

# Relatório - Projeto IA

## Análise de Sentimento na Precificação de Ativos na Bolsa Brasileira

**Matheus Barbosa Ferrari (RA: 10400837), Felipe da Silva Morishita Garbi (RA: 10401362)**

Graduação em Sistemas de Informação – Faculdade de Computação e Informática (FCI) –  
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{matheus.ferrari, felipe.garbi}@mackenzie.br

**Resumo.** *Por meio da análise de notícias, o presente trabalho busca explorar e aplicar técnicas de Processamento de Linguagem Natural (PLN) para avaliar o impacto do sentimento da mídia sobre o comportamento dos investidores e a volatilidade dos preços das ações na Bolsa de Valores Brasileira (B3). O estudo envolve o desenvolvimento de um modelo de PLN para classificar as notícias em sentimentos positivos, negativos ou neutros, e investigar a correlação entre essas variações sentimentais e os movimentos dos preços no mercado. Para validar a eficácia do modelo, será aplicada uma análise quantitativa, utilizando modelos de regressão e outras metodologias avançadas para identificar padrões de influência das notícias nas flutuações do mercado acionário brasileiro. O objetivo do projeto é fornecer uma visão mais aprofundada de como fatores subjetivos, como o sentimento midiático, podem influenciar o comportamento do mercado financeiro, contribuindo tanto para a tomada de decisões mais precisas por parte dos investidores quanto para a ampliação do conhecimento acadêmico sobre o impacto da mídia no mercado financeiro.*

### Introdução

Nos últimos anos, o uso de técnicas de Inteligência Artificial, em especial o Processamento de Linguagem Natural (PLN), tem crescido exponencialmente em diversas áreas, incluindo a financeira. A análise de notícias e a extração de sentimentos se tornaram ferramentas valiosas para a compreensão do comportamento dos investidores e as oscilações dos preços no mercado. No contexto do mercado brasileiro, onde a volatilidade é uma característica marcante, compreender a relação entre o conteúdo divulgado pela mídia e as reações dos investidores pode proporcionar insights cruciais para tomadas de decisão mais assertivas. A forma como essas informações são interpretadas pode gerar variações significativas no sentimento dos investidores, o que, por sua vez, influencia diretamente os movimentos de compra e venda no mercado de ações (Tetlock, 2015; Barber e Odean, 2007).

O presente projeto justifica-se pela necessidade de ferramentas que auxiliem a prever as reações do mercado, oferecendo uma abordagem que alia a tecnologia com as ciências econômicas. Segundo Liu (2012), a análise de sentimento visa extrair e classificar opiniões subjetivas em textos, distinguindo emoções positivas, negativas ou neutras. No mercado financeiro, essa técnica vem sendo aplicada com o objetivo de captar antecipadamente o impacto de eventos informacionais sobre o comportamento dos investidores. Estudos como o de Bollen, Mao e Zeng (2011) evidenciam que indicadores de sentimento derivados de mídias sociais e notícias podem atuar como preditores relevantes para a volatilidade dos mercados. Com a crescente complexidade dos mercados e a influência da mídia sobre os investidores, utilizar técnicas de PLN para analisar grandes volumes de notícias é uma forma eficiente de identificar tendências e padrões de comportamento. Ao automatizar o processo de classificação de sentimentos, podemos oferecer uma análise robusta que complementa as abordagens tradicionais de estudo de mercado.

O objetivo deste trabalho é desenvolver um modelo de PLN capaz de identificar o sentimento predominante em notícias financeiras e correlacioná-lo com a volatilidade dos preços das ações na Bolsa de Valores Brasileira (B3). Isso permitirá prever possíveis oscilações no mercado com base na percepção dos investidores em relação aos acontecimentos econômicos. A opção de utilizar técnicas de Machine Learning e Deep Learning, aplicadas ao contexto da análise de sentimentos, foi escolhida por sua eficiência em lidar com grandes volumes de dados textuais e pela capacidade de produzir resultados precisos e acionáveis.

## **O Problema**

O foco central deste estudo consiste em entender como o sentimento veiculado pelas notícias financeiras influencia o comportamento dos investidores e, consequentemente, a volatilidade dos preços das ações na B3. As oscilações de mercado frequentemente refletem reações emocionais e decisões impulsionadas por informações divulgadas pela mídia. Assim, a dificuldade central do problema reside na extração e classificação automática de sentimentos em textos jornalísticos, e sua posterior quantificação de forma robusta. Essa tarefa envolve desafios técnicos, como ambiguidade linguística, ironia e subjetividade do discurso (Pang & Lee, 2008). Além disso, correlacionar esses sentimentos com dados numéricos de mercado requer o uso de técnicas estatísticas e de aprendizado de máquina que considerem não apenas a acurácia da classificação, mas também a significância da associação com os movimentos observados. O avanço nessa área pode contribuir para a criação de ferramentas preditivas mais precisas e fundamentadas, beneficiando tanto analistas financeiros quanto investidores institucionais e individuais.

