

# Plataforma para simulação de tomada de decisão no escopo do mercado financeiro

Antônio Sidnei M. Cirqueira<sup>1</sup>, Deborah S. A. Fernandes<sup>2</sup>,  
Márcio Giovane C. Fernandes<sup>3</sup>

<sup>1</sup>Estudante, Instituto de Informática (UFG), tonisidneimc@discente.ufg.br

<sup>2</sup>Orientador, Instituto de Informática (UFG), deborah.fernandes@ufg.br

<sup>3</sup>Co-Orientador, Sistemas de Informação - UEG, Anápolis-GO, marcio.giovane@ueg.br

**Abstract.** *This paper proposes to carry out a practical study on the connection between the social mood arising from publications on social networks and the characteristic variability of asset prices in the Brazilian financial market. To this end, the study and development of a platform to simulate the decision-making process of buying and selling shares was used. Technical and sentiment analysis methods were adopted in order to obtain indicators that can help in this process.*

**Keywords**—decision making, stock market, prediction, microblog, sentiment analysis, social mood

**Resumo.** *Este trabalho propõe-se à realizar um estudo prático sobre a conexão entre o humor social advindo das publicações em redes sociais e a variabilidade característica nos preços dos ativos do mercado financeiro brasileiro. Para tal, utilizou-se do estudo e desenvolvimento de uma plataforma para simulação de tomada de decisão para compra e venda de ações. Métodos de análise técnica e de sentimentos foram adotados para a obtenção de indicadores que auxiliassem no processo de compra e venda de ações realizados pelo simulador implementado.*

**Palavras-chave**—tomada de decisão, mercado financeiro, predição, microblog, análise de sentimento, humor social

## 1. Apresentação

O elevado fluxo de dados não-estruturados advindo dos diversos meios de comunicação social online, tais como redes sociais, sites de notícias e buscadores de internet, carrega juntamente consigo inúmeros padrões e tendências, cuja identificação faz-se crucial ao auxílio de governos, empresas, acadêmicos ou negócios em geral, no que tange à previsão e/ou tomada de decisão em eventos de considerável impacto socio-econômico. Ademais, técnicas de aprendizado de máquina, sobretudo de processamento de linguagem natural, são amplamente utilizadas visando tal propósito de extração do valor agregado presente em tais publicações.

No escopo do mercado financeiro em especial, o entendimento à respeito do humor coletivo corrente, que possui reconhecido impacto na movimentação de seus

preços, é de grande interesse por parte de seus investidores, que almejam não só obter boa rentabilidade em seus ativos frente às condições hostis do mercado, como ter uma compreensão sólida a respeito do cenário econômico que dita tal setor. Para tal, aliado à técnicas de análise sentimental, modelos de simulação computacional podem ainda ser utilizados à fim de se expressar tais condições de mercado, dispondo-se ainda da facilidade de poder-se observar o comportamento ao longo do tempo das diversas estratégias de negociação adotadas, sem contudo ter de aplicá-las ao sistema físico em si.

A simulação computacional de sistemas caracteriza-se como uma imitação aproximada da realidade, ou seja, a operação de um processo ou sistema que representa sua execução ao longo do tempo [Banks 2005]. Simulações podem ser empregadas como auxílio em processos decisórios, baseando-se em cenários extraídos do mundo real, aonde as possíveis alternativas são tratadas como variáveis por um modelo matemático que expressa o cenário analisado [Melo et al. 2011],. A simulação pode englobar também todo o processo experimental subsequente, que tem por objetivo descrever o comportamento desse sistema e construir teorias e hipóteses que consideram as observações efetuadas. Através da simulação, é possível ainda avaliar e comparar diferentes cenários quanto às suas contribuições à longo prazo [Salam and Khan 2016].

Diversas são as vantagens da aplicação de modelos simulados. Segundo [Santos et al. 2014], decisões baseadas apenas em experimentos tendem a resultar somente em gargalos, com deficiências no dimensionamento, alocações ineficazes de recursos e problemas operacionais. Os autores de [Robinson et al. 2012] apontam que as ações realizadas sobre os objetos de interesse de um sistema possuem variabilidade, dependência e são complexas, o que torna a análise da evolução do desempenho de um sistema de difícil prevenção sem uma simulação. Já [Reid et al. 2016] salienta que o ambiente virtual, permite ao participante e observadores, controlar o tempo simulado, desenvolver mudanças interativas, coletar informações e avaliar mudanças em um contexto seguro para realizar testes através de uma análise sistêmica.

## **2. Metodologia**

Esta seção apresenta os passos realizados para a construção de um simulador para compra e venda de ações para o mercado brasileiro de bolsa de valores, que leva em conta o referido humor social. Primeiramente foi realizado um estudo sobre os trabalhos realizados na área, em seguida desenvolveu-se um desenho do experimento. Para a implementação foram estudados os conceitos e técnicas necessárias que serão apresentadas nas subseções seguintes.

### **2.1. Trabalhos relacionados**

Diversos autores, anteriormente, estudaram a relação entre o humor social proveniente de sites de notícias, microblogs e/ou redes sociais em geral com a flutuação dos

preços de ações no mercado financeiro. Este é um problema de difícil abordagem, uma vez que a série temporal que expressa a variação no preço dos ativos financeiros é de uma natureza não-linear, como aponta [Keshwani et al. 2018].

