

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Engenharia de Software

# **Ferramenta para análise do Ibovespa em relação ao sentimento da mídia**

**Autor: Eugênio Sales Siqueira**  
**Orientador: Prof. Dr. John Lenon Cardoso Gardenghi**

Brasília, DF  
2022





Eugênio Sales Siqueira

## **Ferramenta para análise do Ibovespa em relação ao sentimento da mídia**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Orientador: Prof. Dr. John Lenon Cardoso Gardenghi

Brasília, DF

2022

---

Eugênio Sales Siqueira

Ferramenta para análise do Ibovespa em relação ao sentimento da mídia/  
Eugênio Sales Siqueira. – Brasília, DF, 2022-  
60 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. John Lenon Cardoso Gardenghi

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA , 2022.

1. Análise de sentimento. 2. Ibovespa. I. Prof. Dr. John Lenon Cardoso Gardenghi. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. Ferramenta para análise do Ibovespa em relação ao sentimento da mídia

CDU 02:141:005.6

---

Eugênio Sales Siqueira

## **Ferramenta para análise do Ibovespa em relação ao sentimento da mídia**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Trabalho aprovado. Brasília, DF, 12 de maio de 2022:

---

**Prof. Dr. John Lenon Cardoso  
Gardenghi**  
Orientador

---

**Prof. Dr. Glauco Vitor Pedrosa**  
Convidado 1

---

**Prof. Dr. Daniel Sundfeld Lima**  
Convidado 2

Brasília, DF  
2022



# Agradecimentos

Primeiramente gostaria de agradecer aos meus familiares e amigos que estiveram junto comigo durante todo esse percurso, mas em especial, a minha mãe Lourdes e minha tia de consideração Célia. Por fim, agradeço ao professor orientador John Gardenghi pela disponibilidade e esclarecimentos para a realização dessa monografia.





# Resumo

Nos últimos anos, a disponibilidade de notícias diárias publicadas tem sido crescente. Além disso, o consumo dessas notícias em tempo real pelo investidor tem se tornado cada vez mais acessível, tendo em vista as redes sociais e portais *online*. Esse volume de notícias, por sua vez, pode ter um impacto potencialmente significativo no movimento das cotações do mercado brasileiro, haja vista que essas notícias são rapidamente consumidas pelos investidores, que por sua vez, influenciam preços com movimentos em massa. Nesse contexto, foi proposto um simulador para suporte em decisões de compra e venda a partir dos métodos de análise técnica MME e MACD junto ao índice de sentimento diário da mídia sobre determinadas empresas que compõem o Ibovespa. Através do processamento de linguagem natural, as notícias foram rotuladas como positivas, negativas e neutras. Dentre os resultados mais relevantes, que inclusive superaram o retorno do Ibovespa no mesmo período, foram obtidos um lucro acumulado de 11,4% e 8,00%, respectivamente nos ativos ITUB4 e BRDC4, utilizando a técnica MME tomando o índice de sentimento dessas empresas como base de cálculo.

**Palavras-chaves:** análise de sentimento. notícia. Ibovespa.



# Abstract

In recent years, the availability of daily published news has been increasing. In addition, the consumption of this news in real time by investors has become increasingly accessible, given the social networks and portals *online*. This volume of news, can have a potentially significant impact on the movement of prices in the Brazilian market, given that this news is quickly consumed by investors, which can influence prices with massive movements. In this context, a simulator was proposed to support buying and selling decisions based on the MME and MACD technical analysis methods along with the daily media sentiment index on certain companies that make up the Ibovespa. Through natural language processing, the news was labeled as positive, negative and neutral. Among the most relevant results, which even surpassed the Ibovespa return in the same period, an accumulated profit of 11.4% and 8.00% were obtained, respectively, in ITUB4 and BRDC4 assets, using the MME technique taking the index of sentiment of these companies as a basis for calculation.

**Key-words:** sentiment analysis. news. Ibovespa.



# Lista de ilustrações

|                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Arquitetura da solução.<br>Autoria própria. . . . .                                                                    | 38 |
| Figura 2 – Volume de tweets por ação 17 de fevereiro de 2022 e 03 de abril de 2022.<br>Fonte: Autoria própria. . . . .            | 42 |
| Figura 3 – Tweet do Valor Econômico envolvendo a Petrobras e a guerra na Ucrânia.<br>Conta do Valor Econômico no Twitter. . . . . | 42 |
| Figura 4 – Volume de tweets por conta no twitter.<br>Autoria própria. . . . .                                                     | 43 |
| Figura 5 – MME (20, 5) com valores de sentimento para ITUB4.<br>Autoria própria. . . . .                                          | 51 |
| Figura 6 – Ordens realizadas para ITUB4 com MME (20,5).<br>Autoria própria. . . . .                                               | 51 |
| Figura 7 – Balanço consolidado para ITUB4 com MME (20,5).<br>Autoria própria. . . . .                                             | 51 |
| Figura 8 – MACD com valores de sentimento para ITUB4.<br>Autoria própria. . . . .                                                 | 52 |
| Figura 9 – Ordens realizadas para ITUB4 com MACD.<br>Autoria própria. . . . .                                                     | 52 |
| Figura 10 – Balanço consolidado para ITUB4 com MACD.<br>Autoria própria. . . . .                                                  | 52 |
| Figura 11 – MME (20, 5) com valores de sentimento para BBDC4.<br>Autoria própria. . . . .                                         | 53 |
| Figura 12 – Ordens realizadas para BBDC4 com MME (20,5).<br>Autoria própria. . . . .                                              | 54 |
| Figura 13 – Balanço consolidado para BBDC4 com MME (20,5).<br>Autoria própria. . . . .                                            | 54 |
| Figura 14 – MACD com valores de sentimento para BBDC4.<br>Autoria própria. . . . .                                                | 55 |
| Figura 15 – Ordens realizadas para BBDC4 com MACD.<br>Autoria própria. . . . .                                                    | 55 |
| Figura 16 – Balanço consolidado para BBDC4 com MACD.<br>Autoria própria. . . . .                                                  | 55 |



# Lista de tabelas

|                                                                 |    |
|-----------------------------------------------------------------|----|
| Tabela 1 – Questões de pesquisa . . . . .                       | 27 |
| Tabela 2 – Critérios de Inclusão . . . . .                      | 28 |
| Tabela 3 – Critérios de Exclusão . . . . .                      | 28 |
| Tabela 4 – Critérios de Qualidade . . . . .                     | 28 |
| Tabela 5 – Total de publicações encontradas por base . . . . .  | 29 |
| Tabela 6 – Total de publicações selecionadas por base . . . . . | 29 |
| Tabela 7 – Publicações destacadas . . . . .                     | 30 |
| Tabela 8 – Requisitos Funcionais . . . . .                      | 33 |
| Tabela 9 – Requisitos Não Funcionais . . . . .                  | 34 |
| Tabela 10 – Resultado das simulações . . . . .                  | 56 |





# Lista de abreviaturas e siglas

|          |                                                                     |
|----------|---------------------------------------------------------------------|
| API      | Interface de Programação de Aplicativos                             |
| B3       | Brasil, Bolsa, Balcão                                               |
| Batch    | Lote de informações                                                 |
| BM&F     | Bolsa de Mercadorias e Futuros                                      |
| CETIP    | Central de Custódia e Liquidação Financeira de Títulos              |
| Juros DI | Juros de Depósito Interbancário                                     |
| MACD     | Convergência/Divergência de Média Móvel                             |
| MME      | Médias Móveis Exponenciais                                          |
| HTTP     | Protocolo de Transferência de Hipertexto                            |
| IPO      | Oferta Pública Inicial                                              |
| ISN      | Índice de Sentimento de Notícias                                    |
| NLP      | Processamento de linguagem natural                                  |
| PICOC    | Estratégia para definição da pergunta de pesquisa                   |
| Pooling  | Verificação contínua de outros programas para consulta de um estado |
| Selic    | Taxa básica da economia                                             |
| Token    | Coleções de caracteres                                              |



# Sumário

|          |                                          |           |
|----------|------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>                        | <b>19</b> |
| 1.1      | Contextualização                         | 19        |
| 1.2      | Definição do problema                    | 20        |
| 1.3      | Objetivos                                | 20        |
| 1.4      | Organização do trabalho                  | 21        |
| <b>2</b> | <b>CONCEITOS E REVISÃO BIBLIOGRÁFICA</b> | <b>23</b> |
| 2.1      | Mercado de ações no Brasil               | 23        |
| 2.2      | Redes Sociais e Sentimento da mídia      | 24        |
| 2.3      | Análise de Sentimento                    | 25        |
| 2.4      | Revisão Sistemática de Literatura        | 26        |
| 2.4.1    | Planejamento                             | 27        |
| 2.4.1.1  | Questões de pesquisa                     | 27        |
| 2.4.1.2  | String de Busca                          | 27        |
| 2.4.1.3  | Bases de Busca                           | 27        |
| 2.4.1.4  | Critério de inclusão e exclusão          | 28        |
| 2.4.1.5  | Avaliação de qualidade                   | 28        |
| 2.4.2    | Execução                                 | 28        |
| 2.4.3    | Resultado e Discussão                    | 29        |
| <b>3</b> | <b>ARQUITETURA DO SISTEMA</b>            | <b>33</b> |
| 3.1      | Requisitos                               | 33        |
| 3.1.1    | Funcionais                               | 33        |
| 3.1.2    | Não Funcionais                           | 34        |
| 3.2      | Coletor de notícias                      | 34        |
| 3.2.1    | Twitter e veículos de comunicação        | 34        |
| 3.3      | Analisador de sentimento                 | 35        |
| 3.3.1    | Pré-processamento                        | 35        |
| 3.3.2    | Abordagem Lexical                        | 35        |
| 3.3.3    | Índice de Sentimento                     | 36        |
| 3.4      | Arquitetura da solução                   | 37        |
| <b>4</b> | <b>ESTRATÉGIA PARA NEGOCIAÇÕES</b>       | <b>41</b> |
| 4.1      | Modelagem do experimento                 | 41        |
| 4.2      | Dados da bolsa e Indicadores de Análise  | 44        |
| 4.2.1    | Análise Fundamentalista                  | 44        |

|            |                                                     |           |
|------------|-----------------------------------------------------|-----------|
| 4.2.2      | Análise Técnica . . . . .                           | 45        |
| 4.2.2.1    | Médias móveis exponenciais . . . . .                | 46        |
| 4.2.2.2    | Convergência/Divergência de médias móveis . . . . . | 47        |
| <b>5</b>   | <b>RESULTADOS . . . . .</b>                         | <b>49</b> |
| <b>5.1</b> | <b>Dados e ferramentas . . . . .</b>                | <b>49</b> |
| <b>5.2</b> | <b>Resultados da simulação . . . . .</b>            | <b>50</b> |
| 5.2.1      | Itaú com MME . . . . .                              | 50        |
| 5.2.2      | Itaú com MACD . . . . .                             | 51        |
| 5.2.3      | Bradesco com MME . . . . .                          | 53        |
| 5.2.4      | Bradesco com MACD . . . . .                         | 54        |
| <b>5.3</b> | <b>Considerações finais . . . . .</b>               | <b>55</b> |
| <b>6</b>   | <b>CONCLUSÃO . . . . .</b>                          | <b>57</b> |
|            | <b>REFERÊNCIAS . . . . .</b>                        | <b>59</b> |

# 1 Introdução

## 1.1 Contextualização

O mercado de ações brasileiro, nos últimos anos, tem crescido como nunca visto outrora. A bolsa brasileira, B3, registra 3,8 milhões de contas cadastradas no 1º semestre de 2021, respectivo a 43% de pessoas físicas a mais que no semestre anterior de acordo o site oficial da B3<sup>1</sup>. Com o crescimento dessa demanda, há uma necessidade crescente de se estudar o mercado financeiro brasileiro, buscando melhor compreender suas oscilações e o comportamento do investidor.

A extração e análise da opinião pública é um tema que tem adquirido notoriedade na comunidade científica e no mundo dos negócios, uma vez que é possível extrair informações desses dados e utilizá-los em tomadas de decisão, em estratégias de empresas, análise de mercado, dentre outros. Na literatura científica, é possível encontrar trabalhos que exploram questões relacionadas à oscilação de preços decorrente da influência de notícias nos movimentos dos investidores, mesmo que de forma estritamente especulativa.

A cotação de uma determinada ação é formada com base em ordens de negociação emitidas pelos investidores do mercado. Dessa forma, o preço justo da ação é estipulado tendo em vista o fluxo de oferta e procura de cada papel. O processo de valorização ou desvalorização de uma ação está relacionada ao comportamento histórico dos preços e às perspectivas futuras de desempenho da empresa emissora da ação. Nesse sentido, essas perspectivas podem ser influenciadas por notícias pertinentes ao mercado de atuação da empresa, ou seja, essas notícias consumidas pelo investidor podem causar uma influência no valor de tanto de um ativo quanto em todo mercado.

Com o crescimento da produção de conteúdo na web, indivíduos e organizações estão usando, cada vez mais, o conteúdo nesses meios de comunicação para a tomada de decisões. No entanto, encontrar e monitorar a opinião em sites na *Web* e destilar as informações contidas neles permanece uma tarefa formidável devido à proliferação de diversos locais. Cada site normalmente contém um grande volume de texto de opinião que nem sempre é facilmente decifrado em longos *blogs* e postagens em fóruns. Dessa forma, entende-se que sistemas automatizados de análise de sentimento são necessários (LIU, 2012).