## Aspectos Éticos no uso da IA

A utilização de Inteligência Artificial demanda cuidados éticos significativos:

- Caso sejam utilizados dados sensíveis ou provenientes de fontes que possam identificar indivíduos ou empresas, estes devem ser anonimizados para preservar a privacidade.
- Os modelos de classificação podem apresentar vieses decorrentes dos dados de treinamento. É fundamental realizar uma análise crítica e transparente dos resultados, indicando limitações e possíveis distorções.
- As previsões e análises geradas pelo sistema devem ser interpretadas como uma ferramenta auxiliar e não substituem a avaliação crítica e especializada dos investidores.
- Considerando o potencial impacto econômico, o projeto deve enfatizar a importância de práticas éticas e responsáveis na implementação e divulgação dos resultados.

## O Dataset

O projeto utilizará um dataset composto por notícias financeiras coletadas a partir de fontes públicas e confiáveis (Valor Econômico, InfoMoney e Exame), por meio de técnicas de web scraping ou RSS Feeds quando disponíveis. Além disso, os dados passarão por um processo de pré-processamento que incluirá:

- Remoção de duplicidades, normalização de texto e eliminação de stopwords.
- Definição semi-automatizada dos sentimentos (positivo, negativo, neutro) para criação de um conjunto de dados rotulado.
- A análise exploratória, ou seja utilização de bibliotecas como pandas, matplotlib e seaborn para entender a distribuição dos dados, identificar padrões e possíveis inconsistências.
- E armazenamento seguro, para garantia de que os dados sensíveis estejam devidamente anonimizados e armazenados de acordo com as diretrizes da LGPD.

## Metodologia

1. **Coleta de Dados:** Extração de notícias financeiras a partir de APIs ou web scraping de portais de notícias.
2. **Pré-processamento:** Limpeza e normalização dos textos, além da tokenização e remoção de stopwords.
3. **Anotação dos Dados:** Definição dos sentimentos por meio de um processo de anotação manual ou semiautomático.
4. **Desenvolvimento do Modelo:** Utilização de técnicas de Machine Learning (por exemplo, SVM, Random Forest) ou Deep Learning para classificação de sentimentos. Além disso, Implementação de modelos de regressão para correlacionar os índices sentimentais com a volatilidade dos preços.
5. **Validação e Testes:** Divisão dos dados em conjuntos de treino e teste, avaliação de métricas como acurácia, precisão, recall e F1-score para a classificação, e análise estatística para a regressão.
6. **Análise dos Resultados:** Interpretação dos resultados obtidos, discussão sobre a eficácia do modelo e avaliação da correlação entre o sentimento e os movimentos do mercado.

## Resultados

A análise das notícias coletadas entre 1º e 26 de maio de 2025 revela um cenário claramente dominado pelo teor negativo. No gráfico de distribuição de sentimentos, observa-se que as notícias classificadas como negativas correspondem à maior parte do corpus, seguidas pelas neutras e, em menor proporção, pelas positivas. Essa predominância de conteúdo negativo sugere que o noticiário financeiro, nesse período, privilegiou alertas e avisos de riscos, reforçando a sensação de incerteza no mercado.

Apesar desse viés pessimista, os scores médios de sentimento para todas as três categorias permanecem acima de 0,4, indicando que mesmo as notícias atribuídas a “neutro” e “positivo” continham um grau razoável de confiança pelo modelo FinBERT-PT-BR. Esse resultado mostra que o classificador atribui com segurança seu julgamento, ainda que nuances emocionais pudessem ser sutis em manchetes mais descritivas.

Quando observamos a evolução diária do volume de notícias, fica claro que o número de manchetes negativas oscila de forma marcante, com picos em dias de grande repercussão (por

exemplo, 1º e 2 de maio, com mais de 40 notícias negativas em um único dia). As flutuações no número de notícias neutras e positivas, embora menores, acompanham de perto essas variações, sugerindo que, em momentos de crise ou divulgação de dados importantes, o mercado não só aumenta sua cobertura jornalística, mas também intensifica a tonalidade emocional.