Em [Bourezk et al. 2020] é proposta a extração do humor do investidor do mercado financeiro de Marrocos através da investigação de entradas públicas em sites online de notícias em francês, relacionadas às ações com o maior volume de negociação. A abordagem baseada em lexicons (dicionários) foi comparada com a de um classificador probabilístico Bayesiano, identificando que o primeiro demonstrou-se mais sensível à variações de contexto e à figuras de linguagem usadas para qualificar diferentes ativos, enquanto o segundo obteve um melhor desempenho na polarização frente à sutileza e diversidade da língua francesa. Correlacionando-se o preço das ações em questão com os resultados obtidos, o autor foi capaz de concluir que o humor negativo exerceu um impacto considerável sobre a flutuação dos índices da bolsa.

Semelhantemente, o problema de se quantificar dados textuais de Microblogs associados à movimentação no mercado financeiro, foi tratado por [Keshwani et al. 2018, Li et al. 2018, Ruan et al. 2018, Onwuegbuche et al. 2019, Fernandes et al. 2019, Lazzini et al. 2021], que escolheram este tipo de rede social - mais especificamente o Twitter - como fonte primordial do humor coletivo. Microblogs como este, são característicos pelas suas postagens curtas e diretas, atualizações rápidas e constantes de informação em tempo real, bem como pela possibilidade de comunicação bidirecional entre os seus usuários.

Os autores de [Ruan et al. 2018] estudaram a rede de confiança de usuários do Twitter, que é baseada na reputação individual, à fim de se estruturar um mecanismo de filtragem e amplificação do microblog, capaz de melhor identificar a correlação deste com os dados financeiros do mercado de ações. Ao invés de simplesmente considerar cada postagem carregando parcela igual do sentimento público, o autor ponderou cada usuário pelo seu poder de influência e reputação na comunidade, sendo que usuários com elevado grau de confiança tendem à exercer influência bem maior. Tal medida foi obtida através da análise da interação em menções, comentários, e repostagens das publicações originais, e de interações indiretas provenientes da propagação dessa rede de confiança entre terceiros. Através do coeficiente de correlação de Pearson, o autor estabeleceu uma forte relação linear entre esses índices de influência e o retorno anormal das ações em questão.

Em num cenário de profunda incerteza e imprevisibilidade, como o foi durante a primeira fase da pandemia de COVID-19 na Itália, os autores de [Lazzini et al. 2021] em meados de 2020, demonstraram que foi possível realizar previsões quanto ao mercado financeiro, através da análise do humor social contido nas postagens das redes sociais. Durante esse período emergencial, diversas medidas sanitárias restritivas fo-

ram aplicadas visando conter o avanço do vírus, o que limitou as reuniões em público e fortaleceu as comunicações diretas através das diversas mídias sociais online. O autor destaca o Twitter pelo o que chama de hiperrealidade - a representação subjetiva da realidade, presente no corpo das mensagens publicadas nesse microblog, sob a ótica de quem as escreve - e como o agregado de milhares dessas postagens podem ser um poderoso indicativo do humor e sentimento geral do público, capaz de influenciar opiniões, pensamentos e comportamentos individuais e coletivos. Em sua análise, o autor identificou o predomínio de sentimentos de medo, desaprovação, raiva e tristeza no conjunto de dados levantado. Foi estudada ainda a correlação entre essas emoções extraídas e os preços de fechamento das ações da bolsa de Milão, através da análise de causalidade de Granger e correlação de Pearson, que estabeleceram um forte relacionamento entre este e as tendências do mercado financeiro.

Em [Li et al. 2018, Onwuegbuche et al. 2019, Mehta et al. 2021] há a comparação de diversas técnicas comumente utilizadas para aprendizado de máquina voltadas à mineração de sentimentos. Classificadores Naïve Bayes Bernoulli, Máquinas de Vetor de Suporte multiclasse, Árvore de Decisão, Máxima Entropia, Memórias Curtas de Longo-Prazo (LSTM) e Redes Neurais Artificiais foram aplicados à tweets para obtenção do humor social. A abordagem de aprendizado supervisionado destacou-se em [Bourezk et al. 2020] quanto à previsão dos preços dos títulos do mercado financeiro em relação àquela que se utiliza somente de dicionários léxicos, justamente porque esta segunda é sensível ao contexto em si.

Por fim, [Fernandes et al. 2019] estruturou e desenvolveu um simulador de tomada de decisões, que visa auxiliar investidores em suas estratégias de investimento. O simulador aplica um conjunto de regras para a compra e venda de ações utilizando indicadores obtidos à a partir de análise de sentimentos de tweets, combinada ou não com a análise técnica da bolsa. O experimento a ser apresentado nas subseções seguintes foi baseado neste.

## **2.2. Experimento**

Para a realização do experimento foram estudadas as características relacionadas ao desenvolvimento de um simulador para, à partir destas, desenvolver a plataforma do experimento. Essas características serão apresentadas na subseção seguinte. Nas demais subseções são descritas as atividades restantes relacionadas ao seu desenvolvimento.

### **2.2.1. Definição dos passos para o desenvolvimento do simulador**

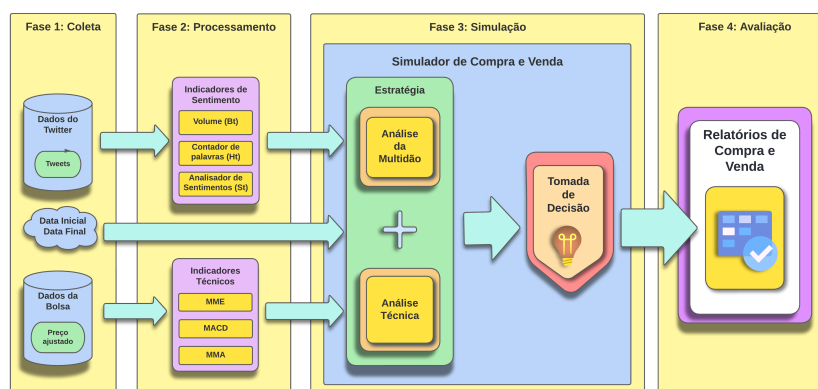
Para que a simulação ocorra de uma forma estruturada e organizada, uma série de etapas precisam ser desenvolvidas. Os autores de [Law 2015] e [Banks 2005] sugerem uma sequência de passos para o estudo e aplicação da simulação:

- (i) **Formulação e estudo do problema:** São definidos os objetivos gerais, questões específicas à serem respondidas na execução do modelo, escopo do projeto, quais configurações do sistema real serão modeladas, as medidas de performance que serão usadas para validar a eficácia das diferentes configurações do sistema, a janela de tempo à se considerar, e os recursos requeridos para a simulação.
- (ii) **Coleta de dados:** São coletadas informações sobre a estrutura do sistema e procedimentos operacionais que especificam os parâmetros do modelo. Também podem ser reunidos dados referentes à performance de simuladores semelhantes, visando a validação do modelo atual.
- (iii) **Definição do modelo:** Constrói-se o modelo conceitual que procura representar as características essenciais do sistema real. Este pode ser enriquecido até que se tenha uma aproximação mais útil e adequada da realidade. O nível de detalhe do modelo dependerá sobretudo dos objetivos fixados, medidas de performance adotadas, disponibilidade de dados e restrições computacionais impostas.
- (iv) **Tradução do modelo:** O modelo conceitual é traduzido para linguagem computacional, gerando um modelo operacional. A utilização de uma linguagem de programação nessa etapa, ao invés de um software de simulação específico, oferece ao analista um maior controle sobre o comportamento do programa e melhor performance durante a execução do modelo.
- (v) **Verificação:** O modelo operacional é ponderado quanto à ser ou não, uma boa representação para a lógica do sistema, e se, através da depuração do programa o modelo está funcionando conforme o esperado.
- (vi) **Validação:** As respostas do modelo simulado são comparadas com as do sistema real quanto à sua coerência. As discrepâncias encontradas e o conhecimento adquirido podem então ser utilizados para o aperfeiçoamento do modelo.
- (vii) **Delineamento Experimental:** São definidos como os testes no modelo devem ser conduzidos considerando parâmetros como o tempo de duração e número de replicações da simulação para o estudo.
- (viii) **Experimentação e análise estatística dos dados de saída:** O modelo é executado através das várias simulações e seus resultados são analisados, e retidos como parâmetro para investigações futuras.
- (ix) **Documentação, apresentação e utilização dos resultados.** Os resultados obtidos só serão suficientemente confiáveis ao processo de tomada de decisão, se forem, ao mesmo tempo, válidos e verossímeis.

### 2.2.2. Definição do modelo conceitual

A Figura 1 apresenta o modelo conceitual do simulador desenvolvido. Tem-se por objetivo a construção de um sistema para compra e venda de ações baseado em

análise técnica, análise de sentimentos e volume de tweets.



**Figura 1. Modelo Conceitual do Simulador**

As configurações à serem modeladas neste simulador serão a compra de ações que se dá pela aplicação de uma ou mais estratégias de negociação no sistema real. A medida de performance adotada será analisar o retorno obtido através da aplicação da estratégia de investimento nas referidas ações, que tem por base as saídas do simulador.

Para a análise técnica, será utilizada a técnica do cruzamento de médias móveis, que são indicadores estatísticos da existência de tendência do mercado ao longo do tempo, bem como suas forças e direções. A escolha destes indicadores se dá pelo fato destes conseguirem expressar a configuração atual do mercado (seja em alta ou em baixa), e por conseguinte, o seu humor, que é justamente o que se busca aliar à análise do humor proveniente do Twitter. As médias móveis adotadas para este estudo serão as seguintes:

- **Média Móvel Exponencial (MME)**: Utiliza-se da ponderação de médias móveis simples, dando maior importância às contribuições mais recentes dos preços. Com isso, é capaz de reagir mais rapidamente às mudanças de tendências de mercado. A média móvel simples (MMS) é calculada obtendo-se o preço médio de um título ao longo de um determinado número de períodos. A fórmula para o cálculo da MMS para  $\tau$  períodos, é dada pela Eq.1 na qual  $x$  é a série temporal do ativo em questão,  $\tau$  é o período utilizado para o cálculo da MMS, e  $t$  é o instante observado. Semelhantemente, define-se a MME pela Eq.2, com  $\alpha$  definido pela Eq.3 sendo o peso atribuído aos instantes no tempo. Como  $\alpha$  varia entre 0 e 1, tem-se que, quanto mais próximo de 1 for este valor, maior será o peso dado às informações mais recentes, enquanto que o contrário será válido se mais próximo de 0. Para a análise técnica, sugere-se a utilização de duas médias móveis e atentar-se aos momentos de cruzamento entre elas. Aqui, os parâmetros adotados serão de 5 e de 13 dias para as médias curta e longa, respectivamente. Sendo assim, quando a média de 5 dias cruzar acima

da linha da média de 13 dias, sinaliza-se um momento de compra, e quando este cruzamento ocorrer de cima para baixo, um sinal de venda é gerado.