---

<sup>1</sup> [https://www.b3.com.br/pt\\_br/noticias/porcentagem-de-investidores-pessoa-fisica-cresce-na-b3.htm](https://www.b3.com.br/pt_br/noticias/porcentagem-de-investidores-pessoa-fisica-cresce-na-b3.htm)

## 1.2 Definição do problema

Dada a possível relação existente entre o sentimento da mídia e o comportamento do investidor influenciado por essas notícias, ressalta-se o trabalho de [Tetlock \(2007\)](#) que concluiu que altos valores de pessimismo da mídia induzem pressão negativa sobre os preços de mercado e valores altos ou baixos de pessimismo incomuns como, por exemplo, em crises levam a um volume de negociação de mercado temporariamente alto.

Ademais, em finanças comportamentais, os pesquisadores intensificaram na última década seus esforços para entender como o sentimento afeta os tomadores de decisão, instituições e mercados individuais ([KEARNY; LIU, 2013](#)). Dessa forma, a disponibilização de ferramentas para análise da influência da mídia sobre os leitores, e logo sobre a oscilação de preços, tornaram-se necessárias.

Nesse contexto, esse trabalho se orienta no sentido de automatizar essas análises através de um índice de sentimento de mídia diário e proposição de um simulador de compra e venda de ativos baseados em métodos de análise técnica junto ao índice de sentimento proposto. Para isso, a partir da extração de tweets e análise de sentimento, técnica de NLP que busca analisar opiniões, sentimentos e emoções presentes em dados, esse trabalho utilizará uma abordagem baseada em léxico, que envolve o cálculo da polaridade de um texto a partir da orientação semântica das palavras.

## 1.3 Objetivos

O objetivo geral desse trabalho consiste no provisionamento de uma ferramenta web, que a partir da extração diária de tweets de contas verificadas, gere pontuações de sentimento para os tweets e, conseqüentemente, um índice com valores diários desses tweets pontuados para os ativos ITUB4 (Itaú), PETR4 (Petrobras), BBDC4 (Bradesco) e B3SA3 (B3). Esses ativos foram escolhidos pelo fato de terem grande participação na composição do Ibovespa. Com a obtenção desse índice de sentimento diário, será proposto um simulador que utiliza métodos de análise técnica junto ao índice proposto para indicações de compra e venda dos ativos elicitados. A partir do objetivo geral podem ser delimitados os seguintes objetivos específicos:

- Desenvolver um processo automático para extração, tratamento, pontuação e persistência para os tweets;
- Desenvolver um analisador de sentimentos de texto;
- Desenvolver uma interface *web* para exibição das indicações de compra e venda, além de gráficos referentes aos métodos de análise técnica;

## 1.4 Organização do trabalho

Este trabalho foi organizado em 6 capítulos, que cada capítulo é descrito, detalhadamente, a seguir:

- Capítulo 2: São apresentados os conceitos fundamentais para a concepção desse trabalho, além do procedimento adotado para realização da Revisão Sistemática de Literatura;
- Capítulo 3: Apresenta a metodologia e tecnologias utilizadas no trabalho;
- Capítulo 4: Apresenta as estratégias de negociação utilizadas no simulador, bem como discorre sobre o conjunto de dados coletado;
- Capítulo 5: Demonstra o resultado das 4 simulações realizadas, onde em cada simulação será apresentado os cruzamentos nos gráficos MME e MACD, além das ordens realizadas e o balanço geral consolidado de cada simulação;
- Capítulo 6: Apresenta as conclusões dos trabalho, discussões sobre o capítulo de resultados e comentários para trabalhos futuros.





## 2 Conceitos e Revisão Bibliográfica

### 2.1 Mercado de ações no Brasil

A bolsa de valores é um local onde as repartições percentuais de empresas listadas são negociadas. A B3 é a única bolsa de valores no Brasil, sendo proveniente da fusão da Bovespa, BM&F e Cetip. Segundo seu site oficial<sup>1</sup>, a instituição promove a intermediação do mercado de capitais, onde dentre as principais responsabilidades implanta e provê mecanismos para: Negociação de ações, Derivativos de ações, Títulos de renda fixa, Derivativos financeiros e Commodities segundo seu site oficial. Na bolsa de valores, há especificações no que tange o mercado de ações. E de acordo com a autoridade reguladora, o mercado de ações pode ser subdividido em primário e secundário.

O mercado primário compreende o lançamento de novas ações no mercado e é uma forma de captação de recursos para o caixa da empresa<sup>2</sup>. Como mercado primário, o mercado de ações permite que as empresas emitam e vendam suas ações ao público comum pela primeira vez por meio do processo de ofertas públicas iniciais (IPO). Essa atividade ajuda as empresas a levantar o capital necessário de investidores.

Uma vez que os novos títulos tenham sido vendidos no mercado primário, eles são negociados no mercado secundário, onde ocorre a troca de propriedade de título. O mercado secundário abrange todas as operações futuras dessas ações, compradas pelo primeiro investidor, denominado subscritor. E por conseguinte, pessoas físicas e jurídicas podem negociar os ativos em que tenham posse, realizando assim, movimentos de compra e venda. Dessa forma, o mercado secundário garante liquidez aos papéis emitidos no mercado primário.

Segundo o site oficial da B3<sup>3</sup>, o Índice Bovespa ou Ibovespa é o principal indicador de desempenho das ações negociadas na B3 e reúne as empresas mais importantes do mercado de capitais brasileiro. Criado em 1968, o índice é reavaliado a cada quatro meses e é resultado de uma carteira teórica de ativos. É composto pelas ações e *units* de companhias listadas na B3 que atendem aos critérios descritos na sua metodologia, o que corresponde a cerca de 80% do número de negócios e do volume financeiro do mercado de capitais do Brasil.

---

<sup>1</sup> [https://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/mercado-de-acoes/mercado-de-acoes-no-brasil.htm](https://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/mercado-de-acoes/mercado-de-acoes-no-brasil.htm)

<sup>2</sup> <https://blog.ativainvestimentos.com.br/mercado-primario-e-mercado-secundario>

<sup>3</sup> [https://www.b3.com.br/pt\\_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm](https://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm)

## 2.2 Redes Sociais e Sentimento da mídia

Uma rede social pode ser entendida como sendo um grupo de pessoas que se interagem através de qualquer mídia social, gerando vínculos e conteúdo para engajamento. Segundo [Boyd e Ellison \(2007\)](#), a rede social é um serviço web que permite indivíduos construir perfis públicos ou semi-públicos dentro de um sistema, articular uma lista de outros usuários com os quais compartilham conexões e visualizar e percorrer suas listas de conexões e outras feitas por outros no sistema. Nessa conjuntura, é possível extrair uma imensidão de opiniões, notícias e tendências que afetam várias áreas da sociedade.

Nesse contexto, foi proposto a utilização da rede social Twitter<sup>4</sup>. Segundo sua página de central de ajuda oficial<sup>5</sup>, o Twitter é um serviço para amigos, familiares e colegas de trabalho se comunicarem e permanecerem conectados por meio da troca de mensagens rápidas e frequentes. As pessoas publicam tweets, que podem conter fotos, vídeos, links e texto. Essas mensagens são postadas em seu perfil, enviadas para seus seguidores e podem ser pesquisadas na pesquisa do Twitter.

Dentro desse contexto, além da utilização do Twitter pelo público geral, o Twitter tem sido utilizado por diversos veículos de comunicação e jornalistas. Dessa forma, a rede social tem sido utilizada para propagação de notícias que por sua vez, são consumidas pelos usuários e que também podem emitir opiniões sobre os eventos ocorridos. Neste contexto, a extração de notícias publicadas por emissores de notícias é uma das temáticas exploradas nessa monografia.

Segundo [Kearny e Liu \(2013\)](#), o sentimento expresso pela mídia é caracterizado pela positividade ou negatividade contida nas notícias, além de comentários ou relatórios de analistas. Estes textos são relevantes para as condições econômicas gerais, para condições gerais do mercado financeiro e para indústrias e empresas individuais. O sentimento em artigos da mídia sobre as condições econômicas gerais é uma escolha apropriada para estudar os padrões e atividades de preços em todo o mercado, enquanto notícias e relatórios de analistas específicos da indústria ou da empresa são mais apropriados para analisar os preços das ações no nível da empresa, retornos, volumes de negociação e outros indicadores de desempenho e atributos.

Ademais, a utilização de fontes de notícias amplas, em vez de um pequeno número de jornais específicos, são mais recomendadas, tendo em vista a atenuação da probabilidade de obtenção de sentimento tendencioso de algumas fontes que não representam o mercado geral. E com o Twitter, essa amplitude torna-se possível, tendo em vista que os maiores veículos de comunicação possuem contas no Twitter.

---

<sup>4</sup> <https://about.twitter.com/en>

<sup>5</sup> <https://help.twitter.com/en/resources/new-user-faq>

## 2.3 Análise de Sentimento

Segundo Liu (2012), a análise de sentimento, também chamada de mineração de opinião, é o campo de estudo que analisa as opiniões das pessoas, sentimentos, avaliações, atitudes e emoções em relação a entidades como produtos, serviços, organizações, indivíduos, questões, eventos, tópicos e seus atributos. A análise de sentimento se concentra principalmente na obtenção de opiniões que expressam ou implicam sentimentos positivos ou negativos a partir de texto não estruturado.

No campo de finanças comportamentais, a avaliação da influência do sentimento dos investidores em suas decisões para realização de operações foi uma linha de pesquisa seguida por Barberis, Shleifer e Vishny (1998). No trabalho deles, foram modelados o sentimento dos investidores de maneira a mostrar como são formadas suas crenças baseados em evidências da Psicologia e de reações exageradas (*overreaction* ou *underreaction*).

A evidência de *underreaction* mostra que em horizontes de 1 a 12 meses, os preços dos títulos reagem pouco às notícias. Como consequência, as notícias são incorporadas apenas lentamente aos preços, que tendem a apresentar autocorrelações positivas nesses horizontes. A evidência de *overreaction* mostra que em horizontes mais longos, talvez de 3 a 5 anos, os preços dos títulos reagem exageradamente a padrões consistentes de notícias que apontam na mesma direção. Ou seja, títulos que tiveram um longo histórico de boas notícias tendem a ficar superfaturados e ter retornos médios baixos posteriormente (BARBERIS; SHLEIFER; VISHNY, 1998).

Quanto às abordagens para análise de sentimento, existem duas abordagens gerais, o léxico e as abordagens baseadas em aprendizado de máquina. A abordagem baseada em dicionário depende de listas predefinidas de palavras com cada palavra classificada como positiva, negativa ou neutra. A abordagem de aprendizado de máquina prevê o sentimento de qualquer conjunto de texto após modelos de treinamento com um grande conjunto de texto classificado manualmente (NGUYEN; CAVA, 2020).

De acordo com Liu (2012), a classificação de texto pode ser feita com base em 3 níveis de granularidade:

- Nível do documento: A análise a nível de documento classifica se um documento expressa um sentimento positivo ou negativo em um contexto geral. Este nível de análise assume que cada documento expressa sentimento sobre uma única entidade ou tema. Assim, não é aplicável a documentos que avaliem ou compare vários temas;
- Nível de frase: A tarefa neste nível vai para as sentenças e determina se cada frase expressou uma opinião positiva, negativa ou neutra. Este nível de análise não assume um nível de entidade único;

- **Nível de entidade:** Em vez de olhar para construções de linguagem (documentos, parágrafos, sentenças, cláusulas ou frases), o nível de entidade responsabiliza-se pela opinião. É baseado na ideia de que uma opinião consiste em um sentimento (positivo ou negativo) e um alvo (de opinião). Por exemplo, embora a frase “Ibovespa futuro sobe e sinaliza recuperação após pior semana do ano, mas a expectativa de subida dos juros DI trazem incertezas ao mercado” tenha um tom positivo, ela não é inteiramente positiva. Isto é, a frase é positiva sobre o Ibovespa, mas negativa sobre o juros DI, evidenciando mais de uma entidade nessa sentença.

## 2.4 Revisão Sistemática de Literatura

A revisão de literatura é atividade essencial no desenvolvimento de trabalhos acadêmicos e científicos. A realização de uma revisão de literatura evita a duplicação de pesquisas ou, quando for de interesse, o reaproveitamento e a aplicação de pesquisas em diferentes escalas e contextos. Além disso, permite observar possíveis falhas nos estudos realizados; conhecer os recursos necessários para a construção de um estudo com características específicas; desenvolver estudos que cubram brechas na literatura; propor temas, problemas, hipóteses e metodologias inovadoras de pesquisa.

A revisão sistemática de literatura é uma técnica de revisão de literatura que tem como intuito coletar e avaliar estudos relevantes acerca de um objeto de pesquisa. É uma modalidade de pesquisa, que segue protocolos específicos. Está focada no seu caráter de reprodutibilidade por outros pesquisadores, apresentando de forma explícita as bases de dados bibliográficos que foram consultadas, as estratégias de busca empregadas em cada base, o processo de seleção dos artigos científicos, os critérios de inclusão e exclusão dos artigos e o processo de análise de cada artigo. O processo é delimitado em 3 fases principais:

- **Planejamento:** Nesta etapa é definido as questões de pesquisa e o protocolo da revisão;
- **Execução:** Nesta etapa é realizado o procedimento especificado na seção de planejamento, seleção, avaliação da qualidade e síntese das publicações encontradas;
- **Resultado e Discussão:** Nesta última etapa é apresentado o resultado e discussão das publicações.

