

Aluno(a): André Furlanetti

Orientador(a): Marisa Gomes da Costa

Curso: MBA em Economia, Investimentos e Banking

Correlação entre Sentimento de Notícias e Retornos do ETF BOVA11 no Mercado Acionário Brasileiro

Introdução

O mercado de capitais brasileiro vive uma expansão, marcada pelo aumento de investidores individuais (Beltramini, 2024). Diante do volume de informações de portais de notícias especializados, compreender o impacto da mídia nas decisões de investimento e nas flutuações de preços tornou-se relevante (Igarashi et al., 2021; Hayashi, 2017). Para investidores e gestores que utilizam o ETF BOVA11 como referência do mercado, entender esses efeitos é importante para estratégias informadas.

Nesse cenário, a Análise de Sentimentos [AS], campo do Processamento de Linguagem Natural [PLN], oferece ferramentas computacionais para extrair automaticamente a polaridade (positiva, negativa ou neutra) de textos (Santos et al., 2023). Sua aplicação no domínio financeiro visa transformar dados qualitativos em indicadores quantitativos, auxiliando na compreensão e na estimativa de movimentos do mercado (Schmitz et al., 2022).

Embora a pesquisa internacional mostre correlações entre sentimento e índices de mercado (Souza e Martins, 2024), o contexto brasileiro ainda carece de estudos aprofundados. Conforme Januário et al. (2022), modelos de aprendizado de máquina como Naive Bayes e redes neurais são eficazes para classificar notícias financeiras em português. Assim, a problemática central desta pesquisa é propor uma metodologia que construa, a partir da AS de notícias, um indicador de sentimento correlacionado com os retornos de curto prazo do BOVA11, oferecendo uma ferramenta de análise potencialmente útil para os agentes do mercado.

Objetivo

O objetivo geral deste trabalho é propor e validar uma metodologia para transformar o sentimento extraído de notícias financeiras brasileiras em um indicador quantitativo, avaliando sua correlação e potencial preditivo sobre os retornos diários do ETF BOVA11. A validação da metodologia será realizada por meio de testes estatísticos de correlação (Pearson) e pela avaliação da acurácia dos modelos de classificação de sentimento, utilizando métricas como Precisão, Recall e F1-Score.

Metodologia

Este estudo será conduzido como uma pesquisa de caráter aplicado, com abordagem quantitativa para coleta e análise de dados. A metodologia estrutura-se nas seguintes fases: Obtenção e Tratamento dos Dados

A obtenção dos dados textuais será feita por meio de *web scraping*, utilizando bibliotecas da linguagem Python. O foco da coleta serão manchetes e resumos de notícias de portais financeiros especializados de grande circulação no Brasil, filtrando por termos-chave como "Ibovespa" e "Bolsa" (Schmitz et al., 2022). A coleta abrangerá o período de janeiro de 2022 a dezembro de 2024, visando capturar diferentes cenários de mercado. Os dados históricos de preços do ETF BOVA11 serão coletados de fontes de dados secundárias, como a plataforma Yahoo! Finance (Souza e Martins, 2024).

Após a coleta, os textos passarão por pré-processamento, procedimento essencial para a modelagem dos dados. Este incluirá conversão para minúsculas, remoção de pontuação, URLs e *stopwords*, conforme práticas recomendadas na literatura para preparação de textos para algoritmos de aprendizado de máquina (Januário et al., 2022).

Análise de Sentimentos e Construção do Indicador

Para classificar os sentimentos das notícias (positivo, negativo, neutro), serão implementados e comparados dois algoritmos de aprendizado de máquina supervisionado. A escolha baseia-se na eficácia comprovada destes métodos para o contexto financeiro brasileiro por Januário et al. (2022), que demonstraram seu desempenho superior em relação a abordagens lexicais:

- **Naive Bayes [NB]:** Classificador probabilístico robusto, frequentemente utilizado como baseline na classificação textual.
- **Multilayer Perceptron [MLP]:** Rede neural artificial que apresentou os melhores resultados de acurácia em estudos comparativos com notícias financeiras em português.

A partir das classificações diárias, será construído um Índice de Sentimento Diário [ISD], seguindo abordagens consolidadas (Souza e Martins, 2024; Santos et al., 2023). O índice será calculado pela diferença entre o número de notícias positivas e negativas, normalizada pelo total de notícias do dia.