$$M_{\tau}(t) = \frac{1}{\tau} \sum_{i=0}^{\tau-1} x_{(t-i)} \quad (1)$$

$$E_{\alpha}(t) = \alpha \sum_{i=0}^{\tau} (1 - \alpha)^i x_{(i+1)} \quad (2)$$

$$\alpha = \frac{2}{\tau + 1} \quad (3)$$

- **Convergência e Divergência de Médias Móveis (MACD):** Este indicador é gerado subtraindo-se uma média móvel exponencial de outra. Os valores clássicos são as médias curtas e longas, referentes aos preços de 26 e 12 períodos, tal como na Eq.4. Um resultado maior que zero, indica que as expectativas mais recentes são mais favoráveis para alta do que as anteriores, enquanto que um valor menor do que zero indica um cenário mais favorável à uma situação de baixa. A linha sinal do MACD, útil para identificar pontos de entrada e saída do mercado, é calculada através da média móvel exponencial de 9 períodos dos valores do próprio MACD conforme Eq.5. A estratégia técnica do MACD é definida como sendo o momento em que sua linha cruza a linha de sinalização, seja de cima para baixo indicando um momento de entrada (compra) ou de baixo para cima indicando um momento de saída (venda).

$$MACD = MME_{12} - MME_{26} \quad (4)$$

$$Sinal = MME_9(MACD) \quad (5)$$

- **Média Móvel Adaptativa de Kauffman (MMA):** Esta é uma variação da média móvel exponencial, porém menos sujeita à ruídos nos dados (whipsaw). A constante  $\alpha$  é recalculada conforme Eq.6, sendo  $E_r$  o raio de eficiência, que identifica a força de uma tendência, definido pela Eq.7. Nesta, O numerador representa a mudança de preço total para o período, e o denominador a soma da variação absoluta de preço para cada barra (volatilidade). Já  $S_f$  e  $S_s$ , são constantes de suavização de uma média móvel exponencial rápida e uma lenta, definidas conforme Eq.8. Os valores padrão para  $a$ ,  $e$  e  $n$  são, respectivamente, 30, 2 e 10. A geração do sinal do MMA ocorre no momento em que sua linha cruza a linha dos preços, seja de cima para baixo indicando um momento de entrada (compra) ou de baixo para cima indicando um momento de saída (venda).

$$\alpha = [(E_r \times (S_f - S_s)) + S_s]^2 \quad (6)$$

$$E_r = \frac{|p_{i+1} - p_{i+n}|}{\sum_{j=1}^{n-1} |p_{i+j} - p_{i+j+1}|} \quad (7)$$

$$S_f = \frac{2}{e+1} \quad \text{e} \quad S_s = \frac{2}{a+1}, \quad \text{com} \quad a > e \quad \text{e} \quad a, e \in \mathbb{N}^*. \quad (8)$$

Com relação à análise de sentimentos, alguns indicadores ingênuos foram adotados baseados no trabalho de [Fernandes et al. 2019]:

- **Volume  $B_t$ :** Também chamado de burburinho, é o fluxo não-estruturado de dados gerado pelos usuários do microblog, que representa o volume de publicações produzido no dia  $t$ . Um volume alto (acima da média), será tomado para indicar que o mercado está aquecido (otimista), enquanto que um volume baixo indicará o oposto (pessimista).
- **Contador de Palavras  $H_t$ :** É o volume de publicações contendo os radicais de palavras em português que fazem menção à compra ou venda de ações, podendo indicar tendências e/ou o humor diário do mercado, e definido na Eq.9.  $BUY_t$  representa a quantidade de tweets que contém o radical "COMP" e  $SELL_t$  a quantidade de tweets que contém o radical "VEND" para o dia em questão. Um  $H_t$  positivo indica um mercado otimista para compra, bem como negativo indique pessimismo para a venda, e mercado neutro caso igual à zero.

$$H(t) = BUY_t - SELL_t \quad (9)$$

- **Analisador de Sentimentos  $S_t$ :** É o sentimento diário extraído da rede social através de análise de sentimentos, e definido como na Eq.10. Nesta,  $N_{pos_t}$  é a quantidade de tweets com polaridade positiva e  $N_{neg_t}$  a quantidade de tweets com polaridade negativa, considerando-se o dia  $t$  em questão. Caso a quantidade de tweets pessimistas seja superior à de otimistas,  $S_t$  será negativo, enquanto que positivo, caso contrário, e neutro se elas forem iguais:

$$S(t) = N_{pos_t} - N_{neg_t} \quad (10)$$

Visando a polarização automática, o modelo estatístico de Regressão Logística foi adotado, por ter obtido uma boa performance nos trabalhos anteriores supracitados. Esse modelo linear generalizado (também conhecido como modelo logit ou classificador de máxima entropia) é útil quando necessita-se modelar a probabilidade de um evento ocorrer como combinação linear de uma ou mais variáveis independentes (preditores), e quando a variável dependente (alvo) é categórica. Ele analisa dados binomialmente distribuídos da forma expressa pela Eq.11 aonde os ensaios de Bernoulli  $n_i$  são conhecidos, enquanto que as probabilidades de êxito  $p_i$  não o são. A modelagem é então realizada com base



nas contribuições, quanto à probabilidade final, de cada ensaio e do conjunto de variáveis independentes. Para tal, Uma transformação logit é aplicada sobre a chance de ocorrência dos eventos em questão, sendo definida conforme Eq.12. A estimativa dos parâmetros desconhecidos  $\beta_j$  é comumente realizada através do método de máxima verossimilhança (MLE). O resultado final do modelo é um hiperplano em um espaço  $m$ -dimensional, que separa pontos de dados com  $m$  recursos em suas respectivas classes.

$$Y_i \sim B(p_i, n_i), \quad \text{para } i = 1, \dots, m \quad (11)$$

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \beta_0 + \sum_{i=1}^k \beta_i x_i \quad (12)$$

### 2.2.3. Coleta e Processamento dos Dados

Tendo em vista o microblog do Twitter como fonte do humor social, bem como o mercado financeiro como sistema-alvo à ser modelado na simulação, faz-se necessário ainda definir como os dados serão coletados e processados à fim de se extrair informação útil ao simulador. As seguintes rotinas foram adotadas:

- **Coleta:**

- Definiu-se a janela de tempo a ser considerada para o levantamento dos dados históricos como sendo o ano de 2021.
- A API do Twitter<sup>1</sup> foi utilizada à fim de se realizar a coleta de tweets referentes à este período, resultando em um total de 6.008.361 publicações.
- Definiu-se que os ativos de interesse para a análise seriam as ações referentes às empresas da *Petrobrás* (PETR4.SA), *Vale do Rio Doce* (VALE3.SA), *Magazine Luiza* (MGLU3.SA) e *Itaú Unibanco* (ITUB4.SA).
- Dados históricos sobre o preço de fechamento ajustado das referidas ações, para a janela de tempo em questão, foram coletados utilizando-se da API do Yahoo Finance<sup>2</sup>.

- **Pré-Processamento:** Após a limpeza dos dados, a etapa de Pré-Processamento resultou em um total de 4.473.230 tweets. Abaixo as rotinas realizadas sobre os dados do Twitter:

- Foram removidas as publicações em língua estrangeira.
- Removidos os caracteres de tabulação, nova linha, espaços repetidos e de pontuação.
- Removidos os tweets publicados em dias não-úteis ou feriados, nos quais não ocorre negociação na bolsa de valores.

<sup>1</sup>disponível em: <https://developer.twitter.com/en/docs/twitter-api/tools-and-libraries/v2>

<sup>2</sup>disponível em: <https://github.com/ranaroussi/yfinance>

- Substituiu-se os emojis pela sua representação semântica. Esta técnica, também conhecida como demojização, transforma o caractere '😊' em "rosto sorridente com olhos sorridentes", por exemplo.
- Removidos os links, hashtags e menções (não sem antes ter separado aquelas que referenciavam às empresas sob consideração).
- Converteu-se todo o corpo textual para letras minúsculas.
- Foram removidas as stopwords do idioma português presentes nos tweets. Estas são palavras que não agregam sentimento em si e tampouco contribuem para o sentido da frase, tais como preposições, artigos e conjunções.
- Um dicionário personalizado foi utilizado para a tradução de gírias e expressões idiomáticas.
- Aplicou-se a técnica de lematização ao corpo textual, que busca reduzir uma palavra à sua raiz, retirando todas as inflexões e resultando em uma outra palavra que pertença à gramática.

• **Processamento:**

- Foram calculados os indicadores de sentimentos  $B_t$  e  $H_t$ , com base nos tweets resultantes da etapa de pré-processamento anterior.
- Realizou-se o treinamento do modelo de Regressão Lógica com base em um outro conjunto de tweets pré-rotulados<sup>3</sup>, ao qual foi aplicada a técnica de vetorização, com base em um modelo de skipgrama de 300 dimensões, pré-treinado em língua portuguesa<sup>4</sup>. Esse conjunto de tweets, devidamente pré-processado, dispunha-se de somente 200 entradas - sobretudo devido à dificuldade na construção de um léxico e indisponibilidade de conjuntos semelhantes na internet - dos quais então, 70% foram separados para treino e 30% para teste.

O modelo de Regressão Logística utilizado para a polarização dos tweets, proveniente da biblioteca Scikit-learn do Python<sup>5</sup>, foi treinado com os parâmetros de penalidade L1, solucionador lib-linear e inverso da força de regularização  $C = 10$ , resultando em uma acurácia média de 0.81, que foi obtida através de validação cruzada do f1-score.

- Foram selecionados e agrupados os tweets que citam ou referenciam as empresas cujos ativos estão sob consideração. Com isso, esse conjunto foi vetorizado, semelhantemente ao anterior, e o modelo de Regressão Logística treinado anteriormente foi utilizado para polarizar cada sentença em otimista (positiva) ou pessimista (negativa). Por fim, a eq. (10) foi aplicada para a identificação do sentimento generalizado de cada dia, o que gerou os dados da fig. 2.

<sup>3</sup>publicado em: <https://conpeexonline.ufg.br/evento/trabalhos/trabalho/623>

<sup>4</sup>disponível em: <http://nilc.icmc.usp.br/nilc/index.php/repositorio-de-word-embeddings-do-nilc>

<sup>5</sup>[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)

	VALE	PETR	MGLU	ITAU
<b>otimista</b>	101	1	76	46
<b>pessimista</b>	130	238	153	180
<b>neutro</b>	8	0	10	13

Figura 2. Quantidade de dias por sentimento

- Realizou-se o cálculo dos indicadores técnicos propostos nas eqs. (2) e (4) à (6), utilizando-se dos dados históricos coletados. Para cada abordagem, a estratégia de cruzamento das médias móveis foi utilizada à fim de se gerar os sinais de compra ou venda, que estão amostrados nas figs. 3 à 5.