Na seção a seguir, aplica-se as etapas do processos de revisão sistemática de literatura no contexto da aplicação desse trabalho, em que identifica-se os trabalhos mais relevantes da área.

## 2.4.1 Planejamento

### 2.4.1.1 Questões de pesquisa

Nesta etapa do trabalho foram definidas as questões centrais correlatas ao objeto de estudo: Análise de sentimento da mídia e a relação com o Ibovespa. Disso, decorre-se as seguintes questões que norteiam a revisão na Tabela 1.

Tabela 1 – Questões de pesquisa

| Identificador | Questão                                               |
|---------------|-------------------------------------------------------|
| Q01           | Como quantificar e avaliar o sentimento da mídia?     |
| Q02           | A correlação entre notícias e os movimentos da bolsa? |

Fonte: Autoria própria.

### 2.4.1.2 String de Busca

Nesta etapa do trabalho foram definidas as palavras-chaves e sinônimos utilizados para a construção da string de busca. Os idiomas escolhidos foram inglês, devido à abrangência maior de pesquisa, e o português, dada a relevância de busca por publicações nacionais acerca do objeto de estudo.

Dessa maneira, as seguintes palavras-chaves em inglês foram definidas: *sentiment analysis, news, stock e market*. Enquanto as seguintes palavras-chaves em português foram definidas: “análise de sentimento”, “notícia”, “mercado”, “ações”, “ibovespa”, “bovespa”.

Disso, foi definido a string de busca, usando PICOC, correlata ao objeto de estudo. Além disso, junto aos termos citados e restrições de busca em título e resumo (TITLE-ABS-KEY), foram adicionados os operadores lógicos AND e OR, foi obtida a seguinte string de busca.

**TITLE-ABS-KEY(( “sentiment analysis” OR “análise de sentimento”) AND (“notícia” OR “news” OR “twitter”) AND (“ibovespa” OR “bovespa” OR “mercado” OR “ações” OR “stock market”) AND NOT (“forecasting” OR “prediction”))**

### 2.4.1.3 Bases de Busca

Com a finalidade de garantir a qualidade necessária para as referências bibliográficas da pesquisa, foram selecionadas as bases mais relevantes para a área de pesquisa. Foram selecionada as seguintes bases:

- Scopus;
- Scielo;

- BDM UnB;
- Google Acadêmico.

#### 2.4.1.4 Critério de inclusão e exclusão

A partir de busca efetuada, definida na seção anterior, foram criados critérios de inclusão e exclusão com o propósito de garantir a seleção dos artigos mais apropriados ao objeto de estudo. Os critérios de inclusão são detalhados nas Tabelas 2 e 3.

Tabela 2 – Critérios de Inclusão

| Identificador | Critério                                                            |
|---------------|---------------------------------------------------------------------|
| CI01          | Possuir DOI                                                         |
| CI02          | Palavras-chave existam nos campos: título, resumo ou palavras chave |
| CI03          | Relacionado à bolsa de valores de qualquer país (Ex.: B3, NYSE)     |

Fonte: Autoria própria.

Tabela 3 – Critérios de Exclusão

| Identificador | Critério                                                 |
|---------------|----------------------------------------------------------|
| CE01          | Relacionados à previsão de cotação de ativos financeiros |
| CE02          | Publicações pagas                                        |
| CE03          | Publicações repetidas                                    |

Fonte: Autoria própria.

#### 2.4.1.5 Avaliação de qualidade

Após o filtro das publicações segundo os critérios de inclusão e exclusão, foi realizada uma avaliação de qualidade nessas publicações, seguindo alguns dos critérios definidos por [Gegentana \(2011\)](#), para checagem dos seguintes critérios na Tabela 4.

Tabela 4 – Critérios de Qualidade

| Identificador | Critério                                                |
|---------------|---------------------------------------------------------|
| CQ01          | O resumo é relevante para o campo de estudo?            |
| CQ02          | A introdução afirma claramente a questão e o resultado? |
| CQ03          | A conclusão foi feita de maneira apropriada?            |

Fonte: Autoria própria.

## 2.4.2 Execução

Nessa etapa foi executada a revisão sistemática de literatura conforme planejado na fase anterior. Dessa forma a Tabela 5, apresenta as publicações encontradas nas bases:

Tabela 5 – Total de publicações encontradas por base

| <b>Base</b>      | <b>Número de publicações</b> |
|------------------|------------------------------|
| Scopus           | 28                           |
| Scielo           | 1                            |
| BDM UNB          | 1                            |
| Google Acadêmico | 11.300                       |

Fonte: Autoria própria.

Após a seleção preliminar, as publicações foram incluídas e descartadas conforme os critérios de inclusão e exclusão definidos nas Tabelas 3 e 4, além dos critérios de qualidade na Tabela 4. Após essa segunda avaliação, a Tabela 6 apresenta o resultado das 6 publicações selecionadas.

Tabela 6 – Total de publicações selecionadas por base

| <b>Base</b>      | <b>Número de publicações</b> |
|------------------|------------------------------|
| Scopus           | 2                            |
| Scielo           | 1                            |
| BDM UNB          | 1                            |
| Google Acadêmico | 2                            |

Fonte: Autoria própria.

### 2.4.3 Resultado e Discussão

Após a realização dos procedimentos da revisão sistemática de literatura pode-se encontrar trabalhos correlatos. Disso, foram destacadas as publicações subsequentes que seguiram uma linha de pesquisa análoga ao objeto de estudo de análise de sentimento da mídia na Tabela 7. Além disso, foram apresentados os resultados principais desses trabalhos destacados.

O artigo de [Nguyen e Cava \(2020\)](#) foi publicado pelo departamento de pesquisa do Banco Central da Austrália com o objetivo de medir a atividade macroeconômica da economia australiana em tempo real. O estudo segue a abordagem baseada em dicionário para construir Índice de Sentimento de Notícias. Esse índice mede o saldo líquido de palavras usadas por jornalistas que são consideradas “positivas” e “negativas”. Os dados brutos usados na construção do índice consistem em notícias diárias extraídas do Dow Jones Newswires Archive (DNA). Cada artigo listado no banco de dados inclui metadados como horário de publicação, idioma, região e categoria. Para medir o sentimento de um conjunto de textos, ou seja, se a notícia é positiva ou negativa, utiliza-se o dicionário de [Loughran e McDonald \(2011\)](#).

Tabela 7 – Publicações destacadas

| Identificador | Publicação                                                                                     | Fonte            |
|---------------|------------------------------------------------------------------------------------------------|------------------|
| P01           | <i>News sentiment and the economy</i>                                                          | Google Acadêmico |
| P02           | <i>Measuring news sentiment</i>                                                                | Google Acadêmico |
| P03           | Análise de sentimentos do twitter como suporte aditivo para a previsão da volatilidade bitcoin | BDM UNB          |
| P04           | Pessimismo e incerteza das notícias e o comportamento dos investidores no Brasil               | Scielo           |
| P05           | <i>Analyzing stock market movements using Twitter sentiment analysis</i>                       | Scopus           |
| P06           | <i>Social Media and the stock markets: an emerging market perspective</i>                      | Scopus           |

Fonte: A autoria própria.

O trabalho realizado por [Shapiro, Sudhof e Wilson \(2017\)](#) foi publicado pelo departamento de pesquisa do Banco Central local de São Francisco. Nesse artigo, é demonstrado um índice construído a partir de pontuações de sentimento para artigos de notícias relacionados à economia usando uma abordagem lexical. Essa nova medida de série temporal de sentimento econômico é derivada de artigos de jornais econômicos e financeiros de janeiro de 1980 a abril de 2015. Na elaboração do índice foi levado em conta que as flutuações diárias na medida de sentimento tendem a ser ruidosas e com isso, foi construído um índice diário suavizado como uma média ponderada posterior dos dados brutos, com pesos que diminuem geometricamente com o tempo desde a publicação do artigo. Essa média ponderada é análoga a como os estoques de capital são geralmente medidos a partir de safras anteriores de investimento: os investimentos mais antigos contribuem menos de acordo com uma taxa de depreciação presumida. Assumiu-se uma taxa de depreciação de 5%; ou seja, a cada dia que passa, os artigos se tornam 5% menos relevantes para o sentimento de hoje.

No trabalho de [Neto \(2018\)](#), a partir do princípio de que existe uma relação entre o sentimento presente em publicações de redes sociais e o preço do *Bitcoin*, é proposto uma ferramenta que analisa o sentimento de *tweets* em tempo real e apresenta um índice de sentimento. Essa ferramenta tem o objetivo de apoiar a previsão do preço da criptomoeda, bem como especulação do seu valor. A análise de sentimento foi realizada a partir da coleta e filtro de publicações relacionadas ao tema, além da utilização do léxico de *Vader*. Como conclusão, pode-se observar que o sentimento acompanhou os movimentos nos preços da moeda, indicando os anseios, medos e desejos dos usuários e investidores, além de indicar também, movimentos de alta e de baixa nos preços.

No trabalho de [Galdi e Gonçalves \(2018\)](#), é estudado a relação entre o teor das edições diárias da mídia financeira especializada no Brasil, capturado por uma métrica de tom textual, e a rentabilidade e volatilidade dos índices de mercado. A base de dados estu-



dada contém 1.237 edições diárias do jornal Valor Econômico, compreendendo o período entre 02 de janeiro de 2012 e 30 de dezembro de 2016. Os resultados do trabalho indicam que o mercado avalia com maior peso palavras de incerteza e negativas divulgadas nas notícias.

No artigo de [Tushar e Saket \(2012\)](#), foi investigado a complexa relação entre a literatura do fórum de tweets com os instrumentos do mercado financeiro (como volatilidade, volume de negociação e preços de ações). Foram analisados os sentimentos de mais de 4 milhões de tweets entre junho de 2010 e julho de 2011 para DJIA, NASDAQ-100 e 13 outras ações tecnológicas de grande capitalização. Os resultados obtidos mostram alta correlação (até 0,88 para retornos) entre os preços das ações e os sentimentos do twitter. Além disso, usando a Análise de Causalidade de Granger, foi validado que o movimento dos preços das ações e índices são muito afetados no curto prazo pelas discussões no Twitter.

De acordo com o trabalho de [Agarwal, Kumar e Goel \(2021\)](#), foi discutido que existem inúmeros estudos que examinam o impacto das mídias sociais no desempenho do mercado de ações, mas há uma escassez de tais evidências nas economias emergentes. E nesse trabalho, foi realizado uma análise de sentimento dos tweets sobre as empresas indianas que fazem parte do Nifty50 ou de qualquer índice setorial, por um período de 15 meses. Os resultados dos testes de Causalidade de Granger indicam que os sentimentos do Twitter têm uma relação significativa com os índices relacionados aos setores bancário e financeiro das bolsas indianas. Além disso, os resultados da Função de Resposta ao Impulso revelam que, nos retornos do índice, o impacto dos sentimentos negativos permanece por mais tempo do que os sentimentos positivos.

Com esses trabalhos destacados, é possível responder algumas questões levantadas na RSL no que se refere ao método de classificação e a possível existência de uma correlação entre notícias e as flutuações do mercado de ações. Assim como no trabalho de [Nguyen e Cava \(2020\)](#), optou-se pela utilização do dicionário de domínio econômico de [Loughran e McDonald \(2011\)](#) para obtenção das pontuações de sentimento dos tweets, tendo em vista os resultados positivos e a linha de pesquisa ser análoga ao objetivo do presente trabalho. E além disso, através dos trabalhos de [Galdi e Gonçalves \(2018\)](#), [Tushar e Saket \(2012\)](#) e [Agarwal, Kumar e Goel \(2021\)](#), foi possível identificar que existem correções entre as notícias e a flutuação de preços, o qual o presente trabalho irá explorar através da proposição de simulações envolvendo compras e vendas baseadas no indicar de sentimento, e exposição dos resultados obtidos.



## 3 Arquitetura do Sistema

Neste capítulo, descreve-se a arquitetura do sistema. Na Seção 3.1 serão apresentados os Requisitos, na Seção 3.2 será apresentado o Coletor de notícias, na Seção 3.3 será apresentado o Analisador de sentimento e na Seção 3.4 será apresentado a Arquitetura da solução.

### 3.1 Requisitos

Nesse trabalho, os requisitos foram divididos em duas categorias, sendo requisitos funcionais e requisitos não funcionais. Os requisitos funcionais descrevem o comportamento do sistema. Essencialmente, eles descrevem o que o sistema deve ou não fazer e podem ser considerados em termos de como o sistema responde às entradas. Enquanto os requisitos não funcionais tratam como premissas e restrições técnicas de um projeto devem ser abordadas. Os requisitos não funcionais não devem afetar o comportamento básico do sistema.

#### 3.1.1 Funcionais

Conforme descrito anteriormente, os requisitos funcionais devem apresentar o comportamento básico do sistema. Dessa forma, os requisitos elicitados são respectivos aos módulos de coletor de notícias, analisador de notícias e interface, onde cada módulo será abordado em seções subsequentes. Por conseguinte, na Tabela 8 encontram-se descritas as funcionalidades principais do sistema.

