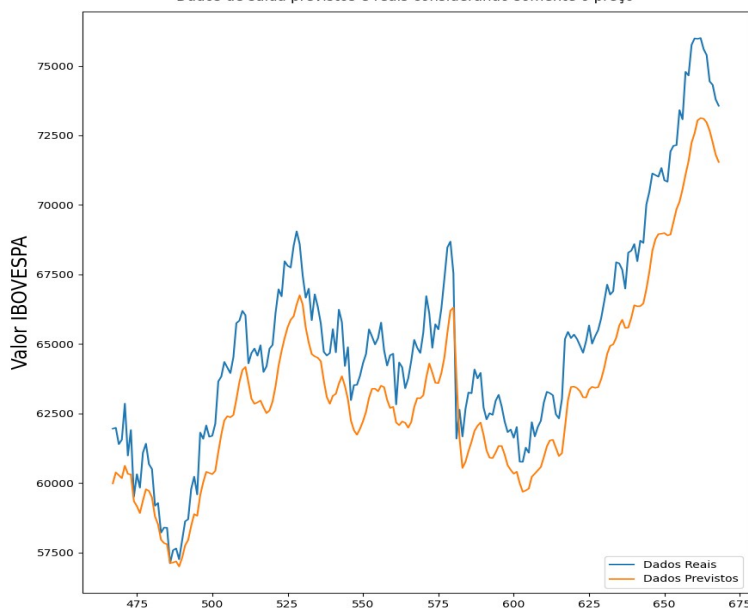
Fonte: Elaborada pelo autor

4.2.2 Conjunto de Dados Univariado Completo

Assim como para a criação do modelo multivariado, utilizou-se as 670 amostras que se obteve de dados de preço de fechamento, dividindo em 70% para treino e 30% para teste. Na Figura 21, podemos observar como a previsão dos dados de testes se comportou em relação aos dados reais. Nessa situação, o modelo aparenta prever a variação do índice de forma precisa, porém acontece um deslocamento da curva de previsão para baixo em relação a curva real. Esse comportamento se repetiu mesmo após a rede ser retreinada com o mesmo conjunto de dados.

Figura 21 - Previsão para a LSTM da seção 4.2.2

Dados de saída previstos e reais considerando somente o preço



Fonte: Elaborada pelo autor

A Tabela 8 dispõe as métricas SMAPE e MSE obtidas para esse modelo. Para as redes LSTM, o MSE foi calculado com os valores normalizados.

Tabela 8 - Resultados para rede LSTM da seção 4.2.2

Resultados LSTM 4.2.2	
MSE	0,003
SMAPE	7,08%

Fonte: Elaborada pelo autor

4.2.3 Conjunto de Dados Multivariado Dividido em Cinco

Para a criação desse modelo, dividiu-se as 670 amostras que se obteve do preço de fechamento e da polaridade de notícias em cinco, dividindo cada conjunto em 70% para treino e 30% para teste. A rede foi retreinada com cada conjunto de dados de treino e as previsões estão dispostas na Figura 22.

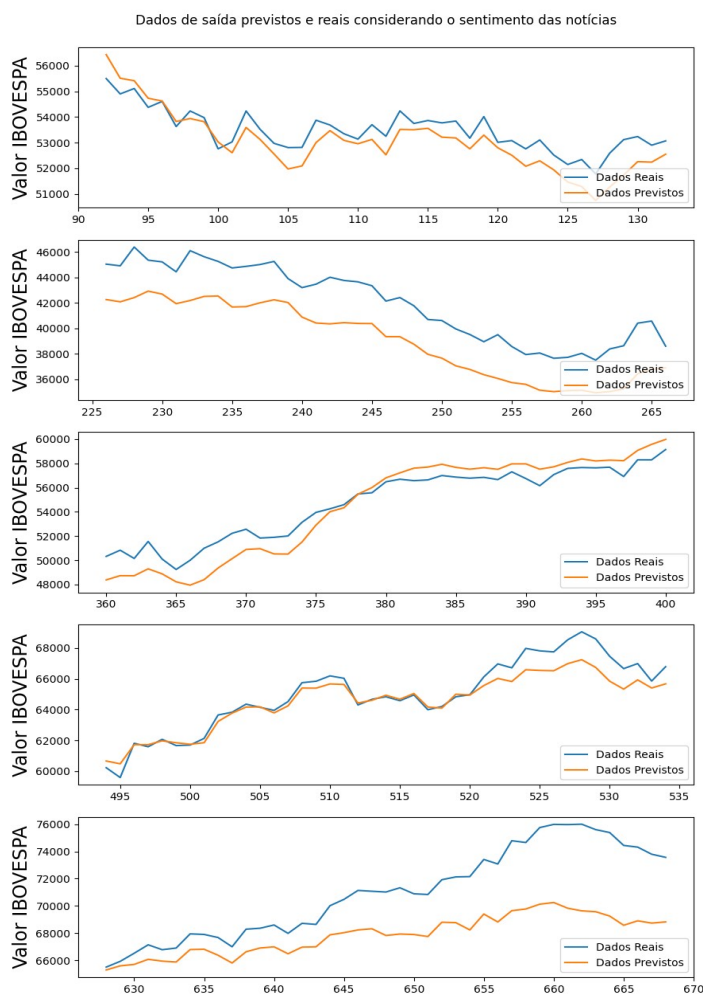
De maneira geral, a previsão da rede conseguiu acompanhar as variações muito bem, no entanto, houve um deslocamento vertical em 2 dos 5 conjuntos de dados testados. No 1º, 3º e 4º conjuntos tivemos uma previsão consistente com a curva real. A Tabela 9 dispõe as métricas SMAPE e MSE obtidas para cada conjunto de teste e o valor médio correspondente. Para as redes LSTM, o MSE foi calculado com os valores normalizados.

