O Boletim de Conjuntura (BOCA) publica ensaios, artigos de revisão, artigos teóricos e empíricos, resenhas e vídeos relacionados às temáticas de políticas públicas.

O periódico tem como escopo a publicação de trabalhos inéditos e originais, nacionais ou internacionais que versem sobre Políticas Públicas, resultantes de pesquisas científicas e reflexões teóricas e empíricas.

Esta revista oferece acesso livre imediato ao seu conteúdo, seguindo o princípio de que disponibilizar gratuitamente o conhecimento científico ao público proporciona maior democratizacão mundial do conhecimento.



BOLETIM DE CONJUNTURA

BOCA

Ano VII | Volume 21 | Nº 62 | Boa Vista | 2025

http://www.ioles.com.br/boca

ISSN: 2675-1488

https://doi.org/10.5281/zenodo.15092801

BOLETIM DE CONJUNTURA

www.ioles.com.br/boca

PROCESSO DE EXTRAÇÃO DE SENTIMENTO NO MERCADO FINANCEIRO

Fernando Salvino da Silva[1]

Laura Alves Pacífico da Silva[2]

Mauro Margalho Coutinho[3]

Jucimar Casimiro de Andrade[4]

Jamille Queiroz Leite[5]

Resumo

Este estudo investiga a relação entre o sentimento, expresso em notícias financeiras, e a movimentação do mercado, utilizando notícias extraídas do Yahoo Finance entre setembro de 2020 e setembro de 2023. A metodologia empregou técnicas de Processamento de Linguagem Natural (PLN) para análise de sentimento, utilizando as bibliotecas NLTK (VADER) e TextBlob. As notícias foram coletadas via web scraping e analisadas quanto à polaridade (positiva, neutra ou negativa), sendo posteriormente comparadas com as variações do índice S&P 500. Os resultados indicaram que o sentimento das notícias foi, em média, levemente positivo, com pequenas discrepâncias entre os métodos de análise. Foi observado que, em determinados períodos, como no início de 2021, um aumento na quantidade de notícias positivas coincidiu com a recuperação do mercado, mas a relação entre sentimento e desempenho do mercado não se mostrou linear, sugerindo a influência de fatores macroeconômicos e geopolíticos. Conclui-se que a análise de sentimento pode ser um indicador complementar na previsão de tendências do mercado financeiro, mas sua aplicação deve considerar múltiplas fontes de informação e modelos mais avançados para aprimorar a precisão das previsões.

Palavras-chave: Análise de Sentimento; Mercado Financeiro; Previsão de Tendências; Processamento de Linguagem Natural.

Abstract

This study investigates the relationship between sentiment expressed in financial news and market movements, using news articles extracted from Yahoo Finance between September 2020 and September 2023. The methodology applied Natural Language Processing (NLP) techniques for sentiment analysis, employing the NLTK (VADER) and TextBlob libraries. News articles were collected through web scraping and analyzed for polarity (positive, neutral, or negative), then compared with fluctuations in the S&P 500 index. The results indicated that financial news sentiment was, on average, slightly positive, with minor discrepancies between analytical methods. In certain periods, such as early 2021, an increase in positive news coincided with market recovery, but the relationship between sentiment and market performance was not linear, suggesting the influence of macroeconomic and geopolitical factors. It is concluded that sentiment analysis can serve as a complementary indicator for predicting financial market trends, but its application should consider multiple sources of information and more advanced models to improve forecast accuracy.

[1] Doutorando em Administração pela Universidade da Amazônia (UNAMA). E-mail: administracaoead@sereducacional.com

[2] Doutoranda em Ciências da Computação pela Universidade Federal de Pernambuco (UFPE). E-mail: laura.silva@sereducacional.com

[3] Doutor em Engenharia Elétrica pela Universidade Federal do Pará (UFPA). E-mail: mauro.margalho@unama.br

[4] Mestre em Administração pela Universidade Federal Rural de Pernambuco (UFRPE). E-mail: jucimar.casimiro@sereducional.com

[5] Doutoranda em Administração pela Universidade da Amazônia (UNAMA). E-mail: jamille.leite@sereducacional.com

Keywords: Financial Market; Natural Language Processing; Sentiment Analysis; Trend Forecasting.

INTRODUÇÃO

Nos últimos anos, a análise de sentimento emergiu como uma ferramenta essencial para interpretar as percepções do mercado financeiro. Com o avanço da tecnologia e o crescimento das fontes digitais de informação, investidores e analistas passaram a recorrer a algoritmos de inteligência artificial para extrair insights a partir de grandes volumes de dados textuais. Esse processo permite captar o tom geral das notícias, publicações em redes sociais e relatórios financeiros, auxiliando na tomada de decisão sobre investimentos.

A crescente volatilidade do mercado e a influência direta da percepção coletiva nas oscilações de preços reforçam a necessidade de metodologias robustas para interpretar sentimentos e tendências. No mercado financeiro, a análise de sentimento adquire uma importância particularmente crítica por dois principais motivos:

- Tomada de Decisão Baseada em Percepção: O mercado financeiro não é movido apenas por métricas e números. As percepções, opiniões e sentimentos dos investidores desempenham um papel fundamental na direção que os preços dos ativos tomam. Se os investidores se sentem otimistas sobre a economia ou sobre uma ação específica, isso pode levar a um aumento na compra de ativos, elevando seu preço. Por outro lado, um sentimento predominante de pessimismo pode resultar em uma venda massiva de ativos, fazendo o preço cair.
- Análise em Tempo Real: O sentimento do mercado é volátil e pode mudar rapidamente com base em notícias, relatórios financeiros, eventos globais e outros fatores. A capacidade de analisar o sentimento em tempo real ou em períodos curtos pode oferecer uma vantagem competitiva aos investidores, permitindo-lhes tomar decisões informadas rapidamente.

Diante desse cenário, este estudo busca responder à seguinte questão: como a análise de sentimento pode ser utilizada para prever movimentos do mercado financeiro? A resposta para essa pergunta é fundamental para aprimorar as estratégias de investimento e mitigar riscos associados a decisões baseadas unicamente em métricas quantitativas. A pesquisa se fundamenta na premissa de que a percepção do público, quando corretamente analisada, pode fornecer sinais preditivos relevantes para a movimentação dos ativos.

Para investigar essa questão, adotamos um modelo metodológico baseado em técnicas de processamento de linguagem natural (PLN) e aprendizado de máquina. A análise foi realizada utilizando dados coletados do *Yahoo Finance*, um dos principais portais de notícias financeiras, e processados por ferramentas de extração de sentimento, como NLTK e TextBlob. A abordagem

metodológica inclui a extração, tratamento e classificação dos textos, bem como a correlação entre os sentimentos identificados e os movimentos do mercado.

O artigo está estruturado da seguinte forma: após esta introdução, a Revisão da Literatura apresenta os principais conceitos e estudos sobre a análise de sentimento no mercado financeiro, abordando diferentes abordagens metodológicas e o impacto dessas técnicas na previsão de tendências. Em seguida, na Metodologia, detalhamos os procedimentos adotados para a coleta, processamento e análise dos dados, bem como as ferramentas empregadas. A seção de Resultados e Discussão apresenta as descobertas da pesquisa, analisando os padrões de sentimento identificados e suas correlações com os movimentos do mercado. Por fim, a Conclusão sintetiza os principais achados, destaca as limitações do estudo e sugere direções para pesquisas futuras.

REVISÃO DA LITERATURA

A análise de sentimento tem se consolidado como uma ferramenta essencial no mercado financeiro, permitindo a interpretação de emoções, opiniões e atitudes presentes em textos financeiros, como notícias, postagens em mídias sociais e relatórios de analistas. Essas análises oferecem *insights* sobre as tendências do mercado e auxiliam na tomada de decisões de investimento (LIU; SHI, 2024). Com o avanço das técnicas de Processamento de Linguagem Natural (PLN) e aprendizado profundo, modelos de linguagem de grande porte (LLMs) têm sido aplicados para aprimorar a análise de sentimento financeiro. Estudos recentes demonstram que modelos como o ChatGPT, quando ajustados especificamente para o domínio financeiro, podem superar modelos tradicionais, oferecendo maior precisão na interpretação de textos financeiros (HUANG *et al*., 2023).

