

Relatório Projeto IA

Análise de Sentimento de Notícias Financeiras e seu Impacto nos Preços das Ações na Bolsa de Valores Brasileira

Matheus B. Ferrari (RA: 10400837)

Felipe S. Garbi (RA: 10401362)

Resumo

Por meio da análise de notícias financeiras, 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 financeiras 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.

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. 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 reside em extrair, classificar e quantificar esses sentimentos de forma automatizada, correlacionando-os com os movimentos do mercado, o que pode oferecer uma ferramenta valiosa para a previsão e análise financeira.

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 de fontes públicas e confiáveis, como portais de economia e seções financeiras de jornais digitais. Os dados passarão por um processo de pré-processamento que incluirá:

- Remoção de duplicidades, normalização de texto e eliminação de stopwords.
- Definição manual ou 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

Coleta de Dados: Extração de notícias financeiras a partir de APIs ou web scraping de portais de notícias.

Pré-processamento: Limpeza e normalização dos textos, além da tokenização e remoção de stopwords.

Anotação dos Dados: Definição dos sentimentos por meio de um processo de anotação manual ou semiautomático.

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.

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.

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 Esperados

Os resultados esperados deste projeto incluem a classificação eficiente de notícias financeiras com alta acurácia, permitindo a identificação clara de sentimentos positivos, negativos e neutros. Espera-se também encontrar uma correlação significativa entre as variações sentimentais e a volatilidade dos preços das ações na B3, oferecendo insights sobre o comportamento do mercado em resposta a eventos midiáticos. Além disso, o projeto visa desenvolver uma ferramenta de apoio à decisão que auxilie investidores e analistas financeiros a interpretar o impacto de notícias no mercado de ações. Por fim, o estudo contribuirá para o avanço acadêmico ao explorar o uso de técnicas de Processamento de Linguagem Natural na análise financeira, potencialmente gerando novos caminhos de pesquisa.

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.

SOHANGIR, Sahar *et al.* Big Data: Deep Learning for financial sentiment analysis. **Journal of Big Data**, v. 5, n. 1, 25 jan. 2018. Disponível em: <https://doi.org/10.1186/s40537-017-0111-6>. Acesso em: 17 de Novembro 2024.

DAY, Min-Yuh; LEE, Chia-Chou. Deep learning for financial sentiment analysis on finance news providers. **International Conference on Advances in Social Networks Analysis and Mining (ASONAM)**, 2016. Disponível em: <https://doi.org/10.1109/asonam.2016.7752381>. Acesso em: 17 de Novembro 2024.

DEVLIN, J. et al. BERT: Pre-training of deep bidirectional Transformers for language understanding. 2018. Disponível em: <http://arxiv.org/abs/1810.04805>. Acesso em: 17 de Novembro 2024.

KUMAR, G. Ravi. et al. Stock Market Analysis Using Deep Learning. **Journal of Financial Markets**, 2023. Disponível em: https://www.researchgate.net/publication/376811635_Stock_Market_Analysis_Using_Deep_Learning. Acesso em: 17 de Novembro 2024.

Github do Projeto:

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