



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE DO CAMPUS ARARANGUÁ
CURSO DE GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Jimmy Li Yan

**Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e
Aprendizado Profundo**

Araranguá
2023

Jimmy Li Yan

**Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e
Aprendizado Profundo**

Trabalho de Conclusão de Curso submetido ao Curso de Graduação em Engenharia de Computação do Centro de Ciências, Tecnologias e Saúde do Campus Araranguá da Universidade Federal de Santa Catarina como requisito parcial para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Alexandre Leopoldo Gonçalves, Dr.

Araranguá
2023

**Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.**

Yan, Jimmy Li

Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e Aprendizado Profundo / Jimmy Li Yan ; orientador, Alexandre Leopoldo Gonçalves, 2023.

40 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Campus Araranguá, Graduação em Engenharia de Computação, Araranguá, 2023.

Inclui referências.

1. Engenharia de Computação. 2. Predição de mercado financeiro. 3. Análise de sentimentos. 4. Classificação de sentimentos. 5. Rede neural profunda. I. Gonçalves, Alexandre Leopoldo. II. Universidade Federal de Santa Catarina. Graduação em Engenharia de Computação. III. Título.

Jimmy Li Yan

**Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e
Aprendizado Profundo**

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia de Computação e aprovado em sua forma final pelo Curso de Graduação em Engenharia de Computação.

Araranguá, 30 de Novembro de 2023.

Prof. Jim Lau, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. Alexandre Leopoldo Gonçalves, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof^a. Andréa Sabedra Bordin, Dra.
Avaliador(a)
Universidade Federal de Santa Catarina

Prof^a. Olga Yevseyeva, Dra.
Avaliador(a)
Universidade Federal de Santa Catarina

Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e Aprendizado Profundo

Stock Market Prediction Based on Sentiment Analysis and Deep Learning

Jimmy Li Yan¹

Alexandre Leopoldo Gonçalves²

2023, Novembro

Resumo

O preço do mercado financeiro representa o resultado da disputa entre compradores e vendedores em relação às expectativas futuras. Soma-se a isto o fato que atualmente as pessoas expressam frequentemente suas opiniões, julgamentos e expectativas sobre o mercado financeiro nas redes sociais. Com a evolução da área de Análise de Sentimentos, em especial, utilizando dados de redes sociais, cada vez mais as percepções dos usuários são utilizadas na previsão de tendências do mercado de ações. Neste sentido, este trabalho propõe um método visando explorar a eficácia da Análise de Sentimentos em dados obtidos a partir de uma rede social como ferramenta para prever movimentos de mercado acionário. Este método é composto por algumas etapas, iniciando pela coleta de mensagens relacionadas ao mercado financeiro de uma rede social, seguida pela aplicação de um modelo de linguagem pré-treinado para a classificação de sentimentos em positivos, neutros e negativos. Os dados são então integrados quantitativamente aos dados do mercado de ações, com foco nos preços de fechamento para servir de base para o modelo de previsão. Para a etapa de previsão desenvolveu-se uma rede neural profunda capaz de lidar com séries temporais e lembrar de informações por longos períodos, sendo vista como ideal para identificar padrões nos preços de determinado ativo, por exemplo, ações ou índices financeiros. Para a etapa de avaliação do método proposto, foram consideradas algumas métricas comumente utilizadas no Aprendizado de Máquina. Foram estabelecidos três experimentos (previsão de toda a série temporal, um dia e cinco dias) gerando modelos de aprendizado com e sem dados de sentimentos. Os resultados sugerem um incremento na precisão das previsões quando são incluídos dados de análise de sentimentos, especialmente para o curto prazo. As representações visuais dos resultados preditivos ressaltam a capacidade do modelo de aprendizado em espelhar os valores reais de determinado ativo, ressaltando o potencial da análise de sentimentos como um complemento significativo aos modelos convencionais de previsão de mercado.

Palavras-chave: Predição de mercado financeiro. Análise de sentimentos. Classificação de sentimentos. Rede neural profunda.

¹ jimyan273@gmail.com

² a.l.goncalves@ufsc.br

Predição de Preços de Mercado Financeiro baseada na Análise de Sentimentos e Aprendizado Profundo

Stock Market Prediction Based on Sentiment Analysis and Deep Learning

Jimmy Li Yan³

Alexandre Leopoldo Gonçalves⁴

2023, November

Abstract

The price in the financial market represents the outcome of the contest between buyers and sellers that can also be considered as a psychological battle concerning future expectations. Added to this is the fact that nowadays people frequently express their opinions, judgments, and expectations about the financial market on social networks. With the evolution of the field of Sentiment Analysis, particularly using social media data, the perceptions of people/users are increasingly being used to predict stock market trends. In this context, this work proposes a method to explore the effectiveness of Sentiment Analysis on data from a social network as a tool to predict stock market movements. This method consists of several stages, starting with the scraping of messages related to the financial market from a social network, followed by the application of a pre-trained language model for classifying sentiments into positive, neutral, and negative. The data are then quantitatively integrated into the stock market data, focusing on closing prices to serve as the basis for the prediction model. For the prediction stage, a deep neural network capable of handling time series and remembering information for long periods was developed, seen as ideal for identifying patterns in the prices of a given asset, such as stocks or financial indices. For evaluating the proposed method, some commonly used Machine Learning metrics were considered. For this, three experiments were established (prediction of the entire time series, one day, and five days) generating learning models with and without sentiment data. The results suggest an increase in the accuracy of predictions when sentiment analysis data is included, especially for short-term forecasts. The visual representations of the predictive results highlight the learning model's ability to mirror the actual values of a given asset, underscoring the potential of sentiment analysis as a significant complement to conventional market prediction models.

Keywords: Stock market prediction. Sentiment analysis. Sentiment classification. Deep neural network.

³ jimyan273@gmail.com

⁴ a.l.goncalves@ufsc.br

1. INTRODUÇÃO

O mercado de ações se baseia no conceito de negociações de compra e venda de ativos, sendo realizadas por indivíduos comuns ou instituições que buscam obter lucros a partir do crescimento da empresa que recebeu determinado investimento (BUSTOS; POMARES-QUIMBAYA, 2020). Essas transações do mercado têm se tornado cada vez mais um assunto discutido no cotidiano da população e, com o crescimento das economias, o número de investidores não profissionais aumentou consideravelmente (ORTMANN; PELSTER; WENGEREK, 2020).

Sendo assim, existe a percepção por parte da população que, ao realizar investimentos não profissionais durante um período de crescimento do mercado de ações, os resultados podem ser positivos, o que tem promovido aumento no número de investidores pessoa física. Tais investimentos possuem uma pequena influência em índices de bolsas de valores, como por exemplo, o S&P500 (um dos principais índices de mercado de ações dos Estados Unidos). Esses índices, funcionam como um indicador do desempenho do mercado de ações ajudando investidores na tomada de decisão à medida que estes variam ao longo do tempo (JIANG, 2021; PARNES, 2019).

As movimentações do mercado financeiro ocorrem de maneira aleatória e, desta forma, prever seu comportamento é uma tarefa complexa, principalmente para pessoas com pouca experiência no mercado de ações. Segundo Carosia, Coelho e Silva (2021), é possível obter algum conhecimento a partir do histórico dos preços das ações, assim como a partir dos sentimentos das notícias, visto que estes influenciam o mercado financeiro, sendo possível então trabalhar na ideia combinada para realizar previsões de mercado.

Mais recentemente, com o crescimento das redes sociais, essas passaram a ser consideradas como um meio para auxiliar na previsão de movimentações do mercado financeiro, entre elas o X[®] (antigo Twitter[®]) se destaca (ARAUJO; FERNANDES, 2021). Tais mídias possibilitam a busca de postagens de diversos assuntos, por exemplo, sobre economia, que por sua vez, tem potencial para afetar diretamente o mercado. Isto se justifica visto que investidores podem tomar decisões baseando-se em mensagens publicadas nas redes sociais, desencadeando um efeito borboleta, já que as notícias se espalham rapidamente, conseguindo influenciar outras pessoas (REED, 2016).

As mensagens de texto podem ser classificadas como positivas, negativas ou neutras, de acordo com a análise de sentimentos baseada nas opiniões e atitudes em relação às entidades, sejam elas pessoas, instituições ou produtos (LIU, 2015). Assim, esses dados constituem-se em uma valiosa fonte para a área de Análise de Sentimentos (do inglês *Sentiment Analysis* - SA) sobre o mercado de ações, uma vez que possuem características como grande volume, velocidade e heterogeneidade, além de serem criados por usuários e conter um significado subjetivo (POUROMID *et al.*, 2021). A utilização de SA, considerando textos de mídias sociais, pode ser verificada no trabalho realizado por Ji *et. al* (2023), em que os autores utilizaram um modelo de linguagem para analisar o sentimento de textos de mídias sociais, assim como uma arquitetura de rede neural para prever a movimentação de 28 ações com resultados promissores.

Neste sentido, para realizar uma análise de sentimentos, pode-se utilizar algoritmos de Aprendizado de Máquinas (do inglês *Machine Learning* - ML) e de Aprendizado Profundo (do inglês *Deep Learning* - DL). Como um ramo do ML, o DL consiste na utilização de algoritmos, também referenciados como arquiteturas, de redes neurais de múltiplas camadas para construir modelos voltados à tomada de decisão (HAO; ZHANG; MA, 2016). Por exemplo, arquiteturas que utilizam DL possuem grande capacidade de extrair informações relevantes de séries temporais de dados financeiros, tanto atuais como passados, viabilizando

a construção de modelos voltados à realização de previsões no mercado de ações (NABIPOUR *et al.*, 2020).

Entre as arquiteturas mais utilizadas para a realização de previsões encontra-se a arquitetura de Rede Neurais Recorrentes (do inglês *Recurrent Neural Networks* - RNN), mas especificamente as de Memória de Longo e Curto Prazo (do inglês *Long Short-Term Memory* - LSTM). Esta arquitetura consegue fazer uma distinção entre dados recentes e antigos, atribuindo pesos diferentes para cada um deles, de modo a esquecer aqueles que considerar irrelevante para realizar a previsão da próxima saída, bem como processar longas sequências de dados de entrada quando comparada com outras arquiteturas (ARAÚJO; FERNANDES, 2021).

Por exemplo, no trabalho de Jia *et al.* (2019), os autores utilizaram uma arquitetura LSTM com uma camada a fim de obter melhores previsões do mercado financeiro, resultando em uma precisão moderada. Já Pan *et al.* (2023), utilizaram dados públicos de mercado e registros de transações para avaliar os movimentos dos preços de ações através de uma rede LSTM bidirecional. Os resultados foram considerados adequados quando comparados a outros modelos, realizando uma boa previsão dos preços das ações e servindo de base para a tomada de decisões financeiras.

Considerando os aspectos acima mencionados, este trabalho tem por objetivo a proposição de um método que integre a SA e o DL no cenário de previsão de mercado financeiro por meio de Redes Neurais Artificiais (do inglês *Artificial Neural Network* - ANN), mais especificamente a arquitetura LSTM. Ademais, se utiliza de modelos de linguagem baseados na arquitetura *transformers* para a análise de sentimentos. O método visa auxiliar investidores na tomada de decisão sobre a compra, a venda ou a manutenção de determinado ativo.

Este documento, para além da presente seção, é composto por mais cinco seções. A segunda seção se refere à Fundamentação Teórica promovendo suporte para o método proposto neste trabalho no que tange aos principais conceitos, seguida pela Seção 3 que apresenta os trabalhos correlatos. Na Seção 4 é detalhado o método proposto com suas quatro etapas e, na Seção 5, são detalhados os materiais e métodos utilizados no trabalho, bem como são discutidos os resultados experimentais. Por último, a Seção 6 contém as considerações finais e trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. MERCADO DE AÇÕES

O mercado de ações é uma parte fundamental de qualquer economia representando o local onde acontecem negociações de empresas de capital aberto (suas ações podem ser negociadas sem que haja a necessidade de uma escrituração de propriedade), cujos compradores podem ser pessoas físicas, bancos e até mesmo outras empresas, que visam ter alguma participação nos lucros de determinada empresa ou nos ganhos com o seu crescimento. É importante ressaltar que a rentabilidade do mercado de ações é variável, uma vez que não há uma previsibilidade dos lucros e, normalmente, as negociações de compra e venda de ações ocorrem em uma bolsa de valores (BUSTOS; POMARES-QUIMBAYA, 2020).

Os conceitos de mercado de ações e bolsa de valores não são novos, visto que já na época das grandes navegações começaram a surgir empresas de capital aberto. A primeira bolsa em que ocorriam negociações de empréstimos foi criada em 1531 na Bélgica, e foi denominada de Bolsa da Antuérpia. Porém, a primeira ação comercializada registrada pertenceu a Companhia das Índias Orientais, criada em 1599, com mais ou menos 3 mil

acionistas com ações no valor de 3 milhões de libras, sendo negociada na Bolsa de Amsterdã somente em 1602 (KISHTAINY, 2012). Em busca de proteger sua economia, em 1571 a Inglaterra criou a Bolsa de Londres, e em 1680 corretores (pessoas que realizam transações, operações de compra e venda na bolsa de valores) de Londres se reuniram na *Jonathan's Coffee House* para efetuarem suas negociações. Todavia, somente no ano de 1792, mais precisamente no dia 17 de maio, é que surgiu a Bolsa de Nova York (do inglês *New York Stock Exchange* - NYSE), cujas transações aconteciam no *Tontine Coffee House*, no distrito de Manhattan. Em 1817, devido ao crescimento da bolsa, esta passou a localizar-se na *Wall Street*, nome que, atualmente, também designa o mercado financeiro dos Estados Unidos da América (EUA) (PETRAM, 2014; GEISST, 2004).

Com o passar dos anos, o mercado financeiro americano vivenciou seus altos e baixos e também seus avanços tecnológicos, como por exemplo, a adoção de telégrafos no ano de 1867 para registrar as oscilações diárias nos preços das ações. Depois vieram os telefones no ano de 1878 e, mais recentemente, a chegada dos computadores. Em 1929 ocorreu a *Black Friday*, em que ocorreu um grande volume de vendas de ações, ocasionando à época a quebra da Bolsa de Nova York e conduzindo os EUA para uma severa recessão econômica com impactos que perduraram até 1941. Já no século XXI, a NYSE avançou para se tornar a maior bolsa do planeta, com a compra de outras bolsas espalhadas pelo globo, como a *American Stock Exchange* e a *Archipelago Exchange*. Atualmente, grandes empresas fazem parte do mercado americano, como a Coca-Cola®, Google®, Meta® e Apple®, representando uma expressiva relevância financeira (NATIONS, 2017; RAPP, 2015).