Historicamente, a análise financeira convencional concentrou-se em métricas quantitativas, como balanços, fluxos de caixa e índices econômicos. No entanto, com a ascensão das redes sociais, blogs, fóruns e outras plataformas digitais, uma vasta quantidade de dados textuais relacionados ao mercado tornou-se disponível. Esses dados refletem as opiniões, sentimentos e expectativas de uma ampla gama de participantes do mercado, desde investidores amadores até profissionais da área. Dada a influência dos sentimentos nas decisões de investimento, uma análise efetiva desse vasto conjunto de dados textuais pode revelar padrões e *insights* que as métricas quantitativas tradicionais não conseguem capturar. Por exemplo, um aumento súbito no sentimento negativo em relação a um ativo específico pode preceder uma

queda em seu preço, mesmo que os indicadores quantitativos pareçam favoráveis (KIRTAC; GERMANO, 2024).

O conceito de extração de sentimento nos mercados financeiros evoluiu significativamente ao longo do tempo, refletindo a crescente complexidade da dinâmica do mercado e o advento de técnicas analíticas avançadas. Inicialmente, a necessidade de prever comportamentos futuros do mercado surgiu do desejo humano de compreender os possíveis desdobramentos de eventos econômicos e políticos (TETLOCK, 2007). Isso levou ao desenvolvimento da previsão baseada em séries temporais, que estabeleceu as bases para a integração da análise de sentimento nos contextos financeiros. Com o avanço tecnológico, o desenvolvimento de ferramentas de PLN, especialmente modelos como BERT (Bidirectional Encoder Representations from Transformers), revolucionou a forma como a análise de sentimento é conduzida. Ao contrário dos modelos tradicionais, que processavam textos de forma linear, a capacidade do BERT de compreender o contexto completo das palavras dentro de uma sentença permitiu interpretações mais sofisticadas de notícias financeiras, relatórios e conteúdos de mídia social (BOLLEN; MAO; ZENG, 2011). Essa inovação possibilitou a identificação de tons emocionais complexos que influenciam os movimentos do mercado, aumentando a precisão das previsões de sentimento.

A ascensão das plataformas de mídia social, especialmente o Twitter, também desempenhou um papel crucial no processo de extração de sentimento. Mao *et al*. (2023) demonstraram uma correlação significativa entre o sentimento expresso no Twitter e as flutuações dos preços das ações, evidenciando a eficácia da análise de sentimento em mídias sociais como um indicador preditivo das tendências de mercado. Ao aproveitar grandes volumes de dados não estruturados de várias fontes, os analistas conseguem derivar insights acionáveis que informam decisões de negociação e estratégias de gerenciamento de risco (COSTOLA *et al.,* 2023). Além disso, estudos como os de Riso e Vacca (2024) indicam que o sentimento influencia diretamente a volatilidade dos índices acionários e que modelos de aprendizado de máquina podem prever a volatilidade de longo prazo, auxiliando investidores na estabilização do sentimento e na redução da incerteza do mercado.

A literatura recente confirma que o sentimento capturado de notícias tem poder explanatório e preditivo sobre os mercados. Liu e Shi (2024) analisaram ações do índice Dow Jones entre 2019 e 2023 e descobriram que tanto notícias negativas quanto positivas impactam significativamente a volatilidade intradiária dos retornos. A substituição da análise tradicional por um classificador de sentimento baseado em GPT-4 resultou em medidas mais precisas e

revelou um impacto ainda maior dessas notícias na dinâmica do mercado. Outros estudos, como os de Jeon *et al*. (2022), reforçam essa hipótese, analisando milhões de notícias de empresas e concluindo que certos eventos noticiosos funcionam como gatilhos para saltos abruptos nos preços das ações.

Os métodos tradicionais de análise de sentimento, baseados em dicionários de palavras positivas e negativas, apresentam limitações, como a incapacidade de capturar contexto, ironia e negações. Por isso, técnicas modernas de PLN, como FinBERT e GPT-4, demonstraram desempenho superior na compreensão de textos financeiros complexos (HUANG *et al*., 2023). A adaptação desses modelos ao domínio financeiro melhora significativamente sua capacidade de análise de sentimento, permitindo uma compreensão mais profunda das nuances presentes nos textos financeiros.

As redes sociais são amplamente utilizadas pela população e desempenham um papel significativo no mercado financeiro. As informações divulgadas *online* podem influenciar positiva ou negativamente o sentimento dos investidores, impactando suas expectativas e estratégias especulativas. Isso ocorre porque essas informações ajudam a atualizar crenças sobre a geração futura de fluxos de caixa dos ativos financeiros, afetando diretamente as oscilações do mercado acionário (KÖCHE *et al*, 2023).

Além disso, a análise de sentimento não se restringe às notícias e redes sociais. Estudos recentes destacam a importância da avaliação de sentimento em relatórios financeiros e conferências de resultados corporativos. Trabalhos como os de Dahal *et al*. (2023) mostram que a inclusão de sentimentos extraídos de relatórios financeiros melhora significativamente a precisão das previsões de preço em comparação com indicadores tradicionais. Essa descoberta reforça a ideia de que informações qualitativas, quando devidamente analisadas, podem complementar e aprimorar modelos preditivos financeiros.

Com a crescente digitalização da sociedade e a proliferação de expressões de opinião *online*, a análise de sentimento continuará a ser uma ferramenta indispensável para entender as percepções e sentimentos do público em relação a uma ampla variedade de tópicos e domínios financeiros. Estudos recentes demonstram que a combinação de aprendizado de máquina com análise de sentimento pode aprimorar previsões de volatilidade e retornos do mercado, fornecendo insights valiosos para investidores e gestores de risco (JAWALE *et al.,* 2023). A incorporação dessas técnicas na tomada de decisão tem se tornado cada vez mais comum, evidenciando a evolução e a sofisticação da análise de sentimento no mercado financeiro.

Processo de web scraping

O *web scraping*, também conhecido como extração ou colheita web, é uma técnica usada para coletar dados da World Wide Web (WWW) e armazená-los para análises futuras. Esses dados podem ser extraídos manualmente por um usuário ou automaticamente por um bot ou rastreador web. O processo de coleta envolve duas etapas principais: a aquisição dos recursos da web e, posteriormente, a extração das informações desejadas desses dados coletados. Inicialmente, o programa de *web scraping* forma uma solicitação HTTP para acessar um site-alvo, podendo ser realizada por meio de consultas GET ou POST. Após a resposta bem-sucedida do servidor, os dados são adquiridos e podem estar em diversos formatos, como HTML, XML, JSON ou mesmo multimídia, incluindo imagens, áudios e vídeos. A segunda etapa do processo consiste na interpretação, reformatação e organização dessas informações em um conjunto estruturado de dados (ZHAO, 2017).

Com o aumento exponencial da quantidade de informações disponíveis na WWW, o *web scraping* tornou-se uma ferramenta essencial para a coleta de dados em larga escala. As técnicas utilizadas evoluíram desde procedimentos *ad hoc*, auxiliados por humanos, até sistemas altamente automatizados capazes de converter sites inteiros em bases de dados organizadas. Além de analisar linguagens de marcação e arquivos JSON, ferramentas modernas de *web scraping* integram-se a técnicas de visão computacional e Processamento de Linguagem Natural (PLN), permitindo simular a navegação humana na web e interpretar o conteúdo extraído de maneira mais sofisticada (DOE *et al*., 2020; LIU, 2022; YU; GURUSWAMY, 2007).

Dada a crescente digitalização do mercado financeiro e a disponibilidade de dados em tempo real, o *web scraping* tem sido amplamente aplicado para coletar informações financeiras a partir de fontes como mídias sociais, fóruns de investidores e portais de notícias. Redes sociais como Twitter, Reddit e LinkedIn são plataformas que concentram opiniões de investidores individuais e especialistas, tornando-se fontes ricas para a análise de sentimento do mercado. Estudos recentes indicam que há uma forte correlação entre o sentimento público expresso nessas plataformas e flutuações nos preços das ações. Por exemplo, análises de dados do Reddit mostram que o humor agregado dos usuários frequentemente se correlaciona com a volatilidade de ações, fornecendo insights preditivos valiosos (VTECHWORKS.LIB.VT.EDU). No Twitter, pesquisas recentes também demonstram que o volume e o tom das postagens influenciam diretamente a variação de preços de ativos financeiros (ARXIV.ORG). Conforme Mulakala *et al*. (2023), integrar postagens de fóruns como o r/WallStreetBets, notícias do *Yahoo Finance* e

séries temporais de preços de ações permite capturar movimentos de curto prazo e antecipar mudanças no sentimento de investidores institucionais e de varejo (ARXIV.ORG).