Tabela 8 – Requisitos Funcionais

| Identificador | Requisito                                                              |
|---------------|------------------------------------------------------------------------|
| RF01          | O sistema deve coletar notícias                                        |
| RF02          | O sistema deve armazenar notícias e metadados                          |
| RF03          | O sistema deve realizar a análise de sentimento de cada notícia        |
| RF04          | O sistema deve apresentar o índice de sentimento como série temporal   |
| RF05          | O sistema deve apresentar um gráfico de <i>candlestick</i> como opção  |
| RF06          | O sistema deve ser capaz de lidar com seleção de intervalo de tempo    |
| RF07          | O sistema deve ser capaz de lidar com seleção de determinados ativos   |
| RF08          | O sistema deve permitir a seleção de uma estratégia de análise técnica |

Fonte: Autoria própria.

### 3.1.2 Não Funcionais

Como dito anteriormente, os requisitos não funcionais têm a função de especificar diretrizes relacionados ao comportamento e restrições do sistema. Dessa forma, a especificação subsequente apresenta requisitos que tratam questões acerca do sistema operacional do servidor, restrições de tempo e usabilidade. Na Tabela 9 são apresentados os requisitos não funcionais do sistema.

Tabela 9 – Requisitos Não Funcionais

| Identificador | Requisito                                          |
|---------------|----------------------------------------------------|
| RNF01         | O sistema deve coletar as notícias a cada 1 hora   |
| RNF02         | O sistema deve armazenar as notícias a cada 1 hora |
| RNF03         | O sistema deve analisar as notícias a cada 1 hora  |
| RNF04         | O servidor do sistema deve funcionar 24/7          |
| RNF05         | A interface do sistema deve ser responsiva         |
| RNF06         | A interface do sistema deve ser <i>Web</i>         |

Fonte: Autoria própria.

## 3.2 Coletor de notícias

### 3.2.1 Twitter e veículos de comunicação

Tendo em vista que o trabalho visa a análise de sentimento de notícias publicadas por veículos de comunicação, foi necessário realizar a escolha de veículos de comunicação confiáveis e com o volume de notícias adequado. Dito isso, optou-se pela extração de notícias a partir da API oficial do Twitter<sup>1</sup>. Essa escolha tem como base o fato de os veículos de comunicação mais relevantes publicarem as notícias de seus respectivos portais também no Twitter. Dessa forma, foi possível obter um volume considerável de notícias a partir de uma variedade de contas verificadas por meio de uma interface de comunicação unificada.

Ademais, os tweets foram filtrados pelo código e nome respectivo a cada empresa e, apenas, tweets em língua portuguesa e contas verificadas. Além disso, retweets e tweets referentes à contas oficiais das empresas em questão foram removidos do filtro. Vale a pena ressaltar que os tweets foram extraídos, apenas, de contas verificadas com o intuito de obter mais notícias de contas de veículos de comunicação do que a opiniões do público em geral como demonstrado no Capítulo 4. Do Twitter, foram extraídos os seguintes dados:

- id: Identificador exclusivo do tweet;

<sup>1</sup> <https://developer.twitter.com/en/docs/twitter-api/tweets/search/api-reference/get-tweets-search-recent>

- `text`: Conteúdo do tweet;
- `created_at`: Hora de criação do tweet;
- `author_id`: Identificador único da conta que publicou o tweet.

## 3.3 Analisador de sentimento

### 3.3.1 Pré-processamento

Para obtenção de melhores resultados na fase de análise de notícias, é necessário, previamente, a realização de um pré-processamento de dados. Dessa forma, é possível abstrair a estrutura da língua com o intuito de obter apenas informações relevantes. Esse pré-processamento remove ruídos, tornando, assim, os dados menos esparsos. A seguir, seguem técnicas de pré-processamento que serão utilizadas.

- **Tradução**: Uma vez que o método de classificação utilizado encontra-se na língua inglesa, é necessário a tradução para inglês dos tweets obtidos previamente para atenuar a perda da carga semântica de cada tweet;
- **Tokenização**: O processo de tokenização tem como objetivo separar sentenças em *tokens*. A normalização é importante para obtenção de unidades lexicais que serão utilizados nas fases subsequentes;
- **Normalização**: Refere-se às tratativas como a remoção de caracteres especiais, remoção de *links* e transformação de letras maiúsculas para minúsculas;
- **Remoção de *stopwords***: Uma das tarefas muito utilizadas no pré-processamento de textos é a remoção de palavras irrelevantes. Esse método consiste em remover palavras muito frequentes, mas sem nenhum viés semântico, tais como artigos, preposições e conjunções. Dessa forma, são removidos informações irrelevantes para a análise a ser realizada posteriormente;
- **Remoção de numerais e pontuação**: Por fim, outra remoção necessária diz respeito à remoção dos numerais e pontuações presentes no texto. Analogamente à técnica da seção anterior, esses caracteres não agregam valor semântico à análise de sentimento.

### 3.3.2 Abordagem Lexical

A abordagem baseado em léxico utiliza recursos que podem ser chamados de léxicos de opinião, que associam palavras à sua orientação de sentimento. A justificativa de uso de dicionários de palavras para análise de sentimento consiste na premissa de que palavras isoladas podem ser consideradas como uma unidade de informação e, assim, podem gerar

indicações do sentimento do texto. Ademais, para a classificação de texto, são utilizados dicionários léxicos, que consistem em um conjunto de palavras previamente polarizadas.

Nesse trabalho, para medir o sentimento das notícias extraídas, ou seja, se a notícia é positiva ou negativa, será utilizado o dicionário de Loughran-McDonald<sup>2</sup>. Este léxico consiste em uma lista de palavras específicas para o domínio da economia e finanças. De acordo com Li, Wu e Wang (2020), com uma lista de palavras gerais não é aplicável a textos financeiros, no entanto, um dicionário de sentimento específico de finanças modela melhor os sentimentos do que outros dicionários de domínio geral.

O dicionário fornece um meio de classificar *tokens* como palavras reais, que é importante para a consistência na contagem de palavras e identifica listas de palavras de sentimento. No trabalho de Loughran e McDonald (2011) foram criados seis listas de palavras diferentes (negativa, positiva, incerteza, litigiosa, modal forte e modal fraca) examinando o uso de palavras em uma grande amostra de mais de 10.000 termos durante 1994-2008. A lista de palavras de Loughran e McDonald contém 354 palavras positivas e 2.337 palavras negativas.

Embora o dicionário seja apropriado ao domínio do trabalho, o idioma das notícias e do dicionário são diferentes. As palavras do dicionário de Loughran-McDonald são pertencentes a língua inglesa, enquanto que as notícias a serem consideradas no presente trabalho são pertencentes à língua portuguesa. No entanto, para contornar o problema relacionado a questão de idiomas distintos entre as notícias e o dicionário, será empregado um pré-processamento adicional de tradução. Esse pré-processamento consiste em realizar a tradução do dicionário ou das notícias mediante testes prévios para observar qual método possui a melhor performance. Para realizar a tradução, foi utilizado o recurso AWS Translate<sup>3</sup>.

### 3.3.3 Índice de Sentimento

Como mencionado na seção anterior, para medir o sentimento das notícias, será utilizado o dicionário de Loughran e McDonald (2011) e a análise de sentimento abordada será em nível de frase. Dentre as listas disponibilizadas no dicionário, serão utilizados apenas as listas de palavras positivas e negativas tendo vista que o índice proposto busca indicar, apenas, o grau de otimismo e pessimismo da mídia. Além disso, uma determinada notícia será dita positiva ou negativa de acordo com o saldo obtido no somatório do peso das palavras da notícia em questão.

Tendo em vista um plano cartesiano para a construção da série temporal do Índice de Sentimento de Notícias (ISN), o eixo das abscissas será correspondente ao período de tempo ( $t$ ), enquanto que o eixo das ordenadas será equivalente ao quociente entre a

<sup>2</sup> <https://sraf.nd.edu/textual-analysis/resources/>

<sup>3</sup> <https://aws.amazon.com/pt/translate/>

diferença entre o total de notícias positivas e negativas e o total de notícias analisadas, num instante de tempo ( $t$ ). O cálculo do índice é dado por:

$$ISN_t = \left( \frac{\sum p_{it} + \sum n_{it}}{pc_t + nc_t} \right) \times 10,$$

onde  $p_{it}$  e  $n_{it}$  são os pesos totais da  $i$ -ésima notícia positiva e negativa, respectivamente, no instante de tempo  $t$  e  $pc_t$  e  $nc_t$  são os total de notícias positivas e negativas no instante de tempo  $t$ . Vale mencionar que o índice  $ISN_t$  abrange valores contínuos no intervalo compreendido entre  $-1$  e  $1$ . No entanto para facilitar a visualização, os valores foram multiplicados por  $10$  e dessa forma, foi alterada, apenas, a escala dos valores de forma que a variação ficará compreendida entre  $-10$  e  $10$ .

### 3.4 Arquitetura da solução

A obtenção de notícias diárias foi realizada através de uma mecanismo de agendamento de tarefas. Por meio desse mecanismo, foi possível programar as tarefas de obtenção de tweets. Para isso, foram configuradas  $4$  *rules*<sup>4</sup> no EventBridge<sup>5</sup> responsáveis por disparar as Lambdas<sup>6</sup> a cada  $1$  hora. Cada Lambda, por sua vez, ficou responsabilizada pela extração automática de notícias do Twitter, pela persistência dessas notícias em um banco de dados e por enviar o *batch* de tweets coletados em uma fila gerenciada pela AWS, denominada SQS<sup>7</sup>.

Essa fila por sua vez é consumida por outra Lambda que realiza um *pooling* dessa fila e toda vez que há tweets disponíveis, realiza o processamento de tradução, limpeza, análise de sentimento, persistência dos dados tratados e valores de sentimento obtidos. O banco de dados utilizado é o MySQL<sup>8</sup> na versão  $8.0.26$  hospedado no RDS<sup>9</sup>.

Para realização das tarefas abordadas nas seções anteriores, foi adotada uma arquitetura orientada à eventos. No diagrama da Figura 1, é demonstrado, sequencialmente, a responsabilidade de cada processo e como eles se comunicam. Além disso, são descritos a sequência de eventos presentes no diagrama.

1. Para cada empresa, foi criada uma Lambda, cujo acionamento é realizado pelo EventBridge que possui  $4$  *rules* responsáveis por executar as lambdas a cada hora;
2. Cada Lambda busca os  $100$  tweets mais recentes, respectivos à cada empresa;

<sup>4</sup> <https://docs.aws.amazon.com/eventbridge/latest/userguide/eb-rules.html>

<sup>5</sup> <https://aws.amazon.com/pt/eventbridge/>

<sup>6</sup> <https://aws.amazon.com/pt/lambda/>

<sup>7</sup> <https://aws.amazon.com/pt/sqs/>

<sup>8</sup> <https://www.mysql.com/>

<sup>9</sup> <https://aws.amazon.com/pt/rds/>

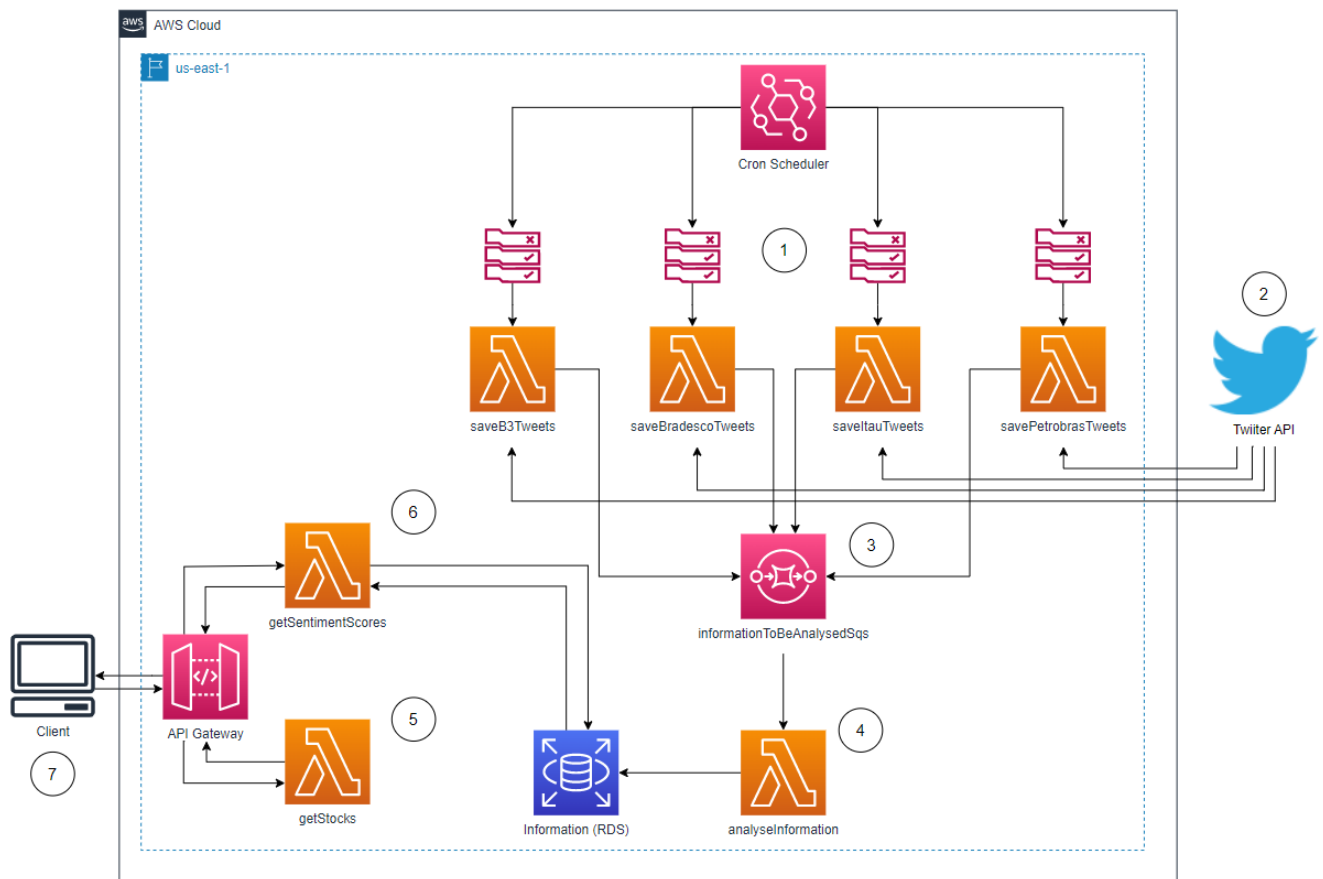


Figura 1 – Arquitetura da solução.  
Fonte: Autoria própria.