Considerando determinado mercado financeiro, existem índices que servem para mensurar o seu desempenho, os quais podem ajudar na tomada de decisões e no entendimento sobre a situação do mercado. Entre estes índices encontra-se o S&P500 (do inglês *Standard & Poor's 500*). Criado em 1957, rastreia as ações tanto na NYSE quanto na NASDAQ, outra bolsa americana bastante voltada às empresas de tecnologia e *internet*. Este índice inclui diversas companhias das mais variadas áreas para representar a economia americana e é considerado um dos mais importantes, sendo um referencial sobre a situação econômica dos EUA (PARNES, 2019). De modo geral, avalia o desempenho das 500 maiores empresas listadas nas bolsas de valores dos EUA (NYSE e NASDAQ), sendo considerado um dos melhores indicadores do desempenho do mercado de ações dos EUA e, por conseguinte, da economia americana, devido à sua abrangência e diversidade. As empresas no S&P500 são selecionadas com base em seu tamanho de mercado, liquidez e setor da indústria, refletindo assim um amplo espectro da economia dos EUA.

O índice é ponderado pelo valor de mercado, o que significa que empresas com maior capitalização têm maior peso no índice (JAIN, 1987). Ou seja, seu valor é calculado com base nos preços das ações das 500 empresas que o compõem, ajustado pelo número total de ações disponíveis para negociação (capitalização de mercado flutuante). Uma mudança no preço das ações de uma empresa com maior peso no índice terá um impacto maior no valor do índice do que uma mudança no preço das ações de uma empresa com menor peso. É, portanto, uma referência sendo frequentemente utilizado como um indicador para a saúde geral do mercado de ações e como referência para o desempenho de carteiras de ações e fundos de investimento (SCHNITZLER, 2018).

De modo geral, assuntos políticos, aspectos sociais e psicológicos, acontecimentos globais e políticas governamentais, desempenham um papel crucial e representam fatores que podem influenciar os movimentos do mercado de ações. Dessa forma, identificar esses fatores, assim como estabelecer suas relações com os investimentos no mercado, representa tarefa desafiadora, enfatizada pelo fato de que muitos investidores estão propensos a realizar suas negociações utilizando aspectos de sua vida, tais como a intuição, a condição financeira e as experiências pessoais. Assim, a associação do mercado financeiro à inteligência

computacional visando melhorar a previsibilidade do mercado é uma área promissora, visto que a partir de um modelo de previsão, é possível também aos usuários seguir determinadas recomendações de investimento, com a expectativa de obtenção de algum lucro (THAKKAR; CHAUDHARI, 2021).

2.1.1 AS ESTRATÉGIAS DE OPERAÇÃO NO MERCADO FINANCEIRO

No mercado financeiro, existem diversas estratégias de operação, cada uma adequada a diferentes perfis de investidores e horizontes de investimento. Entre as estratégias constam o *day trade*, o *swing trade*, o *position trade* e o *buy and hold* (MURPHY, 1999).

O *day trade* é uma estratégia que envolve a compra e venda de ativos financeiros dentro do mesmo dia, buscando lucrar com pequenas oscilações de preço. É uma técnica que exige muito tempo, atenção e conhecimento do mercado, além de ser considerada de alto risco. O *swing trade*, por outro lado, tem um horizonte de tempo um pouco mais longo, geralmente de alguns dias ou semanas. Os investidores que adotam essa estratégia procuram aproveitar tendências de curto e médio prazo nos preços dos ativos, sem a necessidade de acompanhamento constante do mercado como no *day trade*.

Já o *position trade* é uma estratégia de prazo mais longo, onde as posições são mantidas por semanas, meses ou até anos. Essa abordagem se concentra mais na análise fundamentalista, avaliando os fundamentos econômicos das empresas ou setores em que se investe, e é menos influenciada pelas flutuações de curto prazo do mercado. Por fim, a estratégia de *buy and hold* é mais passiva, consistindo em comprar ativos e mantê-los por um longo período, geralmente anos. O foco reside no potencial de crescimento a longo prazo dos ativos. Os investidores que adotam essa estratégia geralmente não se preocupam com as oscilações de curto e médio prazo, confiando na valorização dos ativos ao longo do tempo.

Anexo às estratégias, a análise de tendência desempenha um papel relevante na avaliação de padrões de movimento dos preços dos ativos, o que pode incluir a análise técnica que se concentra em padrões gráficos e indicadores. Adicionalmente, existe a análise fundamentalista que avalia os fatores econômicos e financeiros que influenciam o valor de um ativo. Essa compreensão ajuda os investidores a tomarem decisões mais assertivas sobre quando comprar, vender ou manter ativos, maximizando potenciais lucros e minimizando riscos. Independentemente da estratégia adotada, a análise de tendências fornece *insights* cruciais que podem definir o sucesso ou o fracasso nas operações de mercado. Portanto, a integração efetiva entre a escolha da estratégia operacional e a análise de tendências é um aspecto chave para alcançar resultados positivos no mercado financeiro (EIAMKANITCHAT; MOONTUY; RAMINGWONG, 2016).

2.1.2 ANÁLISE DE TENDÊNCIA EM MERCADOS DE AÇÕES

A análise de tendência em mercados de ações é constituída por técnicas e abordagens utilizadas por investidores e analistas financeiros para identificar padrões e tendências no preço e no volume de negociação de ações. A ideia vinculada à análise de tendência é que os movimentos de preço e volume ao longo do tempo possam ser utilizados para prever o comportamento futuro (PRING, 2021).

Segundo Murphy (2020), para realizar a análise de tendência, os analistas não medem o valor intrínseco, mas utilizam gráficos de preço e outras ferramentas técnicas, como linhas de tendência, médias móveis e osciladores. Tais ferramentas podem ser usadas individualmente ou em combinação para ajudar na identificação de oportunidades de negociação e na tomada de decisões sobre compras e vendas de ações.

Existem diferentes tipos de tendência que podem ser identificadas através da análise técnica, sendo elas: tendência de alta (do inglês *bullish*), tendência de baixa (do inglês *bearish*) e tendência lateral (do inglês *sideway*). Ao identificar essas tendências, os analistas podem tomar decisões sobre quando comprar, vender ou manter ações com base nas expectativas futuras de preço (KAHN, 2018).

Neste contexto, a Hipótese de Mercado Eficientes (do inglês *Efficient Market Hypothesis* - EMH) é uma teoria que sugere que os mercados financeiros são eficientes na incorporação de todas as informações disponíveis em seus preços. De acordo com a EMH, é impossível para os investidores alcançarem retornos consistentes acima da média do mercado, pois todas as informações já estão refletidas nos preços das ações, ou seja, é inútil tentar prever os movimentos futuros dos preços das ações ou encontrar oportunidades de investimento subvalorizadas. A EMH é dividida em três formas: fraca, semiforte e forte. A forma fraca da EMH afirma que os preços e os retornos passados das ações não são úteis para prever os preços ou os retornos futuros das ações. Já a forma semi-forte da EMH afirma que todas as informações disponíveis publicamente já estão refletidas nos preços das ações. Por fim, a forma forte da EMH afirma que todas as informações, incluindo informações confidenciais, já estão refletidas nos preços das ações (LIMA, 2020).

Nesse sentido, a EMH tem sido alvo de muitos debates e críticas, com alguns argumentos de que ela não leva em conta o papel dos participantes do mercado na determinação dos preços, além de alguns apontamentos para casos de comportamento irracional do mercado que parecem contrariar a teoria. Apesar das críticas, a EMH permanece uma teoria amplamente aceita na área financeira e é usada como base para muitas estratégias de investimento, como a de fundos de índice.

Entretanto, a hipótese de mercado eficiente é um conceito teórico e nem sempre é possível observá-lo na prática. Os vieses e fatores comportamentais dos participantes do mercado, bem como a manipulação de mercado e o *insider trading*, podem de fato causar desvios dos preços (ROGERS, 2021). Por exemplo, autores como Carosia *et al.* (2021) acreditam que a análise de sentimento a partir de mídias sociais e de notícias pode fornecer opiniões valiosas que não estão totalmente integradas nos preços das ações. Com base nos resultados, os autores concluíram que a análise de sentimentos consegue ser uma ferramenta importante no auxílio à tomada de decisões a respeito de investimentos.

2.2. ANÁLISE DE SENTIMENTO

A Análise de Sentimento (do inglês *Sentiment Analysis* - SA), também denominada por Mineração de Opinião, é uma área de estudo que analisa opiniões, atitudes e emoções das pessoas sobre os mais variados assuntos expressos por meio de textos. O primeiro estudo de SA surgiu possivelmente em 2003, realizado por Nasukawa e Yi (2003), e a mineração de opinião teve início possivelmente pelo trabalho dos autores Dave *et al.* (2003). No entanto, há relatos de trabalhos que se iniciaram na década de 90, como análise de subjetividade, ponto de vista, extração de descrições sentimentais e cálculo de afetividade, mas somente no século 21 é que houve um maior desenvolvimento dos mesmos e uma normalização dos termos para SA (LIU, 2015).

Com o século 21 também ocorreu o surgimento e crescimento das mídias sociais, resultando em um aumento expressivo na quantidade de dados. Tais dados são entendidos como ativos valiosos quando o intuito é entender as opiniões e comportamentos dos indivíduos. Segundo Melo e Meinerz (2019), a SA permite a geração de informações adicionais a partir desses dados, sendo entendida como uma técnica de mineração de dados, especialmente dados não estruturados, sendo necessária a aplicação de técnicas capazes de

lidar com a linguagem natural para se obter o entendimento e podendo ser aplicada à diversas redes sociais como o Meta[®] e o X[®] (antigo Twitter[®]).

Em geral, textos retirados de diferentes fontes, como redes sociais, blogs e sites de notícias são classificados em 3 categorias: positivos, negativos ou neutros. Esta classificação representa algo básico em SA, podendo ser simplesmente categorizada como uma polarização. Estudos como o de Araújo, Gonçalves e Benvenuto (2013) demonstram estruturas bem mais complexas na avaliação do sentimento de determinado texto. Todavia, para este trabalho, o foco reside nas 3 categorias básicas.

Ademais, esta classificação também se aplica à opiniões relacionadas aos preços dos mercados de ações, podendo indicar a eventuais relações entre, por exemplo, as notícias a respeito de uma empresa e as oscilações em seus preços no mercado, tornando passível o uso da SA na incorporação de informações importantes com o intuito de realizar previsões no mercado de ações (ARAÚJO; FERNANDES, 2021).

Além disso, o artigo do autor Hussein (2018) aborda sobre os desafios enfrentados pela análise de sentimentos, destacando a crescente geração de conteúdos opinativos na Internet como conteúdo valioso de informações para empresas, mas que no entanto passa por desafios em relação a detecção da polaridade de sentimentos e compreensão do significado preciso das expressões. Ele enfatiza que a análise de sentimentos não se limita a identificar expressões explícitas, mas também deve levar em conta as implicações implícitas e o contexto geral dos textos. Ao final, o artigo propõe direções para pesquisas futuras, sugerindo a expansão do escopo de comparação para incluir novas pesquisas e técnicas continuamente.

O Ortigosa *et al.* (2014) explora a análise de sentimentos em mensagens do Meta[®] (antigo Facebook[®]) e sua aplicação no e-learning. Utilizando um método híbrido que combina técnicas de léxico e aprendizado de máquina, o estudo desenvolveu a aplicação SentBuk, que classifica mensagens de acordo com a polaridade emocional. Essa ferramenta é usada para entender o estado emocional dos alunos, auxiliando sistemas de aprendizado adaptativos a oferecer recomendações personalizadas e fornecendo feedback valioso para professores em ambientes de aprendizagem online. A SentBuk respeita a privacidade e as políticas de dados do Meta[®], permitindo que os usuários visualizem análises dos seus próprios sentimentos e dos amigos. O estudo demonstra a eficácia da aplicação na análise de sentimentos e destaca sua importância para o contexto educacional.

2.3. APRENDIZADO PROFUNDO

A Inteligência Artificial (do inglês *Artificial Intelligence* - AI) pode ser definida como a criação de modelos computacionais que objetivam reproduzir aspectos do intelecto humano, tais como, a capacidade de raciocinar e tomar decisões na resolução de problemas (RUSSEL; NORVIG, 2016). Dessa forma, a AI tem sido amplamente utilizada em diversas áreas, como por exemplo, a financeira, sendo que uma parcela de seu sucesso é atribuída ao Aprendizado Profundo (do inglês *Deep Learning* - DL) e às Redes Neurais Artificiais (do inglês *Artificial Neural Network* - ANN) (LUDERMIR, 2021).

Uma ANN é um modelo inspirado no sistema biológico de aprendizado desempenhado pelo cérebro humano, a fim de realizar uma tarefa específica permitindo a tomada de decisão com base no seu próprio aprendizado considerando determinada tarefa. Baseia-se em uma rede de células computacionais simples, também chamadas de neurônios ou unidades de processamento (HAYKIN, 2001). Podem ser utilizadas para resolver problemas considerados complexos, para os quais o comportamento das variáveis não é conhecido. Sendo assim, as ANNs conseguem aprender através de exemplos e, dessa forma, armazenar o que foi aprendido para posterior aplicação, gerando um modelo não-linear, com potencial para análises voltadas ao mercado de ações (SPÖRL *et. al.*, 2011).

De maneira geral, o DL tem permitido avanços significativos em várias tarefas realizadas pelos seres humanos. Por exemplo, no trabalho de Zheng *et al.* (2023) os autores trabalharam na análise postural de seres humanos utilizando DL, sendo esta decisão fundamental graças à capacidade do DL de processar e analisar grandes conjuntos de dados complexos. O DL tem sido fundamental também em áreas como a análise de movimentos no esporte, o monitoramento de pacientes na área da saúde e na interface entre humanos e computadores.

Já Menghani (2023), abordou a otimização de modelos de DL para que estes sejam mais eficientes, particularmente em termos de uso de recursos computacionais e energia. Isso é crucial para a implementação em dispositivos com recursos limitados, como *smartphones* e outros dispositivos de Internet das Coisas (do inglês *Internet of Things* - IoT). A eficiência de modelos de DL não apenas expande sua aplicabilidade em vários dispositivos e plataformas, mas também abre caminho para avanços em áreas como processamento de linguagem natural, visão computacional e análise preditiva em tempo real.

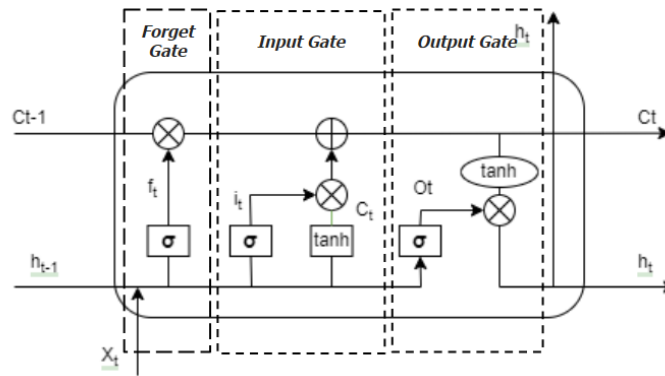
No contexto de DL constam diversas arquiteturas de ANN, identificadas como Redes Neurais Profundas (do inglês *Deep Neural Networks* - DNN), cada uma com características únicas e adequadas para diferentes tipos de tarefas. Por exemplo, as Redes Neurais Convolucionais (do inglês *Convolutional Neural Network* - CNN) são amplamente utilizadas em visão computacional, adequadas para processar imagens e vídeos devido à sua habilidade em capturar padrões espaciais (YIN *et al.* 2017). Já as Redes Neurais Recorrentes (do inglês *Recurrent Neural Network* - RNN) são projetadas para lidar com dados sequenciais, como textos ou séries temporais, sendo eficazes na modelagem de dependências temporais (ALZUBAIDI *et al.*, 2021). Existem algumas arquiteturas de RNNs, sendo uma delas apresentada na próxima seção.