Além das redes sociais, portais de notícias como *Yahoo Finance*, Bloomberg e Reuters desempenham um papel crítico na disseminação de informações que influenciam decisões de investimento. O *web scraping* tem sido empregado para extrair e quantificar o tom dessas manchetes, criando índices de sentimento de notícias que auxiliam na previsão da volatilidade do mercado. Kaur *et al*. (2022) desenvolveram um sistema de *scraping* de notícias em tempo real combinado com análise de sentimento, utilizando o algoritmo VADER para estudar o impacto do tom da mídia sobre as ações (SEMANTICSCHOLAR.ORG). Outras abordagens incluem a criação de índices agregados, como o *Daily News Sentiment Index* do Fed de São Francisco, que utiliza contagem de palavras em notícias econômicas para gerar um indicador diário do sentimento do mercado (FRBSF.ORG). Essas iniciativas evidenciam o valor da combinação de dados não estruturados – como tweets e manchetes de notícias – com métricas quantitativas tradicionais para uma visão mais abrangente do humor dos mercados financeiros.

Embora bibliotecas como BeautifulSoup e Selenium sejam amplamente utilizadas no *web scraping* para extrair conteúdo textual de páginas da web, muitas pesquisas recentes têm adotado APIs (Application Programming Interface) financeiras para simplificar a coleta de dados e garantir maior organização. APIs oficiais do Twitter e Reddit permitem a obtenção de posts públicos em formatos estruturados (JSON), filtrando informações por palavras-chave, hashtags ou período de tempo específico. Em pesquisas recentes, Mulakala *et al*. (2023) combinaram scraping e APIs, extraindo comentários do Reddit e notícias financeiras via scraping, enquanto utilizaram a API do Twitter para capturar menções relevantes a eventos do mercado (ARXIV.ORG). Além dessas, serviços como Alpha Vantage e Finnhub oferecem APIs que fornecem não apenas manchetes de notícias, mas também escores de sentimento pré-calculados, permitindo a criação de indicadores em tempo real sobre o humor do mercado financeiro (WEALTH-LAB.COM; REDDIT.COM).

O próximo desafio, após a coleta dos textos brutos via scraping ou API, consiste na análise de sentimento, etapa que se beneficia de avanços em PLN. Tradicionalmente, muitos estudos empregavam abordagens baseadas em léxicos, como o dicionário Loughran e McDonald (2011), amplamente utilizado para análise de documentos financeiros, dada sua capacidade de capturar nuances específicas do jargão econômico (ARXIV.ORG). Entretanto, técnicas baseadas apenas na contagem de palavras positivas e negativas apresentam limitações, especialmente diante de ironias e ambiguidades linguísticas. Por isso, estudos recentes passaram a empregar

modelos supervisionados de Machine Learning, como Naïve Bayes, Support Vector Machines (SVM) e redes neurais profundas. Koukaras *et al.* (2022) demonstraram que um classificador SVM alimentado com dados de sentimento extraídos de tweets conseguiu prever movimentos diários do mercado com um F-score de 76,3%, superando abordagens puramente lexicais (MDPI.COM).

Uma das evoluções mais relevantes da última década é o uso de modelos *transformers* treinados para PLN no domínio financeiro. O FinBERT, por exemplo, é uma adaptação do modelo BERT ajustada especificamente para analisar relatórios corporativos, discussões de mercado e manchetes econômicas (RESEARCHGATE.NET). Estudos como o de Huang *et al*. (2022) apontam que o FinBERT supera modelos genéricos e léxicos especializados na classificação de sentimento financeiro (RESEARCHGATE.NET). Com o avanço dos Large Language Models (LLMs), pesquisas recentes exploram o uso de modelos como ChatGPT para análise de sentimento em tempo real. Fatouros *et al*. (2023) compararam a eficácia do ChatGPT-3.5 e FinBERT na classificação de sentimento em notícias de câmbio, concluindo que o ChatGPT apresentou 35% mais precisão e correlação 36% maior com os retornos reais do mercado (ARXIV.ORG). Esses avanços consolidam a importância de modelos de PLN e web scraping na construção de sistemas preditivos robustos para análise de tendências financeiras.

Portanto, a união de web *scraping*, APIs financeiras e PLN tornou-se uma ferramenta essencial para extrair, estruturar e interpretar dados financeiros não estruturados. O estado da arte atual aponta para uma convergência dessas tecnologias, permitindo que investidores e instituições monitorem em tempo real o sentimento do mercado e antecipem movimentos financeiros com maior precisão. O desafio contínuo reside na adaptação dessas ferramentas ao dinamismo dos mercados, refinando a extração de sentimento diante da evolução do vocabulário e das estratégias de comunicação financeira.

Análise de sentimento

A análise de sentimento, também conhecida como mineração de opiniões, é uma subárea da mineração de dados e da análise de dados que busca identificar e extrair opiniões ou sentimentos de fontes de dados, como textos, imagens ou áudios. Suas aplicações são vastas e incluem áreas como marketing, finanças, saúde e política. Dada a explosão de dados não estruturados disponíveis na forma de resenhas online, tweets, blogs e postagens em redes sociais,

a capacidade de analisar automaticamente tais dados para extrair *insights* úteis tornou-se cada vez mais valiosa para empresas e pesquisadores (REIS *et al*. 2015).

A técnica emprega vários métodos e ferramentas, muitas das quais têm raízes em técnicas de processamento de linguagem natural (NLP) e aprendizado de máquina. Algoritmos, como máquinas de vetores de suporte (SVM) e redes neurais, são comuns na classificação de sentimentos (KANAKARAJ; GUDDETI, 2015). Contudo, a tarefa não é isenta de desafios. A ambiguidade linguística, o uso de sarcasmo, a presença de entidades múltiplas em um texto e os diferentes contextos culturais e demográficos são barreiras que tornam a análise de sentimento uma área de pesquisa em constante evolução.

A análise de sentimentos é amplamente empregada no setor financeiro, onde textos de notícias, comentários em redes sociais e relatórios corporativos frequentemente refletem o humor dos investidores e as tendências do mercado (INFORMATIONMATTERS.ORG, 2023). Investidores institucionais e de varejo tomam decisões estratégicas influenciadas pelo sentimento do mercado, tornando essencial a utilização de ferramentas avançadas para interpretar e quantificar essas informações. Estudos indicam que oscilações no humor dos investidores, detectadas em plataformas como Twitter e Reddit, podem antecipar flutuações nos preços de ativos (NYAKURUKWA; SEETHARAM, 2024). No contexto financeiro, essa técnica se baseia na hipótese de que o sentimento coletivo pode influenciar preços de ações e até mesmo prever movimentos de curto prazo no mercado (FATOUROS *et al*., 2023).

A análise de sentimentos pode ser realizada utilizando quatro abordagens principais: métodos baseados em léxicos, aprendizado de máquina tradicional, aprendizado profundo (deep learning) e modelos baseados em transformers. Cada uma dessas abordagens apresenta vantagens e limitações que influenciam sua aplicabilidade no setor financeiro.

Os métodos léxicos empregam dicionários de palavras associadas a polaridades positivas ou negativas. No setor financeiro, o dicionário de Loughran e McDonald é amplamente utilizado para quantificar o tom de relatórios e notícias corporativas (ARXIV.ORG, 2023). No entanto, essa abordagem enfrenta desafios significativos, como a necessidade de atualização contínua dos léxicos e a dificuldade de capturar ironia e contexto. Recentemente, pesquisadores propuseram métodos híbridos que combinam léxicos especializados com aprendizado de máquina para melhorar a precisão da classificação (RIZINSKI *et al*., 2023).

O aprendizado de máquina tradicional trata a análise de sentimentos como um problema de classificação de texto. Modelos como Naive Bayes, Support Vector Machines (SVM) e regressão logística são treinados em grandes volumes de textos rotulados para aprender padrões

de associação entre palavras e sentimentos (KIRTAC; GERMANO, 2024). No entanto, esses modelos têm dificuldades em capturar nuances semânticas e dependem fortemente da qualidade dos dados de treinamento.