Por fim, ao confrontar o sentimento médio diário com o retorno percentual da ação PETR4, é possível encontrar indícios de que variações no tom das notícias podem acompanhar ou mesmo anteceder oscilações nos preços. Embora o padrão não seja estritamente linear, há trechos em que um aumento do pessimismo diário coincide com quedas no retorno da ação, assim como momentos de sentimento mais positivo se alinham a movimentos de valorização. Esses achados sugerem que, mesmo de forma sutil, o conteúdo emocional das notícias exerce influência sobre o comportamento de curto prazo no pregão.

## **Conclusão**

Os resultados indicam que o sentimento presente nas notícias pode desempenhar um papel relevante na dinâmica do mercado acionário. A relação entre o sentimento das notícias e o retorno da ação PETR4, ainda que não perfeitamente linear, sugere potencial de uso da análise de sentimento como ferramenta complementar na análise de investimentos. Em especial, os picos de sentimento coincidem com variações relevantes no preço da ação, o que reforça a ideia de que o conteúdo emocional das notícias pode influenciar as decisões dos investidores.

Este estudo, portanto, contribui para a compreensão do impacto qualitativo das informações financeiras sobre o mercado, apontando para futuras aplicações de modelos de Processamento de Linguagem Natural (PLN) como suporte na tomada de decisão de analistas, investidores e sistemas automatizados. Trabalhos futuros poderão ampliar a base de dados, incluir outras ações, e explorar modelos preditivos mais robustos, como LSTM ou Transformers fine-tunados para o mercado financeiro brasileiro.

## **Endereço GitHub e Endereço do vídeo no Youtube**

<https://github.com/Matheus-B-Ferrari/AnaliseDeSentimentosB3>

<https://youtu.be/O67WE3y0SyY>

## Bibliografia

LIU, Bing. Sentiment analysis and opinion mining. **Synthesis Lectures on Human Language Technologies**, v. 5, n. 1, p. 1-167, 2012. Disponível em <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>. Acesso em: 21 de Setembro de 2024

BOLLEN, Johan; MAO, Huina; ZENG, Xiao-Jun. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1-8, 2011. Disponível em <https://arxiv.org/abs/1010.3003>. Acesso em: 21 de Setembro de 2024

PANG, B.; LEE, L. Opinion mining and sentiment analysis. **Foundations and Trends in Information Retrieval**, v. 2, n. 1-2, p. 1-135, 2008. Disponível em <https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>. Acesso em: 21 de Setembro de 2024

TETLOCK, Paul C. The role of media in finance. **Handbook of media economics**, Vol. 1. North-Holland, 2015. p. 701-721. Disponível em <https://doi.org/10.1016/B978-0-444-63685-0.00018-8>. Acesso em: 21 de Setembro de 2024

BARBER, Brad M.; ODEAN, Terrance. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. **Review of Financial Studies**, v. 21, n. 2, p. 785-818, 24 dez. 2007. Disponível em: <https://doi.org/10.1093/rfs/hhm079>. Acesso em: 17 de Novembro 2024.

SANTOS, Lucas L.; BIANCHI, Reinaldo A. C.; COSTA, Anna H. R.. FinBERT-PT-BR: Análise de Sentimentos de Textos em Português do Mercado Financeiro. **BRAZILIAN WORKSHOP ON ARTIFICIAL INTELLIGENCE IN FINANCE (BWAIF)**, 2023. Disponível em: <https://sol.sbc.org.br/index.php/bwaif/article/view/24960>. Acesso em: 15 de Maio de 2025.

IGARASHI, W.; VALDEVIESO, G. S.; IGARASHI, D. C. C. Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações / Sentiment analysis and technical indicators: a correlation analysis of asset prices and polarity of stock market news. **Brazilian Journal of Business**. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BJB/article/view/24482>. Acesso em: 24 de Março de 2025.

FALCÃO, Matheus Cabral dos Santos. Análise de Sentimento de Notícias do Mercado Financeiro, **Universidade Federal do Rio de Janeiro, Escola Politécnica**, 2020. Disponível em: <http://repositorio.poli.ufrj.br/monografias/monopoli10031707.pdf>. Acesso em: 24 de Março de 2025.

FERREIRA, Emanuel Elias. Análise do Impacto de Notícias na Previsão do Índice Bovespa. **Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais**, 2023. Disponível em: <https://repositorio.ifmg.edu.br/server/api/core/bitstreams/8437cbde-1d14-479f-8218-7b0e82aaf4e7/content>. Acesso em: 24 de Março de 2025.

LIU, C. et al. Large language models and sentiment analysis in financial markets: a review, datasets, and case study. **IEEE Access**, 2024. Disponível em: <https://ieeexplore.ieee.org/document/10638546>. Acesso em: 6 abr. 2025.