Tabela 9 - Resultados para rede LSTM da seção 4.2.3

Resultados LSTM 4.2.3					
MSE por conjunto	0,0005	0,005	0,001	0,0007	0,009
MSE Médio	0,003				
SMAPE por conjunto	4,21%	117,21%	7,93%	2,96%	10,04%
SMAPE Médio	28,47%				

Fonte: Elaborada pelo autor

Figura 22 - Previsões para a LSTM da seção 4.2.3



Fonte: Elaborada pelo autor

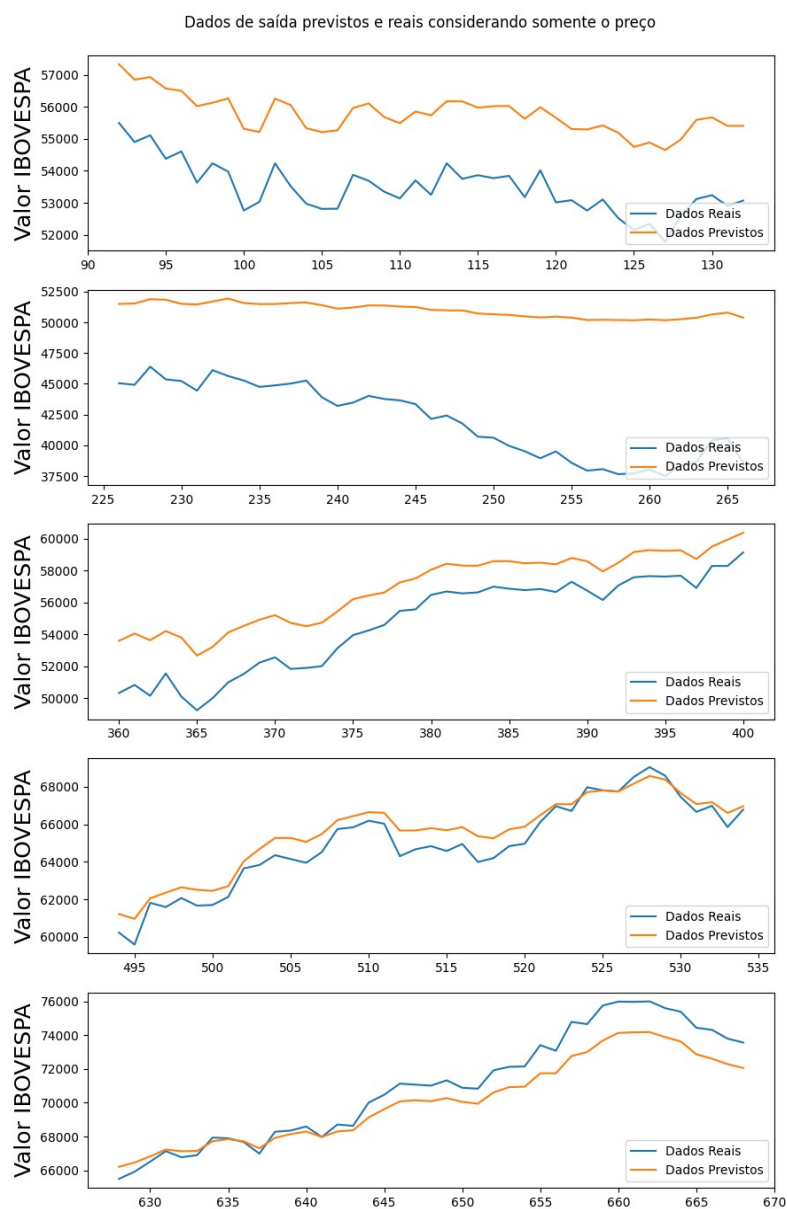
4.2.4 Conjunto de Dados Univariado Dividido em Cinco

Para a criação desse modelo, dividiu-se as 670 amostras que se obteve do preço de fechamento de notícias em cinco, dividindo cada conjunto em 70% para treino e 30% para teste. A rede foi retreinada com cada conjunto de dados de treino e as previsões estão dispostas na Figura 23.

De maneira geral, a previsão da rede conseguiu acompanhar as variações reais, no entanto, houve um deslocamento vertical em 2 dos 5 conjuntos de dados testados. O

segundo conjunto não obteve previsões boas. No 4º e 5º conjuntos tivemos uma previsão consistente com a curva real. A Tabela 10 dispõe as métricas SMAPE e MSE obtidas para cada conjunto de teste e o valor médio correspondente. Para as redes LSTM, o MSE foi calculado com os valores normalizados.

Figura 23 - Previsões para a LSTM da seção 4.2.4



Fonte: Elaborada pelo autor

Tabela 10 - Resultados para rede LSTM da seção 4.2.4

Resultados LSTM 4.2.4					
MSE por conjunto	0,004	0,06	0,003	0,0005	0,01
MSE Médio	0,013				
SMAPE por conjunto	14,00%	111,92%	11,46%	2,63%	3,37%
SMAPE Médio	28,67%				

Fonte: Elaborada pelo autor

4.2.5 Discussão dos Resultados Obtidos para LSTM

Como pode ser observado, a introdução das polaridades diárias das notícias trouxe uma melhoria geral nos modelos de previsão do índice Ibovespa. Tanto no caso em que consideramos o conjunto de dados completos, quanto no caso em que dividimos o conjunto em cinco, foi possível notar uma melhoria no desempenho das redes quando introduzimos essa segunda variável. Outro fator que deve ser ressaltado é o melhor desempenho obtido se utilizando de um conjunto de dados completos em comparação ao aprendizado contínuo com cinco conjuntos de dados.