O aprendizado profundo revolucionou a análise de sentimentos ao permitir que redes neurais recorrentes (RNN), redes LSTM (Long Short-Term Memory) e redes convolucionais (CNN) capturassem contextos complexos em textos financeiros (ABOWATH *et al*., 2025). Modelos mais avançados, como o FinBERT, uma versão adaptada do BERT para o setor financeiro, demonstraram desempenho superior na interpretação de relatórios de mercado e transcrições de conferências de resultados (FATOUROS *et al*., 2023).

A recente ascensão dos modelos baseados em transformers, como GPT-3 e ChatGPT, trouxe avanços significativos para a análise de sentimentos. Esses modelos utilizam mecanismos de atenção que processam grandes quantidades de texto e identificam padrões linguísticos sutis. Estudos indicam que o ChatGPT-3.5 superou o FinBERT em 35% na precisão da classificação de sentimentos de notícias financeiras, mostrando alta correlação entre suas previsões e os retornos reais do mercado (FATOUROS *et al.,* 2023).

Assim, não há abordagem única que seja universalmente superior – a escolha depende do caso de uso. Métodos léxicos oferecem rapidez e clareza, mas requerem manutenção e perdem nuance. Métodos de ML tradicionais equilibram simplicidade e adaptabilidade, mas necessitam dados rotulados e têm poder limitado. Deep learning e transformers fornecem máxima acurácia e capturam nuances, mas ao custo de recursos computacionais e menor transparência. Na prática, muitas soluções híbridas emergem: por exemplo, usar um modelo transformer para sugerir palavras a um léxico (como no XLex mencionado), ou combinar outputs de um modelo de deep learning com lógica de negócio. O importante é compreender esses trade-offs para aplicar a análise de sentimentos de forma eficaz no setor financeiro.

METODOLOGIA DA PESQUISA

Descrição do conjunto de dados

A presente pesquisa fundamenta-se na análise de sentimentos de textos financeiros extraídos do portal *Yahoo Finance*. O *Yahoo Finance* foi escolhido devido à sua ampla influência no setor financeiro, sendo uma das principais fontes de informação para investidores, analistas e pesquisadores de mercado. O portal disponibiliza cotações de ativos, índices de

mercado e análises aprofundadas, tornando-o um ambiente relevante para a extração e interpretação do sentimento predominante no mercado financeiro ao longo do período analisado. A coleta de dados abrange um período de três anos, compreendido entre 01/09/2020 e 01/09/2023, permitindo uma análise detalhada das tendências de sentimento ao longo do tempo.

Especificidade dos dados

Os dados utilizados na pesquisa consistem em textos extraídos diretamente da página inicial do *Yahoo Finance*. Estes dados incluem principalmente notícias, artigos, análises e comentários relacionados ao mercado financeiro global.

Fonte de Coleta: A obtenção dos textos foi realizada por meio de web scraping, utilizando a biblioteca BeautifulSoup para extrair os conteúdos textuais contidos nos elementos de parágrafo (<p>) da página. Esse processo permitiu a coleta estruturada das notícias apresentadas no portal durante o período analisado.

Amplitude dos Dados: Como a coleta foi limitada à página inicial, os dados representam um instantâneo do sentimento do mercado em cada momento específico. O *Yahoo Finance* exibe as notícias mais relevantes e atuais, garantindo que os textos coletados refletem os temas de maior impacto no período de extração.

Natureza dos Dados: Os textos abordam uma ampla gama de tópicos, incluindo oscilações de ações, eventos geopolíticos, políticas econômicas e outras informações que impactam diretamente os mercados financeiros. Dado esse escopo variado, espera-se que os sentimentos expressos nos textos apresentem variações significativas, dependendo do contexto econômico do momento.

Justificativa para escolha dos dados

A opção por coletar notícias do *Yahoo Finance* se justifica por diversos fatores metodológicos. Primeiro, o *Yahoo Finance* oferece acesso **gratuito e público** a notícias atualizadas do mercado financeiro, sem barreiras de paywall, o que viabiliza a extração contínua de dados. Além disso, trata-se de um portal amplamente utilizado por investidores e analistas; portanto, as notícias ali veiculadas tendem a impactar as expectativas do mercado, tornando-as fontes adequadas para analisar sentimento e possíveis efeitos sobre preços de ativos. A

abrangência do *Yahoo Finance* – cobrindo desde notícias corporativas até indicadores macroeconômicos – garante um volume considerável de dados para análise estatística.

Do ponto de vista científico, utilizar notícias financeiras como dados de entrada é coerente com pesquisas anteriores que relacionam sentimento da mídia e comportamento do mercado. Por exemplo, Sharma *et al*. (2022) empregaram manchetes do *Yahoo Finance* para prever movimentos de ações, categorizando-as em positivas, negativas ou neutras através de análise de sentimento. Essa adoção por estudos recentes reforça a confiabilidade e relevância dos dados do *Yahoo Finance* para este tipo de investigação. Em resumo, escolhemos esses dados por serem facilmente acessíveis, relevantes ao problema proposto e por já terem demonstrado utilidade em trabalhos semelhantes na literatura.

Apresentação das bibliotecas NLTK e TextBlob

Python é amplamente adotada em análise de dados e processamento de linguagem natural (NLP) devido à sua sintaxe intuitiva, facilitando a escrita e compreensão de códigos. A comunidade ativa e global proporciona um ecossistema de bibliotecas e recursos ricos, tornando a solução de problemas mais acessível. A análise foi feita em Google Colab está disponível em: https://colab.research.google.com/drive/1bGTq7Z_jYgOI80_z_Qhl-Ix1-eefM3ef?usp=sharing

Para conduzir a análise de sentimento, utilizamos duas bibliotecas de Processamento de Linguagem Natural (PLN) em Python: NLTK e TextBlob. A NLTK (Natural Language Toolkit) é uma plataforma amplamente utilizada para tarefas de PLN, oferecendo módulos para tokenização, remoção de stopwords, análise gramatical e ferramentas lexicon-based para sentimento. Em especial, a NLTK integra o VADER (*Valence Aware Dictionary and sEntiment Reasoner*), um léxico e modelo de regras voltado à análise de sentimento em textos curtos. O VADER, fornecido pelo NLTK, é reconhecido por ser ajustado à linguagem informal de redes sociais e funcionar bem também em outros domínios de texto. Dessa forma, a NLTK provê a base necessária tanto para pré-processar as notícias (por exemplo, segmentar frases, normalizar palavras) quanto para calcular um escore de sentimento através do VADER.

Por sua vez, o TextBlob é uma biblioteca construída sobre o NLTK (e other recursos como *Pattern*) que simplifica operações comuns de PLN com uma API intuitiva. O TextBlob oferece funcionalidades de tokenização, *part-of-speech tagging* (etiquetagem morfossintática) e um módulo embutido de análise de sentimentos lexicon-based. Diferentemente de modelos complexos de última geração, o TextBlob utiliza um dicionário de polaridades para avaliar o

texto, retornando diretamente métricas de polaridade e subjetividade. A polaridade varia de -1 a 1 e indica o teor sentimental do texto (negativo, neutro ou positivo), enquanto a subjetividade varia de 0 a 1 e reflete o grau de conteúdo subjetivo (opinião) versus conteúdo objetivo. Por exemplo, uma frase com polaridade próxima de -1 seria altamente negativa, ao passo que subjetividade próxima de 1 indica que o texto expressa mais opinião do que fato. Em suma, o TextBlob se destaca por fornecer de forma fácil medidas quantitativas de sentimento, o que complementa as ferramentas do NLTK.

Metodologia de análise usando as duas bibliotecas

A análise de sentimento proposta utiliza duas bibliotecas robustas e amplamente reconhecidas para processamento de linguagem natural (NLP) em Python: NLTK e TextBlob. O objetivo é avaliar o sentimento geral das notícias apresentadas na página inicial do *Yahoo Finance*. A seguir, delineamos passo a passo a metodologia empregada:

Preparação e Ambiente

A implementação da análise foi realizada no Google Colab, uma plataforma de computação em nuvem que permite a execução de scripts Python sem a necessidade de configuração local.