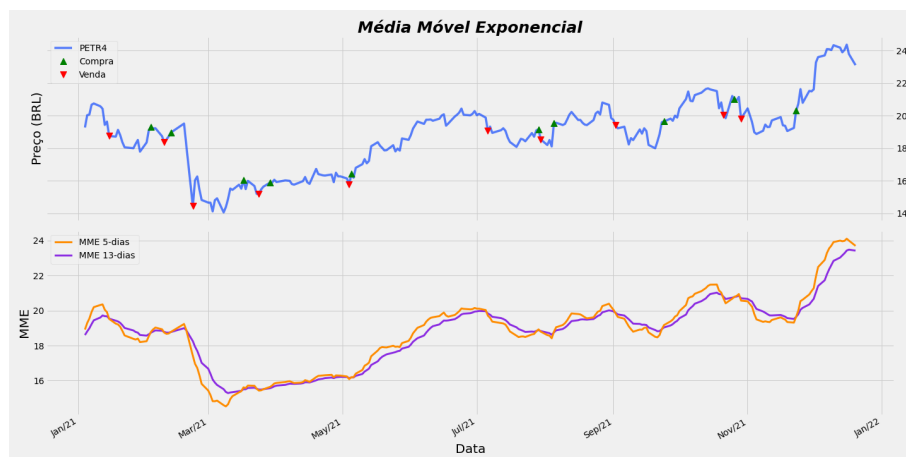


Figura 3. MME - Petrobrás

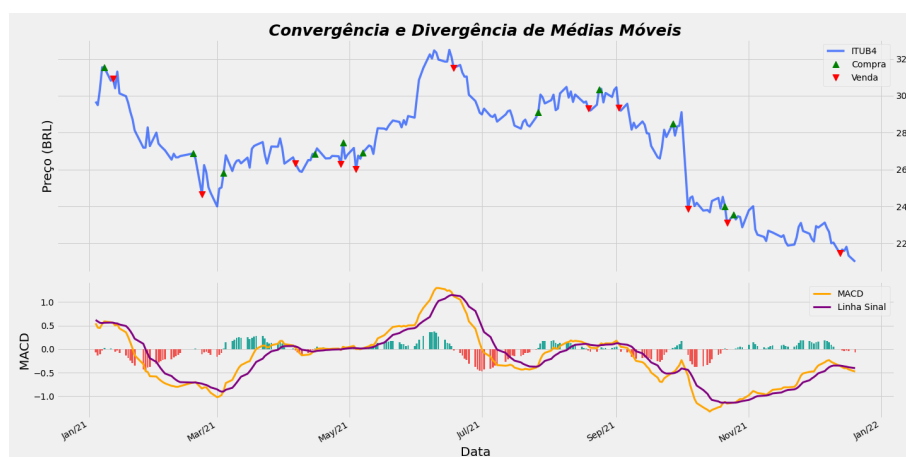


Figura 4. MACD - Itaú

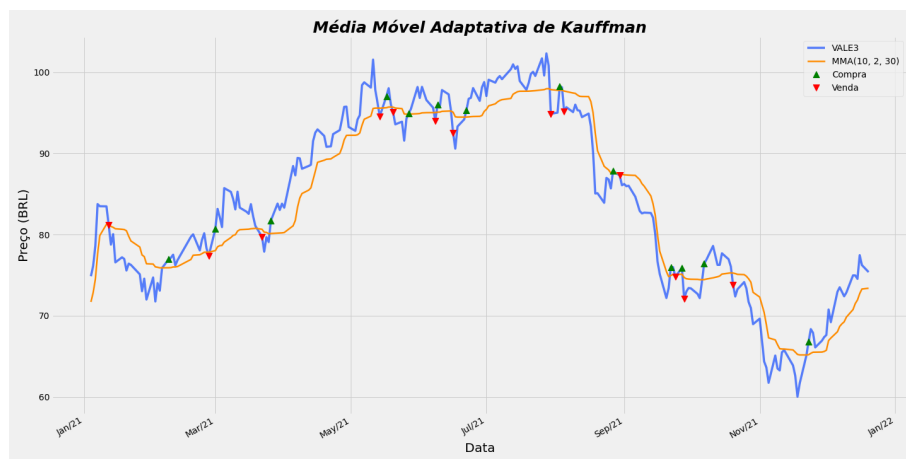


Figura 5. MMA - Vale do Rio Doce

#### 2.2.4. Implementação do modelo

A linguagem de programação Python<sup>6</sup> foi escolhida à fim de se realizar a tradução do modelo e execução das rotinas do simulador. Com os indicadores técnicos, de sentimento e de volume de tweets, o simulador combina as saídas através das seguintes regras, à fim de se gerar novos sinais de compra e/ou venda:

- **Módulo 1- Análise Técnica:** Se qualquer uma das estratégias de análise técnica, isto é, MME, MACD ou MMA, gerar um sinal de compra ou venda, tal que as outras duas não o contradigam de alguma forma (indicando compra e venda ao mesmo tempo, por exemplo), então este sinal será propagado no simulador. Caso contrário, nenhum sinal será enviado deste módulo.
- **Módulo 2 - Análise Sentimental:** Se  $S_t$  for otimista e/ou  $H_t$  indicar compra, acompanhados de um  $B_t$  alto, um sinal de compra será gerado. O contrário ocorrerá para  $S_t$  pessimista e/ou  $H_t$  indicando venda juntamente com  $B_t$  baixo, o que resultará em um sinal de venda. Considera-se ainda o sentimento de até dois dias anteriores ao atual como sendo válidos para a análise. Caso o sentimento/humor do mercado no dia não esteja muito claro (neutro ou contradizente), nenhum sinal será enviado deste módulo.
- **Módulo 3 - Análise Combinada:** Combina-se as saídas dos módulos de Análise Técnica e de Análise de Sentimentos, tal que, se os dois concordarem entre-si, este sinal será propagado pelo simulador.
- **Módulo 4 - Tomada de Decisão:** A técnica de *buy-and-hold*, que consiste em comprar ou vender uma ação tão logo esta esteja disponível (baseando-se nas saídas diárias obtidas do módulo 3), foi utilizada para produzir a saída final de cada ativo. Quando uma ação é comprada, o simulador muda seu estado para "COMPRADO", enquanto que quando essa ação é vendida, seu estado é alterado para "VENDIDO", indicando entrada e saída do mercado,

<sup>6</sup><https://www.python.org/>

respectivamente. Só é possível efetuar uma compra quando nenhum ativo tiver sido adquirido anteriormente ou o simulador estiver em estado de VENDIDO.