Análise da Relação com o Mercado

Os retornos diários do BOVA11 serão calculados com base no logaritmo natural da variação dos preços de fechamento. A relação entre o ISD e os retornos do ETF será investigada por meio de duas técnicas estatísticas:

- **Correlação de Pearson:** Para medir a associação linear entre o sentimento e os retornos, tanto no mesmo dia (contemporânea) quanto com um dia de defasagem (avaliando a relação temporal) (Igarashi et al., 2021).
- **Regressão Linear Simples:** Para testar formalmente se o sentimento do dia anterior (ISD_{t-1}) possui capacidade explicativa sobre os retornos do dia corrente ($Retorno_t$) (Souza e Martins, 2024), analisando-se também a significância estatística do modelo através do p-valor.

Validação da Metodologia

A robustez da etapa de classificação de sentimentos será verificada a partir de um conjunto de teste. Uma amostra aleatória, correspondente a aproximadamente 10% do corpus total, será rotulada manualmente para servir como este conjunto. O desempenho dos modelos NB e MLP será avaliado quantitativamente por meio das métricas de Precisão, Recall e F1-Score, além da análise da Matriz de Confusão (Medeiros e Borges, 2019; Beltramini, 2024).

Resultados Esperados

Com a aplicação da metodologia proposta, espera-se alcançar os seguintes resultados, que serão verificados ao longo da pesquisa:

- A verificação de que os modelos de aprendizado de máquina, especialmente a rede neural MLP, apresentam alta acurácia na classificação de sentimentos de notícias financeiras em português, em linha com os achados de Januário et al. (2022).
- A construção de um Índice de Sentimento Diário [ISD] original, derivado de um amplo conjunto de notícias brasileiras.
- A identificação de uma correlação estatisticamente significativa entre o sentimento das notícias e os retornos de curto prazo do BOVA11.
- O desenvolvimento de um modelo quantitativo que possa servir como um indicador auxiliar para a análise e tomada de decisões de investimento, complementando as ferramentas tradicionais.

Cronograma de Atividades

Atividades planejadas	Mês										
	Out/25	Nov/25	Dez/25	Jan/26	Fev/26	Mar/26	Abr/26	Mai/26	Jun/26	Jul/26	Ago/26
Revisão da literatura e consolidação teórica	X	X									
Entrega: Projeto de Pesquisa		X									
Coleta e tratamento dos dados		X	X	X							
Análise dos dados e modelos				X	X						
Entrega: Resultados Preliminares					X						
Discussão dos resultados e Redação						X	X				
Entrega: TCC e Agendamento da Defesa							X				
Preparação da apresentação e materiais								X	X		
Período de Defesa									X	X	X

Referências

Beltramini, J.V. 2024. Análise Comparativa de Grandes Modelos de Linguagem na Avaliação de Sentimentos Sobre o IBOVESPA em Redes Sociais. Monografia em Sistemas de Informação. Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

Hayashi, A.H. 2017. Big Data e computação preditiva para suporte a decisões no mercado financeiro. In: CONTECSI - International Conference on Information Systems and Technology Management, 2017, São Paulo, SP, Brasil. Anais...

Igarashi, W.; Valdevieso, G.S.; Igarashi, D.C.C. 2021. Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. Brazilian Journal of Business 3(1): 470-486.

Imai, B.Y.L.; Garcia, C.M.; Rocha, M.V.; Koerich, A.L.; Britto Jr, A.S.; Barddal, J.P. 2024. Is it Fine to Tune? Evaluating SentenceBERT Fine-tuning for Brazilian Portuguese Text Stream

Classification. In: 2024 IEEE International Conference on Big Data (BigData), 2024. Anais... p. 1765-1774.

Januário, B.A.; Carosia, A.E.O.; Silva, A.E.A.; Coelho, G.P. 2022. Sentiment Analysis Applied to News from the Brazilian Stock Market. IEEE Latin America Transactions 20(3): 512-518.

Medeiros, M.C.; Borges, V.R.P. 2019. Tweet Sentiment Analysis Regarding the Brazilian Stock Market. In: VIII Brazilian Workshop on Social Network Analysis and Mining, 2019, Gramado, RS, Brasil. Anais... p. 71-82.

Santos, L.L.; Bianchi, R.A.C.; Costa, A.H.R. 2023. FinBERT-PT-BR: Análise de Sentimentos de Textos em Português do Mercado Financeiro. In: II Brazilian Workshop on AI in Finance, 2023, João Pessoa, PB, Brasil. Anais... p. 144-155.

Schmitz, M.; Immich, R.; Pessin, G.; Rocha Filho, G.P. 2022. Towards the Categorization of Brazilian Financial Market Headlines. IEEE Latin America Transactions 20(2): 344-351.

Souza, D.M.S.; Martins, O.S. 2024. Brazilian stock market performance and investor sentiment on Twitter. Revista de Gestão 31(1): 18-33.