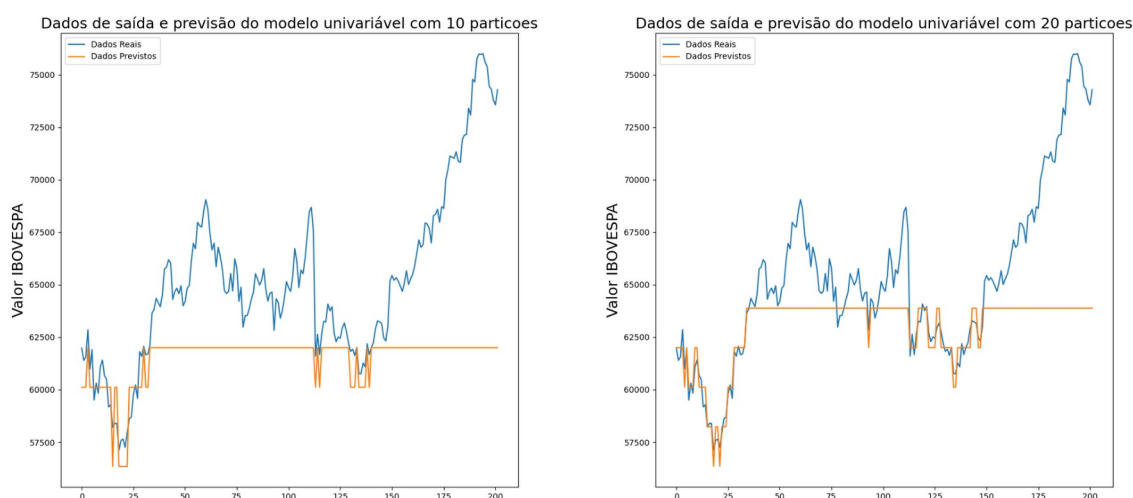
4.3 SÉRIES TEMPORAIS NEBULOSAS

Nessa sessão apresentaremos os resultados obtidos para a previsão utilizando STN. Conforme explicado na seção 3.3.2, testou-se seis conjuntos de dados diferentes: um conjunto completo contendo os dados de fechamento do mercado e a polaridade das notícias; um conjunto completo contendo somente os dados de fechamento do mercado; o primeiro dos conjuntos dividido em cinco e em seteparte; e o segundo conjunto dividido em cinco e em sete partes. Para cada conjunto, variou-se as quantidades de partições do modelo em 10 e 20. Para os dados multivariados, utilizamos dois modelos de STN diferentes, um com peso e um sem.

4.3.1 Modelo Univariado

Para a construção desse modelo, consideramos somente os preços de fechamento. Na Figura 24, temos os dados previstos e reais quando consideramos o conjunto de dados completos para 10 e 20 partições.

Figura 24 - Previsões FTS univariado com conjunto de dados completos



Fonte: Elaborada pelo autor

Como podemos observar, o fenômeno de corte explicado na seção 3.3.2 ocorreu, por isso tivemos que dividir o conjunto de dados em cinco partes. As previsões estão dispostas na Figura 25. A divisão dos dados ainda não foi o suficiente para impedir os cortes e, por conta disso, optou-se por dividir o conjunto de dados em sete partes. As previsões para essa situação estão dispostas na Figura 26.

Apesar da divisão dos dados, ainda foi possível verificar a ocorrência de regiões de corte do modelo. Em todas as situações, foi possível verificar que utilizar 20 partições entrega um resultado superior ao uso de 10. A Tabela 11 dispõe os resultados das métricas para todos os modelos univariados vistos nessa seção. O melhor modelo foi obtido quando dividimos os dados em 7 partes.

Tabela 11 - Resultados modelos univariado

Resultados Modelos Univairados						
Divisões do Conjunto	Completo		5		7	
Partições	10	20	10	20	10	20
MSE Médio	5088,6	3880,8	3114,8	3037,1	1885	1933,5
SMAPE Médio	5,59%	3,57%	4,77%	4,49%	2,76%	2,72%