3. Os tweets coletados são enviados para a fila `informationToBeAnalysedSqs`;
4. A Lambda `analyseInformation` recebe o *batch* de tweets da fila, realiza o pré-processamento, aplica o analisar do sentimento e persiste a pontuação do tweet na tabela *Information* do MySQL hospedada em uma instância do RDS. Além da pontuação, é persistido o identificador único, a data, o autor e o identificador para cada empresa;
5. A Lambda `getStocks` disponibiliza dados sobre as cotações provenientes da biblioteca `Yahooquery`<sup>10</sup>;
6. A Lambda `getSentimentScores` é responsável por realizar a listagem dos tweets com suas respectivas pontuações na tabela *Information*;
7. Os dados das ações quanto das pontuações dos tweets são requisitados através de chamadas HTTP pelo *Frontend* ao API Gateway, que por sua vez endereça as requisições as Lambdas `getSentimentScores` e `getStocks`.

<sup>10</sup> <https://pypi.org/project/yahooquery/>



Após o processamento do *Backend*, os cálculos dos métodos de análise técnica, bem como as indicações de compra e venda, são realizados no *Frontend*. Os repositórios do projeto podem ser encontrado no Github no endereço: Organização do Github<sup>11</sup>. A organização dos repositórios pode ser descrita como:

- *Interface*: Hospeda o *Frontend* do simulador, cujo desenvolvimento foi feito com o framework React.js<sup>12</sup>;
- *Collector*: Hospeda o código responsável pela coleta, tratamento, persistência e listagem dos tweets pontuados como abordado na Figura 1. O código desse repositório foi desenvolvido com o Node.js<sup>13</sup>;
- *Stocks*: Hospeda o código responsável pela listagem dos dados respectivo à cotação das ações. O código desse repositório foi desenvolvido com o Python<sup>14</sup>;

---

<sup>11</sup> <https://github.com/News-Sentiment-Analisy-Tool>

<sup>12</sup> <https://pt-br.reactjs.org/>

<sup>13</sup> <https://nodejs.org/en/>

<sup>14</sup> <https://www.python.org/>



## 4 Estratégia para negociações

Neste capítulo, será apresentado o módulo de estratégia de negociações empregado a partir dos dados coletados e analisados. Essa apresentação abrange o desenvolvimento do simulador de compra e venda de determinados ativos para auxílio na estratégia do investidor. Os resultados obtidos a partir de experimentos realizados com ativos específicos em janelas de tempo pré-estipuladas serão apresentados no capítulo subsequente.

### 4.1 Modelagem do experimento

Para a experimentação com o simulador de compra e venda, foram utilizadas os ativos BBDC4 do banco Bradesco, ITUB4 do banco Itaú e B3SA3 da B3 devido, principalmente, a maior visualização de sinais propícios para entradas e saídas obtidas a partir do indicador de análise técnica MACD baseado nos valores diários do índice de sentimento para cada ativo, que será abordado em detalhes ainda nesse capítulo. Essas empresas estão entre as que possuem maior participações na bolsa de valores brasileira e por consequência, um volume significativo de negociações e volume de notícias diárias na mídia suficientes para realização dos experimentos.

O período de tempo para a análise é iniciado em 17 de fevereiro de 2022 e finalizado em 03 de abril de 2022. Para a realização da coleta, os tweets foram filtrados pelo código do ativo e o nome das respectivas empresas, além da exclusão de retweets e tweets de contas oficiais das próprias empresas, sendo elas: itaucultural, itau, itaucorretora, itauSocial, Bradesco, bradescoseguros e B3\_Oficial. Vale a pena ressaltar, também, que os tweets foram coletados, apenas, de contas verificadas com a prerrogativa de buscar mais notícias em detrimento de opiniões do público geral.

Os tweets foram coletados na medida em que foram publicados no Twitter a cada hora e além disso, esses tweets coletados no dia  $t$  contribuíram para afirmar um valor que representará a tendência ou o sentimento do dia  $t$ . Por fim, para cada dia desse intervalo, foram coletados, a partir do histórico de preços, os preços de abertura, mínimo, máximo e fechamento para as ações em questão.

Ademais, vale destacar que a janela de tempo para coleta não foi maior devido à restrição de acesso imposta pela API do twitter, o qual houve uma negativa quanto ao acesso de nível acadêmico, impossibilitando assim, a obtenção de tweets criados a partir de 2006.

No gráfico de barras da Figura 2, verifica-se que o volume de tweets referente à Petrobras foi superior em relação às outras empresas. Dentre alguns motivos que levaram a

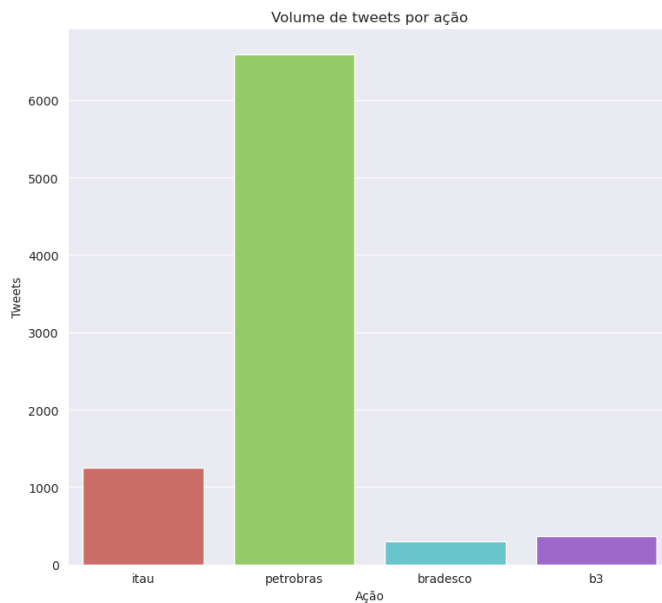


Figura 2 – Volume de tweets por ação 17 de fevereiro de 2022 e 03 de abril de 2022.  
Fonte: Autoria própria.

esse volume superior, ressalta-se principalmente à questão da guerra na Ucrânia iniciada em 24 de fevereiro de 2022 e todo impacto nos preços dos combustíveis no Brasil, levantando assim, um volume maior de tweets e notícias de forma geral sobre a empresa nesse período. Embora, o volume considerável, a Petrobras não apresentou sinais propícios para entrada e saída, e portanto, não foi considerada para a realização do experimento.



Figura 3 – Tweet do Valor Econômico envolvendo a Petrobras e a guerra na Ucrânia.  
Fonte: Conta do Valor Econômico no Twitter.

Tendo em vista que uma das prerrogativas do presente trabalho é a questão da análise de sentimento da mídia, optou-se pela obtenção das notícias por meio da API

do Twitter em detrimento da utilização de crawlers. Essa escolha foi tomada haja vista que os veículos de comunicação como Valor Econômico, Infomoney, Estadão Economia, Exame, Uol Economia, dentre outros, publicam notícias de seus sites no twitter.

Dessa forma, foi possível coletar um volume maior de notícias tanto dessas fontes quanto de outras. No gráfico da Figura 4, é apresentado percentualmente a participação das trintas contas do twitter que mais tiveram participação na obtenção das notícias desse trabalho e como esperado, a maioria dessas contas são provenientes de veículos de comunicação. Ademais, observa-se que o percentual de volume de notícias por perfil do Twitter se encontra diversificado, o que conseqüentemente, atenua o risco de enviesamento dos valores do índice de sentimento gerado a partir dos tweets.

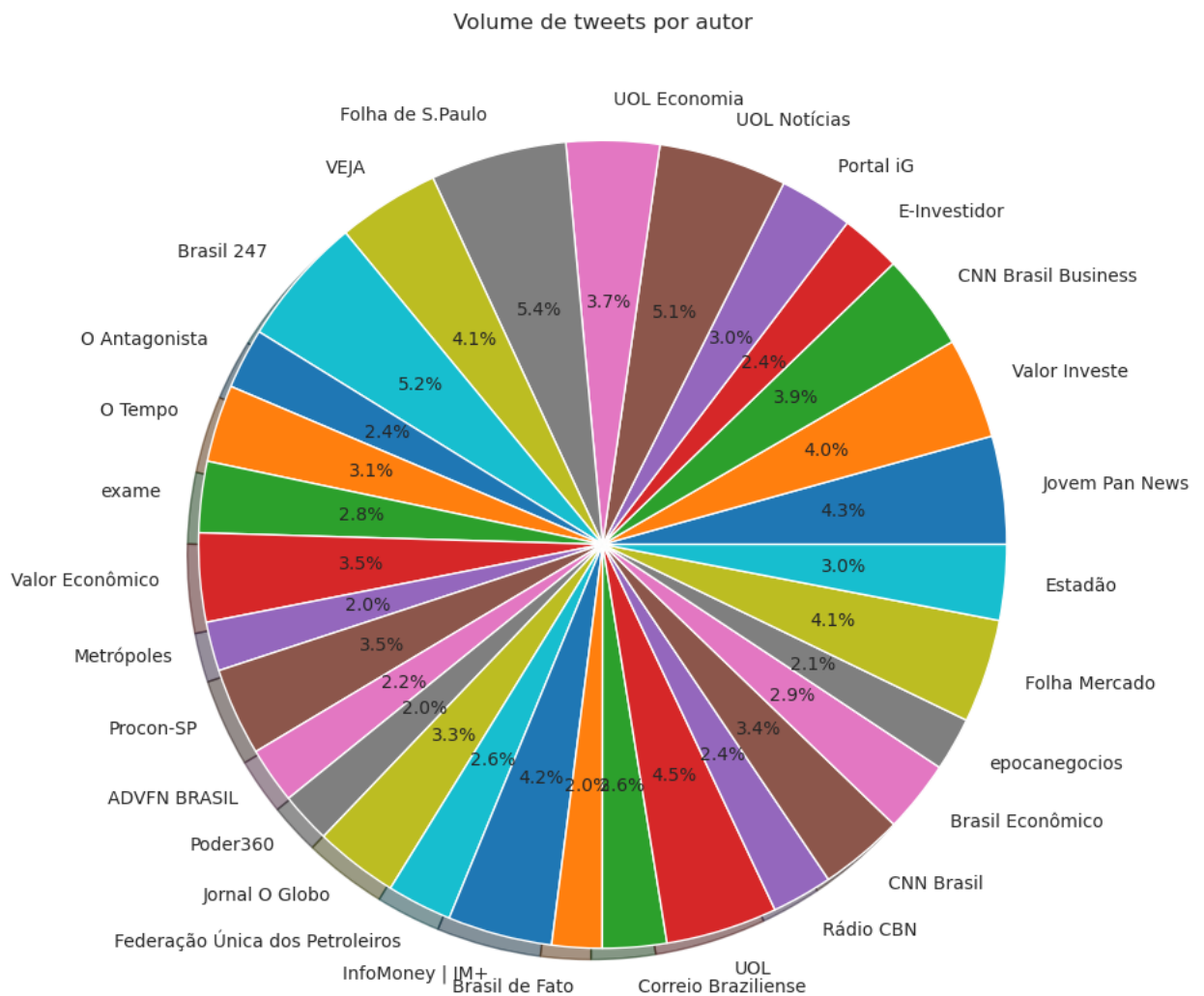


Figura 4 – Volume de tweets por conta no twitter.

Fonte: Autoria própria.

## 4.2 Dados da bolsa e Indicadores de Análise

Os dados diários dos ativos pré-determinados do mercado de ações do Brasil foram obtidos através da biblioteca Yahooquery<sup>1</sup>. Essa biblioteca é uma interface python para *endpoints* não oficiais da API do Yahoo Finance<sup>2</sup>, isto é, o pacote disponibiliza dados coletados a partir do *front-end* do Yahoo Finance. A justificativa para utilização desse recurso deve-se, principalmente, ao fato das interfaces oficiais do Yahoo Finance e Google Finance terem sido descontinuadas pelas respectivas empresas.