2.3.1 LSTM

Redes neurais podem apresentar problemas ao usar informações passadas na realização de determinada tarefa. Todavia, em alguns casos é necessário apenas usar informações mais recentes para desempenhar um trabalho e, nesta circunstância, em que a distância entre a informação importante a ser aprendida e o momento temporal em que a tarefa executada é pequena, uma rede neural pode aprender a usar a informação passada. Porém, quando esta distância é grande, pode haver um comprometimento no aprendizado, ou seja, quando as informações importantes são muito antigas em relação ao momento da tarefa, a rede pode ter problemas em capturar e aplicar dados antigos no processo de aprendizado. Esse desafio é particularmente notável em tarefas que exigem a compreensão de dependências de longo prazo, onde o aprendizado e a memória da rede são cruciais (STAUEMEYER; MORRIS, 2019).

A partir disso, surge a arquitetura de Memória de Longo e Curto Prazo (do inglês *Long Short-Term Memory* - LSTM), sendo esta um caso particular de RNNs que não apresenta problemas em relação ao aprendizado à medida que o espaço entre a informação relevante e a execução da tarefa aumenta. Para isso, apresentam uma estrutura mais complexa, característica comum em relação às ANNs de DL quando comparadas a arquiteturas mais antigas. Enquanto arquiteturas convencionais apresentam apenas uma única camada de rede neural, a LSTM poderá apresentar várias camadas interagindo de maneira a capturar dependências de longo prazo nos dados (YU *et al.*, 2019). Esta arquitetura é particularmente útil na previsão de séries temporais, pois podem aprender e reter informações ao longo do tempo, tornando-as ideais para analisar e prever padrões em dados financeiros, sequências de linguagem natural e outras possibilidades. A Figura 1 apresenta a estrutura padrão de uma LSTM.

Figura 1 - Estrutura padrão de uma arquitetura LSTM



Fonte: Adaptado de Yu *et al.* (2019)

A linha horizontal que está na parte superior do diagrama é identificada como o **estado da célula**, percorrendo toda a cadeia com poucas interações, permitindo que a informação transite livremente sem que ocorra qualquer alteração na mesma. A LSTM tem a capacidade de remover ou adicionar informações ao **estado da célula** que será regulado por estruturas chamadas de “porta”, que geralmente são 3, *input gate*, *output gate* e *forget gate*, sendo utilizadas para proteger e controlar o **estado da célula**.

Na sequência existe uma camada chamada de sigmóide, também identificada como **porta de esquecimento**. Esta tem como função decidir quais informações poderão passar através de uma sequência de números 0 ou 1. No primeiro caso indica que nenhuma informação deverá passar e, no segundo caso, toda e qualquer informação poderá passar. Posteriormente, existe uma etapa de decisão sobre qual informação será armazenada, sendo constituída de duas partes: a primeira é a **porta de entrada**, na qual são decididos quais valores serão atualizados e, a segunda, é uma camada *tanh*, em que é criado um vetor de novos possíveis valores que poderão ser adicionados ao **estado da célula**.

Seguindo, existe uma combinação dessas duas partes para que haja a atualização, e para tal, é necessário multiplicar o resultado da **porta de entrada** com o resultado da **porta de esquecimento**, a fim de descartar informações antigas de tal forma que se possa adicionar valores candidatos à atualização. Por último, é necessário que haja um filtro para a obtenção do resultado que será executado através de uma camada sigmóide que decidirá qual parte será produzida, h_t ou *tanh*, para a saída O_t . A partir disso, a informação será adicionada no **estado de célula** através de uma camada *tanh* (determinando que os valores estejam entre -1 e 1), para adicionar apenas as partes que foram aprendidas anteriormente.

Este é um modelo básico de LSTM, mas podem haver variações tais como “conexões tipo olho mágico”, possibilitando que os portões olhem o **estado da célula**, ou um acoplamento das portas de entrada e esquecimento, em que haverá uma tomada de decisão em conjunto, esquecendo uma informação somente quando outra é atualizada em seu lugar e adicionando uma nova informação somente quando outra é esquecida (SMAGULOVA; JAMES, 2019).

2.3.2 BERT

Nos últimos anos, as pesquisas em AI têm evoluído muito, principalmente a partir do desenvolvimento de novas arquiteturas de redes neurais, mais especificamente arquiteturas bidirecionais a partir de transformadores (do inglês *transformers*). Treinamentos com a

característica bidirecional podem ser aplicados a modelos de linguagem habilitando um melhor entendimento do contexto e fluxo da linguagem quando comparado aos modelos tradicionais, ou seja, de sentido único.

Uma boa representação desse tipo de modelo é o BERT (*Bidirectional Encoder Representation from Transformers*), proposto por pesquisadores da Google AI® em 2018 (DEVLIN *et al.*, 2019), se utiliza de estruturas que aprendem a relação contextual entre as palavras de um texto. Uma característica importante desta arquitetura é sua composição baseada em dois mecanismos separados: um codificador (do inglês *encoder*) e um decodificador (do inglês *decoder*). O primeiro irá ler o texto, já o último irá produzir uma previsão para a tarefa que deve ser realizada (XU *et al.*, 2019).

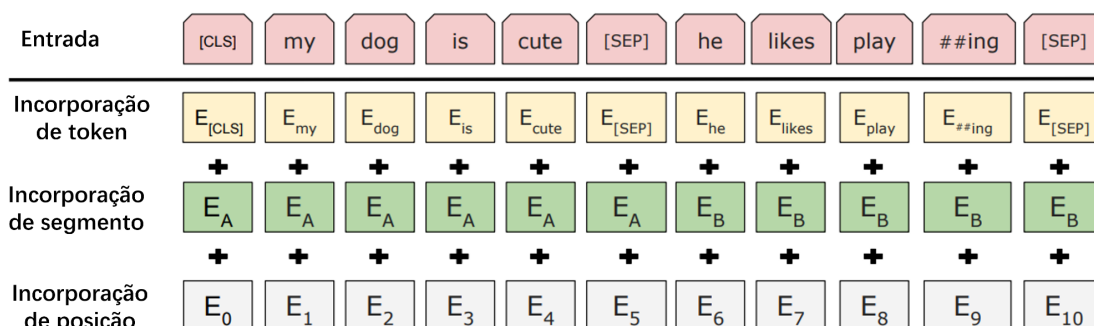
De modo geral, o codificador *transformer* irá ler uma sequência inteira de palavras de uma vez só, possibilitando que o modelo aprenda o contexto de uma palavra com base apenas nas palavras que estão ao seu redor. Entretanto, assim como outros modelos de linguagem, o BERT tem como desafio realizar determinadas previsões em determinada língua, e para superar isso, se utiliza de duas estratégias (XU *et al.*, 2019).

A primeira estratégia é o *Masked LM* (MLM), que consiste em ter uma porcentagem das palavras mascaradas por um símbolo, para que o modelo possa prever qual o valor original da palavra apenas com base no contexto das palavras que estão em sequência e não estão mascaradas. O modelo técnico é composto pela adição de uma camada de classificação sobre o codificador, seguido de uma multiplicação dos vetores de saída através de uma matriz de incorporação, originando a um vocabulário. Por último, irá realizar o cálculo da probabilidade de cada palavra desse vocabulário por meio da função *softmax*.

Já a segunda estratégia é o *Next Sentence Prediction* (NSP), em que o modelo irá receber pares de sentenças como entrada de modo a prever se a segunda sentença do par é de fato a sentença que viria logo em seguida a primeira sentença no documento original. Assim, durante o treinamento, metade das entradas são as sentenças subsequentes que estão no documento original e a outra metade são sentenças aleatórias para treinar o modelo que entende as relações entre as frases.

Para que haja uma distinção entre as sentenças durante o processo de treinamento, o processamento foi realizado de modo que um *token* CLS esteja disposto no início da primeira frase e um *token* SEP seja inserido no final de cada uma das frases (Figura 2). Em cada *token* é introduzida uma frase que indica se fosse a frase A ou frase B, assim como há uma incorporação de posição (do inglês *position embeddings*) a cada *token*, a fim de indicar sua posição na sequência.

Figura 2 - Representação do modelo BERT para determinada sentença de entrada



Fonte: Adaptado de Devlin *et al.* (2019)

Ao final, para saber se há uma conexão entre as frases é necessário que todas as sequências de entrada passem por um *transformer* e a saída de um *token* CLS deverá ser

transformada em um vetor 2x1, usando uma simples camada de classificação. Assim, calcula-se a probabilidade com a função *softmax* (DEVLIN *et al.*, 2019).

O modelo BERT serviu, desde sua criação, como base para vários modelos de linguagem. Entre eles, consta o modelo RoBERTa proposto pelo Yinhan Liu *et al.* (2019). RoBERTa possui, essencialmente, um comportamento parecido com o BERT. Todavia, possui como diferença principal a forma como o processo de mascaramento é executado. Enquanto o BERT utiliza uma abordagem estática, RoBERTa, por outro lado, se utiliza de uma abordagem dinâmica, ou seja, diferentes fragmentos de uma sentença são mascarados em cada época de treinamento do modelo. Esta característica, além de promover vantagem quanto ao tempo de treinamento, permite com que o modelo seja utilizado de maneira mais eficiente em diferentes cenários, por exemplo, na análise de sentimentos.

3. TRABALHOS CORRELATOS

Através de buscas na literatura científica, foram selecionados trabalhos visando identificar métodos e abordagens de análise de sentimento e de predição de mercado de ações. A pesquisa foi realizada nas bases de artigos acadêmicos ScienceDirect[®], Scopus[®], Web of Science[®], Springer Link[®] e IEEEExplore[®]. Foram considerados artigos científicos em língua inglesa publicados entre 2018 e 2023 com a seguinte *string* de busca: (S&P500 OR S&P OR "S&P 500" OR "Standard & Poor's 500") AND ("sentiment analysis") AND (Forecasting OR Prediction). A Tabela 1 apresenta os resultados obtidos na busca.

A inclusão do termo "S&P500" e suas variações na *string* de busca foi considerada devido à sua relevância no mercado financeiro, visto que representa um amplo espectro da condição econômica e das tendências do mercado de ações nos Estados Unidos. Portanto, ao incluir "S&P500" na busca, promoveu-se uma revisão mais direcionada a estudos relevantes e significativos que abordam a análise de sentimentos e previsões de mercado financeiro no contexto de um importante índice de mercado. Ao utilizar tal estratégia, a pesquisa garantiu uma concentração de estudos relacionados a um índice representativo no mercado de ações, ao invés de dispersar o foco em múltiplos mercados financeiros de diferentes países ou em ações de empresas específicas.

Tabela 1 - Resultado da revisão de literatura considerando a busca inicial

Base de Artigos Acadêmicos	Quantidade de Artigos Resultantes
ScienceDirect [®]	3
Scopus [®]	27
Web of Science [®]	16
Springer Link [®]	163
IEEEExplore [®]	57
Total de resultados	266

Fonte: Elaborado pelo autor (2023)

Face a quantidade de artigos, foram definidos alguns critérios para se chegar aos trabalhos mais relacionados, entre eles, artigos escritos em inglês, na categoria "Article", e na área de "Computer Science", em um intervalo de tempo de 2018 a 2023 (últimos 5 (cinco) anos completos), resultando inicialmente em um total de 266 artigos. Após isso, foi realizada uma análise do título dos artigos recuperados em todas as bases. Cabe mencionar que, no caso específico da base Springer Link[®], verificou-se que após os 20 primeiros artigos, a

relevância diminuía significativamente, assim como a presença das palavras-chave utilizadas na *string* de busca. Sendo assim, após este filtro inicial pelo título e integração de artigos repetidos, restaram 39 trabalhos para a leitura completa. Durante a leitura, cada artigo recebeu uma nota visando identificar os 10 artigos mais correlatos/relevantes para o presente trabalho que são discutidos a seguir.

Segundo Carosia *et. al* (2021), dentro das inúmeras aplicações de análise de sentimentos, o mercado financeiro é o que vem recebendo mais atenção, sendo que, a utilização de notícias torna-se útil para prever as movimentações do mercado de ações, uma vez que textos financeiros influenciam as decisões de investidores. A SA serve para analisar os textos e classificá-los em positivos e negativos, podendo ser utilizada conjuntamente com o Aprendizado de Máquinas. Dessa forma, os autores propuseram uma estratégia de investimento baseado em SA de notícias financeiras e sua aplicação no mercado financeiro brasileiro. Para tal, realizaram uma comparação de diferentes ANNs de forma a determinar a melhor arquitetura para o mercado brasileiro, além do teste de causalidade de Granger, para verificar a relação entre o mercado financeiro e os sentimentos das notícias. Como resultado, foi possível comprovar que utilizar a SA impacta na elevação de lucros quando comparada simplesmente a outras estratégias como *buy & hold*, em cenários em que havia algum investimento mensal e investimentos acumulados por mais de 6 meses.

No trabalho de Bozanta *et. al* (2021) avalia diferentes modelos de aprendizado de máquina para análise de sentimentos em *tweets* financeiros da plataforma StockTwits®. Foram empregados alguns modelos, tais como modelos tradicionais de classificação como regressão logística (logística regressiva-LR) e floresta aleatória, e modelos de DL como LSTM e GRU, assim como modelos *transformers* como BERT, DistillBERT, RoBERTa e XLNet. Para alcançar seus objetivos, foram usados 100.000 mensagens de 5 empresas/índices, sendo Apple®, Amazon®, Boeing®, Walt Disney® e SPDR S&P 500 ETF *Trust*, obtendo os seguintes resultados: os modelos pré-treinados baseados em *transformers* foram os que precisaram de mais tempo para serem treinados, tais como RoBERTa, BERT e XLNet. Ademais, para avaliar o desempenho dos modelos empregados foram usados parâmetros de acurácia, precisão, e um *score* métrico, sendo que o modelo *transformer* RoBERTa superou os demais na classificação de sentimentos como "*bearish*" ou "*bullish*". Portanto, os autores concluem que os modelos pré-treinados baseados em *transformers* são melhores em relação aos modelos tradicionais para classificar mensagens. Ademais, deixam um adendo sobre a necessidade de se utilizar um conjunto de dados maior, bem como a necessidade de incluir outras ações e índices, coletar dados de outras redes sociais, além de aumentar o período de coleta.