### 2.2.5. Validação

Para a avaliação da estratégia de negociação adotada, bem como das operações realizadas pelo simulador, o cálculo do retorno total acumulado (lucro) e do risco frente ao investimento, foram utilizados, considerando-se a contribuição somente dos dias em que o simulador permaneceu em estado de COMPRADO. Os retornos percentuais diários, que vêm de uma distribuição log-normal, são definidos como sendo conforme a Eq.13. Nesta,  $p_t$  é o valor do preço de fechamento ajustado para os dividendos no dia  $t$ . O retorno cumulativo do  $k$ -ésimo dia é dado pelo somatório dos retornos diários expostos na Eq.14. Desta forma, o retorno cumulativo total será dado para  $k = N$ , que é a quantidade de dias observados na simulação.

Define-se ainda o índice de *sharpe*, que indica o quanto um retorno compensa o risco assumido pelo investidor, conforme Eq.15, na qual  $E[.]$  é o valor esperado de retorno para o  $i$ -ésimo ativo,  $r_f$  a taxa livre de risco com relação à um ativo de referência e  $\sigma_i$  o seu desvio padrão. Com isso, as métricas apresentadas na Figura 6 foram geradas.

$$r_t = \ln(p_t) - \ln(p_{t-1}) \quad (13)$$

$$C_k = \sum_{i=1}^k r_i, \quad \text{para } k = 1, 2, \dots, N \quad (14)$$

$$SR = \frac{E[r_i - r_f]}{\sigma_i} \quad (15)$$

	VALE3	PETR4	MGLU3	ITUB4
<b>Compras</b>	11	5	10	7
<b>Vendas</b>	10	5	9	7
<b>Retorno cumulativo</b>	56.59%	21.66%	56.13%	33.55%
<b>Índice de Sharpe</b>	2.562	1.701	2.768	2.655

Figura 6. Relatório final

## 3. Resultados e Discussão

Numa análise de negociação usual, um índice de *sharpe* abaixo de 1.0 é considerado ruim, enquanto que entre 1.0 e 1.99 é considerado adequado, de 2.0 à 2.99 é considerado muito bom, e acima de 3.0 é considerado excelente. Dentre os ativos

em questão, as ações da Petrobrás se destacaram devido ao seus índice de sharpe, retorno cumulativo e quantidade de operações realizadas inferior às demais. Isso se deve sobretudo à identificação de um forte sentimento pessimista proveniente dos tweets que a referenciavam, conforme a fig. 2 aponta. Este fato foi notado durante a execução deste estudo e que supõe-se ser fruto dos diversos conflitos políticos que envolveram a empresa no ano em questão, que podem ter desencadeado uma profunda rejeição social para com a entidade em si. Por outro lado, ações como as da Magazine Luiza - que estava em uma tendência de queda à longo prazo - e da Vale do Rio Doce, não sofreram tamanho impacto devido ao sentimento presente no microblog se apresentar relativamente balanceado entre otimismo e pessimismo.

#### **4. Conclusão**

Através dos diversos resultados obtidos neste trabalho, observa-se que o investidor que escolhesse a estratégia apresentada seria capaz de obter um lucro positivo para todos os ativos em questão, o que corrobora na indicação do humor coletivo como um relevante e poderoso indicador da dinâmica operacional deste mercado.

Um dos desafios encontrados durante a execução deste trabalho, foi a indisponibilidade nos meios de publicação de um grande conjuntos de dados pré-rotulados para a língua portuguesa, contextualizados ao mercado financeiro, com o qual o modelo de Regressão Logística pudesse ser treinado mais amplamente. Contudo, esse modelo é caracterizado por comportar-se bem mesmo para um conjunto de dados relativamente pequeno, pois requer um pequeno número de suposições para tal.

Trabalhos futuros podem preocupar-se em construir um léxico específico ao mercado financeiro ou ainda utilizar-se de outros indicadores técnicos, como as Bandas de Bollinger ou Índice de Força Relativa (RSI), à fim de se evitar a multicolinearidade advinda do uso de indicadores de uma única categoria, como acontece comumente com os indicadores de tendência adotados. Também podem ser estudadas outras métricas de avaliação de performance, bem como técnicas de extração do humor social além das adotadas.

#### **Referências**

- Banks, J. (2005). *Discrete-event system simulation*. Prentice Hall, 4th edition.
- Bourezk, H., Raji, A., Acha, N., and Barka, H. (2020). Analyzing moroccan stock market using machine learning and sentiment analysis. In *2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, pages 1–5. IEEE.
- Fernandes, D. S., Fernandes, M. G., Borges, G. A., and Soares, F. A. (2019). Decision-making simulator for buying and selling stock market shares based on twitter indicators and technical analysis. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 2626–2632. IEEE.