As bibliotecas necessárias para o estudo foram instaladas e importadas no ambiente, incluindo NLTK, TextBlob, BeautifulSoup e Matplotlib. Para o NLTK, foi necessário o download dos recursos específicos:

- nltk.download('punkt') (tokenização);
- nltk.download('vader_lexicon') (lexicon VADER);
- nltk.download('stopwords') (remoção de stopwords).

Coleta de Dados

Os dados foram coletados por meio de web scraping utilizando BeautifulSoup e requests para acessar a página principal do *Yahoo Finance* e extrair o conteúdo dos parágrafos:

- Dados Primários: Notícias coletadas diretamente do *Yahoo Finance*.

- Dados Secundários: Informações de contexto para validação dos resultados, como preços do índice S&P 500 e eventos econômicos relevantes.

Análise de Sentimento com NLTK

Iniciamos o *SentimentIntensityAnalyzer* do NLTK, uma ferramenta projetada para avaliar a intensidade do sentimento em textos. (VADER, 2024).

Para cada notícia ou texto coletado, calculamos uma pontuação de sentimento usando essa ferramenta. A pontuação composta, que varia de -1 (extremamente negativa) a 1 (extremamente positiva), é então armazenada em uma lista chamada *nltk_results*. A categorização seguiu a metodologia padrão:

- Escore **> 0.05** → Sentimento positivo
- Escore **< -0.05** → Sentimento negativo
- Escore **entre -0.05 e 0.05** → Sentimento neutro

Análise de Sentimento com TextBlob

Para cada notícia ou texto, também utilizamos o *TextBlob* para calcular uma pontuação de sentimento.

A polaridade, que também varia de -1 a 1, é extraída e armazenada na lista *textblob_results* seguindo a mesma categorização de sentimentos, com a possibilidade de identificar subjetividade nos textos analisados (SHAH; MUKHERJEE; PATEL, 2022).

A escolha dessas bibliotecas fundamenta-se na eficiência comprovada em estudos anteriores, que destacam a eficácia do VADER para análise de sentimentos de notícias financeiras e a capacidade do TextBlob de capturar nuances linguísticas em textos mais longos (WILSON; HOFFMAN; PIWOWARSKI, 2023).

Consolidação dos Resultados

Calculamos a média das pontuações de sentimento obtidas tanto pelo NLTK quanto pelo *TextBlob*, resultando em dois valores distintos: *media_nltk* e *media_textblob*.

Após aplicar ambas as técnicas de análise de sentimento, NLTK/VADER e TextBlob, procedemos à comparação dos resultados. Primeiramente, examinamos a distribuição de classes

de sentimento obtidas por cada abordagem separadamente, observando a frequência de notícias classificadas como positivas, neutras ou negativas. De modo geral, observamos uma correlação entre os dois métodos nas tendências principais, com ambos refletindo picos de notícias negativas ou positivas em determinados dias. No entanto, discrepâncias pontuais foram observadas. O TextBlob, por exemplo, tendeu a classificar mais frases como neutras em comparação ao VADER, que por vezes detectou leve positividade ou negatividade nesses casos. Para consolidar os resultados, focamos na análise quantitativa da média de sentimentos gerada por cada método. Calculamos a média do escore composto (compound) do VADER e a média da polaridade do TextBlob para todas as notícias analisadas. Em seguida, comparamos as médias obtidas, buscando identificar tendências gerais de sentimento (positivo, negativo ou neutro) e a concordância entre os métodos. Uma alta concordância entre as médias sugere robustez na detecção da tendência geral de sentimento, independentemente da ferramenta utilizada. Divergências significativas, por outro lado, podem indicar vieses específicos de cada método e demandariam uma análise mais aprofundada.

Visualização

Para visualizar os resultados da análise de sentimento, empregamos duas abordagens principais. Primeiramente, construímos um gráfico de barras que compara a média de sentimentos obtida com as ferramentas NLTK e TextBlob. Esse gráfico permite observar diretamente a diferença entre as médias calculadas por cada método, indicando possíveis variações na sensibilidade de cada ferramenta na detecção da polaridade das notícias. Em segundo lugar, geramos um gráfico de linhas que sobrepõe a contagem de notícias positivas ao longo do tempo com os preços do índice S&P 500, abrangendo o período de 01/09/2020 a 01/09/2023. Essa visualização tem como objetivo explorar a relação temporal entre o sentimento expresso nas notícias financeiras e o desempenho do mercado, permitindo identificar potenciais padrões de correlação ou divergência entre esses dois conjuntos de dados. Estudos anteriores demonstraram que a análise de sentimentos pode ser correlacionada com movimentos de mercado, sugerindo que uma variação expressiva no tom das notícias pode preceder oscilações no índice S&P 500 (SHAH *et al.,* 2022). Adicionalmente, embora não implementada no código atual, consideramos o uso de nuvens de palavras para visualizar os termos mais frequentes associados a cada classe de sentimento (positivo, negativo e neutro). Essa técnica, se aplicada, permitiria uma análise mais aprofundada do conteúdo das notícias, revelando os principais temas

que impulsionam o sentimento em cada categoria.

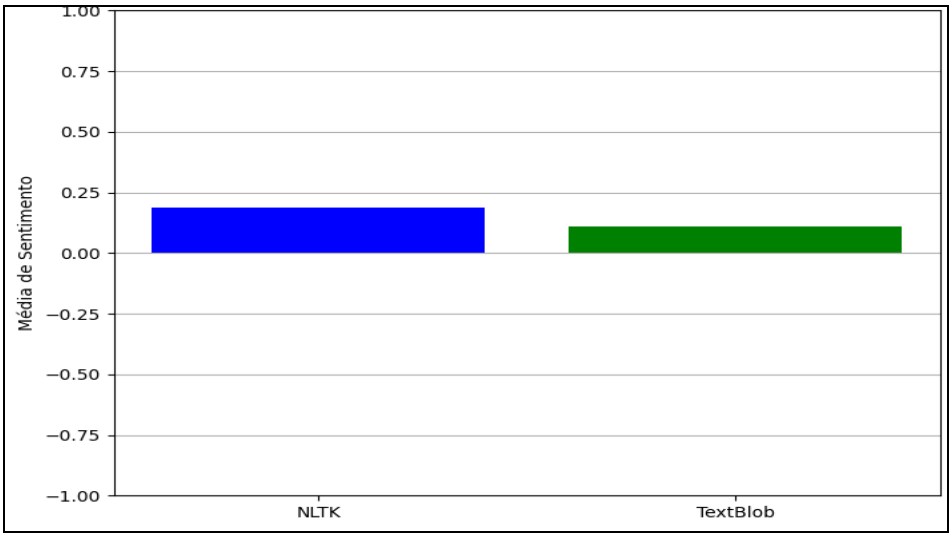
RESULTADOS E DISCUSSÃO

A análise de sentimento das notícias coletadas da página inicial do *Yahoo Finance* revelou uma tendência geral levemente positiva, com variações dependendo da metodologia utilizada. As bibliotecas de análise de sentimento NLTK (VADER) e TextBlob, ambas amplamente reconhecidas na literatura, corroboraram essa observação, embora com pequenas discrepâncias em suas avaliações. A média de sentimento calculada pelo NLTK foi de 0,187, enquanto o TextBlob registrou 0,1112, sugerindo que o primeiro interpretou o tom das notícias com um viés ligeiramente mais otimista do que o segundo.

O gráfico 1 apresenta um comparativo das médias de sentimento obtidas por ambas as bibliotecas. A diferença entre os resultados, embora sutil, destaca nuances metodológicas importantes, como os diferentes léxicos e abordagens computacionais de cada ferramenta. O VADER, utilizado no NLTK, foi projetado para capturar sentimentos em textos informais e curtos, enquanto o TextBlob utiliza uma análise baseada em lexicons predefinidos, que podem não estar totalmente adaptados ao contexto financeiro.

Essa discrepância não é incomum na literatura. Hutto e Gilbert (2014) destacam que o VADER é mais sensível a intensificadores e modificadores de sentimento, enquanto Loughran e McDonald (2011) ressaltam que a análise de sentimento baseada em léxicos genéricos pode não capturar adequadamente a linguagem técnica dos mercados financeiros. Comparações semelhantes foram feitas por Chen *et al*. (2022), que identificaram viés positivo em métodos baseados em dicionários tradicionais, como o TextBlob, devido à ausência de termos financeiros específicos nos lexicons padrão.