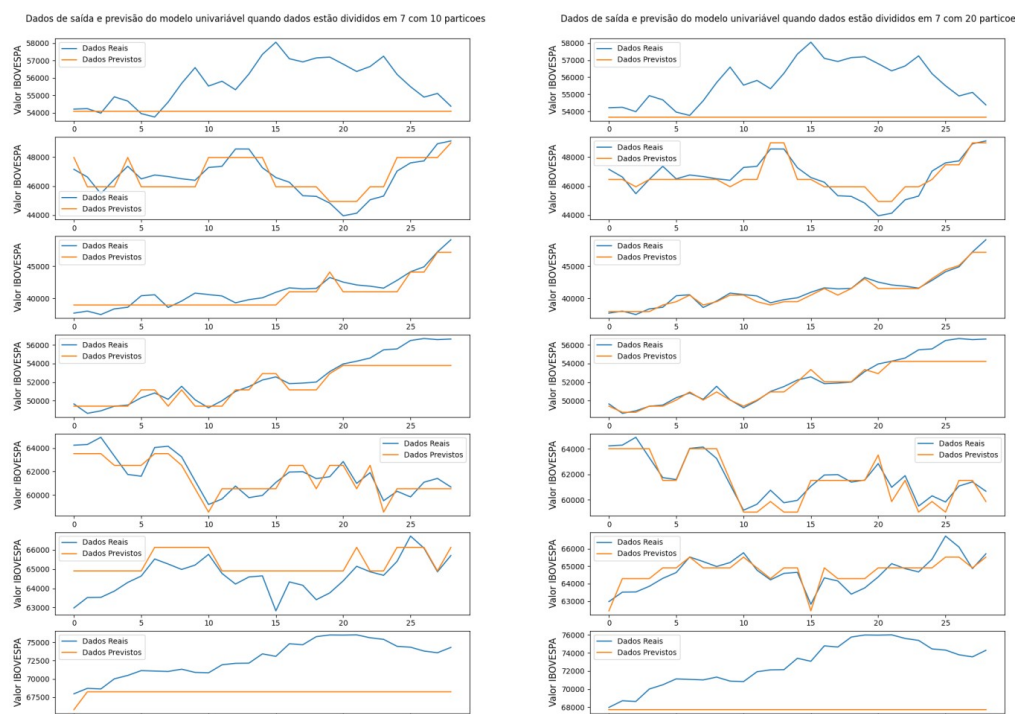
Fonte: Elaborada pelo autor

Figura 25 - Previsões FTS univariado com conjunto de dados dividido em cinco



Fonte: Elaborada pelo autor

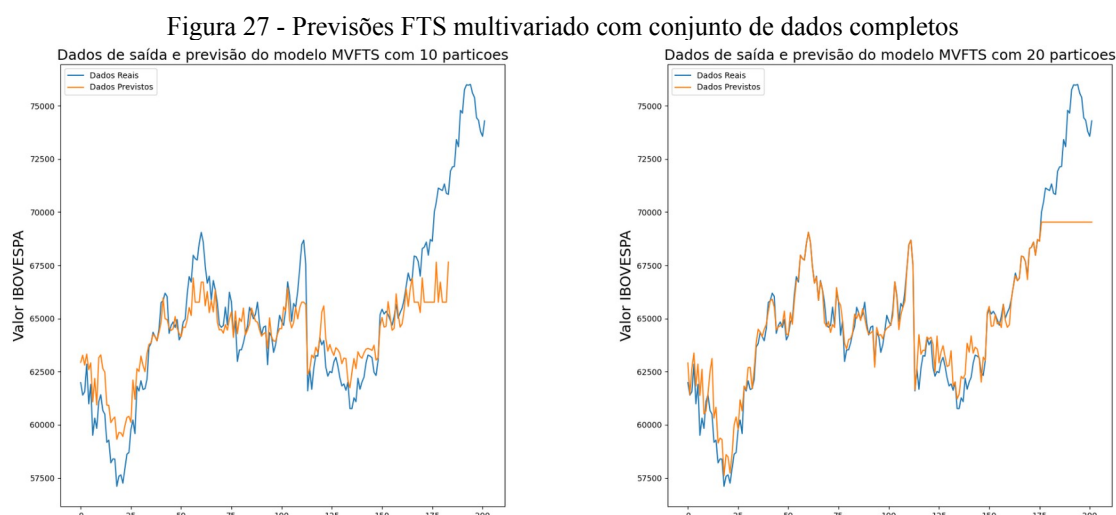
Figura 26 - Previsões FTS univariado com conjunto de dados dividido em sete



Fonte: Elaborada pelo autor

4.3.2 Modelo Multivariado

Construiu-se os modelos multivariados variando as partições e o tamanho do conjunto de dados, da mesma forma que o modelo univariado. A diferença está na introdução das polaridades das notícias na entrada do modelo. Na Figura 27, temos os dados previstos e reais quando consideramos o conjunto de dados completos para 10 e 20 partições.



Fonte: Elaborada pelo autor

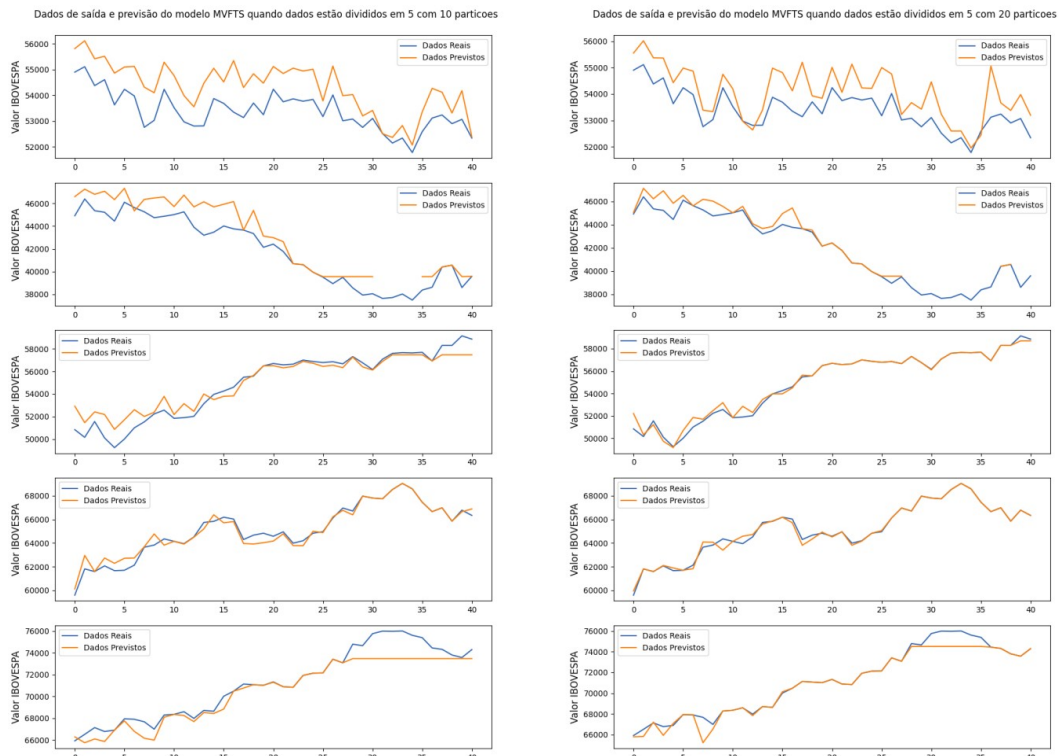
Podemos verificar que a introdução das polaridades das notícias já apresentou uma melhoria na previsão da FTS univariada e não houve tantos cortes como anteriormente. As Figuras 28 e 29 apresentam as previsões para a divisão dos dados em cinco e sete partes, respectivamente. Podemos perceber que, para todas os conjuntos, dividir o espaço da série temporal em 20 partições resultou em uma acurácia melhor que quando se divide em 10. A Tabela 12 dispõe os resultados das métricas para todos os modelos multivariados vistos nessa seção. Podemos notar que o melhor modelo foi obtido quando dividimos os dados em 5 partes e utilizamos 20 partições.

Tabela 12 - Resultados modelos multivariado

Resultados Modelos Multivariados						
Divisões do Conjunto	Completo		5		7	
Partições	10	20	10	20	10	20
MSE Médio	1732,4	1811,5	1193	900,8	1314,4	998
SMAPE Médio	2,00%	1,70%	1,80%	1,32%	1,92%	1,44%