A partir dessa biblioteca, foram extraídos os preços de abertura, fechamento, mínimo e máximo para as ações ITUB4, BBDC4 e B3SA3 nos dias úteis o qual ocorrem os pregões na bolsa de valores do Brasil. As indicações de compra e vendas foram baseadas no método de análise técnica de convergência/divergência de médias móveis (MACD), utilizando o índice e sentimento como base ao invés do preço de fechamento. Vale ressaltar, que tradicionalmente, na análise de ações, existem duas ramificações principais que se dividem em análise fundamentalista e análise técnica.

### 4.2.1 Análise Fundamentalista

A análise fundamentalista pode ser subdividida em *top down* e o *bottom-up*. No *top down*, os movimentos dos preços dos ativos são regidos pelas variáveis macroeconômicas, taxa de juros, oferta de crédito, preço de commodities, dentre outros e resumidamente, primeiro analisa-se a economia como um todo, na sequência o setor de atuação da empresa e por último a saúde da empresa. No *bottom-up*, é o oposto, a variação de um determinado ativo é dada por variáveis microeconômicas, sendo priorizadas as expectativas e comportamento de cada empresa (MALTA; CAMARGOS, 2016).

A análise fundamentalista, de acordo com Sanvicente e Mellagi (2008), consiste no processamento de informações econômicas e políticas, além de variáveis que possam determinar o valor futuro de um ativo. A partir da análise, pode-se constatar determinado ativo encontra-se superavaliado ou subavaliado. O objetivo, normalmente, é adquirir o ativo quando o mesmo encontra-se subavaliado, ou seja, comprar um ativo que vale mais do que o mercado pensa traz a expectativa de bons retornos.

De acordo com Materlanc (2005), em um mercado eficiente, o preço de mercado das ações reflete os lucros potenciais da empresa e seus dividendos, os riscos do negócio, os riscos financeiros decorrentes da estrutura de capital da empresa e valor dos ativos, bem como as variáveis ambientais e outros fatores intangíveis que possam afetar o valor da empresa. Ou seja, reflete o valor presente do fluxo de caixa para os sócios.

Ademais, o valor intrínseco de cada ação é calculado por analistas fundamentalis-

<sup>1</sup> <https://pypi.org/project/yahooquery/>

<sup>2</sup> <https://finance.yahoo.com/>

tas por meio de relatórios de auditoria, demonstração de resultados, balanços trimestrais, registros de dividendos, registros de vendas, capacidade de gerenciamento e situação competitiva da empresa (MATARAZZO, 2009). Assim tem-se o valor intrínseco de cada ação com base na previsão do fluxo de caixa para os próximos anos, podemos dizer, que se o preço de mercado de uma ação é menor do que o seu valor intrínseco, então seu preço possui elevada probabilidade de subir, atraindo o interesse dos investidores (CHAVES; ROCHA, 2004).

Os indicadores da análise fundamentalista podem ser classificados nos seguintes grupos, conforme Lagioia (2011) que são:

- Mercado: tem a finalidade voltada para a análise de ações da empresa;
- Liquidez: mensuram a capacidade de uma empresa a arcar com o pagamento de suas obrigações a partir de seus ativos;
- Endividamento: medem a participação de recursos de terceiros com relação ao capital próprio da empresa;
- Rentabilidade: tem a função de evidenciar a rentabilidade dos capitais investidos.

#### 4.2.2 Análise Técnica

Segundo Chaves e Rocha (2004), a análise técnica se baseia em três princípios que auxiliam no desenvolvimento das demais técnicas, que são:

- Desconto: O primeiro diz que o preço desconta tudo e que esse desconto leva em consideração vários fatores como, por exemplo, econômicos e políticos, ou seja, para o autor todos estes fatores estarão contidos no valor/preço da ação;
- Tendência: Já o segundo princípio, relata que o preço age por tendência, isto é, o preço acompanha as decisões dos acionistas, tanto positiva, quanto negativamente, ou vice-versa, originando as tendências ascendentes e descendentes;
- Emocional: O terceiro princípio abrange a parte emocional dos agentes que são movidos por perdas e ganhos, sendo representados por padrões que vão se repetindo no decorrer do tempo.

A Análise técnica busca compreender o comportamento dos preços, a partir de modelos matemáticos e os preços históricos. Dentre os indicadores mais comuns de análise técnica, podem ser destacados: médias móveis exponenciais (MME), bandas de Bollinger, convergência/divergência de médias móveis(MACD), índice de força relativa (IFR), dentre outros.

De forma complementar, a Análise Gráfica é pautada na análise de padrões em gráficos que implicam na observação de tendências, pontos de inflexão e conseqüentemente, períodos de reversões sob uma ação. Por meio desta técnica, é possível observar o momento exato para compra/venda de determinado ativo.

Por conseguinte, segundo Vidotto (2009), entre os gráficos mais utilizados para análise gráfica de ativos, podem ser citados: gráfico de volume, gráfico de linha, gráfico de barra e gráfico de *candlesticks*. Sendo que este último foi utilizado no simulador deste trabalho. Esses gráficos tem o intuito de orientar a tendência, mostrando a direção em que o mercado se encontra, sendo esta: de baixa, de alta e lateral. Além disso, pode-se agrupar as tendências em dois seguimentos principais, sendo, de continuidade e reversão. A continuidade diz respeito ao mantimento da tendência no mesmo sentido que se encontrar e contrariamente, a reversão diz respeito à mudança de direção da tendência.

Embora Teixeira e Oliveira (2010), afirmar que o preço de fechamento de um certo ativo ou o volume de negociações devem ser utilizados exclusivamente na formulação de determinados indicadores técnicos, foi utilizado nesse trabalho o valor de sentimento diário de cada ativo para realização do experimento. Essa escolha tem como referência o trabalho de Alves (2015), que embora tenha aplicado os valores de sentimento diário como suporte aditivo ao MACD, e como resultado, tenha obtido mais de 200% de lucro acumulado na janela de análise proposta, no presente trabalho objetivou-se a realização do experimento de compras e vendas baseados diretamente no índice de sentimento diário proposto, que por sua vez, foi utilizado como base na formulação das técnicas MME e MACD ao invés do preço de fechamento dos ativos.

Na sequência é apresentado, detalhadamente, os indicadores de análise técnica: Médias móveis exponenciais e Convergência/Divergência de médias móveis. Esses indicadores para rastreamento de tendência/reversão foram os escolhidos para a realização do experimento.

#### 4.2.2.1 Médias móveis exponenciais

A média móvel exponencial é uma extensão da média móvel simples, fazendo uma ponderação de observações passadas, utilizando a suavização da mesma para reduzir a quantidade de sinais de compra ou venda<sup>3</sup>. De acordo com Vidotto (2009), as médias móveis são utilizadas para suavizar os ruídos das flutuações dos preços de curto prazo, de maneira que seja fácil identificar e definir tendências.

Tradicionalmente na análise técnica, a MME é utilizada tomando o preço de fechamento como base de cálculo, mas no presente trabalho essa técnica utilizará o valor de sentimento diário como base. Para o seu cálculo, será utilizado o valor de sentimento

<sup>3</sup> <https://br.advfn.com/educacional/analise-tecnica/media-movel-exponencial>



diário e o período que representa a quantidade de dias a serem considerados. A MME pode ser definida como:

$$MME_t = (1 - k) \times MME_{t-1} + k \times ISN_t,$$

onde  $n$  é a quantidade de dias considerados,  $t$  o dia em questão e  $ISN_t$  o valor do sentimento dos tweets no dia em questão e  $k = 2/(n + 1)$ .

A média móvel exponencial atribui maior peso aos dados mais recentes e reage com mais rapidez às mudanças do que a MM simples. Ao mesmo tempo, a MME não salta em resposta a velhos dados (VIDOTTO, 2009). Nesse trabalho, optou-se por utilizar os valores mais convencionais para realização do experimento que são, respectivamente, 20 para média lenta e 5 dias para a média curta.

Além disso, vale ressaltar que os sinais de compra/venda são emitidos a medida que ocorrem cruzamentos entre as médias curta e lenta. Isto é, um sinal de compra é caracterizado quando a média curta ultrapassa a média longa e um sinal de venda é caracterizado quando ocorre o inverso.

#### 4.2.2.2 Convergência/Divergência de médias móveis

O indicador de análise técnica MACD foi desenvolvido por Gerald Appel em 1960. O indicador é composto de duas linhas: uma linha denominada MACD e outra linha denominada linha de sinal. A linha MACD é composta pela diferença de duas médias móveis exponenciais. Ela responde às mudanças nos preços com relativa rapidez. A linha de sinal é composta pela linha MACD, ajustada por outra MME. Ela responde às mudanças nos preços com mais lentidão (ELDER, 2004).

Os sinais de compra e venda são emitidos quando a linha MACD mais rápida cruza acima ou abaixo da linha de sinal mais lenta. Neste caso, quando a linha MACD cruza para baixo a linha de sinal, é indicativo de venda da ação analisada, enquanto que, quando a linha MACD cruza para cima a linha de sinal, é indicativo de compra desta mesma ação.

Segundo Elder (2004), este indicador pode ser construído da seguinte forma:

- Cálculo da MME de 12 dias dos preços de fechamento;
- Cálculo da MME de 26 dias dos preços de fechamento;
- Subtração da MME de 26 dias da MME de 12 dias para obtenção da linha de MACD (rápida);
- Cálculo da MME de 9 dias da linha rápida para obtenção da linha de sinal (lenta).

No entanto, o indicador desse trabalho foi utilizado o valor diário do índice de sentimento para cada ativo como base de cálculo ao invés do preço de fechamento. Por conseguinte, o MACD pode ser definido como:

$$\text{MACD} = \text{MME}_{12} - \text{MME}_{26}.$$

A linha de sinal, obtida pela média móvel exponencial de nove dias dos valores obtidos para o MACD, pode ser definida como:

$$\text{Sinal} = \text{MME}_{9\text{MACD}}.$$

Dentre os diversos indicadores técnicos que compõem o grupo de indicadores denominados rastreadores de tendência, o MACD foi escolhido pela simplicidade no cálculo dos indicativos de compra e venda de ações. O MACD é dos indicadores técnicos mais precisos para captação de sinais de compra e venda de ativos, além de ser um dos melhores recursos de análise técnica para captar tendências e reversões ([ELDER, 2004](#)).

## 5 Resultados

Neste capítulo, serão apresentados os resultados obtidos a partir do processamento das entradas pré-determinadas, discutidas as saídas do simulador e apresentado uma comparação com o retorno do Índice Bovespa (Ibovespa) para o mesmo período. Além disso, o simulador desenvolvido pode ser encontrado no seguinte endereço: Simulador<sup>1</sup>.

### 5.1 Dados e ferramentas

Para a concepção do experimento foram adotados os seguintes parâmetros de entrada: janela de tempo, empresa e um indicador técnico. E a seguir, é apresentado detalhadamente os dados utilizados para a simulação.

- Janela de tempo: O intervalo de tempo adotado inicia-se em 17 de fevereiro de 2022 até 03 de abril de 2022, ou seja, 45 dias. Essa janela de tempo não foi maior devido à limitação de obtenção de tweets mais antigos devido ao nível de acesso concedido pela plataforma, como suscitado na seção 4.1;
- Quantidade total de tweets: Durante o período destacado acima, foram coletados 8496 tweets. Desses, 1253 são referentes ao Itaú (ITUB4), 6573 à Petrobras (PETR4), 298 ao bradesco (BBDC4) e 372 à B3 (B3SA3), conforme a Figura 2. Para a simulação, serão utilizados, apenas, os tweets referentes as empresas Itaú e Bradesco;
- Dados da bolsa de valores: Foram utilizados os valores de preço de abertura para as ações ITUB4 e BBDC4 para os dias em que o pregão esteve aberto. Vale ressaltar que os preços de abertura foram utilizados ao invés do preço de fechamento com o intuito de realizar uma operação sempre no dia seguinte da constatação de um sinal, ou seja, a primeira data do pregão após um dos cruzamentos ter ocorrido.
- Estratégias adotadas: Foram utilizados o MACD e o MME, calculados a partir do valor de sentimento diário obtidos da análise de sentimento dos tweets. Para o MACD, foram utilizados os valores de 26 dias para média lenta, 12 dias para a média rápida e 9 dias para a linha de sinal. Já para o MME, foram utilizados 20 dias para a média lenta e 5 dias para a média rápida.

Para cada empresa foi analisado a movimentação do MME e MACD conforme a janela de tempo destacada com intuito de observar com o MACD, o cruzamento das

<sup>1</sup> <https://macd-news-sentiment.netlify.app/>

linhas (MACD e a linha de sinal) e com o MME, o cruzamento das médias móveis (lenta e curta). Com isso, foi possível detectar sinais de tendência e reversão, e operar de acordo esses sinais. E a partir do processamento do simulador, serão apresentadas as seguintes saídas:

- Ordens: Demonstração de todas as operações realizadas na janela de tempo pré-determinada. Nessa demonstração, são apresentados a operação realizada (compra ou venda), o preço de abertura do ativo em reais, o lucro líquido obtido nas operações de venda em reais, a variação percentual entre o valor de compra e venda e a data da operação;
- Balanço consolidado: Demonstração do resultado final obtido com as operações. Nessa demonstração constam o lucro líquido em reais, o lucro acumulado percentualmente, o número de operações de compra e o número de operações de venda.