Já o trabalho realizado por Chatziloilos, Gunopulos e Konstantinou (2021) enfatiza que as pesquisas envolvendo mercado de ações vêm aumentando recentemente, principalmente àquelas aplicadas à área de tecnologia como investigar a previsão de tendências no mercado de ações utilizando aprendizado profundo com dados financeiros e textuais. Dessa forma, os autores tiveram como objetivo a utilização das técnicas de SA e análise técnica para verificar a possibilidade de prever ou não prever de maneira eficiente o mercado de ações. Foram utilizados 3 modelos de DL e ML, sendo, LSTM, KNN e Árvore de decisão, juntamente com dados numéricos e dados textuais obtidos a partir da rede social X® (antigo Twitter®). Por fim, os autores concluem que o melhor método de aprendizagem foi o LSTM quando utilizado com dados numéricos, viabilizando a sua utilização no cenário de bolsa de valores, gerando lucros significativamente maiores do que a estratégia tradicional de Buy & Hold.

Kumar *et. al* (2021) propuseram um modelo híbrido de DL, utilizando o LSTM e o algoritmo de colônia de abelhas artificiais (ABC), que foi escolhido devido a sua maior facilidade de implementação e maior capacidade em lembrar padrões de comportamento. É

visto também como um bom algoritmo de busca local, contribuindo para a solução de otimização de problemas, sendo aplicados na bolsa NASDAQ®, para prever as ações de empresas como Apple®, Microsoft® e Intel®, utilizando um período de 10 anos. Como resultados, os autores relatam que a utilização do LSTM otimizado com o ABC promoveu uma melhora na precisão das previsões do mercado, tendo margens de erro aceitáveis. Entretanto, os autores concluem que a volatilidade do mercado não depende apenas dos preços antigos de suas ações, mas também depende da situação governamental do país, da economia mundial, da presença de políticas públicas e privadas, entre outras.

Outro estudo realizado por Ji *et. al* (2023) também disserta sobre a importância das mídias sociais na previsão dos mercados de ações. Dessa forma, os autores tiveram como objetivo implementar um modelo de Memória de Longo e Curto Prazo baseada em Atenção (do inglês *Attention-based Long-Short Term Memory* - ALSTM) para prever o mercado de ações, juntamente com o modelo BERT para realizar a classificação de sentimento, considerando 28 ações em um período de 3 anos. A partir disso, foram obtidos os seguintes resultados: no quesito de classificação de sentimento, o modelo BERT obteve um bom desempenho (85,9% de acurácia). Já para a previsão de ações, o ALSTM foi o que apresentou o melhor desempenho para as 28 ações quando comparado a outros modelos. Ao final, os autores concluem que a utilização combinada de diversas fontes, melhora a acurácia das previsões, todavia, também é necessário levar em consideração a situação financeira da empresa e notícias disponíveis em jornais.

Consequente, na pesquisa de Swathi, Kasiviswanath e Rao (2022), os autores afirmam que opiniões negativas e positivas influenciam nos preços das ações de mercado e, com base nisso, utilizaram dados do X® para prever o mercado por meio da arquitetura LSTM juntamente com o algoritmo Otimização baseada em ensino e aprendizagem (do inglês *Teaching and Learning based Optimization* - TLBO). Ademais, o otimizador Adam foi utilizado para verificar o nível de aprendizado. Com isso, os autores finalizam enfatizando que a proposta realiza previsões utilizando-se de 4 etapas: pré-processamento, classificação baseada em LSTM, verificação de aprendizado baseado no otimizador Adam e otimização baseada na TLBO. Tanto a TLBO quanto o Adam, contribuíram positivamente para a eficiência das previsões do LSTM, atingindo um desempenho maior quando comparado a outras técnicas (acurácia de 94,73%).

No trabalho de Wang *et. al* (2023), os autores atestam que a maior parte dos estudos referentes a previsões do mercado de ações se baseia em métodos de classificação, desconsiderando inúmeros fatores que podem afetar diferentes fases dos modelos propostos. Portanto, estes propõem um modelo híbrido de aprendizado, unindo ANNs com medidores técnicos de mídias sociais e SA. Com o modelo, os autores concluíram ser possível realizar previsões de ações utilizando os métodos corretos, e que o modelo híbrido construído por eles apresentou uma boa perspectiva, uma vez que atingiu uma acurácia de 73,41%, demonstrando a validade da incorporação de indicadores técnicos e da SA.

Outro trabalho é o realizado por Gao *et.al* (2022), que reforça a importância das notícias para realizar previsões acerca do mercado de ações. Os autores propuseram um novo modelo de previsão de Índice de Mercado de Ações (do inglês *Stock Market Index* - SMI), utilizando uma combinação de diferentes fontes de notícias para analisar o sentimento dessas, ou seja, construíram o modelo de Raciocínio Evidencial por meio de Rede Neural Recorrente e Algoritmo Genético (do inglês *Evidential Reasoning-Recurrent Neural Network-Genetic Algorithm* - ER-RNN-GA), sendo que a RNN objetiva criar uma base de classificação de notícias, o ER para integrar a capacidade de previsão das diferentes fontes e o GA para otimizar a base de classificação, melhorando os resultados. Sendo assim, os pesquisadores concluíram que esse modelo é superior aos métodos convencionais, confirmando sua validade para o mercado de ações.

Kavya, Kumar e Kusuma (2022) afirmam que a inteligência artificial deve ser utilizada no mercado de ações. Desta forma, os autores utilizaram um sistema de suporte inovador para plataformas interativas, empregando a Análise de Sentimento Baseada em Aspecto (ABSA) para avaliar o estado mental dos usuários através de suas mensagens de texto. Este sistema analisa a polaridade dos sentimentos, com foco especial na identificação e alteração do uso de palavras ofensivas em conversas. A técnica ABSA é utilizada para categorizar sentimentos em positivos, neutros e negativos, considerando o contexto e a complexidade do sarcasmo e do humor. A eficácia do sistema é avaliada usando PyScript, se destacando pela sua capacidade de extrair sentimentos subjacentes para melhorar a comunicação e manter relacionamentos saudáveis. Como resultado foram obtidos os seguintes dados: o algoritmo LinearSVG teve a melhor precisão (90%). O artigo também propõe futuras melhorias, como a inclusão de análise de voz, otimização de código e arquitetura baseada em nuvem, visando aprimorar a precisão e a eficiência do sistema.

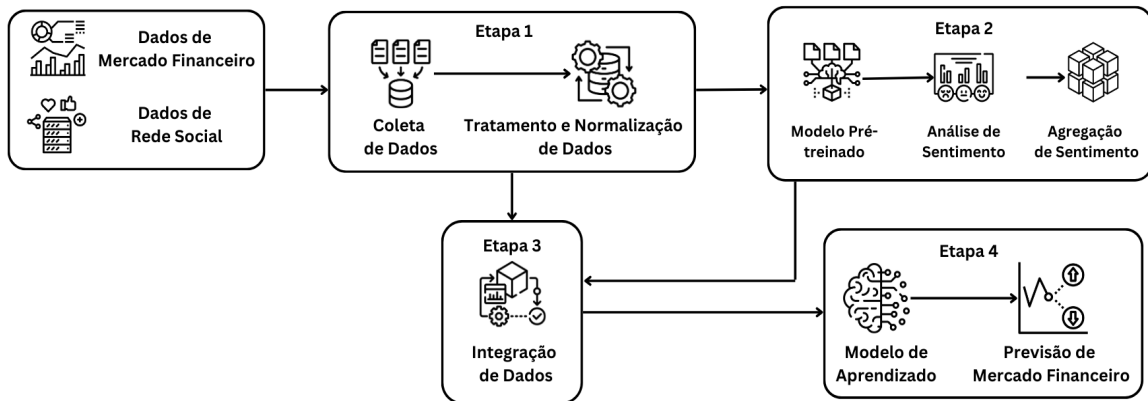
Por fim, Kinyua *et. al* (2021) fizeram um estudo sobre o impacto imediato de tweets postados por Donald Trump no período de 2017 a 2019, no índice S&P500, DJIA e NYSE, utilizando SA e a VADER (do inglês *Valence Aware Dictionary for Sentiment Reasoning*), voltada à análise de textos de mídias sociais e aprendizado de máquinas, por meio de algoritmos de regressão, como a regressão linear, a Regressão por Árvore de Decisão (do inglês *Decision Tree Regression - DTR*) e a Regressão por Florestas Randômicas (do inglês *Random Forest Regression - RFR*). Sendo assim, o estudo concluiu que para S&P500 e DJIA, postagens realizadas pelo presidente nos horários em que o mercado de ações estava funcionando impactavam negativamente os índices.

Os artigos apresentados nesta seção discutem o uso de análise de sentimentos, classificação de textos e aprendizado de máquinas para prever o mercado de ações. A maioria dos autores concorda que esses métodos podem ser eficazes para prever o movimento dos preços das ações, mas que também são necessários estudos adicionais para melhorar a precisão das previsões. Percebe-se ainda que em alguns trabalhos ocorreu a integração de dados de sentimento com dados de movimentação de determinada ação. Todavia, esta integração não fica clara. No presente trabalho, conforme será detalhado nas próximas seções, os sentimentos de mensagem de redes sociais são agregados por dia e então integrados ao preço de fechamento. Outro ponto que pode ser destacado é a configuração nos hiperparâmetros utilizados, assim como a estruturação da arquitetura da rede LSTM, representando um diferencial em relação aos trabalhos correlatos identificados por clarificar como a arquitetura se organiza.

4. MÉTODO PROPOSTO

Nesta seção é detalhado o método proposto voltado à previsão de movimentos no mercado de ações através do uso da Análise de Sentimentos em redes sociais e Redes Neurais. Para tal, quatro etapas principais foram constituídas, sendo: a) coleta e tratamento de dados de mercado de ações; b) análise de sentimento de mensagens de redes sociais por meio de modelos de linguagem pré-treinados; c) integração dos dados de mercado financeiro e o sentimento das mensagens; e d) treinamento do modelo de aprendizado por meio de uma ANN para a realização de previsão de preço de ações (Figura 3). Em seguida, o detalhamento de cada etapa é apresentado, possibilitando o entendimento como estas se interconectam para atender o objetivo deste trabalho, ou seja, a previsão de mercado de ações.

Figura 3 - Detalhamento das etapas do método proposto



Fonte: Elaborado pelo autor (2023)

4.1 ETAPA 1: Coleta e Tratamento de Dados

Nesta primeira etapa, a coleta de dados deve ser realizada a partir de fontes de dados que permitam subsidiar a previsão do mercado de ações se utilizando da análise de sentimento. Sendo assim, mensagens de redes sociais e dados de movimentações no mercado de ações são insumos fundamentais para o método proposto. Para o primeiro caso, deve-se utilizar alguma ferramenta de coleta e raspagem de dados capaz de capturar postagens relevantes baseadas em palavras-chave específicas, assim como, outros filtros, como intervalos de dados, garantindo a relevância do conteúdo para determinado estudo. Paralelamente, os dados financeiros, incluindo preços de ações, volumes de negociação e índices de mercado, também devem ser obtidos. Tais dados são essenciais para estabelecer uma correlação entre as percepções do mercado e as flutuações financeiras. Uma vez coletados, os dados devem passar por um processo de tratamento e limpeza, onde ruídos, dados duplicados e irrelevantes são removidos ou modificados. A normalização é então aplicada para garantir a uniformidade dos dados, o que permite comparações válidas e a aplicação eficaz de técnicas analíticas nas etapas subsequentes do método.

4.2 ETAPA 2: Análise de Sentimentos através de um Modelo Pré-treinado e Agregação de Dados

A segunda etapa do método proposto envolve a utilização de um modelo de linguagem pré-treinado voltado à análise de sentimento em conteúdo textual; no contexto do trabalho representado por mensagens de redes sociais. Este modelo de NLP, extensivamente treinado em vastos conjuntos de dados de texto, é habilmente utilizado para determinar a porcentagem (limitada a 1 (um)) de sentimento positivo, neutro e negativo em cada *tweet*. Após a classificação dos sentimentos, é realizado um cálculo de integração dos tipos de sentimento (positivo, negativo e neutro) para cada data, somando-se as porcentagens de mensagens positivas, neutras e negativas e dividindo pelo número total de mensagens da mesma data, resultando numa média diária para cada tipo de sentimento. Este passo é crucial para padronizar os resultados, permitindo uma análise temporal coerente. Vale mencionar que os dados são armazenados em uma estrutura de dados visando promover suporte à próxima etapa.

4.3 ETAPA 3: Integração de Dados

Na terceira etapa do método proposto, procede-se com a filtragem e integração dos conjuntos de dados. Inicialmente, são retirados os registros correspondentes a dias não úteis, como feriados e fins de semana, uma vez que estes períodos não apresentam atividade no mercado financeiro e, portanto, podem distorcer a análise de correlação entre os sentimentos expressos nos *tweets* e as variações do mercado. Após essa filtragem, a integração dos dados de sentimentos agregados por dia com os dados financeiros obtidos do Yahoo Finance® é realizada. Este processo deve ser executado de maneira cuidadosa visando garantir que os dados de sentimento agregados estejam sincronizados com o correspondente dado financeiro por determinada data.

Segue-se então à normalização dos dados integrados, que é essencial para neutralizar qualquer discrepância de escala entre as variáveis e facilitar o processamento por modelos de ML. A normalização é também uma etapa crítica para permitir que o modelo de aprendizagem alcance a sua máxima eficiência, pois dados em uma escala comum reduzem o risco de enviesamento em direção à variáveis com maior magnitude. Com os dados devidamente preparados e normalizados, estes estão prontos para serem utilizados como entrada para o modelo de aprendizado de máquina na próxima etapa do método.

4.4 ETAPA 4: Previsão de Mercado Financeiro pelo Modelo de Aprendizado

Na última etapa, ocorre a construção da arquitetura de ANN a ser utilizada na predição do mercado de ações. Para este trabalho, uma arquitetura do tipo recorrente é utilizada e representa o núcleo do modelo preditivo. Esta arquitetura foi configurada para processar e aprender as sequências temporais dos dados de mercado, levando em conta não apenas os valores históricos de preços, mas também os dados resultados da análise de sentimentos integrados na etapa anterior.

A partir disso, utilizando o conjunto de dados, a rede neural pode ser treinada utilizando duas abordagens visando a comparação, sendo, na primeira, utilizando exclusivamente dados de mercado de ações para estabelecer uma linha de base e, na segunda, os dados de sentimento foram incorporados, com vistas a quantificar o impacto da análise de sentimentos nas previsões de tendências de mercado. Esta abordagem dualista possibilita uma comparação direta do desempenho preditivo com e sem o auxílio do contexto emocional capturado a partir de mensagens de redes sociais. Ademais, a construção da rede neural envolve a correta definição e otimização de hiperparâmetros. Por fim, esta deve ser submetida a um processo de treinamento, validação cruzada e teste para assegurar sua capacidade de generalização, assim como sua robustez frente a novos dados evitando, por exemplo, problemas de sobreajuste na aprendizagem (do inglês *overfitting*).