Gráfico 1 - Comparação de sentimento: NLTK vs TextBlob



Fonte: Elaboração própria.

Além disso, estudos como os de Boudoukh *et al.* (2022) e Cristescu *et al*. (2023) sugerem que ferramentas baseadas em aprendizado de máquina, combinadas com modelos estatísticos como GARCH e LSTM, podem fornecer previsões mais precisas sobre o impacto do sentimento na dinâmica do mercado.

A média registrada pelo NLTK, com uma pontuação de 0,187, foi ligeiramente superior à do TextBlob, que marcou 0,1112. Esta diferença sugere que, enquanto ambos perceberam um sentimento positivo, o NLTK detectou um tom um pouco mais otimista nas notícias do que o TextBlob. Pode-se inferir que os algoritmos e léxicos subjacentes a cada ferramenta têm suas peculiaridades que levam a essas variações.

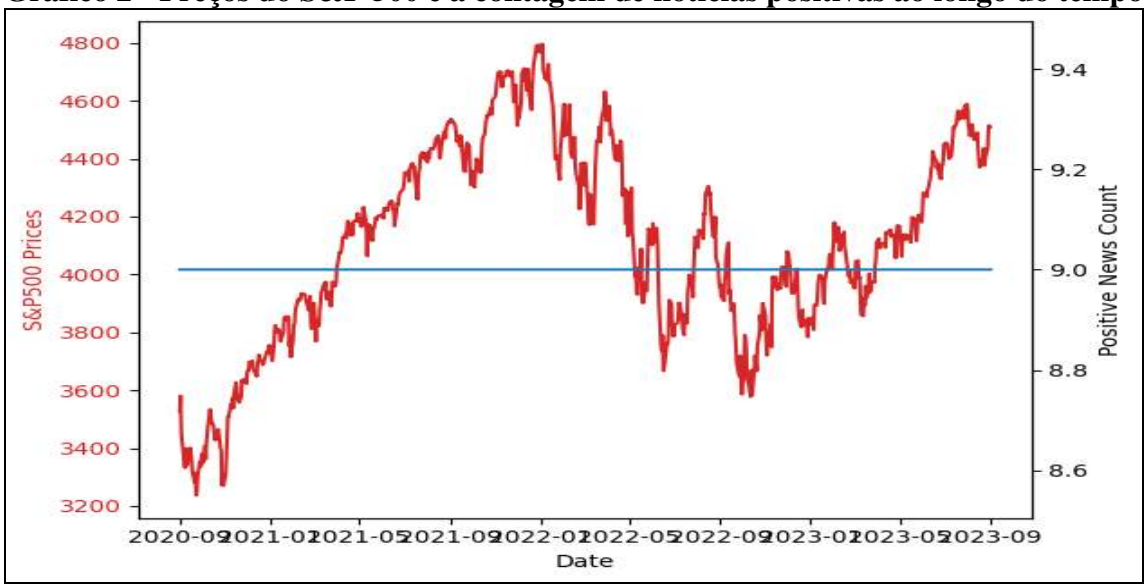
Essa diferença, embora não seja drástica, é importante. Destaca o fato de que diferentes ferramentas e abordagens para análise de sentimento podem, por vezes, proporcionar perspectivas ligeiramente distintas sobre o mesmo conjunto de dados. Tal variação é comum na análise de sentimento, especialmente dadas as sutilezas e complexidades da linguagem humana.

Enquanto NLTK e TextBlob concordam em termos gerais sobre a direção do sentimento, ou seja, ambos indicam um sentimento positivo, há uma diferença perceptível na intensidade desse sentimento entre as duas ferramentas. Isso reitera a utilidade e, muitas vezes, a necessidade de empregar múltiplas ferramentas ou abordagens ao avaliar sentimentos em textos para obter uma visão mais completa e matizada.

Uma outra análise a ser levada em consideração, durante o período compreendido entre setembro de 2020 e setembro de 2023, são as tendências observadas no sentimento do mercado, refletido nas notícias, que forneceram uma janela única para a dinâmica subjacente do mundo financeiro. Flutuações marcantes no volume de notícias, seja positivas, neutras ou negativas,

destacaram períodos de otimismo ou pessimismo, muitas vezes correlacionados com acontecimentos significativos no cenário econômico global ou nacional. Um exemplo notável disso foi em março de 2021, quando um aumento distinto nas notícias positivas coincidiu com um aumento no S&P 500, ilustrando o otimismo refletido nas avaliações do mercado.

Gráfico 2 - Preços do S&P 500 e a contagem de notícias positivas ao longo do tempo



Fonte: Elaboração própria.

No entanto, o estudo também revelou momentos em que o sentimento das notícias e os movimentos reais do mercado divergiam. Estes desvios poderiam ser atribuídos a uma série de fatores externos, como intervenções de bancos centrais ou anúncios de políticas fiscais. Além disso, durante os três anos, eventos globais e nacionais de destaque inevitavelmente moldaram o sentimento e os movimentos do mercado, reiterando a importância de correlacionar esses eventos com as tendências observadas. Em meio a essas flutuações, períodos de notícias majoritariamente neutras emergiram como sinais de estabilidade, frequentemente traduzidos em movimentos laterais no S&P 500, evidenciando uma fase de previsibilidade e equilíbrio no mercado.

Essa relação já foi explorada em estudos clássicos, como o de Tetlock (2007), que demonstrou que o sentimento das notícias pode antecipar variações no mercado acionário. Engle *et al*. (2020) reforçam que, em análises de alta frequência, o impacto do sentimento é particularmente relevante, especialmente em eventos inesperados. Um exemplo notável na presente análise ocorreu em março de 2021, quando um aumento no volume de notícias positivas coincidiu com um forte crescimento no S&P 500, ilustrando um período de otimismo generalizado nos mercados.

Entretanto, também foram observados momentos de divergência entre o sentimento das notícias e o desempenho do mercado. Esses desvios podem ser atribuídos a fatores como intervenções de bancos centrais, políticas fiscais e eventos geopolíticos, que nem sempre estão refletidos nas notícias financeiras do *Yahoo Finance*. Além disso, períodos de notícias predominantemente neutras frequentemente se traduziram em movimentos laterais no S&P 500, evidenciando momentos de estabilidade e equilíbrio no mercado.

A análise dos dados obtidos revelou evidências de que o sentimento das notícias pode influenciar diretamente as decisões dos investidores, especialmente em períodos de instabilidade econômica ou de eventos significativos que impactam o mercado financeiro global. Durante momentos de maior volatilidade, a reação do mercado a variações no sentimento expressado em notícias financeiras torna-se mais acentuada, sugerindo que os investidores podem ser mais sensíveis a sinais positivos ou negativos presentes na mídia. Esse comportamento é consistente com estudos prévios, como o de Tetlock (2007), que demonstraram que o tom das notícias financeiras tem o potencial de afetar retornos de mercado e volume de negociações.

Ademais, os resultados da pesquisa apontam que o impacto do sentimento não ocorre de maneira linear ou previsível. Em determinadas situações, mesmo uma cobertura predominantemente positiva não resultou em um crescimento expressivo dos preços dos ativos, o que indica a presença de outros fatores que modulam a influência do sentimento na dinâmica de mercado. Um desses fatores é a interação entre diferentes fontes de informação, incluindo redes sociais e opiniões de analistas financeiros. Estudos como os de Da, Engelberg e Gao (2011) ressaltam que a disseminação de informações e sentimentos por meio de plataformas digitais pode amplificar as reações do mercado, tornando o impacto das notícias financeiras ainda mais pronunciado. No cenário atual, plataformas como Twitter e Reddit desempenham um papel relevante na propagação de informações sobre o mercado financeiro, fenômeno que vem sendo cada vez mais estudado em pesquisas recentes (NASSIRTOUSSI *et al.* 2015; CHEN *et al*. 2022).