Além disso, para o resultado de cada simulação, é apresentado um *benchmark* com valorização do Índice Bovespa (Ibovespa) no mesmo período. Os valores foram extraídos do site oficial da B3<sup>2</sup>. Dessa forma, é possível avaliar se a valorização dos ativos foi satisfatória, tendo em vista que o Ibovespa é um referencial amplamente utilizado pelos analistas financeiros (VIDOTTO, 2009).

## 5.2 Resultados da simulação

### 5.2.1 Itaú com MME

No gráfico da Figura 5, foram constado 4 sinais, ou seja, houveram 4 cruzamentos da linha com a média móvel lenta (azul) e a a linha de média móvel curta (laranja). Para cada cruzamento, houve uma compra realizada na data do pregão seguinte à ocorrência e dessa forma, foram realizadas 4 operações, sendo 2 de compra e 2 de venda.

Essas operações podem ser vistas, detalhadamente, na tabela da Figura 6. Nessa tabela, é possível visualizar que nas 2 vendas realizadas foi possível obter um margem de lucro de 10,1% e 1,23%, respectivamente, em cada operação de venda. No entanto, vale a ressalva que 2 operações foram realizadas com o mesmo valor no dia 27 de março de 2022, isso deve-se ao fato de terem ocorridos 2 cruzamentos no final de semana que antecedeu essa data, um no sábado e outro no domingo, e quando isso ocorre, são realizadas as operações respectivas no pregão seguinte.

E ao final da realização das operações, foi possível obter um lucro acumulado de 11,4% nas operações realizadas durante esse período, como demonstrado na Figura 7.

<sup>2</sup> [https://www.b3.com.br/pt\\_br/market\\_data\\_e\\_indices/indices/indices\\_amplos/indice\\_ibovespa\\_ibovespa\\_estatisticas\\_historicas.htm](https://www.b3.com.br/pt_br/market_data_e_indices/indices/indices_amplos/indice_ibovespa_ibovespa_estatisticas_historicas.htm)

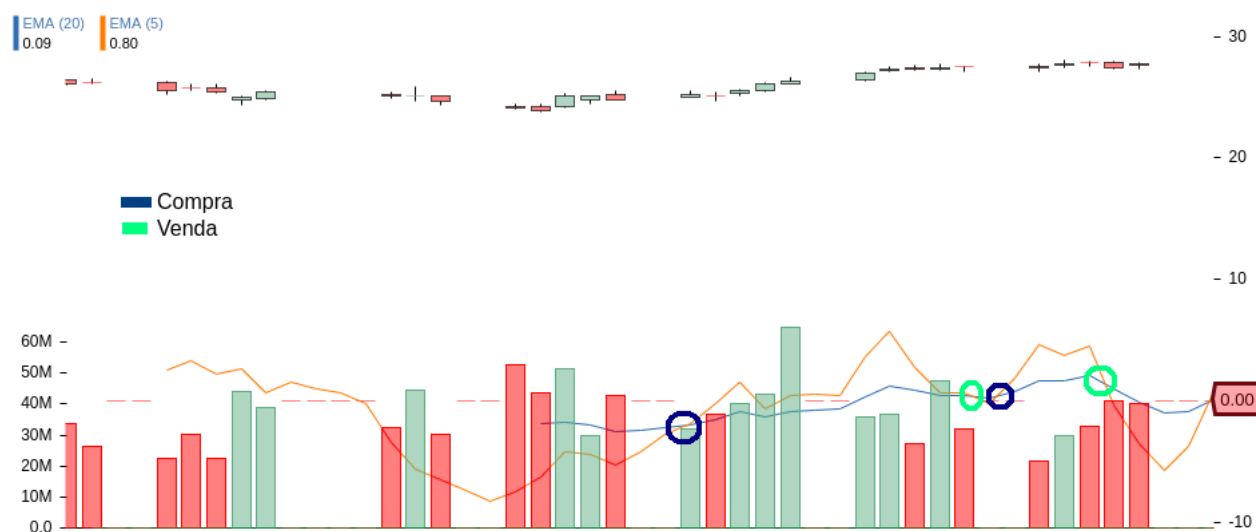


Figura 5 – MME (20, 5) com valores de sentimento para ITUB4.

Fonte: Autoria própria.

| Código | Operação | Preço Abertura (R\$) | Lucro Líquido (R\$) | Varição C/V (%) | Data       |
|--------|----------|----------------------|---------------------|-----------------|------------|
| ITUB4  | compra   | R\$ 25.02            | R\$ 0.0             | 0.00%           | 13-03-2022 |
| ITUB4  | venda    | R\$ 27.54            | R\$ 2.5             | 10.1%           | 27-03-2022 |
| ITUB4  | compra   | R\$ 27.54            | R\$ 0.0             | 0.00%           | 27-03-2022 |
| ITUB4  | venda    | R\$ 27.88            | R\$ 0.34            | 1.23%           | 30-03-2022 |

Figura 6 – Ordens realizadas para ITUB4 com MME (20,5).

Fonte: Autoria própria.

Além disso, vale a pena ressaltar como *benchmark* que o Ibovespa registrou entre 13 de março de 2022 e 30 de março de 2022, 109.927,62 e 120.259,76 pontos, respectivamente. E em valores percentuais, isso significa que o Ibovespa teve um retorno de 9,3%, ficando abaixo do retorno obtido nessa simulação.

| Lucro Líquido (R\$) | Lucro Acumulado (%) | Operações de compra | Operações de venda |
|---------------------|---------------------|---------------------|--------------------|
| R\$ 2.9             | 11.4%               | 2                   | 2                  |

Figura 7 – Balanço consolidado para ITUB4 com MME (20,5).

Fonte: Autoria própria.

### 5.2.2 Itaú com MACD

No gráfico da Figura 8, foram constados 2 sinais, ou seja, houveram 2 cruzamentos da linha da linha MACD com a linha de sinal. E com isso foram realizadas, apenas, uma operação de compra e outra venda.

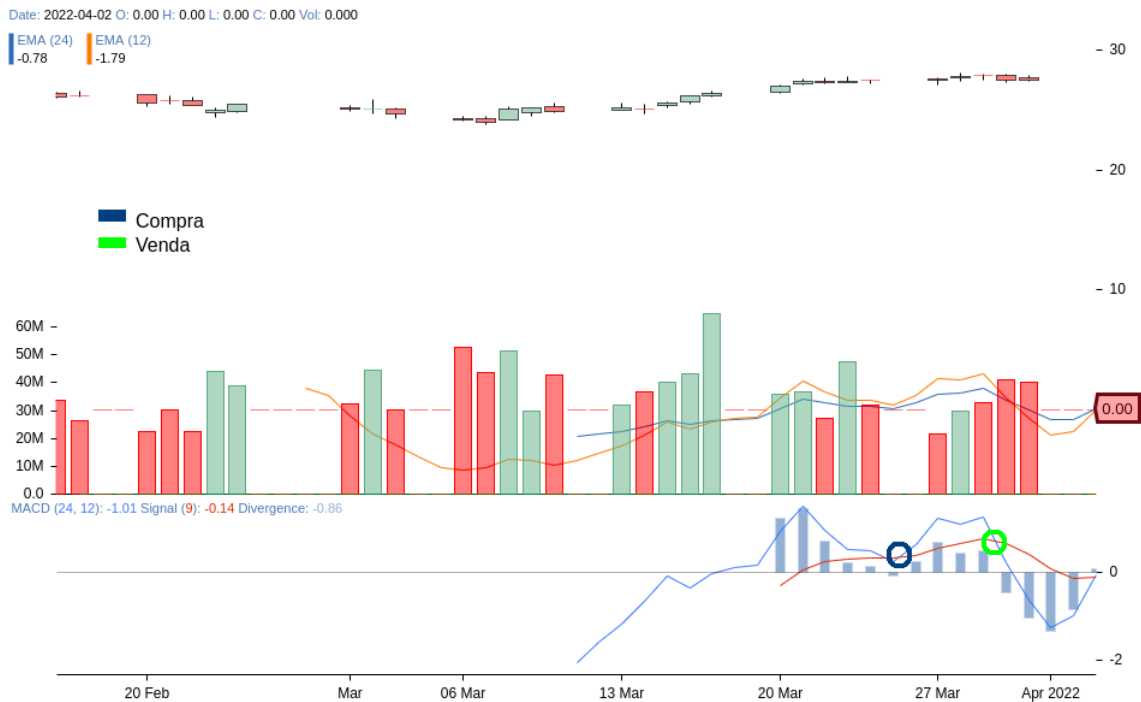


Figura 8 – MACD com valores de sentimento para ITUB4.

Fonte: Autoria própria.

Vale ressaltar que o número de operações ficou baixíssimo, tendo em vista a janela de tempo de 45 dias para a aplicação do MACD, além do número menor de cruzamentos das linhas durante o período. Na tabela da Figura 9, é possível visualizar que foi possível obter uma margem de lucro de 1.23%, onde esse percentual, também representa, o lucro acumulado do balanço consolidado dessa simulação, como demonstrado na Figura 10.

| Código | Operação | Preço Abertura (R\$) | Lucro Líquido (R\$) | Variação C/V (%) | Data       |
|--------|----------|----------------------|---------------------|------------------|------------|
| ITUB4  | compra   | R\$ 27.54            | R\$ 0.0             | 0.00%            | 27-03-2022 |
| ITUB4  | venda    | R\$ 27.88            | R\$ 0.34            | 1.23%            | 30-03-2022 |

Figura 9 – Ordens realizadas para ITUB4 com MACD.

Fonte: Autoria própria.

| Lucro Líquido (R\$) | Lucro Acumulado (%) | Operações de compra | Operações de venda |
|---------------------|---------------------|---------------------|--------------------|
| R\$ 0.34            | 1.23%               | 1                   | 1                  |

Figura 10 – Balanço consolidado para ITUB4 com MACD.

Fonte: Autoria própria.

Além disso, como *benchmark*, o Ibovespa registrou entre 27 de março de 2022 e 30 de março de 2022, 118.737,78 e 120.259,76 pontos, respectivamente. E em valores

percentuais, isso significa que o Ibovespa teve um retorno de 1,28%, ficando um pouco acima do retorno obtido nessa simulação.

### 5.2.3 Bradesco com MME

No gráfico da Figura 11, é possível observar que houveram 12 cruzamentos entre a linha laranja (média rápida) e a linha azul (média lenta). E assim como na simulação para o Itaú, para cada cruzamento, houve uma compra realizada na data do pregão seguinte à ocorrência. E resumidamente, houveram 6 sinais de venda e 7 sinais de compra dentro da janela de tempo pré-estabelecida.

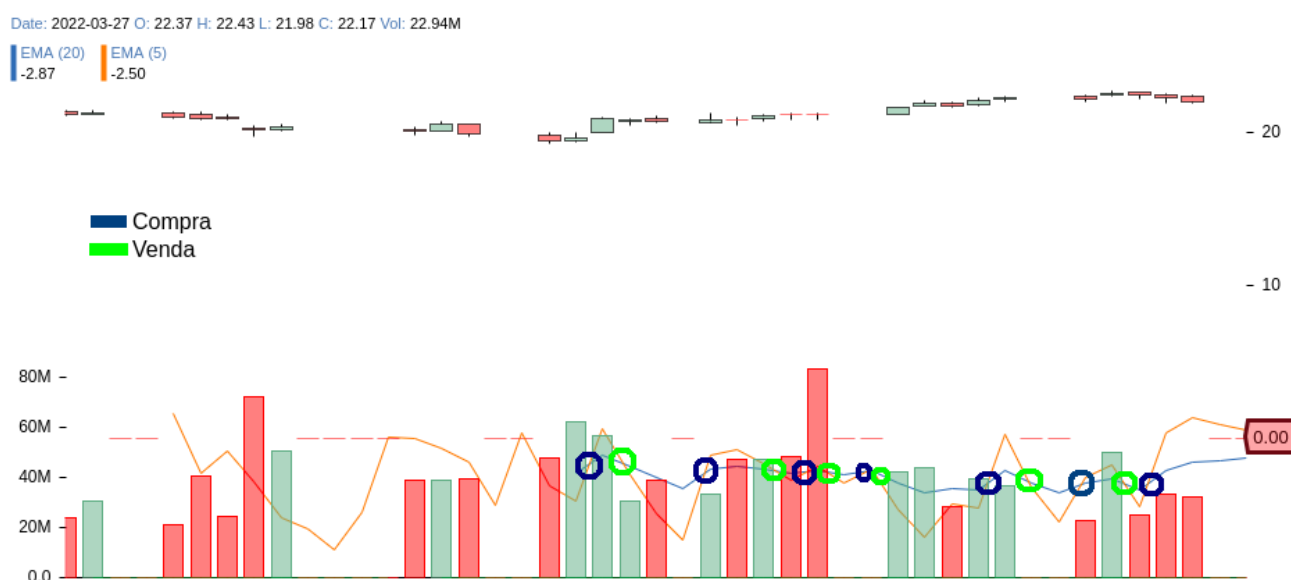


Figura 11 – MME (20, 5) com valores de sentimento para BBDC4.

Fonte: Autoria própria.

Essas 13 operações são descritas na tabela da Figura 12. Nessa tabela, é possível visualizar que houveram 5 operações de venda com margem de lucro positiva, sendo: 3.39%, 2.41%, 0,14%, 0.81% e 1.03%, respectivamente. Além disso, assim como na simulação do Itaú, houve uma venda com margem de lucro de 0% devido à cruzamentos ocorridos durante dias sem pregão e com isso, foram efetuadas as respectivas operações no pregão seguinte.