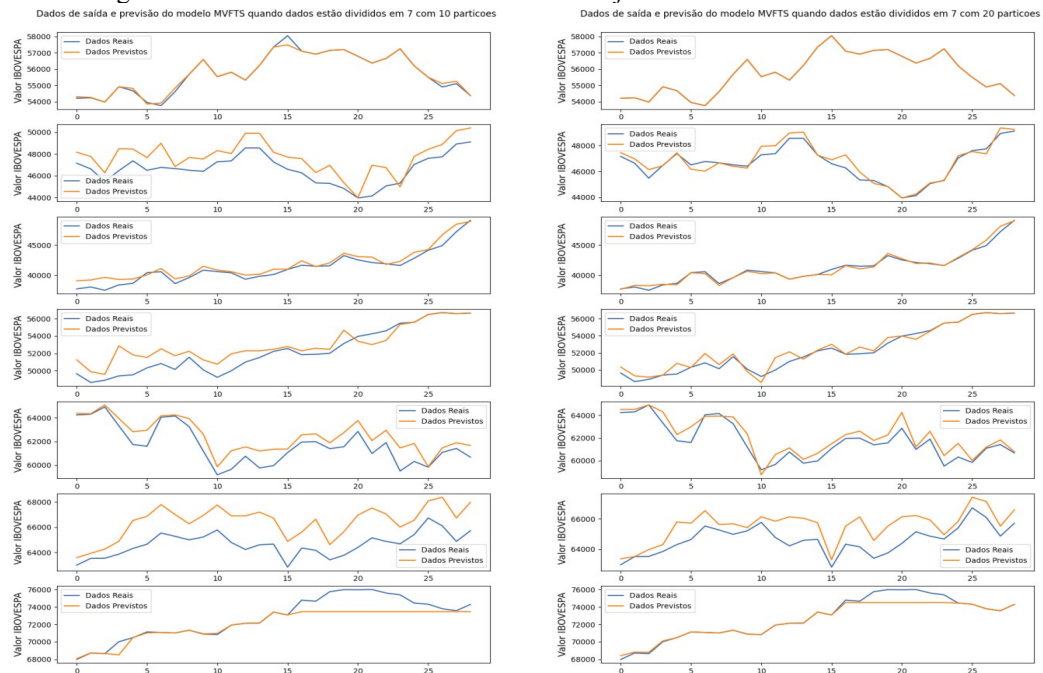
Fonte: Elaborada pelo autor

Figura 28 - Previsões FTS multivariado com conjunto de dados dividido em cinco



Fonte: Elaborada pelo autor

Figura 29 - Previsões FTS multivariado com conjunto de dados dividido em sete

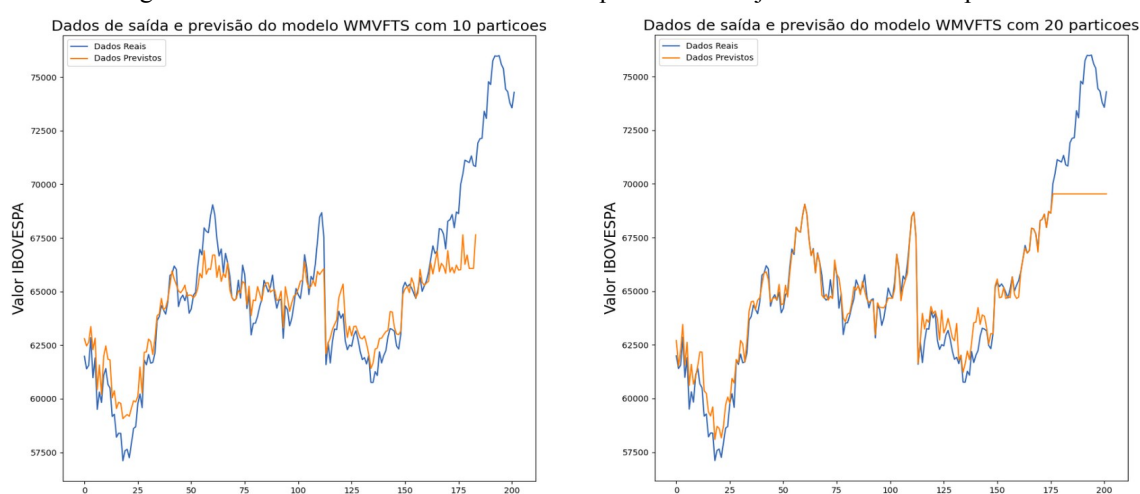


Fonte: Elaborada pelo autor

4.3.3 Modelo Multivariado com Pesos

Por fim, criou-se modelos multivariados com pesos, variando as partições e o tamanho do conjunto de dados, da mesma forma que nos dois modelos anteriores. Na Figura 30, temos os dados previstos e reais quando consideramos o conjunto de dados completos para 10 e 20 partições.

Figura 30 - Previsões FTS multivariado com pesos com conjunto de dados completos



Fonte: Elaborada pelo autor

A introdução de pesos melhorou o desempenho do modelo, mas, ainda assim, no final da série ocorre o fenômeno de corte. As Figuras 31 e 32 apresentam as previsões para a divisão dos dados em cinco e sete partes, respectivamente. Podemos perceber que, assim como para o modelo anterior, dividir o espaço da série temporal em 20 partições resultou em uma acurácia melhor que quando se divide em 10. A Tabela 13 dispõe os resultados das métricas para todos os modelos multivariados com peso vistos nessa seção. Podemos notar que o melhor modelo foi obtido quando dividimos os dados em 5 partes e utilizamos 20 partições.

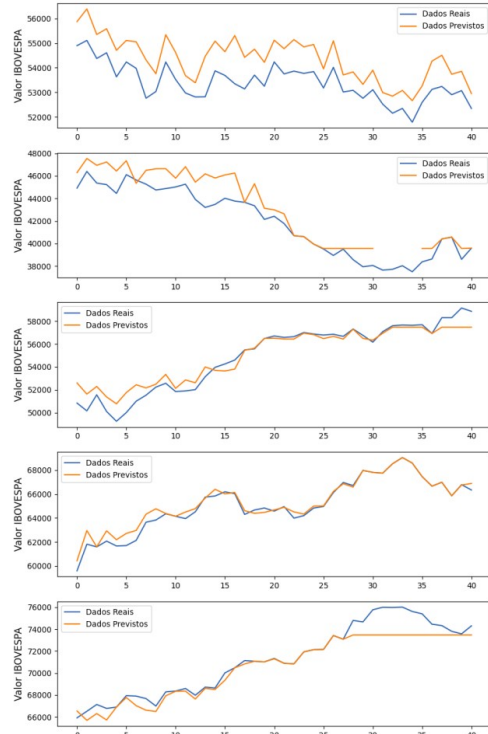
Tabela 13 - Resultados modelos multivariado com pesos

Resultados Modelos Multivariados com Pesos						
Divisões do Conjunto	Completo		5		7	
Partições	10	20	10	20	10	20
MSE Médio	1554,5	1783,8	1154	885,1	1251,9	952
SMAPE Médio	1,71%	1,67%	1,74%	1,31%	1,85%	1,38%