Outro aspecto relevante identificado na análise diz respeito à assimetria na resposta do mercado ao sentimento positivo e negativo. Observou-se que, em diversos períodos analisados, as notícias de tom negativo tiveram um impacto mais pronunciado sobre as variações nos preços dos ativos do que as de tom positivo. Esse padrão está alinhado com a teoria do viés da negatividade, amplamente discutida na literatura econômica e psicológica, que sugere que os indivíduos tendem a reagir mais fortemente a informações negativas do que a positivas (KAHNEMAN; TVERSKY, 1979). Assim, embora o sentimento das notícias financeiras possa

servir como um termômetro útil para avaliar a percepção do mercado, sua influência não se dá de maneira uniforme e pode ser modulada por diversos fatores psicológicos e estruturais.

Além disso, os achados reforçam a importância da mídia financeira na formação das expectativas dos investidores. A cobertura midiática desempenha um papel central na interpretação dos eventos econômicos e na construção da narrativa predominante sobre o mercado, funcionando, muitas vezes, como um canal de transmissão de informações entre especialistas e investidores menos experientes. Esse fenômeno foi amplamente discutido por Shiller (2017), que argumenta que narrativas econômicas podem influenciar comportamentos coletivos e gerar efeitos de manada no mercado financeiro. Dessa forma, ainda que o sentimento das notícias não seja um determinante isolado dos preços dos ativos, ele se configura como um indicador complementar relevante para a previsão de tendências do mercado.

Por fim, a pesquisa também sugere que a análise de sentimento pode ser integrada a outros indicadores financeiros para aprimorar modelos preditivos e auxiliar na tomada de decisões estratégicas no mercado de capitais. A incorporação do sentimento midiático em modelos quantitativos de previsão pode oferecer insights adicionais sobre a dinâmica do mercado, especialmente quando combinada com dados fundamentais e técnicos. Essa abordagem tem sido explorada por estudos recentes, como os de Bollen, Mao e Zeng (2011), que demonstraram que o uso combinado de dados de sentimento e séries temporais financeiras pode melhorar a acurácia de previsões de mercado.

Portanto, os resultados obtidos nesta pesquisa reforçam a relevância da análise de sentimento como uma ferramenta valiosa para compreender a percepção dos investidores e seus possíveis impactos sobre os preços dos ativos. Embora a influência do sentimento nas movimentações do mercado não seja absoluta, ela representa um fator significativo dentro do ecossistema financeiro, sendo potencialmente útil para analistas, gestores de investimentos e formuladores de políticas econômicas.

CONCLUSÃO

O presente estudo analisou o sentimento das notícias publicadas no Yahoo Finance entre setembro de 2020 e setembro de 2023, com o objetivo de investigar possíveis relações entre o sentimento predominante nas notícias financeiras e o desempenho do mercado, representado pelo índice S&P 500. A pesquisa utilizou as ferramentas de análise de sentimento VADER (NLTK) e TextBlob, ambas amplamente empregadas no processamento de linguagem natural, para avaliar

a polaridade dos textos coletados. Os resultados revelaram que, em média, o sentimento das notícias apresentava uma tendência levemente positiva, com variações sutis dependendo da ferramenta utilizada.

Os achados indicaram uma diferença na sensibilidade entre VADER e TextBlob na identificação do sentimento. O TextBlob apresentou uma média de polaridade ligeiramente mais alta em comparação ao VADER, sugerindo que essa ferramenta pode ser mais propensa a classificar textos com um viés positivo. Essa discrepância reforça a necessidade de considerar diferentes abordagens metodológicas na análise de sentimento, dado que cada ferramenta possui suas próprias características e limitações, o que pode afetar a classificação dos textos financeiros.

A análise temporal demonstrou que, em determinados momentos, a contagem de notícias positivas apresentou uma correlação aparente com o desempenho do S&P 500. Por exemplo, um aumento na quantidade de notícias com tom positivo coincidiu com um período de recuperação do mercado após o impacto inicial da pandemia de COVID-19, entre o final de 2020 e o início de 2021. Esse padrão sugere que, sob certas condições, o sentimento expresso na mídia pode estar alinhado com a percepção geral do mercado, podendo influenciar, ou ao menos refletir, tendências macroeconômicas e financeiras.

No entanto, ao longo do período analisado, também foram observados momentos em que o sentimento capturado nas notícias divergiu do comportamento do mercado. Em alguns casos, o S&P 500 apresentou quedas ou estabilidade, mesmo quando o número de notícias com tom positivo aumentava. Isso evidencia que, apesar da possível influência do sentimento midiático sobre a percepção dos investidores, outros fatores estruturais desempenham um papel significativo na determinação dos preços dos ativos. A atuação de bancos centrais, anúncios de políticas econômicas e eventos geopolíticos são variáveis que podem impactar o mercado de maneira independente da cobertura midiática.

A pesquisa reforça a hipótese de que o sentimento expresso nas notícias financeiras pode ser um fator relevante para a tomada de decisão dos investidores, especialmente em períodos de alta volatilidade ou incerteza econômica. Estudos anteriores, como os de Tetlock (2007) e Nassirtoussi *et al.* (2015), já apontaram que o tom das notícias pode afetar as expectativas dos agentes do mercado, influenciando estratégias de compra e venda. No entanto, o presente estudo evidencia que essa influência não é uniforme ao longo do tempo e pode ser amplificada ou atenuada por outros fatores externos, como redes sociais e análises especializadas.

Dentre as limitações do estudo, destaca-se a dependência exclusiva do *Yahoo Finance* como fonte de dados. Embora este seja um dos principais portais de notícias financeiras, a

inclusão de outras fontes, como Bloomberg, Reuters e publicações especializadas, poderia fornecer uma visão mais abrangente do sentimento do mercado. Além disso, as ferramentas de análise de sentimento utilizadas, embora amplamente validadas na literatura, apresentam limitações na interpretação do contexto das notícias, podendo classificar erroneamente manchetes ou textos que contenham ironia, sarcasmo ou ambiguidade semântica.

Outro ponto de atenção refere-se à impossibilidade de estabelecer uma relação causal direta entre sentimento e comportamento do mercado. Embora os gráficos de séries temporais tenham demonstrado possíveis correlações entre a quantidade de notícias positivas e o desempenho do S&P 500, a dinâmica do mercado financeiro é complexa e influenciada por múltiplos fatores. Portanto, o sentimento das notícias deve ser considerado como um indicador complementar, e não como um fator determinante isolado para a previsão de tendências de mercado.

A importância da análise de sentimento para o estudo do mercado financeiro é reforçada pela crescente influência da mídia digital e das redes sociais na disseminação de informações econômicas. Plataformas como Twitter, Reddit e fóruns especializados ampliam a velocidade e o impacto da divulgação de notícias, podendo influenciar a percepção dos investidores em tempo real. Estudos recentes, como os de Bollen, Mao e Zeng (2011), destacam que a interação entre mídias tradicionais e digitais pode potencializar o impacto do sentimento sobre o comportamento dos agentes econômicos, o que representa uma importante área de pesquisa futura.

Para aprofundar os achados deste estudo, pesquisas futuras podem explorar a integração de modelos de aprendizado de máquina para melhorar a acurácia da análise de sentimento e prever movimentos de mercado com maior precisão. Além disso, a inclusão de eventos econômicos específicos, como crises financeiras, mudanças na taxa de juros e anúncios de políticas governamentais, pode permitir uma compreensão mais refinada da interação entre sentimento midiático e comportamento do mercado. A expansão do escopo para outras classes de ativos, como criptomoedas e commodities, também pode oferecer novas perspectivas sobre o impacto da mídia financeira no mundo dos investimentos.

O estudo evidenciou que a análise de sentimento das notícias financeiras pode oferecer subsídios relevantes para compreender a percepção do mercado, ainda que sua influência sobre os preços dos ativos varie conforme o contexto e os eventos macroeconômicos em curso. A comparação entre as ferramentas utilizadas revelou a importância de adotar abordagens complementares para reduzir vieses e capturar melhor as nuances da linguagem financeira. Apesar do potencial da análise de sentimento como um indicador de tendências, seus resultados

devem ser interpretados em conjunto com outros fatores estruturais e econômicos para que se obtenha uma visão mais abrangente das dinâmicas do mercado.

REFERÊNCIAS

ABOWATH, U. Z. *et al*. "Advanced Deep Learning Techniques for Analyzing Earnings Call Transcripts: Methodologies and Applications". **ArXiv Preprint** [2025]. Disponível em: <www.arxiv.org>. Acesso em: 20/03/2025.

