

# Análise de Correlação Cruzada entre Emoções de Notícias e Preços de Ações com Grandes Modelos de Linguagem

Gabriel Marquezan Pozzati da Luz<sup>1</sup>, Joaquim Vinicius Carvalho Assunção<sup>1</sup>

<sup>1</sup>Universidade Federal de Santa Maria (UFSM)  
CEP 97105-900 – Santa Maria – RS – Brasil

marquezan.luz@acad.ufsm.br, joaquim@inf.ufsm.br

**Abstract.** *The impact of emotions on financial market behavior has been widely studied, but recent technological advances have enabled new approaches. This study analyzes the cross-correlation between emotions extracted from news and stock prices on the Brazilian Stock Exchange. Large Language Models were used to classify emotions in news titles about specific companies, generating one-month time series for both emotions and prices. The analysis revealed varying degrees of correlation between market sentiment and price fluctuations depending on the time lag, suggesting that changes in sentiment may precede market movements. The results highlight the potential of LLMs for forecasting strategies and decision support.*

**Resumo.** *O impacto das emoções no comportamento do mercado financeiro tem sido amplamente estudado, mas os avanços tecnológicos recentes têm possibilitado novas abordagens. Este estudo analisa a correlação cruzada entre emoções extraídas de títulos de notícias e preços de ações da Bolsa de Valores do Brasil. Grandes Modelos de Linguagem classificaram emoções em notícias de empresas específicas, gerando séries temporais de um mês para emoções. A análise revelou diferentes graus de correlação entre o sentimento do mercado e a variação dos preços, conforme o delay, sugerindo que mudanças no sentimento podem preceder movimentos de mercado. Os resultados destacam o potencial dos LLMs para estratégias de previsão e apoio à decisão.*

## 1. Introdução

O mercado financeiro é influenciado por uma complexa interação de fatores econômicos, políticos e psicológicos. Entre eles, as emoções dos agentes desempenham papel crucial, impactando decisões de compra e venda e, consequentemente, a formação de preços de ativos [Santos and Barros 2011]. Segundo Chen et al. (2019), o sentimento do mercado, frequentemente capturado em mídias sociais, pode antecipar movimentos nos preços das ações, oferecendo insights valiosos para investidores e gestores. Nos últimos anos, Grandes Modelos de Linguagem (LLMs) ampliaram as possibilidades de análise automatizada de textos e extração de emoções em larga escala, antes realizadas por modelos tradicionais de mineração de texto.

Este trabalho busca medir a correlação entre emoções extraídas de notícias e preços de ações listadas na Bolsa de Valores (B3), além de avaliar a performance dos LLMs na classificação de emoções. Foram selecionadas ações de diferentes portes<sup>1</sup>: Banco do Brasil (BBAS3), Marfrig (MRFG3) e SLC Agrícola (SLCE3).

---

<sup>1</sup>Blue Chip, Mid Cap, Small Cap

## 2. Trabalhos Relacionados

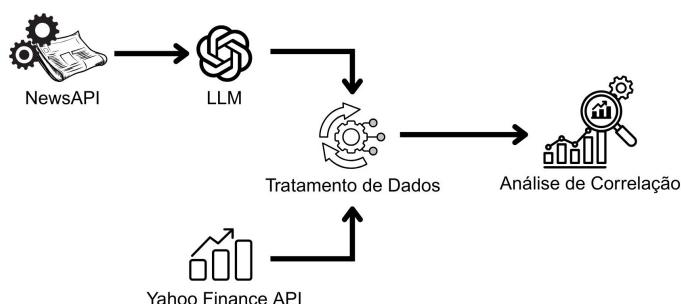
A ideia de que as emoções podem traduzir parte da aleatoriedade do mercado de ações impulsionou o desenvolvimento de trabalhos semelhantes a este. No entanto, a forma como as emoções são modeladas e relacionadas aos preços varia significativamente entre os estudos.