5. RESULTADOS EXPERIMENTAIS

5.1 CENÁRIO DE ESTUDO

Este estudo é situado em um contexto em que dados de mercado financeiro, mais especificamente de mercado de ações, e dados resultantes da análise de sentimentos sobre postagens a respeito de determinado índice de bolsa de valores se interconectam. Desta forma, o objetivo reside em examinar de maneira detalhada a dinâmica da emoção em conteúdos postados por usuários declarando seus sentimentos e como utilizar tais dados de maneira que possam ser incorporados na previsão de tendência no mercado financeiro. Para

tal, dados do YahooFinance[®] e X[®] (antigo Twitter[®]) foram utilizados e coletados por meio das bibliotecas *yfinance* e *snsrape* disponíveis em Python.

No primeiro caso, os dados do YahooFinance[®] foram obtidos considerando o período entre 1 de julho de 2022 a 30 de março de 2023 sem considerar finais de semana e feriados, resultando em 188 registros. Como resultado, 7 (sete) colunas, *Date*, *Open*, *High*, *Low*, *Close*, *Adj Close* e *Volume* foram coletadas e armazenadas em uma estrutura de dados.

Para o segundo caso, os dados do X[®] foram obtidos com o mesmo período dos dados do YahooFinance[®], resultando em 275,800 *tweets* (como uma mensagem na plataforma X[®] é chamada) sendo então armazenados em uma estrutura de dados serializada em Python chamada *pickle*.

5.2 MATERIAIS E MÉTODOS

Para a implementação e execução do método proposto, o equipamento utilizado foi um processador Intel[®] Core™ i7 da 11ª Geração i7-11800H com a frequência de TDP 2.30 GHz de 8 núcleos e 16 GB de memória RAM de 3200 MHz. Além disso, devido a utilização do PLM RoBERTa para a análise de sentimento, é requerida a instalação da biblioteca PyTorch baseada em Torch. Esta pode ser operada em uma GPU NVIDIA compatível com Arquitetura de Dispositivo de Computação Unificada (do inglês *Compute Unified Device Architecture* - CUDA) para otimizar e acelerar o processamento. Neste trabalho, foi utilizada uma placa de vídeo de NVIDIA[®] GeForce[®] RTX 3060, que oferece 6 GB de memória gráfica GDDR6 com um barramento de memória de 192 bits e 3840 núcleos CUDA.

Para a implementação e instanciação do método proposto a linguagem de programação Python[®] foi considerada, assim como diversas bibliotecas também implementadas nesta linguagem. Entre estas bibliotecas, destacam-se: *yfinance*, *pandas*, *snsrape*, *pipeline*, *emoji*, *sklearn*, *keras*. A *yfinance* fornece uma maneira fácil de se obter os dados históricos de mercado de ações diretamente do Yahoo Finance[®]. O pacote *pandas* permite a manipulação e análise de dados de alto desempenho; de fácil utilização, *pandas* é essencial para trabalhar com dados tabulares oferecendo estruturas de dados como *DataFrame* e *Series*, e funcionalidades para rápida edição e visualização de dados. A *snsrape* permite raspar (do inglês *scrape*) dados de várias redes sociais, incluindo X[®], sem a necessidade de usar a API das plataformas. Embora a biblioteca *pipeline* possa se referir a vários pacotes, no contexto de aprendizado de máquina com *sklearn*, um *pipeline* é uma sequência de transformações de dados seguida por um processo de modelagem. Facilita a construção e a gestão de fluxos de processamento de dados e modelagem preditiva. A biblioteca *emoji* foi utilizada para trabalhar com *emojis* em texto. Foi empregada para tratar ou remover *emojis* dos dados textuais coletados do X[®], o que é uma passo fundamental no pré-processamento de dados em análise de sentimentos.

A biblioteca *scikit-learn* é uma das mais utilizadas no contexto de ML. Oferece uma gama de algoritmos para tarefas de classificação, regressão e clusterização, assim como para a redução de dimensionalidade, além de ferramentas para ajuste de modelo utilizadas neste trabalho, pré-processamento de dados, seleção de variáveis e avaliação de modelos. A biblioteca *keras* permite construir redes neurais de alto nível que executam sobre o TensorFlow[®]. É conhecida por sua facilidade de uso e permite a construção e treinamento de modelos de DL, sendo utilizada para construir e treinar uma rede neural do tipo LSTM.

Para a avaliação do método foram utilizadas as métricas de Erro Médio Absoluto (MAE) (Equação 1), Raiz Quadrada do Erro Médio (RMSE) (Equação 2) e Coeficiente de Determinação (R^2) (Equação 3), como segue:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (1)$$

onde y_i é o valor de predição, x_i é o valor real, n é o número total dos dados.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

onde Y_i é o valor observado, \hat{Y}_i é o valor previsto, n é o número total dos dados.

Equação 1 e 2 são fórmulas para calcular o valor de erro ou diferença entre os valores previstos e valores reais em estatística. Já R^2 permite avaliar a qualidade de um modelo de regressão visando explicar a variação dos dados observados.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3)$$

$$SS_{res} = \sum_{i=1}^n (y_i - f_i)^2 \quad (4)$$

$$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \text{ indica a média dos valores observados}$$

onde SS_{res} é a soma residual dos quadrados, sendo os resíduos as diferenças entre os valores observados e os valores estimados pelo modelo. Já SS_{tot} é o quadrado da soma das diferenças entre os valores de observação e os valores médios. R^2 é então entendido como um coeficiente de determinação variando entre 0 e 1. Quando o R^2 está próximo de 1, isso significa que o modelo tem um bom desempenho, pois as previsões são muito semelhantes aos dados reais. Caso contrário, se o R^2 está próximo de 0, as previsões são pouco precisas em relação aos dados reais. Em outras palavras, um R^2 alto indica que o modelo pode explicar uma grande parte da variação dos dados, tornando-o mais confiável para previsões.

5.3 PRÉ-PROCESSAMENTO

A implementação do trabalho foi desenvolvida pela linguagem de programação Python, utilizada frequentemente em áreas como a Inteligência Artificial, o Aprendizado Profundo e a Análise de Dados. Além disso, diversas bibliotecas e modelos de linguagem que servem de suporte estão disponíveis (conforme mencionado na seção anterior), tornando o desenvolvimento de aplicações mais simplificado.

O processo começa com a obtenção dos dados contidos nas postagens feitas nas redes sociais em que o usuário tenha interesse em analisar. Para tanto, foi utilizada a biblioteca Python *snsrape* permitindo a coleta, no contexto deste trabalho, de *tweets* sem nenhum limite ou restrição a partir da rede social X[®]. Neste caso específico, os dados foram coletados definindo palavras-chave relacionadas às ações específicas, setores ou empresas. Para este trabalho a palavra-chave utilizada foi S&P500 com intervalo de tempo entre 01/07/2022 e

30/03/2023, sendo encontrados um total de 275,800 *tweets*. O Quadro 1 apresenta exemplos de *tweets* coletados.

Quadro 1 - Dados de *tweets* do período de 01/07/2022 até 30/03/2023

Índice	Data	Tweet
0	2022-07-01 00:04:25+00:00	NC State gets worked at Clemson #breaking #breakingnews...
1	2022-07-01 00:05:01+00:00	Ancelotti: "An unforgettable year is coming to...
...		
275798	2023-03-30 23:56:56+00:00	Top stocks with TA score trending DOWN (SP500)..
275799	2023-03-30 23:59:43+00:00	USA index futures tame. We churn and churn\n\n...

Fonte: Elaborado pelo autor (2023)

Sendo assim, os dados coletados são salvos em formato *pickle* do Python, sendo esta uma maneira popular de armazenar e recuperar objetos de dados. Basicamente, uma hierarquia de objetos Python é convertida em uma sequência de *bytes*, a qual pode ser gravada em um arquivo ou transferida por uma rede e, posteriormente, reconstruída em uma cópia idêntica da hierarquia original de objetos. Uma das funcionalidades mais amplamente utilizadas objetiva salvar modelos de ML após o treinamento ser concluído evitando que o mesmo seja retreinado frequentemente.

Após a coleta dos dados, é necessário pré-processá-los para remover ruídos e informações irrelevantes. Esta etapa envolve a limpeza, como remoção de caracteres especiais, *hiperlinks* e menções de usuários, substituição de abreviações e falhas ortográficas, padronização para letras minúsculas e entre outras (Quadro 2).

Quadro 2 - Os textos de *tweets* após de retirar os caracteres especiais

Tweet Original	Tweet Limpo
NC State gets worked at Clemson #breaking #breakingnews...	nc state gets worked at clemson
Ancelotti: "An unforgettable year is coming to...	ancelotti an unforgettable year is coming to a...
...	
Top stocks with TA score trending DOWN (SP500)..	top stocks with ta score trending down sp500
USA index futures tame. We churn and churn\n\n...	usa index futures tame we churn and churn spx

Fonte: Elaborado pelo autor (2023)

Seguindo, são necessários alguns processamentos adicionais. Na análise de sentimento, considerando *tweets*, a remoção de palavras *stopwords* é uma etapa importante para filtrar palavras comuns que têm pouco ou nenhum significado contextual ou sentimento.

Stopwords, geralmente, são artigos, preposições, pronomes e outras palavras frequentemente encontradas em um texto. Ao eliminar tais palavras, concentra-se nos termos mais relevantes e que expressam, no caso de um *tweet*, um sentimento. Esse procedimento foi realizado por meio da biblioteca NLTK (*Natural Language Toolkit*) em Python, que fornece listas predefinidas de *stopwords* (*stoplist*) para diversos idiomas como “a”, “an”, “the”, “by”, “on”, entre outras. Sendo assim, cada palavra do *tweet* é comparada com a lista e, caso exista correspondência, esta é excluída. O Quadro 3 apresenta exemplos de *tweets* antes e após o processamento de retirada de *stopwords*.

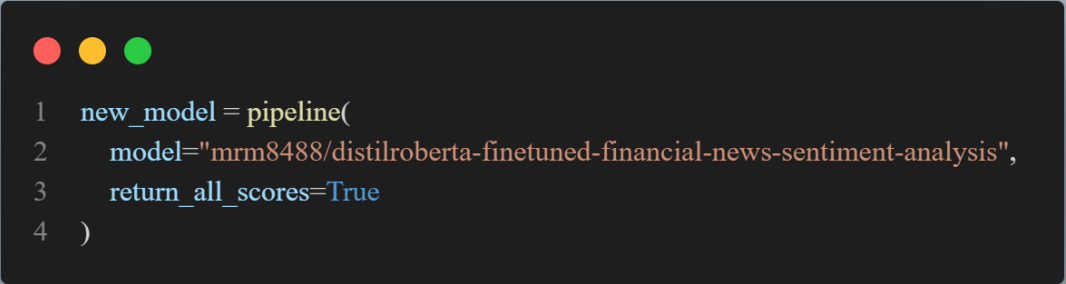
Quadro 3 - Exemplos de *tweets* após a retirada de *stopwords* dos textos

<i>Tweet Limpo</i>	<i>Tweet Processado</i>
nc state gets worked at clemson	nc state gets worked clemson
ancelotti an unforgettable year is coming to a...	ancelotti unforgettable year coming end style
...	...
top stocks with ta score trending down sp500	top stocks score trending down sp500
usa index futures tame we churn and churn spx	usa index futures tame churn churn spx

Fonte: Elaborado pelo autor (2023)

Após a etapa de pré-processamento dos dados, que deve ser criteriosa, sendo realizadas limpezas e normalizações, é então utilizado determinado PLM para executar a análise de sentimentos. Neste trabalho, conforme mencionado anteriormente, o PLM RoBERTa (ver seção 2.3.3) foi utilizado, mais especificamente, uma simplificação deste modelo chamado DistilRoBERTa desenvolvido pelo *Facebook AI Research* (FAIR) como uma melhoria e uma variante do BERT. Assim como o BERT, RoBERTa também é um modelo baseado em *transformers* projetado para entender a complexidade da linguagem humana, assim como o contexto no texto. Esse modelo, conhecido por sua robustez e precisão, foi fundamental para interpretar com eficácia as nuances emocionais contidas nos *tweets* coletados para este trabalho. A Figura 4 apresenta a instanciação do modelo DistilRoBERTa treinado especificamente para a análise de sentimento.

Figura 4 - Modelo Pré-treinado DistilRoBERTa para análise de sentimentos



```

1 new_model = pipeline(
2     model="mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis",
3     return_all_scores=True
4 )

```

Fonte: Elaborado pelo autor (2023)

O PLM utilizado permite calcular o *score* de cada *tweet* retornando a porcentagem de determinado sentimento, ou seja, classificado como positivo, neutro e negativo, presente no conteúdo analisado. Com isso, na sequência, para cada data (dia), somam-se os percentuais de sentimentos positivos, neutros e negativos dos *tweets* analisados dividindo pelo total de *tweets* do respectivo dia. Esta normalização, permite refletir a prevalência de cada sentimento de forma proporcional dia a dia. Ou seja, permite sintetizar as informações para se obter uma perspectiva consolidada e quantitativa dos sentimentos distribuídos ao longo do período de estudo. A Figura 5 apresenta a função que possibilita este caso estabelecendo a média diária por tipo de sentimento.

Figura 5 - Função de cálculo de porcentagem média de sentimentos por dia

```
1 def Sentiment_Analysis(data):
2     sum_neg = 0
3     sum_neu = 0
4     sum_pos = 0
5     df = pd.DataFrame(columns=['Date','Negative', 'Neutral', 'Positive'])
6     for tweet in data['Text_Cleaned']:
7         sum_neg += new_model(tweet)[0][0]['score']
8         sum_neu += new_model(tweet)[0][1]['score']
9         sum_pos += new_model(tweet)[0][2]['score']
10    return sum_neg/len(data), sum_neu/len(data), sum_pos/len(data)
```

Fonte: Elaborado pelo autor (2023)

Com os valores médios dos sentimentos e os dados financeiros adquiridos da plataforma Yahoo Finance®, realizou-se a etapa final de consolidação concatenando os dados por dia. Todavia, antes de serem utilizados para o treinamento do modelo de aprendizagem, deve-se definir as datas como índices dos dados, bem como retirar feriados e dias de final de semana. Isto assegura uma correspondência temporal precisa entre os valores de sentimentos e os movimentos do mercado de ações (Quadro 4).