- Keshwani, K., Agarwal, P., Kumar, D., et al. (2018). Prediction of market movement of gold, silver and crude oil using sentiment analysis. In *Advances in Computer and Computational Sciences*, pages 101–109. Springer.
- Law, A. (2015). *Simulation modeling and analysis*. McGraw-Hill, 5th edition.
- Lazzini, A., Lazzini, S., Balluchi, F., and Mazza, M. (2021). Emotions, moods and hyperreality: social media and the stock market during the first phase of covid-19 pandemic. *Accounting, Auditing & Accountability Journal*.
- Li, M., Yang, C., Zhang, J., Puthal, D., Luo, Y., and Li, J. (2018). Stock market analysis using social networks. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–10.
- Mehta, P., Pandya, S., and Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. *PeerJ Computer Science*, 7:e476.
- Melo, N. A. F., Pessanha, A. M. B., and Rocha Filho, S. M. D. (2011). Estudo da aplicação do software arena em um contrato de prestação de serviço de manutenção de instrumentação. *Exatas & Engenharias*, 1(02).
- Onwuegbuche, F., Wafula, J., and Mungatu, J. (2019). Support vector machine for sentiment analysis of nigerian banks financial tweets. *Journal of Data Analysis and Information Processing*, 07:153–173.
- Reid, J., Stone, K., Huang, L., and Deutsch, E. S. (2016). Simulation for systems integration in pediatric emergency medicine. *Clinical Pediatric Emergency Medicine*, 17(3):193–199.
- Robinson, S., Radnor, Z. J., Burgess, N., and Worthington, C. (2012). Simlean: Utilising simulation in the implementation of lean in healthcare. *European Journal of Operational Research*, 219(1):188–197.
- Ruan, Y., Durresi, A., and Alfantoukh, L. (2018). Using twitter trust network for stock market analysis. *Knowledge-Based Systems*, 145:207–218.
- Salam, M. A. and Khan, S. A. (2016). Value creation through lean management: a case study of healthcare service operations. *International Journal of Services and Operations Management*, 25(3):275–293.
- Santos, J. A. A. d., Ferrão, S. R., de Souza, R. C., Shioya, H. M., and Monteiro, C. I. (2014). Modelagem, simulação e otimização da dinâmica operacional do processo de fabricação de placas de sinalização. *Revista GEINTEC-Gestão, Inovação e Tecnologias*, 4(1):692–703.

## 5. Informações complementares



Certificamos que **Antônio Sidnei Moreira Cirqueira**, participou da palestra **“Pesquisa e inovação: a UFG em diálogo com a sociedade”**, atividade do **Programa Diálogos em Pesquisa e Inovação**, promovida pela **Pró-Reitoria de Pesquisa e Inovação da Universidade Federal de Goiás**, no dia 16 de Fevereiro de 2022 com 02 (duas) horas de duração.

Goiânia, 16 de Fevereiro de 2022

*Prof.ª Dr.ª Helena Carasek Cascudo*

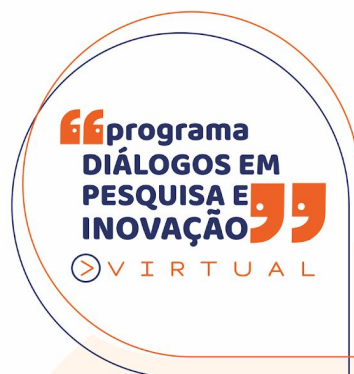
Pró-Reitora de Pesquisa e Inovação

**PRPI**  
PRÓ-REITORIA DE  
PESQUISA E INOVAÇÃO



**UFG**  
UNIVERSIDADE  
FEDERAL DE GOIÁS





Certificamos que **Antônio Sidnei Moreira Cirqueira**, participou da palestra **“Formas de proteção por meio da Propriedade Industrial e a importância de proteger o conhecimento produzido”**, atividade do **Programa Diálogos em Pesquisa e Inovação**, promovida pela **Pró-Reitoria de Pesquisa e Inovação da Universidade Federal de Goiás**, no dia 08 de Março de 2022 com 02 (duas) horas de duração.

Goiânia, 08 de Março de 2022

Prof.ª Dr.ª Helena Carasek Cascudo

Pró-Reitora de Pesquisa e Inovação

**PRPI**  
PRÓ-REITORIA DE  
PESQUISA E INOVAÇÃO



**UFG**  
UNIVERSIDADE  
FEDERAL DE GOIÁS



Certificamos que **ANTÔNIO SIDNEI MOREIRA CIRQUEIRA**, participou da palestra “**Quais as boas práticas que devo observar como jovem pesquisador?**”, atividade do **Programa Diálogos em Pesquisa e Inovação**, promovida pela **Pró-Reitoria de Pesquisa e Inovação da Universidade Federal de Goiás**, no dia 31 de Março de 2022 com 02 (duas) horas de duração.

Goiânia, 31 de Março de 2022

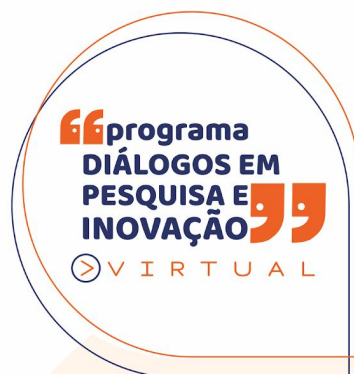
Prof.ª Dr.ª Helena Carasek Cascudo

Pró-Reitora de Pesquisa e Inovação

**PRPI**  
PRÓ-REITORIA DE  
PESQUISA E INOVAÇÃO



**UFG**  
UNIVERSIDADE  
FEDERAL DE GOIÁS



Certificamos que **Antônio Sidnei Moreira Cirqueira**, participou da palestra “**Metaverso: a reconfiguração da realidade em nossas vidas**”, atividade do **Programa Diálogos em Pesquisa e Inovação**, promovida pela **Pró-Reitoria de Pesquisa e Inovação da Universidade Federal de Goiás**, no dia 13 de Abril de 2022 com 02 (duas) horas de duração.

Goiânia, 13 de Abril de 2022

Prof.ª Dr.ª Helena Carasek Cascudo  
Pró-Reitora de Pesquisa e Inovação

**PRPI**  
PRÓ-REITORIA DE  
PESQUISA E INOVAÇÃO



**UFG**  
UNIVERSIDADE  
FEDERAL DE GOIÁS