Em Chen et al. (2021), utilizou-se uma extensão do modelo dinâmico SIRS, inspirado na dinâmica de doenças infecciosas, para simular a propagação do sentimento dos investidores no mercado acionário. A partir desse arcabouço, investigaram-se os mecanismos de acionamento de *circuit breakers* nos Estados Unidos e a forma como a comunicação entre investidores e a difusão de emoções influenciam a volatilidade. Embora esse estudo tenha fornecido evidências relevantes sobre o papel dos investidores emocionais, sua abordagem é essencialmente baseada em modelagem matemática, diferindo do presente trabalho, que aplica LLMs a dados reais de notícias para construir séries temporais e medir empiricamente a correlação entre emoções e preços.

De maneira semelhante, em Zhou et al. (2021), foram coletados textos publicados por usuários das redes sociais Twitter (atualmente chamado "X") e Weibo em relação à bolsa de valores chinesa, classificando emoções entre cinco categorias: Raiva, Desgosto, Tristeza, Medo e Alegria. Após a classificação, os autores aplicaram modelos lineares e não lineares de Aprendizado de Máquina (ML) para prever movimentos do mercado, atingindo cerca de 60% de acurácia. Diferentemente desse estudo, o presente trabalho concentra-se na extração de emoções a partir de notícias jornalísticas, em vez de redes sociais, além de empregar LLMs para diferentes portes de ações da B3, permitindo observar como o impacto emocional varia conforme o perfil do ativo.

## 3. Metodologia

O processo desenvolvido neste trabalho consiste em três partes principais: obtenção dos dados, tratamento dos dados e análise de correlação, como ilustrado na Figura 1.



**Figura 1. Fluxo de trabalho**

### 3.1. Obtenção dos Dados

Visando automatizar a coleta de dados e evitar o viés de confirmação, foram utilizadas duas APIs públicas: NewsAPI para a coleta de títulos de notícias e Yahoo Finance API para a coleta de dados financeiros, como o histórico de cotação das ações, sendo de 12 de julho de 2025 até 12 de agosto de 2025 o período analisado.

As notícias fornecidas pela NewsAPI foram enviadas ao Chat GPT-5 visando a classificação de emoções entre seis classificações possíveis: muito negativa, negativa, neutra, positiva e muito positiva. Para todas as notícias fornecidas, foram armazenados título, data de publicação, emoção e ativo relacionado. De maneira semelhante, para cada cotação, foram armazenados valor no fechamento do mercado, data e ativo relacionado.

### 3.2. Tratamento dos Dados

Devido à ausência de notícias em alguns dias do período, foi necessário tratar valores nulos com duas suposições: (i) quando não havia emoção classificada no primeiro dia, assumiu-se neutralidade; (ii) as emoções persistem até a publicação de nova notícia, refletindo a continuidade do sentimento de mercado.

Considerando o caráter ordinal das emoções, estas foram codificadas de -2 (muito negativa) a 2 (muito positiva). Nos casos de várias notícias em um mesmo dia, utilizou-se a média aritmética das emoções como valor representativo.

Para assegurar a validade da análise de correlação cruzada, foi necessário verificar a estacionariedade das séries. O teste KPSS [Kwiatkowski et al. 1992] indicou que as séries de emoções já eram estacionárias, enquanto as séries de cotações não eram. Para torná-las adequadas à análise, aplicou-se o Método da Primeira Diferença [Hyndman and Athanasopoulos 2018], definido na equação  $y'_t = y_t - y_{t-1}$ .

## 4. Resultado Parciais e Conclusão

O cálculo de correlação foi feito utilizando o módulo Pandas<sup>2</sup> do Python, que suporta dados tabulares e facilita a manipulação. Tendo em vista que diferentes classes de ações podem reagir com diferentes níveis de volatilidade, a análise de correlação foi realizada considerando espaços de tempo (lags) de zero a sete dias entre a cotação e a emoção proveniente da notícia.