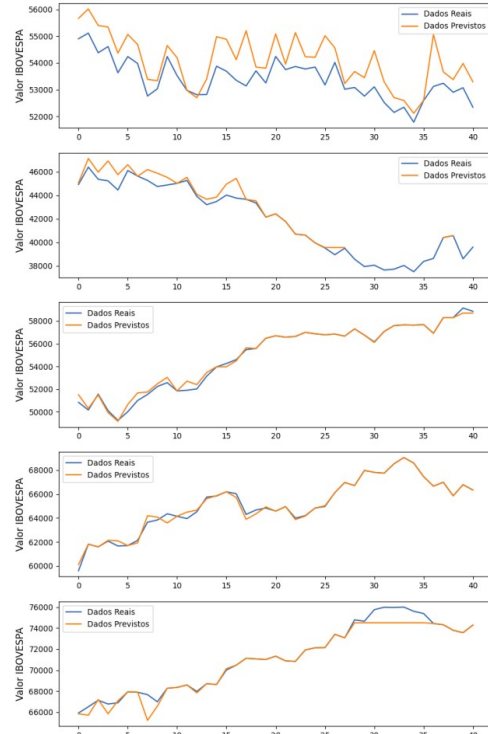
Fonte: Elaborada pelo autor

Figura 31 - Previsões FTS multivariado com pesos com conjunto de dados dividido em cinco

Dados de saída e previsão do modelo WMVFTS quando dados estão divididos em 5 com 10 partições



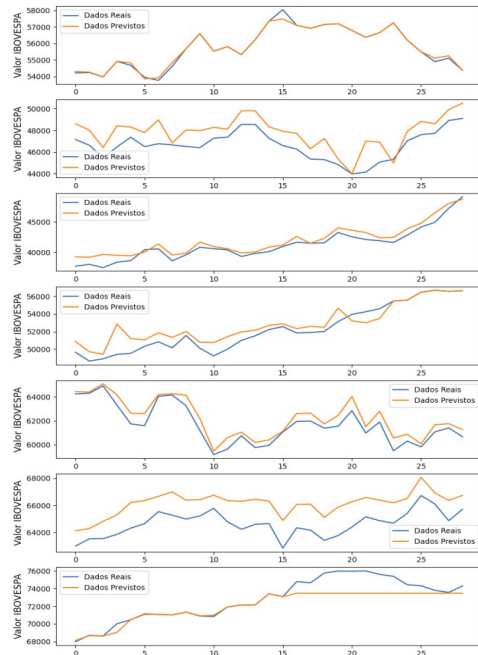
Dados de saída e previsão do modelo WMVFTS quando dados estão divididos em 5 com 20 partições



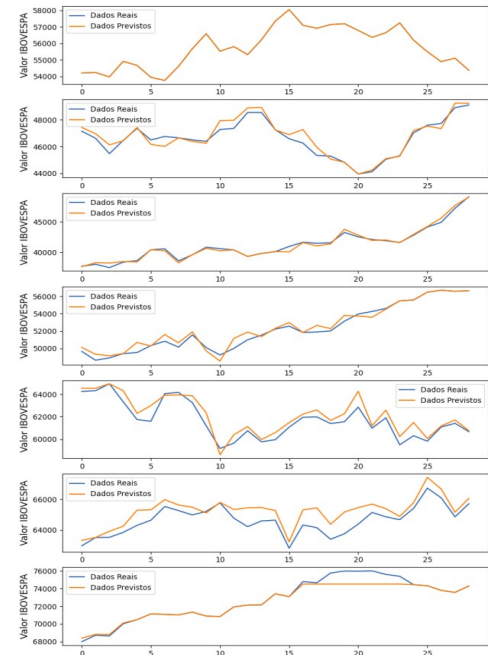
Fonte: Elaborada pelo autor

Figura 32 - Previsões FTS multivariado com pesos com conjunto de dados dividido em sete

Dados de saída e previsão do modelo WMVFTS quando dados estão divididos em 7 com 10 partições



Dados de saída e previsão do modelo WMVFTS quando dados estão divididos em 7 com 20 partições



Fonte: Elaborada pelo autor

4.3.4 Discussão dos Resultados Obtidos para STN

A introdução da polaridade diária das notícias melhorou o desempenho dos modelos que utilizaram a técnica de STN para fazer a previsão do índice Ibovespa. De forma geral, o uso de 20 partições foi o melhor para o conjunto de dados que estamos analisando. Ao realizar a divisão dos dados em cinco partes também notamos resultados significativamente melhores, pois reduziu-se o problema do corte que foi discutido na seção 3.3.2.

4.4 COMPARAÇÃO LSTM E STN

Ambas as técnicas utilizadas possuem vantagens e desvantagens que discutiremos brevemente nessa seção. Por um lado as redes LSTM foram capazes de prever todo o conjunto de dados sem sofrer cortes. Pelo outro, as STN apresentaram um desempenho superior em termos das métricas SMAPE (como o MSE é sensível a escala, optou-se por não usá-lo para comparar um modelo com o outro).

Dentre todos, o melhor desempenho foi para o modelo STN multivariado com pesos. Mesmo apresentando alguns cortes durante a previsão de dados, a performance obtida pelo modelo foi de longe a melhor e, por isso, conquistou o pódio. Um segundo lugar próximo é o modelo multivariado sem pesos. Sua performance foi levemente pior, mas, ainda com os cortes, foi superior aos modelos de LSTM. O modelo multivariado de LSTM fica em terceiro e o modelo univariado da LSTM em quarto.

O modelo univariado de STN teve o pior desempenho. Apesar de suas métricas serem melhores que as métricas de ambos os modelos LSTM, os cortes contantes na previsão desqualificam esse modelo. Com isso, concluímos que os modelos multivariados, de forma geral, possuem um desempenho superior aos modelos univariados estudados. Os três modelos que fizeram uso das polaridades de notícias ficaram acima dos que usaram somente os preços de fechamento do índice.

4.5 ESTRATÉGIA DE INVESTIMENTOS

Por fim, utilizou-se os modelos gerados acima para fazer um *backtesting*, conforme foi descrito na seção 3.4. Os modelos considerados para aplicação da estratégia de investimento foram apenas os modelos que utilizaram do conjunto de dados completo. A Tabela 14 mostra os retornos obtidos ao aplicar a estratégia de investimento para cada modelo, bem como a melhoria em relação ao rendimento obtido pela técnica de *buy and hold*.