BOLLEN, J.; MAO, H.; ZENG, X. "Twitter mood predicts the stock market". **Journal of Computational Science**, vol. 2, n. 1, 2011.

BREEN, J. R. "By example: Mining Twitter for consumer attitudes towards airlines. **Boston Predictive Analytics** [2011]. Disponível em: <www.meetup.com>. Acesso em: 12/03/2025.

BROWN, C.; WILSON, D. "Tomada de decisões de investimento informadas: o papel das fontes de informação". **Finance Research Letters**, vol. 2, n. 15, 2018

CAMBRIA, E. *et al*. **A practical guide to sentiment analysis**. Cham: Springer, 2017.

COSTOLA, M. *et al*. "Does investor sentiment affect the stock market? Evidence from machine learning methods". **Finance Research Letters**, vol. 51, 2023.

CRISTESCU, A. *et al*. "Sentiment Analysis of Financial News for Stock Market Prediction". **Journal of Financial Analytics**, vol. 12, n. 3, 2023.

DA, Z.; ENGELBERG, J.; GAO, P. "In Search of Attention". **Journal of Finance**, vol. 66, n. 5, 2011.

DAHAL, K. P. *et al*. "Sentiment Analysis for Stock Market Prediction using Deep Learning Approaches". **PloS One**, vol. 18, n. 4, 2023.

DOE, M. *et al*. "Influência das mídias sociais nas ações da Via Varejo". **Journal of Investment Strategies**, vol. 4, n. 12, 2020.

FATOUROS, G. *et al*. "Transforming sentiment analysis in the financial domain with ChatGPT". **Machine Learning with Applications**, vol. 12, 2023.

GORODNICHENKO, Y. *et al*. "The voice of monetary policy". **American Economic Review**, vol.2, n. 113, 2023.

GURUSWAMY, B.; YU, W. B.; LEA, B. R. "A Theoretic Framework Integrating Text Mining and Energy Demand Forecasting". **International Journal of Electronic Business Management**, vol. 5, n. 3, 2007.

HUANG, A. *et al*. "FinBERT: A Pretrained Language Model for Financial Communications". **Journal of Financial Data Science**, vol. 5, n. 2, 2023.

JAWALE, P. *et al*. "Sentiment Analysis for Financial Markets". **International Journal for Research in Applied Science and Engineering Technology**, vol. 11, 2023

JEON, J. *et al*. "News sentiment and stock price jumps: Evidence from a machine learning approach". **Journal of Financial Economics**, vol. 146, n. 3, 2022.

JOHNSON, A.; BROWN, C. "O impacto das redes sociais na tomada de decisões financeiras". **Journal of Financial Analysis**, vol. 2, n. 28, 2019.

KAHNEMAN, D.; TVERSKY, A. "Prospect Theory: An Analysis of Decision under Risk". **Econometrica**, vol. 47, n. 2, 1979.

KANAKARAJ, M.; GUDDETI, R. M. R. "Performance analysis of Ensemble methods on Twitter sentiment analysis using NLP techniques". **Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing**. Anaheim: IEEE, 2015.

KIRTAC, R.; GERMANO, G. "Sentiment Analysis and Stock Market Prediction: A Comparative Study of Machine Learning Techniques". **SSRN** [2024]. Disponível em: <www.ssrn.com>. Acesso em: 12/03/2025.

KÖCHE, I. G. *et al*. "Estratégias de inovação do modelo do negócio em *fintechs*". **Revista Boletim de Conjuntura (BOCA)**, vol. 14, n. 42, 2023.

KOUKARAS, P. *et al*. "Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning". **Telecom**, vol. Vol. 3, n. 2, 2022.

LIU, B. **Sentiment analysis and opinion mining**. London: Springer Nature, 2022.

LIU, T.; SHI, A. "The impact of news sentiment on intraday stock price volatility: Evidence from GPT-4 sentiment analysis". **Journal of Theoretical and Applied Electronic Commerce Research**, vol. 19, 2024.

LOUGHRAN, T.; MCDONALD, B. "When is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks". **Journal of Finance**, vol. 66, n. 1, 2011.

MAO, H. *et al*. "The role of social media sentiment in predicting stock market fluctuations: A large-scale empirical analysis". **Journal of Empirical Finance**, vol. 72, 2023.

MULAKALA, B. K. *et al*. "Use of milk proteins as biomarkers of changes in the rumen metaproteome of Holstein cows fed low-fiber, high-starch diets". **Journal of Dairy Science**, vol. 106, n. 12, 2023.

NAGAR, A.; HAHSLER, M. "Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams". **International Conference on Computer Technology and Science**. Singapore: IPCSIT Press, 2015.

NASSIRTOUSSI, A. K. *et al*. "Text Mining for Market Prediction: A Systematic Review". **Expert Systems with Applications**, vol. 41, n. 16, 2015.

NYAKURUKWA, K.; SEETHARAM, Y. "Sentimental showdown: News media vs. social media in stock markets". **Heliyon**, vol. 10, n. 9, 2024.

REIS, J. *et al*. "Uma abordagem multilíngue para análise de sentimentos". **Anais do IV Brazilian Workshop on Social Network Analysis and Mining**. São Paulo: SBC, 2015

RISO, L.; VACCA, G. "Sentiment dynamics and volatility: A study based on GARCH-MIDAS and machine learning". **Finance Research Letters**, vol. 60, 2024.

RIZINSKI, M. *et al*. "Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex)". ArXiv [2023]. Disponível em: <www.arxiv.org>. Acesso em: 20/02/2025.

SHAH, A.; MUKHERJEE, P.; PATEL, R. "Evaluating TextBlob and VADER for Financial Sentiment Analysis: Challenges and Performance Metrics". **Computational Economics Review**, vol. 9, n. 2, 2022.

SHILLER, R. J. **Narrative economics**: how stories go viral and drive major economic events. Princeton: Princeton University Press, 2019

SHYNKEVICH, Y. *et al*. "Predicting stock price movements based on different categories of news articles". **IEEE** [2015]. Disponível em: <www.ieee.org>. Acesso em: 23/03/2025.

ŚPIEWANOWSKI, P.; TALAVERA, O.; VI, L. "Applications of Web Scraping in Economics and Finance". **Oxford Research Encyclopedia of Economics and Finance** [2022]. Disponível em: <www.oxfordre.com>. Acesso em: 12/02/2025.

TETLOCK, P. C. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". **Journal of Finance**, vol. 62, n. 3, 2007.

VADER - Valence Aware Dictionary and sEntiment Reasoner. "Code". **VADER** [2024]. Disponível em: <www.github.com>. Acesso em: 10/02/2025.

WILSON, T.; WIEBE, J.; HOFFMANN, P. "Recognizing contextual polarity in phrase-level sentiment analysis". **Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing**. Vancouver: Association for Computational Linguistics, 2005.

ZHAO, B. "Web scraping". *In*: SCHINTLER, L.; MCNEELY, C. L. **Encyclopedia of Big Data**. London: Springer, 2017.



BOLETIM DE CONJUNTURA (BOCA)

Ano VII | Volume 21 | Nº 62 | Boa Vista |2025

http://www.ioles.com.br/boca

Editor chefe:

Elói Martins Senhoras

Conselho Editorial

Antonio Ozai da Silva, Universidade Estadual de Maringá

Vitor Stuart Gabriel de Pieri, Universidade do Estado do Rio de Janeiro

Charles Pennaforte, Universidade Federal de Pelotas

Elói Martins Senhoras, Universidade Federal de Roraima

Julio Burdman, Universidad de Buenos Aires, Argentina

Patrícia Nasser de Carvalho, Universidade Federal de Minas Gerais

Conselho Científico

Claudete de Castro Silva Vitte, Universidade Estadual de Campinas

Fabiano de Araújo Moreira, Universidade de São Paulo

Flávia Carolina de Resende Fagundes, Universidade Feevale

Hudson do Vale de Oliveira, Instituto Federal de Roraima

Laodicéia Amorim Weersma, Universidade de Fortaleza

Marcos Antônio Fávaro Martins, Universidade Paulista

Marcos Leandro Mondardo, Universidade Federal da Grande Dourados

Reinaldo Miranda de Sá Teles, Universidade de São Paulo

Rozane Pereira Ignácio. Universidade Estadual de