Quadro 4 - Dados normalizados para o treinamento do modelo de aprendizado

Data	Negativo	Neutro	Positivo	Fechamento
01-07-2022	0.320868	0.486771	0.192361	3825.330078
05-07-2022	0.303997	0.495866	0.200137	3831.389893
...				
2023-03-29	0.142104	0.520804	0.337092	4027.810059
2023-03-30	0.159957	0.511287	0.328755	4050.830078

Fonte: Elaborado pelo autor (2023)

A próxima fase envolveu a normalização dos dados, um processo-chave para garantir que a escala dos preços de fechamento do mercado (do inglês *close*) e os percentuais dos sentimentos se mantivessem consistentes, facilitando a comparação e correta análise. A escolha do preço de fechamento é justificada por este representar o preço bruto, ou seja, o valor monetário representando o último preço negociado em um título antes do mercado encerrar oficialmente para as negociações normais. Representa o julgamento comum do mercado sobre o preço e o consenso psicológico dos participantes no mercado, ou seja, é o resultado de decisões e interações entre compradores e vendedores.

5.4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Uma vez integrado e normalizado os dados, é requerida a construção do modelo de aprendizado, ou da arquitetura de rede neural a ser utilizada. Para o presente trabalho, conforme mencionado na Seção 5.2), uma rede LSTM foi desenvolvida utilizando TensorFlow® e *keras*, composta por cinco camadas sequenciais, cada uma contendo 50 unidades de neurônios. A escolha desta configuração visa capturar a complexidade e a temporalidade dos dados através de múltiplas representações hierárquicas. Para mitigar o risco de sobreajuste da memória da rede (*overfitting*), uma preocupação comum em modelos de DL com múltiplos parâmetros. Neste sentido, é necessário aplicar uma taxa de abandono (do inglês *dropout*) de 0.2 em cada camada LSTM. Esse procedimento desativa aleatoriamente um conjunto de neurônios durante o treinamento, forçando a rede neural a aprender padrões mais robustos que não dependem de um único caminho.

Ademais, as camadas retornam as sequências completas de estados ocultos para cada intervalo de tempo em forma de tensores 3D, o que é crucial para garantir a passagem da sequência completa de saídas para a próxima camada, permitindo que a rede mantenha o estado ao longo do tempo. A última camada da arquitetura é uma camada densa, que consolida as informações extraídas pelas camadas anteriores para realizar a predição final. Essa camada densa atua como um decodificador das características aprendidas pelas camadas LSTM e projeta a saída para o espaço desejado, ou seja, a predição do próximo valor na série temporal financeira sendo analisado (Figura 6).

Figura 6 - Estrutura da rede neural LSTM implementada neste trabalho

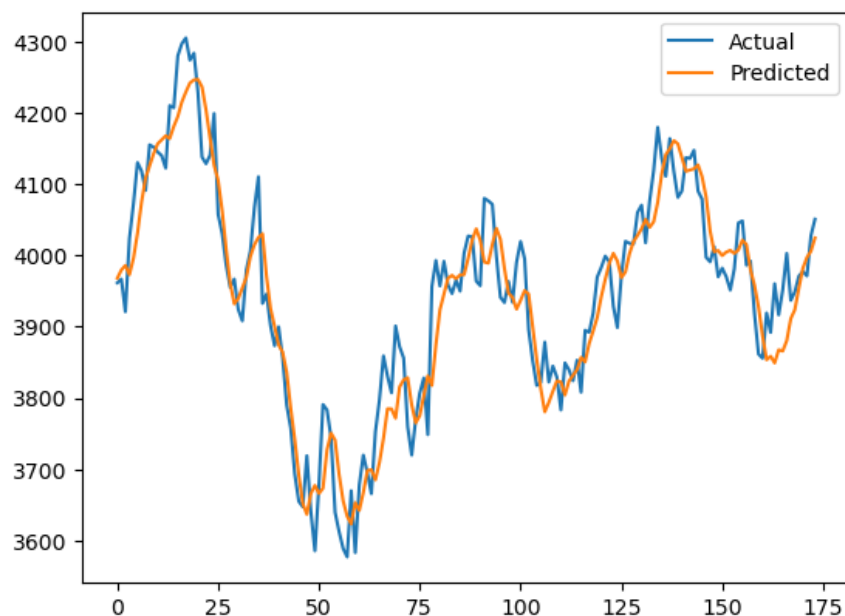
```
1 from keras.models import Sequential
2 from keras.layers import Dense, LSTM, Dropout
3
4 # LSTM model
5 lookback=14
6 model = Sequential()
7 model.add(LSTM(50, return_sequences=True, input_shape=(X_train.shape[1], X_train.shape[2])))
8 model.add(Dropout(0.2))
9
10 model.add(LSTM(units=50, return_sequences=True))
11 model.add(Dropout(0.2))
12
13 model.add(LSTM(units=50, return_sequences=True))
14 model.add(Dropout(0.2))
15
16 model.add(LSTM(units=50))
17 model.add(Dropout(0.2))
18
19 model.add(Dense(y_train.shape[1]))
20 model.compile(loss='mean_squared_error', optimizer='adam')
```

Fonte: Elaborado pelo autor (2023)

Com o intuito de analisar os resultados produzidos pelo modelo de predição, três experimentos foram elaborados, envolvendo tanto os dados somente de bolsa de valores quanto integrados com a análise de sentimentos.

No primeiro experimento, foram utilizados os dados combinados, ou seja, o preço de fechamento e o resultado da análise de sentimentos. Assim, o foco reside na avaliação das diferenças entre as predições da rede LSTM em relação ao preço de fechamento. A Figura 7 ilustra o desempenho do modelo quando integrado com a análise de sentimentos a partir dos *tweets*. Observa-se uma proximidade entre as curvas dos valores reais do mercado (azul) e as previsões do modelo (laranja). Importante ressaltar que este primeiro teste foi executado sobre todo o conjunto de dados (188 dias) objetivando apresentar globalmente o potencial de predição da arquitetura LSTM implementada.

Figura 7 - Previsão do S&P500 com análise de sentimentos (todo o período)

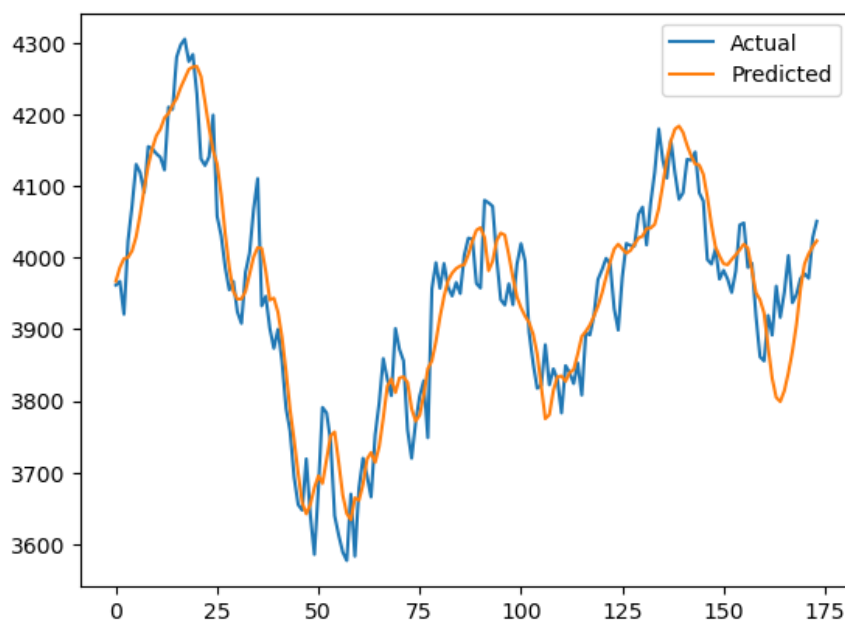


Fonte: Elaborado pelo autor (2023)

O resultado é então analisado através de algumas métricas, conforme apresentado na Seção 5.2. Para o MAE foi obtido um valor de 0.2901 e para o RMSE o valor de 0.3526 foi obtido, sendo considerados valores adequados em tarefas de correlação/previsão de dados. Já o R^2 atingiu o valor 0.8812, indicando um alto grau de explicação da variância dos preços pelo modelo.

Em contraste, a Figura 8 representa o modelo LSTM utilizando apenas os dados de preço de fechamento, sem a análise de sentimentos. A correspondência entre os valores previstos e os reais é ligeiramente menos precisa, o que é refletido por um MAE um pouco maior de 0.3047, um RMSE de 0.3784 e um R^2 de 0.8632.

Figura 8 - Previsão do S&P500 sem análise de sentimentos (todo o período)

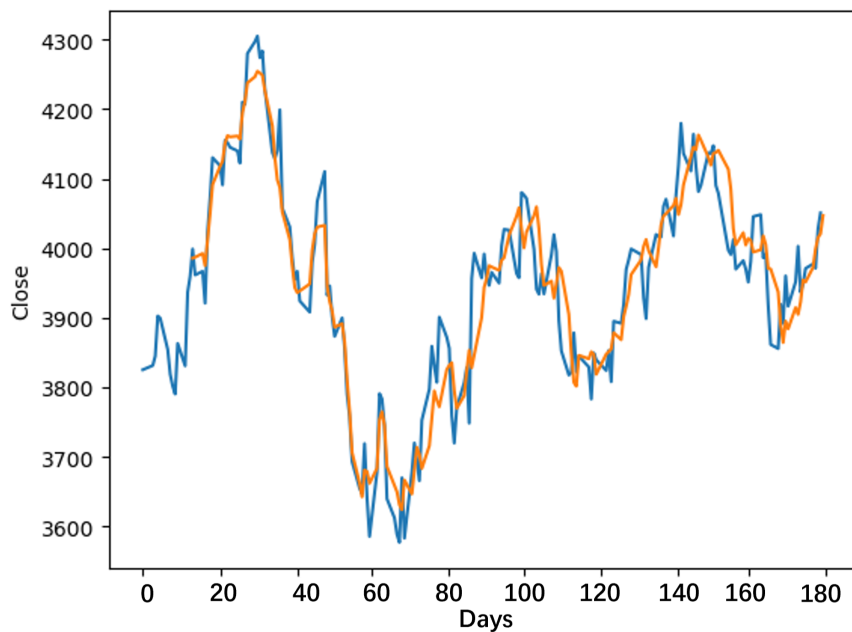


Fonte: Elaborado pelo autor (2023)

No segundo experimento com os resultados apresentados nas Figuras 9 e 10, são comparados o desempenho do modelo LSTM quando os dados de sentimentos são incluídos versus quando apenas os dados de fechamento são utilizados com deslocamento de 1 (um) dia de previsão. Especificamente, o modelo analisa dados dos 174 dias anteriores para fazer suas previsões. A análise dos erros reforça a precisão incrementada pelo modelo que integra sentimentos atingindo um MAE de 41.5139 com sentimentos contra um MAE ligeiramente superior de 44.1186 sem a análise de sentimentos. Além disso, o RMSE foi consideravelmente menor no modelo com dados de sentimentos, valor de 51.6107, em comparação com um RMSE de 53.8987 para o modelo sem dados de sentimentos. Isso sugere que os sentimentos capturados a partir de *tweets* contribuem para a modelagem da volatilidade do mercado. Os valores de MAE e RMSE mais elevados, quando comparados ao primeiro experimento, podem ser atribuídos à maior complexidade e variabilidade inerentes à análise de uma janela de tempo olhando para o passado de 174 dias.

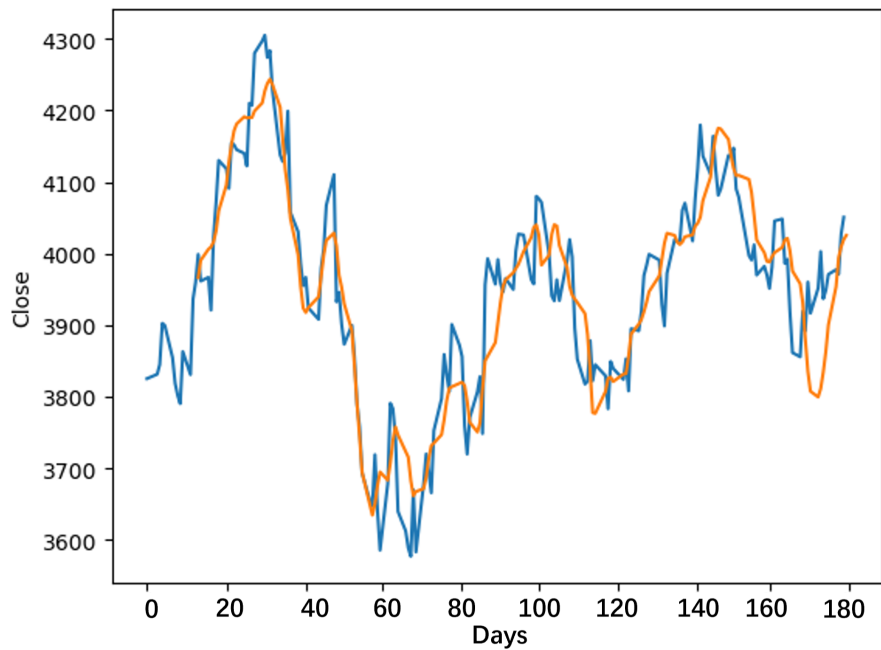
Por fim, o R^2 para o modelo com dados de sentimentos foi de 0.8893, indicando que uma grande proporção da variância dos preços de fechamento é explicada pelo modelo. Em contraste, o modelo sem dados sentimentais teve um R^2 de 0.8615, indicando uma capacidade preditiva um pouco menor.

Figura 9 - Previsão do S&P500 com análise de sentimentos (previsão de 1 dia)



Fonte: Elaborada pelo autor (2023)

Figura 10 - Previsão do S&P500 sem análise de sentimentos (previsão de 1 dia)



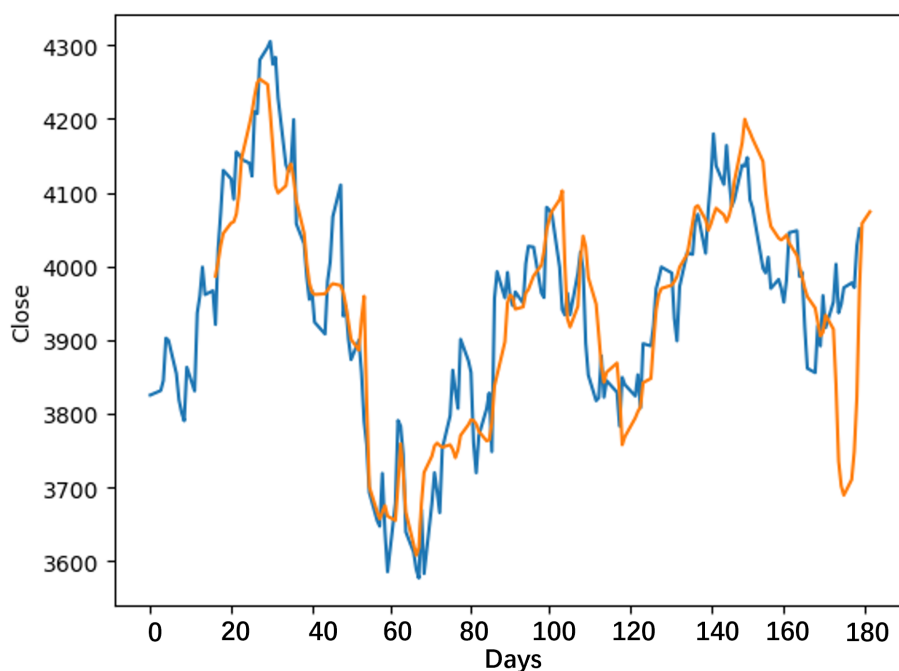
Fonte: Elaborado pelo autor (2023)

Por fim, no terceiro experimento, as previsões foram realizadas com um horizonte temporal um pouco mais extenso, cobrindo cinco dias de negociações futuras, o que promove desafios para modelos de predição. A comparação das previsões geradas pelo modelo LSTM, com e sem a incorporação de dados sentimentais, revela uma dinâmica interessante. As Figuras 11 e 12 correspondentes aos resultados com cinco dias de antecedência ilustram que, embora o modelo enriquecido com dados sentimentais mantenha uma fidelidade razoável com os valores reais de fechamento do mercado, a distância entre as previsões e os valores reais tende a aumentar com o prolongamento do período de previsão.