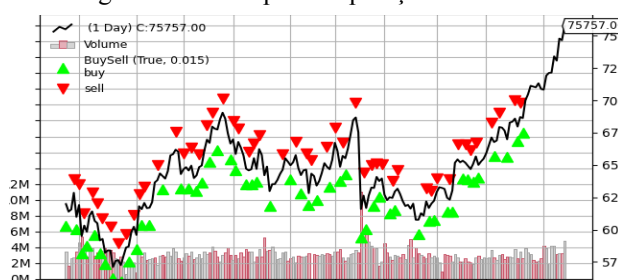
Tabela 14 - Resultado da aplicação de investimento para alguns dos modelos

Modelo	Retorno Obtido	Desempenho comparado ao <i>buy and hold</i>
LSTM Univariado	53,49%	238,97%
LSTM Multivariado	52,83%	234,79%
STN Univariado	26,71%	69,26%
STN Multivariado	56,64%	258,94%
STN Multivariado com Pesos	57,38%	263,62%
<i>Buy and hold</i>	15,78%	0,00%

Fonte: Elaborada pelo autor

Como pode-se perceber, para as redes LSTM o desempenho quando se aplica uma estratégia de investimentos foi semelhante. Os resultados obtidos para as STN multivariadas também foram semelhantes entre si e próximos ao resultado obtido utilizando as LSTM. Para o STN univariado, tivemos uma queda no desempenho, no entanto, ainda assim obteve-se um resultado 69,25% melhor que se tivéssemos aplicado uma estratégia de *buy and hold*. A Figura 33 mostra um exemplo das operações feitas durante o *backtesting* para o STN multivariado com pesos. Em verde, operações de compra, em vermelho, de venda.

Figura 33 - Exemplo de operações realizadas



Fonte: Elaborada pelo autor

5 CONCLUSÃO

O uso de polaridades de notícias na previsão do índice Ibovespa, feita por algoritmos de aprendizado de máquina, tem um impacto positivo em seu desempenho. Em todos os algoritmos estudados, o uso de polaridades de notícias resultou em uma previsão mais próxima dos valores reais. Além disso, os algoritmos que utilizaram a polaridade tiveram um retorno no investimento semelhante ou levemente melhor que os algoritmos que se limitavam ao uso somente do preço de fechamento, quando submetidos a mesma estratégia simples de investimento.

O uso de técnicas de aprendizado de máquina se provou superior ao uso do *buy and hold*, técnica comumente adotada no mercado para investimentos ao longo prazo. A possibilidade do uso de LSTM para operações de médio e longo prazo é mais atrativa, visto que a rede se demonstrou capaz de prever dados em períodos maiores. No curto prazo, o uso de STN é recomendado, visto que apresentaram uma acurácia maior em períodos curtos. Em ambos os casos, as polaridades das notícias poderão ser utilizadas para melhorar o desempenho do algoritmo.

Os achados desse estudo ajudam a comprovar a ideia de que o mercado financeiro brasileiro possui leves ineficiências, pois o investidor capaz de acessar mais informações e fazer uma interpretação dessas notícias poderá obter um retorno superior. Além disso, podemos observar que a tese da economia comportamental não pode ser descartada e que o sentimento do mercado é um fator importante para melhorar sua previsibilidade.

Em estudos futuros, será interessante explorar como a polaridade encontrada em outros agentes de notícias e/ou em redes sociais podem impactar o mercado financeiro. Além disso, o desenvolvimento de um léxico específico para notícias do mercado financeiro e de uma estratégia de investimento para se beneficiar do uso da polaridade de notícias podem ser contempladas. Recomenda-se, também, a realização de estudos com ações diferentes do mercado financeiro e a customização dos parâmetros e das camadas dos algoritmos usados.

Além disso, pode-se propor a investigação de novas arquiteturas de Processamento de Linguagem Natural, como o Attention e Transformers. O uso de

modelos de Series Temporais Nebulosas mais adequados, como o modelos variantes no tempo, e Redes LSTM mais profundas e de maior ordem também devem ser investigados. Por fim, será interessante verificar como que o uso da previsão intervalar e probabilística pode ser utilizada para criação de cenários de riscos, que conversam com uma estratégia de investimento mais robusta.

REFERÊNCIAS

ALMEIDA, R. J. A. **LeIA - Léxico para Inferência Adaptada**. Disponível em: <<https://github.com/rafjaa/LeIA>>. Acesso em: 25 set. 2020.

AMIDI, A.; AMIDI, S. **CS 230 - Recurrent Neural Networks Cheatsheet**. Disponível em: <<https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>>. Acesso em: 26 set. 2020.

APPLE, G. **Technical Analysis Power Tools for Active Investors**. 1ª ed. [s.l.] Ft Press, 2005.

BHANDE, A. **What is underfitting and overfitting in machine learning and how to deal with it**. Disponível em: <<https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76>>. Acesso em: 26 set. 2020.

BROWNLEE, J. **How to Decompose Time Series Data into Trend and Seasonality** *Machine Learning Mastery*, 29 jan. 2017. Disponível em: <<https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>>. Acesso em: 25 set. 2020

BROWNLEE, J. **How to Choose Loss Functions When Training Deep Learning Neural Networks** *Machine Learning Mastery*, 29 jan. 2019. Disponível em: <<https://machinelearningmastery.com/how-to-choose-loss-functions-when-training-deep-learning-neural-networks/>>. Acesso em: 25 set. 2020

BURCHELL, J. **Using VADER to handle sentiment analysis with social media text**. Blog. Disponível em: <<https://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html>>. Acesso em: 25 set. 2020.

BUSHAEV, V. **Understanding RMSprop — faster neural network learning**. Disponível em: <<https://towardsdatascience.com/understanding-rmsprop-faster-neural-network-learning-62e116fcf29a>>. Acesso em: 30 set. 2020.

CAVALCANTE, R. C. et al. Computational Intelligence and Financial Markets: A Survey and Future Directions. **Expert Systems with Applications**, v. 55, p. 194–211, ago. 2016.

CHATRATH, A. et al. Currency jumps, cojumps and the role of macro news. **Journal of International Money and Finance**, v. 40, p. 42–62, fev. 2014.

CHEN, S.-M. Forecasting enrollments based on fuzzy time series. **Fuzzy Sets and Systems**, v. 81, n. 3, p. 311–319, ago. 1996.