Dias	0	1	2	3	4	5	6	7
BBAS3	0,13	0,01	-0,30	0,02	-0,10	-0,14	-0,31	-0,18
MRFG3	0,32	0,06	0,02	0,02	-0,02	0,02	0,02	0,02
SLCE3	0,24	-0,23	-0,27	-0,42	-0,19	-0,19	0,19	0,25

**Quadro 4.1. Correlação entre emoções e cotações**

O Quadro 4.1 apresenta os coeficientes de correlação cruzada entre as séries de emoções extraídas das notícias e as cotações dos ativos selecionados, considerando lags de zero a sete dias. Os resultados evidenciam padrões distintos para cada ação, reforçando a hipótese de que o impacto das emoções varia conforme o perfil do ativo.

Para a ação BBAS3 (Banco do Brasil), classificada como *Blue Chip*, observou-se correlação positiva fraca no lag 0 (0,13), indicando que variações de sentimento não têm impacto imediato significativo. No entanto, a correlação assume valores negativos em lags maiores, alcançando -0,31 no lag 6, sugerindo que emoções positivas podem anteceder correções de preço. Na ação MRFG3 (Marfrig), *Mid Cap*, a maior correlação ocorreu

---

<sup>2</sup>A correlação cruzada pode ser obtida no Pandas utilizando a função *Series.corr()* para séries alinhadas. Para deslocamentos temporais, emprega-se *Series.shift(lag)* antes do cálculo.

no lag 0 (0,32), demonstrando que o preço responde de forma imediata ao sentimento noticiado. Nos lags subsequentes, os valores se aproximaram de zero, sem indicar padrões de antecipação. Já para a SLCE3 (SLC Agrícola), *Small Cap*, a correlação foi mais volátil. Houve correlação positiva em lag 0 (0,24), mas rapidamente se tornou negativa em lag 1 (-0,23) e alcançou o maior valor absoluto no lag 3 (-0,42).

De forma geral, os resultados sugerem que: (i) ativos de maior capitalização, como BBAS3, apresentam correlações mais estáveis e de menor intensidade; (ii) ativos de médio porte, como MRFG3, tendem a reagir de forma imediata, mas perdem relação em prazos maiores; (iii) ativos de pequeno porte, como SLCE3, exibem oscilações bruscas, incluindo inversão de sinal, evidenciando maior vulnerabilidade ao fluxo de informações.

Portanto, o uso de LLMs para a análise de emoções é uma ótima alternativa a algoritmos de mineração de texto, reduzindo custo computacional local. Além disso, é importante ressaltar que este trabalho atua como pilar para trabalhos futuros, como:

- Coleta de notícias do cenário global: inúmeros ativos da B3 sofrem grande impacto do cenário internacional, sendo de suma importância sua análise de emoções e correlação.
- Coleta de outros indicadores financeiros: utilizar outros indicadores financeiros juntamente com as emoções extraídas para realizar a previsão de cotações.
- Análise utilizando estatísticas mais robustas: apesar de ser útil, a Correlação, por si só, não é explicativa o suficiente. Portanto, o uso de outras estatísticas é complementar a este trabalho.
- Comparação de modelos de ML: devido às diferentes naturezas de cada porte de ativo da B3, é ideal encontrar o tipo de modelo que melhor se adequa a cada um.

## Referências

- Chen, M.-Y., Liao, C.-H., and Hsieh, R.-P. (2019). Modeling public mood and emotion: Stock market trend prediction with anticipatory computing approach. *Computers in Human Behavior*, 101:402–408.
- Chen, Y., Zhu, S., and He, H. (2021). The influence of investor emotion on the stock market: Evidence from an infectious disease model. *Discrete Dynamics in Nature and Society*, 2021(1):5520276.
- Hyndman, R. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts, Melbourne, Australia, 2 edition.
- Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178.
- Morettin, P. A. and Toloi, C. M. C. (2018). *Análise de Séries Temporais*, volume 1. Editora Edgard Blücher Ltda., São Paulo, SP, Brasil, 3 edition.
- Santos, J. O. d. and Barros, C. A. S. (2011). O que determina a tomada de decisão financeira: razão ou emoção? *Revista Brasileira de Gestão de Negócios*, 13(38):7–20.
- Zhou, Z., Xu, K., and Zhao, J. (2018). Tales of emotion and stock in china: volatility, causality and prediction. *World Wide Web*, 21:1093–1116.