Por fim no balanço consolidado da Figura 13, foi possível obter um lucro acumulado de 8.00% nas operações realizadas durante esse período. Além disso, como *benchmark*, o Ibovespa registrou entre 08 de março de 2022 e 29 de março de 2022, 111.203,45 e 120.014,17 pontos, respectivamente. E em valores percentuais, isso significa que o Ibovespa teve um retorno de 7,92%, ficando um pouco abaixo do retorno obtido nessa simulação.

| Código | Operação | Preço Abertura (R\$) | Lucro Líquido (R\$) | Varição C/V (%) | Data       |
|--------|----------|----------------------|---------------------|-----------------|------------|
| BBDC4  | compra   | R\$ 20.04            | R\$ 0.0             | 0.00%           | 08-03-2022 |
| BBDC4  | venda    | R\$ 20.72            | R\$ 0.68            | 3.39%           | 9-03-2022  |
| BBDC4  | compra   | R\$ 20.71            | R\$ 0.0             | 0.00%           | 13-03-2022 |
| BBDC4  | venda    | R\$ 21.21            | R\$ 0.50            | 2.41%           | 16-03-2022 |
| BBDC4  | compra   | R\$ 21.17            | R\$ 0.0             | 0.00%           | 17-03-2022 |
| BBDC4  | venda    | R\$ 21.20            | R\$ 0.030           | 0.142%          | 20-03-2022 |
| BBDC4  | compra   | R\$ 21.20            | R\$ 0.0             | 0.00%           | 20-03-2022 |
| BBDC4  | venda    | R\$ 21.20            | R\$ 0.0             | 0.00%           | 20-03-2022 |
| BBDC4  | compra   | R\$ 22.19            | R\$ 0.0             | 0.00%           | 24-03-2022 |
| BBDC4  | venda    | R\$ 22.37            | R\$ 0.18            | 0.811%          | 27-03-2022 |
| BBDC4  | compra   | R\$ 22.37            | R\$ 0.0             | 0.00%           | 27-03-2022 |
| BBDC4  | venda    | R\$ 22.60            | R\$ 0.23            | 1.03%           | 29-03-2022 |
| BBDC4  | compra   | R\$ 22.43            | R\$ 0.0             | 0.00%           | 30-03-2022 |

Figura 12 – Ordens realizadas para BBDC4 com MME (20,5).

Fonte: Autoria própria.

| Lucro Líquido (R\$) | Lucro Acumulado (%) | Operações de compra | Operações de venda |
|---------------------|---------------------|---------------------|--------------------|
| R\$ 1.6             | 8.00%               | 7                   | 6                  |

Figura 13 – Balanço consolidado para BBDC4 com MME (20,5).

Fonte: Autoria própria.

#### 5.2.4 Bradesco com MACD

No gráfico da Figura 14, foram constatados 5 cruzamentos da linha da linha MACD com a linha de sinal. Assim, como na simulação do Itaú para o MACD, foram obtidas poucas operações devido a curta de janela de tempo adotada. Na tabela da Figura 15, é possível visualizar que foi possível obter uma margem de lucro de 0.81% e 1.03%, respectivamente em cada venda.

Por fim, é possível visualizar no balanço consolidado do período um lucro acumulado de 1.85%, como demonstrado na Figura 16. Além disso, como *benchmark*, o Ibovespa registrou entre 24 de março de 2022 e 29 de março de 2022, 119.052,91 e 120.014,17 pontos, respectivamente. E em valores percentuais, isso significa que o Ibovespa teve um retorno de 0,8%, sendo inferior ao retorno obtido nessa simulação.





Figura 14 – MACD com valores de sentimento para BBDC4.

Fonte: Autoria própria.

| Código | Operação | Preço Abertura (R\$) | Lucro Líquido (R\$) | Varição C/V (%) | Data       |
|--------|----------|----------------------|---------------------|-----------------|------------|
| BBDC4  | compra   | R\$ 22.19            | R\$ 0.0             | 0.00%           | 24-03-2022 |
| BBDC4  | venda    | R\$ 22.37            | R\$ 0.18            | 0.811%          | 27-03-2022 |
| BBDC4  | compra   | R\$ 22.37            | R\$ 0.0             | 0.00%           | 27-03-2022 |
| BBDC4  | venda    | R\$ 22.60            | R\$ 0.23            | 1.03%           | 29-03-2022 |
| BBDC4  | compra   | R\$ 22.43            | R\$ 0.0             | 0.00%           | 30-03-2022 |

Figura 15 – Ordens realizadas para BBDC4 com MACD.

Fonte: Autoria própria.

| Lucro Líquido (R\$) | Lucro Acumulado (%) | Operações de compra | Operações de venda |
|---------------------|---------------------|---------------------|--------------------|
| R\$ 0.41            | 1.85%               | 3                   | 2                  |

Figura 16 – Balanço consolidado para BBDC4 com MACD.

Fonte: Autoria própria.

### 5.3 Considerações finais

Sobre os resultados obtidos, a simulação com MME tanto para ITUB4, quanto BBDC4 tiveram retornos maiores em relação a simulação com MADC, como demonstrado na Tabela 10. Além disso, vale ressaltar que, apenas, a simulação feita com MACD para

o ITUB4 obteve um retorno menor do que o Índice Bovespa (Ibovespa) para o mesmo período.

Tabela 10 – Resultado das simulações

| <b>Ativo</b> | <b>Método</b> | <b>Lucro Acumulado</b> | <b>Retorno Ibovespa</b> | <b>Período</b>          |
|--------------|---------------|------------------------|-------------------------|-------------------------|
| ITUB4        | MME           | 11.4%                  | 9,3%                    | 13-03-2022 à 30-03-2022 |
| ITUB4        | MACD          | 1.23%                  | 1,28%                   | 27-03-2022 à 30-03-2022 |
| BBDC4        | MME           | 8.00%                  | 7,92%                   | 08-03-2022 à 29-03-2022 |
| BBDC4        | MACD          | 1.85%                  | 0,8%                    | 24-03-2022 à 29-03-2022 |

Fonte: Autoria própria.

## 6 Conclusão

Nessa monografia, objetivou-se a formulação de um indicador capaz de captar tendências e reversões baseadas no sentimento de notícias. E a partir dos valores diários de notícias extraídas de contas verificadas da rede social Twitter, junto à indicadores de análise técnica, foi possível identificar sinais de operações de compra e venda de determinados ativos par obtenção de lucro.

Essa indicador tem como premissa, assim como em outros trabalhos destacados no Capítulo 2, que a oscilação de preços da bolsa de valores têm a influência de notícias nos movimentos dos investidores, mesmo que de forma estritamente especulativa. E para isso, tendo em vista, que os veículos de mídia publicam as notícias de seus respectivos portais no Twitter, foi possível obter uma diversificação relevante de veículos de mídia no Twitter como demonstrado no Capítulo 4.

Sobre os resultados demonstrados no Capítulo 5, a partir da compra e venda dos ativos ITUB4 e BBDC4 baseados nos sinais obtidos a partir do MACD e MME com valores diários da análise de sentimento, foi possível perceber que as premissas possuíam uma certa coerência tendo em vista as margens de lucro positivas que foram obtidas.

Além disso, como demonstrado na Tabela 10, apenas, uma das simulações realizadas não obteve retorno superior ao Índice Bovespa (Ibovespa) para o mesmo período. No entanto, embora as margens positivas, a janela de tempo para a simulação poderia ter compreendido mais dias para assim, garantir uma maior confiabilidade.

Vale ressaltar também que essa monografia não considerou flutuações dos preços das ações durante o dia. Foram considerados, apenas, os valores de abertura do pregão para cada ativo, além de que, o índice de sentimento foi construído a partir de uma média simples diária, ou seja, uma média de sentimento dos tweets durante o dia. Em cenários cotidianos de análise financeira deveriam ser considerados as flutuações em um espaço de tempo menor.

Sobre a questão do método de classificação utilizado, os resultados obtidos com dicionário do trabalho de [Loughran e McDonald \(2011\)](#) se demonstraram coerentes com os resultados obtidos de forma geral, embora os tweets filtrados estarem na língua portuguesa e o dicionário ser de linguagem inglesa. Por fim, o trabalho cumpriu com o objetivo principal e o método de análise empregado se demonstrou potencialmente eficaz para auxílio na tomada de decisão nas operações de compra e venda.

Em trabalhos futuros, talvez, fosse interessante a filtragem de tweets de ativos da Nasdaq e Nyse, cujas contas fossem de veículos de mídia de língua inglesa para obtenção

de uma análise mais confiável. Além disso, poderia haver uma combinação de outros indicadores de análise técnica para corroboração do método empregado, haja vista que existem outros indicadores que podem garantir auxílios interessantes relacionando volume de negociação e o preço de fechamento ajustado dos ativos. E finalmente, a opção de cancelar ordens, uma vez, para evitar ocorrências de compras e vendas com o mesmo preço, como demonstrado no Capítulo 5.

# Referências

- AGARWAL, S.; KUMAR, S.; GOEL, U. Social media and the stock markets: an emerging market perspective. *Journal of Business Economics and Management*, p. 1614–1632, 2021. Citado na página 31.
- ALVES, D. S. Uso de técnicas de computação social para tomada de decisão de compra e venda de ações no mercado brasileiro de bolsa de valores. Universidade de Brasília, 2015. Citado na página 46.
- BARBERIS, N.; SHLEIFER, A.; VISHNY, R. A model of investor sentiment. *Journal of Financial Economics*, v. 49, p. 307–343, 1998. Citado na página 25.
- BOYD danah m.; ELLISON, N. B. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, v. 13, p. 210–230, 2007. Citado na página 24.
- CHAVES, D. A. T.; ROCHA, K. C. Análise técnica e fundamentalista: divergências, similaridade e complementaridades. 2004. Citado na página 45.
- ELDER, A. Como se transformar em um operador e investidor de sucesso. Rio de Janeiro: Campus, 2004. Citado 2 vezes nas páginas 47 e 48.
- GALDI, F. C.; GONÇALVES, A. M. Pessimismo e incerteza das notícias e o comportamento dos investidores no brasil. 2018. Citado 2 vezes nas páginas 30 e 31.
- GEGENTANA. A systematic review of automated software engineering. University of Gothenburg, 2011. Citado na página 28.
- KEARNY, C.; LIU, S. Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 2013. Citado 2 vezes nas páginas 20 e 24.
- LAGIOIA, U. C. T. Fundamentos do mercado de capitais. São Paulo: Atlas, 2011. Citado na página 45.
- LI, X.; WU, P.; WANG, W. Incorporating stock prices and news sentiments for stock market prediction: A case of hong kong. *Information Processing Management*, 2020. Citado na página 36.
- LIU, B. Sentiment analysis and opinion mining. In: *Morgan & Claypool Publishers*. [S.l.: s.n.], 2012. v. 5, p. 1–167. Citado 2 vezes nas páginas 19 e 25.
- LOUGHRAN, T.; MCDONALD, B. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *Journal of Finance*, Forthcoming, 2011. Disponível em: <<https://ssrn.com/abstract=1331573>>. Citado 4 vezes nas páginas 29, 31, 36 e 57.
- MALTA, T. L.; CAMARGOS, M. A. Variáveis da análise fundamentalista e dinâmica e o retorno acionário de empresas brasileiras entre 2007 e 2014. *REGE Revista de Gestão*, 2016. Citado na página 44.

- MATARAZZO, D. C. Finanças empresariais: Essencial. São Paulo: Atlas, 2009. Citado na página 45.
- MATERLANC, R. Avaliação de empresas: um guia para fusões e aquisições e gestão de valor. São Paulo: Pearson Prentice Hall, 2005. Citado na página 44.
- NETO, N. A. B. Análise de sentimentos do twitter como suporte aditivo para a previsão da volatilidade bitcoin. Universidade de Brasília, Engenharia de Software, 2018. Citado na página 30.
- NGUYEN, K.; CAVA, G. L. News sentiment and the economy. Reserve Bank of Australia, 2020. Disponível em: <<https://www.rba.gov.au/publications/bulletin/2020/jun/news-sentiment-and-the-economy.html>>. Citado 3 vezes nas páginas 25, 29 e 31.
- SANVICENTE, A. Z.; MELLAGI, A. F. Mercado de capitais e estratégias de investimento. São Paulo: Atlas, 2008. Citado na página 44.
- SHAPIRO, A. H.; SUDHOF, M.; WILSON, D. Measuring news sentiment. Federal Reserve Bank of San Francisco Working, 2017. Disponível em: <<https://doi.org/10.24148/wp2017-01>>. Citado na página 30.
- TEIXEIRA, L. A.; OLIVEIRA, A. L. I. A method for automatic stock trading combining technical analysis and nearest neighbor classification. 2010. Citado na página 46.
- TETLOCK, P. C. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62, 1139-1168, 2007. Citado na página 20.
- TUSHAR, R.; SAKET, S. Analyzing stock market movements using twitter sentiment analysis. Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining, 2012. Citado na página 31.
- VIDOTTO, R. S. O moving average convergence-divergence como ferramenta para a decisão de investimentos no mercado de ações. 2009. Citado 3 vezes nas páginas 46, 47 e 50.