CHOLLET, F. **Keras**. Disponível em: <<https://github.com/fchollet/keras>>. Acesso em: 30 set. 2020.

CINTULA, P.; FERMÜLLER, C. G.; NOGUERA, C. Fuzzy Logic. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2017 ed. [s.l.] Metaphysics Research Lab, Stanford University, 2017.

CLOWES, M. **Adaptive-market theory offers investor insights**. Disponível em: <<https://www.investmentnews.com/article/20050207/SUB/502070709/1008/TOC>>. Acesso em: 16 jun. 2019.

DAS, S. R. et al. **Yahoo! for amazon: Sentiment extraction from small talk on the web**. 8th Asia Pacific Finance Association Annual Conference. **Anais...**2001

DRO. **mementum/backtrader**. [s.l: s.n.].

FAMA, E. F. The Behavior of Stock-Market Prices. **The Journal of Business**, v. 38, n. 1, p. 34–150, jan. 1965.

FAMA, E. F. Efficient Capital Markets: A Review of Theory and Empirical Work. **The Journal of Finance**, v. 25, p. 383–417, 1970.

FOLHA DE SÃO PAULO. **Para mercado, nem pacote fiscal evitará degradação financeira - 25/09/2015 - Mercado**. Disponível em: <<http://www1.folha.uol.com.br/mercado/2015/09/1686300-para-mercado-nem-pacote-fiscal-evitara-degradacao-financeira.shtml>>. Acesso em: 28 set. 2020.

GRAVES, A.; MOHAMED, A.; HINTON, G. Speech Recognition with Deep Recurrent Neural Networks. **arXiv:1303.5778 [cs]**, 22 mar. 2013.

GUPTA, R. **Getting started with Neural Network for regression and Tensorflow**. Disponível em: <<https://medium.com/@rajatgupta310198/getting-started-with-neural-network-for-regression-and-tensorflow-58ad3bd75223>>. Acesso em: 25 set. 2020.

HAYES, A. **Correlation**. Disponível em: <<https://www.investopedia.com/terms/c/correlation.asp>>. Acesso em: 1 out. 2020.

HAYES, A. **How to Use the Dow Theory to Analyze the Market**. Disponível em: <<https://www.investopedia.com/terms/d/dowtheory.asp>>. Acesso em: 16 jun. 2019.

HOCHREITER, S.; SCHMIDHUBER, J. Long Short-Term Memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1 nov. 1997.

KHADJEH NASSIRTOUSSI, A. et al. Text mining for market prediction: A systematic review. **Expert Systems with Applications**, v. 41, n. 16, p. 7653–7670, nov. 2014.

LIN, T. C. W. A Behavioral Framework for Securities Risk. **Seattle U. L. Rev**, v. 34, p. 325–380, 2011.

LO, A. W. The Adaptive Markets Hypothesis. **The Journal of Portfolio Management**, v. 5, n. 2004, p. 15–29, 2004.

LO, A. W. *Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis*. p. 44, 2005.

MANNARSWAMY, S. **Everything you need to know about neural networks** **Open Source For You**, 16 mar. 2017. Disponível em: <<https://www.opensourceforu.com/2017/03/neural-networks-in-detail/>>. Acesso em: 16 set. 2020

MARLESSON. **News of the Brazilian Newspaper**. Disponível em: <<https://kaggle.com/marlesson/news-of-the-site-folhauol>>. Acesso em: 29 jun. 2020.

MULLAINATHAN, S.; THALER, R. **Behavioral Economics**. Cambridge, MA: National Bureau of Economic Research, out. 2000. Disponível em: <<http://www.nber.org/papers/w7948.pdf>>. Acesso em: 9 jun. 2019.

OLAH, C. **Understanding LSTM Networks -- colah's blog**. Disponível em: <<http://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Acesso em: 26 set. 2020.

PEIXEIRO, M. **The Complete Guide to Time Series Analysis and Forecasting**. Disponível em: <<https://towardsdatascience.com/the-complete-guide-to-time-series-analysis-and-forecasting-70d476bfe775>>. Acesso em: 25 set. 2020.

PRIANDA, B. G. **Chen and Cheng's Model on The Fuzzy Time Series Method for Forecasting the Number of Honda Motor....** Disponível em: <<https://medium.com/@BAYUGALIH/chen-and-chengs-model-on-the-fuzzy-time-series-method-for-forecasting-the-number-of-honda-motor-468413b4f57a>>. Acesso em: 26 set. 2020.

SCHUMAKER, R. P. et al. Evaluating sentiment in financial news articles. **Decision Support Systems**, v. 53, n. 3, p. 458–464, jun. 2012.

SILVA, P. **Um breve tutorial sobre Séries Temporais Nebulosas**. Disponível em: <<https://medium.com/ensina-ai/um-breve-tutorial-sobre-s%C3%A9ries-temporais-nebulosas-1ef6ab86ce12>>. Acesso em: 24 set. 2020a.

SILVA, P. **Um breve tutorial sobre Séries Temporais Nebulosas — Parte II (com estudo de caso em Energia Solar)**. Disponível em: <<https://medium.com/ensina-ai/um-breve-tutorial-sobre-s%C3%A9ries-temporais-nebulosas-parte-ii-com-estudo-de-caso-em-energia-solar-c87092717c39>>. Acesso em: 26 set. 2020b.

SILVA, P. C. D. L. E. et al. **PYFTS/pyFTS: Stable version 1.6**. [s.l.] Zenodo, 2019.

SONG, Q.; CHISSOM, B. S. Forecasting enrollments with fuzzy time series — Part I. **Fuzzy Sets and Systems**, v. 54, n. 1, p. 1–9, fev. 1993.

SONI, U. **VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text**. Disponível em: <<https://medium.com/@urvisoni/vader-a-parsimonious-rule-based-model-for-sentiment-analysis-of-social-media-text-34380204b96f>>. Acesso em: 25 set. 2020.

VENKATACHALAM, M. **Recurrent Neural Networks**. Disponível em: <<https://towardsdatascience.com/recurrent-neural-networks-d4642c9bc7ce>>. Acesso em: 26 set. 2020.

YU, L.-C. et al. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. **Knowledge-Based Systems**, v. 41, p. 89–97, mar. 2013.