Esta tendência é quantificada pelos indicadores de erro. Com a inclusão dos sentimentos é observado um MAE de 53.3168, enquanto o modelo sem dados de sentimentos apresenta um MAE um pouco maior, de 55.8878. Por sua vez, o RMSE apresenta um aumento substancial em ambos os modelos, sendo 71.3912 com a análise de sentimentos e 78.5504 sem a análise de sentimentos. Isso sugere que, apesar de os dados sentimentais fornecerem uma melhoria nas previsões, a incerteza aumenta com previsões de longo prazo.

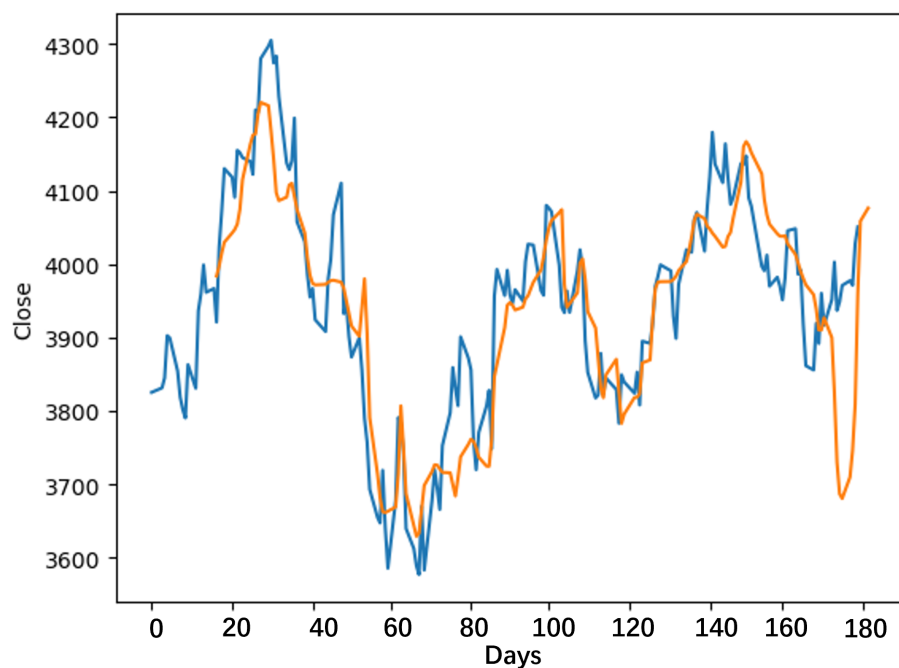
Adicionalmente, o R^2 para o modelo de cinco dias com dados de sentimento é de 0.7906, enquanto o modelo sem dados sentimentais possui um R^2 de 0.7465. Estes resultados indicam que, embora o modelo que inclui análise de sentimentos tenha uma capacidade preditiva superior, ainda há uma margem considerável de variação não explicada, o que é esperado em previsões de médio prazo, principalmente em um ambiente volátil como o mercado financeiro.

Figura 11 - Previsão do S&P500 com análise de sentimentos (previsão de 5 dias)



Fonte: Elaborado pelo autor (2023)

Figura 12 - Previsão do S&P500 sem análise de sentimentos (previsão de 5 dias)



Fonte: Elaborado pelo autor (2023)

Em síntese, os resultados experimentais (Tabela 2) reforçam a relevância da inclusão de dados sentimentais para aprimorar a precisão das previsões do modelo LSTM, mas também destacam os desafios inerentes à previsão de movimentos de mercado em prazos mais longos. Isto pode ser percebido, principalmente, no terceiro cenário em que os valores de erros aumentam consideravelmente. Estes resultados deixam claro que o método implementado e, principalmente, o modelo de previsão, podem atingir um desempenho

adequado no curto prazo, permitindo ganhos em operações em bolsa de valores. Todavia, para períodos maiores, estudos adicionais e, possivelmente, a integração de outras técnicas e modelos possam promover resultados mais adequados. Isto se evidencia com o aumento nos valores de MAE e RMSE para previsões de 5 (cinco) dias em relação a 1 (um) dia. Em determinados cenários um R^2 abaixo de 0,8 até pode ser considerado adequado, mas em um mercado tão volátil como o financeiro, os resultados obtidos no terceiro cenário podem ser altamente negativos.

Tabela 2 - Valores de MAE, RMSE e R^2 para os três cenários

Experimento	MAE	RMSE	R^2
Previsão geral com sentimento	0,2901	0,3526	0,8812
Previsão geral sem sentimento	0,3047	0,3784	0,8632
Previsão de 1 dia com sentimento	41,5139	51,6107	0,8893
Previsão de 1 dia sem sentimento	44,1186	53,8987	0,8615
Previsão de 5 dias com sentimento	53,3168	71,3912	0,7906
Previsão de 5 dias sem sentimento	55,8878	78,5504	0,7465

Fonte: Elaborado pelo autor (2023)

O potencial do método proposto em auxiliar na tomada de decisão de investidores pode ser ilustrado com os resultados obtidos em diferentes cenários de previsão. A partir dos resultados de previsão de 1 (um) dia, o modelo que incluiu a análise de sentimentos obteve um MAE de 41.5130 e um RMSE de 51,6107. Em contraste, o modelo que não incluiu análise de sentimentos teve um MAE um pouco maior de 44.1186 e um RMSE de 53,8987.

Tais resultados indicam que a inclusão de análise de sentimentos melhora a capacidade do modelo em prever movimentos do mercado financeiro, proporcionando aos investidores previsões mais precisas e, conseqüentemente, aumentando as chances de decisões de investimento mais assertivas. Por exemplo, para previsões de um dia o modelo com análise de sentimentos previu um valor de fechamento para o dia 31 de março de 2023 (próximo dia útil considerando o conjunto de *tweets* coletados) de 4,021.60, enquanto o modelo sem análise de sentimentos previu 4,018.99. Comparando com o valor real do dia 31 de março de 2023, que foi de 4,109.31 (conforme dados do Yahoo Finance[®]), o modelo com análise de sentimentos se aproximou mais do valor real.

Adicionalmente, o coeficiente de determinação (R^2) para o modelo com dados de sentimentos foi de 0.8893, indicando que uma parte considerável da variância dos preços de fechamento é explicada pelo modelo. Em comparação, o modelo sem dados de sentimentos teve um R^2 de 0.8614, sugerindo uma capacidade preditiva inferior.

Por fim, os resultados de 5 dias ainda demonstram uma capacidade preditiva superior quando considerada a agregação de sentimentos, embora exista uma variação maior dos preços não explicada pelo modelo. O aumento significativo nos valores de MAE e RMSE para previsões de 5 dias em comparação com previsões de um dia, sugere que a incerteza aumenta com horizontes temporais mais longos, embora o método e o modelo de previsão possam oferecer um desempenho adequado no curto prazo. Desta forma, estudos adicionais e a integração de outras técnicas e modelos podem ser necessários para obter resultados mais precisos em períodos mais longos.

6. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

O objetivo deste trabalho consistiu na proposição de um método voltado à integração de dados de mercado financeiro, mais especificamente, de bolsa de valores com mensagens (postagens) efetuadas em redes sociais. A partir disso, os dados são utilizados como entrada para um modelo de Aprendizado Profundo com o intuito de realizar previsões considerando o índice S&P500.

O trabalho se alicerça em uma pesquisa bibliográfica que abrangeu tanto a fundamentação teórica quanto a identificação de trabalhos correlatos. De modo geral, a literatura destaca a crescente importância da Análise de Sentimentos derivada de redes sociais como uma ferramenta preditiva nos mercados financeiros. Estudos demonstraram a eficácia de modelos de DL e estratégias de investimento baseadas em dados sentimentais extraídos de notícias e mensagens de redes sociais enfatizando o uso de modelos de linguagem como BERT e RoBERTa, assim como a robustez da arquitetura LSTM em comparação com métodos tradicionais de ML.

No desenvolvimento do método proposto foi delineada uma abordagem sequencial que inicia com a coleta de dados de redes sociais, seguida pela aplicação de modelos de linguagem avançados para a classificação de sentimentos. Tais sentimentos são posteriormente integrados aos dados de preços de fechamento do mercado de ações para fornecer um conjunto de dados rico e multifacetado.

Os materiais e métodos empregados incluem conjuntos de dados de *tweets* (mensagens da rede social X[®]) e preços do mercado financeiro, com o suporte de bibliotecas desenvolvidas em Python como *yfinance*, *pandas*, *snsrape* e *keras*. Tais bibliotecas foram fundamentais para a extração, processamento e modelagem de dados, bem como para a implementação e treinamento da rede neural LSTM.

Foram conduzidos três experimentos distintos, em que o primeiro compara a eficácia da arquitetura LSTM com e sem a inclusão de dados de sentimentos. O segundo experimento avalia a capacidade do modelo em prever movimentos do mercado no dia seguinte, utilizando tanto dados sentimentais quanto apenas dados de fechamento do mercado. O terceiro, estende a janela de previsão para cinco dias. Cada experimento foi analisado com base em métricas comumente utilizadas na área de Aprendizado de Máquina como MAE, RMSE e R².

As considerações finais deste trabalho reforçam a validade da inclusão da análise de sentimentos nas previsões de mercado, especialmente em horizontes de curto prazo. De maneira geral, os resultados evidenciam a importância de integrar análise de sentimentos com dados de mercado financeiro. Permitem assim, concluir que o método proposto atingiu seu objetivo ao incorporar a análise de sentimentos de *tweets* na previsão do mercado financeiro, mais especificamente, na bolsa de valores. Ademais, demonstra que as percepções e reações expressas nas redes sociais podem, de fato, ser indicadores relevantes para lidar com as flutuações do mercado.

Embora os modelos de aprendizado produzidos que incluem análise de sentimentos possuam capacidade preditiva superior em relação aos que não consideram, ainda há uma margem considerável de variação não explicada, o que é esperado em previsões de médio prazo quando se considera a volatilidade do mercado financeiro. Isto abre diversas possibilidades de evolução, tanto no tratamento dos dados quanto no modelo de previsão.

Uma direção promissora envolve a implementação do método proposto na forma de serviço, permitindo o acesso por diferentes plataformas visando atender às variadas necessidades de investidores. Esta abordagem promoveria flexibilidade e acessibilidade, permitindo que investidores utilizassem o método em diferentes configurações de acordo com suas preferências e recursos disponíveis. Em tal ambiente, o método poderia processar continuamente fluxos de dados de várias redes sociais e fontes de notícias financeiras,

fornecendo *insights* atualizados e precisos sobre tendências de mercado. Além disso, o serviço poderia ser configurado para enviar alertas e atualizações personalizadas aos investidores, promovendo maior agilidade nas decisões. Além disso, o acesso do serviço por meio de *smartphones* ou *tablets* promoveria flexibilidade e conveniência ao investidor individual ao obter análises sem restrição geográfica.

Neste trabalho foi explorada a análise de sentimentos em *tweets* para prever o mercado financeiro, sendo necessário expandir a investigação para incluir dados de outras redes sociais, como Reddit® e Meta®. Essa expansão tem potencial para diversificar e enriquecer a compreensão em como sentimentos variados afetam o mercado financeiro. Por outro lado, é fundamental avaliar criticamente a autenticidade das mensagens, pois o aumento da geração de conteúdo por Inteligência Artificial eleva o risco de incorporação de dados distorcidos ou falsos, que podem comprometer a integridade das análises. Por outro lado, um desafio emergente é o acesso a essas valiosas fontes de dados, isto é, às redes sociais. As plataformas de mídias sociais estão gradualmente introduzindo taxas para a utilização de suas APIs, o que pode limitar o acesso aos dados necessários para pesquisas como a presente, além de impor um custo adicional pela utilização desses serviços. Tais mudanças no modelo de negócios das plataformas podem representar um obstáculo significativo para a pesquisa contínua nesse campo.

Ademais, é imprescindível a consideração de variáveis macroeconômicas e políticas, que também exercem influência considerável sobre o mercado financeiro. Indicadores econômicos e mudanças regulatórias são fatores que, se integrados à análise de sentimentos e dados financeiros, podem oferecer uma visão mais holística e potencialmente conduzir a previsões de mercado financeiro mais precisas e robustas.

Por fim, é essencial considerar o estudo e implementação de outros modelos de Inteligência Artificial. O teste e a implementação de diversas arquiteturas de AI, como CNNs, sistemas de atenção, novos algoritmos emergentes baseado em arquitetura *transformers* ou Transformadores de Fusão Temporal (do inglês *Temporal Fusion Transformer - TFT*), podem revelar diferentes possibilidades na previsão de tendências do mercado financeiro a partir de grandes volumes de dados sociais e econômicos. Pesquisas futuras nessa direção possuem, não somente potencial de expandir o entendimento atual neste domínio, mas também de fortalecer a robustez e a confiabilidade dos modelos preditivos no campo das finanças comportamentais.

Portanto, trabalhos futuros deverão não apenas ampliar a diversidade e quantidade de dados analisados, mas também desenvolver mecanismos sofisticados para verificar a veracidade das fontes, adaptando-se às novas demandas associadas ao acesso a essas informações. A incorporação de análises multidimensionais que abrangem perspectivas econômicas e políticas pode transformar significativamente a precisão dos modelos preditivos em finanças.

REFERÊNCIAS

ALZUBAIDI, L. *et al.* Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Journal of Big Data**, v. 8, n. 1, 31 mar. 2021. Disponível em: <https://link.springer.com/article/10.1186/s40537-021-00444-8>. Acesso em: 22 nov 2023.

ARAÚJO, M. *et al.* Métodos para análise de sentimentos no twitter. In: **Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)**. sn, 2013. p. 19. Disponível em: <https://homepages.dcc.ufmg.br/~fabricio/download/webmedia13.pdf>. Acesso em: 12 jul 2023.

ARAÚJO, V. L.; FERNANDES, B. J. T. Análise de Sentimentos de Textos do Twitter sobre Mercado de Ações Brasileiro. **Revista de Engenharia e Pesquisa Aplicada**, v. 6, n. 5, p. 18-26, 2021. Disponível em: <http://revistas.poli.br/~anais/index.php/repa/article/view/1750/774>. Acesso em: 07 mar 2022.

BOZANTA, A. *et al.* Sentiment analysis of stocktwits using transformer models. In: **2021 20th IEEE international conference on machine learning and applications (ICMLA)**. IEEE, 2021. p. 1253-1258. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9680251>. Acesso em: 17 out 2023.

BUSTOS, O.; POMARES-QUIMBAYA, A. Stock market movement forecast: A systematic review. **Expert Systems with Applications**, v. 156, p. 113464, 2020. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417420302888>. Acesso em: 7 mar 2022.

CAROSIA, A. E. de O.; COELHO, G. P.; DA SILVA, A. E. A. Investment strategies applied to the Brazilian stock market: a methodology based on sentiment analysis with deep learning. **Expert Systems with Applications**, v. 184, p. 115470, 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417421008812>. Acesso em: 07 mar. 2022.

CHATZILOIZOS, G.-M.; GUNOPULOS, D.; KONSTANTINOU, K. Forecasting Stock Market Trends using Deep Learning on Financial and Textual Data. In: **DATA**. 2021. p. 105-114. Disponível em: <https://www.scitepress.org/Papers/2021/106188/106188.pdf>. Acesso em: 17 out 2023.

DAVE, K.; LAWRENCE, S.; PENNOCK, D. M. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: **Proceedings of the 12th international conference on World Wide Web**. 2003. p. 519-528. Disponível em: <https://dl.acm.org/doi/abs/10.1145/775152.775226>. Acesso em: 7 mar 2022.

DE SOUZA, V. A.; DE SOUZA, Érica Ferreira; MEINERZ, Giovani Volnei. Análise de sentimento em tempo real de notícias do mercado de ações. **Brazilian Journal of Development**, v. 7, n. 1, p. 11084-11091, 2021. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/23959/19224>. Acesso em: 7 mar 2022

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K.. BERT: Pre-training of deep bidirectional transformers for language understanding. 2018. Disponível em: <https://arxiv.org/abs/1810.04805>. Acesso em: 7 mar 2022.

EIAMKANITCHAT, N.; MOONTUY, T.; RAMINGWONG, S. Fundamental analysis and technical analysis integrated system for stock filtration. **Cluster Computing**, v. 20, n. 1, p. 883–894, 30 nov. 2016. Disponível em: <https://link.springer.com/article/10.1007/s10586-016-0694-2>. Acesso em: 11 dez 2023.

FERNANDES, W. R.; DA SILVA, R. C. Aprendizagem profunda de máquinas: conceitos, técnicas e bibliotecas. Dez. 2017. Disponível em: https://www.researchgate.net/publication/324844857_Aprendizagem_profunda_de_maquinas_conceitos_tecnicas_e_bibliotecas. Acesso em: 17 out 2023.

GAO, R. *et al.* Integrating the sentiments of multiple news providers for stock market index movement prediction: A deep learning approach based on evidential reasoning rule. **Information sciences**, v. 615, p. 529-556, 2022. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0020025522011495>. Acesso em: 17 out 2023.

HAO, Y. *et al.* Visualizing and Understanding the Effectiveness of BERT. Disponível em: <https://arxiv.org/abs/1908.05620>. Acesso em: 17 out 2023.

JAIN, P. C. The Effect on Stock Price of Inclusion in or Exclusion from the S&P 500. **Financial Analysts Journal**, v. 43, n. 1, p. 58–65, jan. 1987. Disponível em: <https://www.tandfonline.com/doi/epdf/10.2469/faj.v43.n1.58?needAccess=true>. Access em: 3 dez 2023.

JANUÁRIO, B. A. *et al.* Sentiment analysis applied to news from the Brazilian stock market. **IEEE Latin America Transactions**, v. 20, n. 3, p. 512-518, 2021. Disponível em: <https://ieeexplore.ieee.org/abstract/document/9667151>. Acesso em: 07 mar. 2022.

JI, Z. *et al.* Exploring the impact of investor's sentiment tendency in varying input window length for stock price prediction. **Multimedia Tools and Applications**, p. 1-35, 2023. Disponível em: <https://link.springer.com/article/10.1007/s11042-023-14587-8>. Acesso em: 17 out 2023.

JIA, M. *et al.* Analysis and Research on Stock Price of LSTM and Bidirectional LSTM Neural Network. Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019), 2019. Disponível em: <https://www.atlantispress.com/proceedings/iccia-19/125913164>. Acesso em: 17 out 2023.

JIANG, H. Stock Market Return and Household Financial Investments. In: 2021 5th International Conference on E-Business and Internet, 2021, Singapura, Singapura. **Association for Computing Machinery**, Nova Iorque, Estados Unidos, 2021. Disponível em: <https://doi.org/10.1145/3497701.3497719>. Acesso em: 07 mar. 2022.

KAHN, M. N. An Introduction to Technical Analysis: Charting Techniques and Trading Strategies. Nova Iorque: McGraw-Hill, 2018. Acesso em: 17 out 2023.

KAVYA, K. S.; BP, V. K.; KUSUMA, S. M. Aspect Level Sentiment Analysis Aided Supportive System for Interactive Platforms. In: **2022 Third International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)**. IEEE, 2022. p. 1-8. Disponível em: <https://ieeexplore.ieee.org/abstract/document/10099722>. Acesso em: 17 out 2023.

KINYUA, J. D. *et al.* An analysis of the impact of President Trump's tweets on the DJIA and S&P 500 using machine learning and sentiment analysis. **Journal of behavioral and experimental finance**, v. 29, p. 100447, 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S2214635020303762>. Acesso em: 17 out 2023.

KISHTAINY, N. *et al.* **The economics book**. New York: Dk Pub, 2012.

KUMAR, R.; KUMAR, P.; KUMAR, Y. Integrating big data driven sentiments polarity and ABC-optimized LSTM for time series forecasting. **Multimedia Tools and Applications**, v. 81, n. 24, p. 34595-34614, 2022. Disponível em: <https://link.springer.com/article/10.1007/s11042-021-11029-1>. Acesso em: 17 out 2023.

LEE, C.-C.; GAO, Z.; TSAI, C.-L. BERT-Based Stock Market Sentiment Analysis. In: 2020 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), 2020, Taoyuan, Taiwan. **IEEE**, 2020. Disponível em: [10.1109/ICCE-Taiwan49838.2020.9258102](https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258102). Acesso em: 07 mar. 2022.

LIMA, L. A. de O. Auge e declínio da hipótese dos mercados eficientes. **Brazilian Journal of Political Economy**, v. 23, p. 531-546, 2020. Disponível em: <https://www.scielo.br/j/rep/a/kJw4LjkM8GGQ4RpbTcfqFyN/?lang=pt>. Acesso em: 7 mar 2022.

LIU, B. **Sentiment analysis: mining opinions, sentiments, and emotions**. 2. ed. Reino Unido: Cambridge University Press. 2020. Acesso em: 17 out 2023.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, p. 85-94, 2021. Disponível em: <https://www.scielo.br/ea/a/wXBdv8yHBV9xHz8qG5RCgZd/?lang=pt&format=html>. Acesso em: 7 mar 2022.

MEDEIROS, M. C. Metodologias para análise de sentimentos de tweets sobre o mercado financeiro. 2019. Disponível em: https://bdm.unb.br/bitstream/10483/29207/1/2019_MuriloCerqueiraMedeiros_tcc.pdf. Acesso em: 7 mar 2022.

MENGHANI, G. Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better. **ACM Computing Surveys**, v. 55, n. 259, p. 1-37, 20 jan. 2023. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3578938>. Acesso em: 17 out 2023.

MOGHAR, A.; HAMICHE, M. Stock market prediction using LSTM recurrent neural network. **Procedia Computer Science**, Varsóvia, Polônia, v. 170, p. 1168-1173, abr. 2020. Disponível em: <https://doi.org/10.1016/j.procs.2020.03.049>. Acesso em: 07 mar. 2020.

MURPHY, J. J. Technical analysis of the financial markets : a comprehensive guide to trading methods and applications. **New York: New York Institute Of Finance**, 1999. Acesso em: 5 dez 2023

MURPHY, J. J. The Art of Technical Analysis: The Definitive Guide to Trading Methods and Tactics. 2. ed. **Nova Iorque: Penguin Random House**, 2020. Acesso em: 17 out 2023.

NABIPOUR, M. et al. Deep Learning for Stock Market Prediction. **Entropy**, v. 22, n. 8, p. 840, 30 jul. 2020. Disponível em: <https://www.mdpi.com/1099-4300/22/8/840>. Acesso em: 12 dez 2023.

NASUKAWA, T. ; YI, J. Sentiment analysis: Capturing favorability using natural language processing. In: **Proceedings of the 2nd international conference on Knowledge capture**. 2003. p. 70-77. Disponível em: <https://dl.acm.org/doi/abs/10.1145/945645.945658>. Acesso em: 7 mar 2022.

NIELSEN, M. **Redes Neurais Artificiais: Uma Abordagem Prática**. São Paulo: Novatec, 2015. Acesso em: 17 out 2023.

ORTMANN, R.; PELSTER, M.; WENGEREK, S. T. COVID-19 and investor behavior. **Finance research letters**, [s.l.], v. 37, p. 101717, nov. 2020. Disponível em: <https://doi.org/10.1016/j.frl.2020.101717>. Acesso em: 07 mar. 2022.

PAN, S. *et al.* Nonlinear asset pricing in Chinese stock market: A deep learning approach. **International Review of Financial Analysis**, v. 87, p. 102627, 1 maio 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1057521923001436>. Acesso em: 17 out 2023.

PARNES, D. Exploring economic anomalies in the S&P500 index. **The Quarterly Review of Economics and Finance**, [s.l.], v. 76, p. 292-309, mai. 2020. Disponível em: <https://doi.org/10.1016/j.qref.2019.09.012>. Acesso em: 07 mar. 2022.

POUROMID, M. *et al.* ParsBERT Post-Training for Sentiment Analysis of Tweets Concerning Stock Market. In: 2021 26th International Computer Conference, Computer Society of Iran (CSICC), 2021, Teerã, Irã. **IEEE**, 2021. Disponível em: <10.1109/CSICC52343.2021.9420569>. Acesso em: 07 mar. 2022.

PRING, M. J. **Technical Analysis for Dummies**. 2. ed. Hoboken, NJ: Wiley, 2021. Acesso em: 17 out 2023.

RAUBER, T. W. Redes neurais artificiais. **Universidade Federal do Espírito Santo**, v. 29, 2005. Disponível em: https://www.researchgate.net/profile/Thomas-Rauber-2/publication/228686464_Redес_neurais_artificiais/links/02e7e521381602f2bd000000/Redes-neurais-artificiais.pdf. Acesso em: 07 mar 2022

REED, M. A study of social network effects on the stock market. **Journal of Behavioral Finance**, [s.l.], v. 17, n. 4, p. 342-351, nov. 2016. Disponível em: <https://doi.org/10.1080/15427560.2016.1238371>. Acesso em: 07 mar. 2022.

ROGERS, J. W. H. *et al.* **The Efficient Market Hypothesis: A Review of the Empirical Evidence**. 2021. Acesso em: 17 out 2023.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial: Uma Introdução**. 3. ed. São Paulo: Pearson, 2016. Acesso em: 17 out 2023.

SCHNITZLER, J. S&P 500 inclusions and stock supply. **Journal of Empirical Finance**, v. 48, p. 341–356, 1 set. 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0927539818300537>. Acesso em: 5 dez 2023.

SMAGULOVA, K.; JAMES, A. P. A survey on LSTM memristive neural network architectures and applications. **The European Physical Journal Special Topics**, v. 228, n. 10, p. 2313–2324, out. 2019. Disponível em: <https://link.springer.com/article/10.1140/epjst/e2019-900046-x>. Acesso em: 17 out 2023.

SWATHI, T.; KASIVISWANATH, N.; RAO, A. Ananda. An optimal deep learning-based LSTM for stock price prediction using twitter sentiment analysis. **Applied Intelligence**, v. 52, n. 12, p. 13675-13688, 2022. Disponível em: <https://link.springer.com/article/10.1007/s10489-022-03175-2>. Acesso em: 17 out 2023.

STAUDEMAYER, R. C.; MORRIS, E. R. Understanding LSTM: A tutorial into long short-term memory recurrent neural networks. **arXiv preprint arXiv:1909.09586**, 2019. Disponível em: <https://arxiv.org/pdf/1909.09586.pdf>. Acesso em: 7 mar 2022

THAKKAR, A.; CHAUDHARI, K. A Comprehensive Survey on Deep Neural Networks for Stock Market: The Need, Challenges, and Future Directions. **Expert Systems with Applications**, p. 114800, mar. 2021. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0957417421002414>. Acesso em: 7 mar 2022.

WANG, Z. *et al.* Learning-based stock trending prediction by incorporating technical indicators and social media sentiment. **Cognitive Computation**, v. 15, n. 3, p. 1092-1102, 2023. Disponível em: <https://link.springer.com/article/10.1007/s12559-023-10125-8>. Acesso em: 17 out 2023.

XU, H. *et al.* **BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis**. [s.l: s.n.]. Disponível em: <https://arxiv.org/pdf/1904.02232.pdf>. Acesso em: 17 out 2023.

YIN, W. *et al.* **Comparative Study of CNN and RNN for Natural Language Processing**. 2017. Disponível em: <https://arxiv.org/abs/1702.01923>. Acesso em: 22 nov 2023.

YU, Y. *et al.* A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. **Neural Computation**, v. 31, n. 7, p. 1235–1270, jul. 2019. Disponível em: <https://direct.mit.edu/neco/article-abstract/31/7/1235/8500/A-Review-of-Recurrent-Neural-Networks-LSTM-Cells>. Acesso em: 17 out 2023.

ZAREMBA, A. *et al.* The quest for multidimensional financial immunity to the COVID-19 pandemic: Evidence from international stock markets. **Journal of International Financial Markets, Institutions and Money**, [s.l.], v. 71, p. 101284, mar. 2021. Disponível em: <https://doi.org/10.1016/j.intfin.2021.101284>. Acesso em: 07 mar. 2022.

ZHENG, C. *et al.* Deep Learning-Based Human Pose Estimation: A Survey. **ACM Computing Surveys**, v. 56, n. 11, p. 1–37, 9 jun. 2023. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3603618>. Acesso em: 17 out 2023.