

---

# IMPACTO DOS SENTIMENTOS SOBRE A VACINA DE COVID-19 NO MERCADO FINANCEIRO BRASILEIRO

Karoline Pereira Branco <sup>1</sup>  
Gabriel Rodrigo Gomes Pessanha <sup>2</sup>  
Eleanderson Campos Eugênio Filho <sup>3</sup>

---

▪ Artigo recebido em: 06/10/2021 ▪ Artigo aceito em: 09/07/2023 ▪▪ Segunda versão aceita em: 22/11/2023

## RESUMO

Este trabalho foi construído com o propósito de atender a dois objetivos principais: o primeiro foi a realização da análise de sentimentos visando compreender qual é o sentimento preponderante dos brasileiros usuários do Twitter a respeito da vacina/vacinação contra a Covid-19 no país, e o segundo foi avaliar se existe relação de dependência entre este sentimento e as oscilações do mercado acionário do país. Para a realização da pesquisa, foram usados algoritmos de *machine learning*, tanto para a análise de sentimentos, que foi realizada pelo método *Naive Bayes*, quanto para a previsão do mercado financeiro que usou o método SVM aliado à técnica de validação cruzada para séries temporais. Como resultado verificou-se que o sentimento dominante do grupo analisado a respeito do tema é negativo, sendo assim para todos os dias que compuseram a amostra da pesquisa, e a principal marca das mensagens negativas é a reafirmação da situação de pandemia e a alusão a termos relacionados à política. Ademais, a pesquisa não foi capaz de confirmar a relação de dependência entre o sentimento diário da população sobre a vacina contra a Covid-19 e as oscilações observadas no mercado financeiro. Por fim, pelos resultados encontrados sugere-se uma ação mais incisiva dos entes políticos em campanhas de vacinação e informação, para que a população recupere a credibilidade em seus entes públicos no que diz respeito ao controle da pandemia no país, bem como também sugere-se a aplicação de outros métodos para a averiguação da relação entre sentimento e mercado financeiro.

**Palavras-Chave:** Análise de sentimentos. Imunização. Pandemia de Covid-19. Mercado acionário. Previsão

---

<sup>1</sup>Mestre em estatística aplicada e biometria, Universidade Federal de Alfenas, <https://orcid.org/0000-0002-9289-2194>, Av. Celina Ferreira Ottoni, 4000 - Padre Vitor, Varginha - MG, CEP: 37048-395. Telefone: (35) 99957-0191. E-mail: karoline.branco@unifal-mg.edu.br.

<sup>2</sup>Doutor em Administração, Universidade Federal de Alfenas, <http://orcid.org/0000-0002-6480-357X>, Av. Celina Ferreira Ottoni, 4000 - Padre Vitor, Varginha - MG, CEP: 37048-395. Telefone: (35) 99144-6617. E-mail: gabriel.pessanha@unifal-mg.edu.br.

<sup>3</sup>Doutor em estatística, <http://orcid.org/0000-0002-6244-1237>, Av. Celina Ferreira Ottoni, 4000 - Padre Vitor, Varginha - MG, CEP: 37048-395. Telefone: (35) 99228-3290, E-mail: [eleandersoncampos@gmail.com](mailto:eleandersoncampos@gmail.com).

Editor responsável pela aprovação do artigo: Dr. Ewerton Alex Avelar  
Editor responsável pela edição final do artigo: Dr. Ewerton Alex Avelar

# IMPACT OF SENTIMENTS ABOUT THE COVID-19 VACCINE ON THE BRAZILIAN FINANCIAL MARKET

## ABSTRACT

This paper had two main objectives: the first was to perform a sentiment analysis to detect the dominant feeling of twitter brazilian users about the vaccine/vaccination against Covid-19 in that country, and the second one was to investigate if there is a relationship of dependence between the verified feeling and the oscillations of the intern stock market. To carry out the research, machine learning algorithms were used. For sentiment analysis the method used was *Naive Bayes*, and to execute the forecast of the financial market, it was used *SVM* method combined with the cross-validation technique applied to time series. As a result, it was found that the dominant feeling about the analyzed topic was negative for all days that comprised the research sample and the highlight of the negative messages was the reaffirmation of the pandemic situation and the allusion to terms related to politics. Furthermore, the research was not able to confirm the dependence relationship between the daily sentiment about the Covid-19 vaccine and the oscillations observed in the financial market. Finally, the found results suggest a more incisive action by political entities in vaccination and information campaigns, to regain public credibility with regard to the control of the pandemic in the country. Despite that, it suggests the application of other methods to investigate the relationship between sentiment and the financial market, for example, to analyze this relation in hours and minutes, instead of days.

**Keywords:** Sentiment Analysis. Immunization. Covid-19. Stock Market. Forecast.

## 1 INTRODUÇÃO

Em janeiro de 2020, o mundo recebeu com apreensão a notícia de que a China havia reportado à Organização Mundial da Saúde (OMS) a ocorrência, na cidade de Wuhan, de uma doença que até então era desconhecida. O que a princípio acreditava-se tratar do surgimento de um tipo de pneumonia grave, posteriormente foi identificado como o vírus SARS CoV-2, caracterizado como um novo tipo de Coronavírus (Estevão, 2020). Apelidada de COVID-19, pois os primeiros casos já haviam sido constatados no final do ano de 2019, a doença se espalhou de forma veloz pelo mundo, fazendo com que a OMS a elevasse ao status de pandemia em 11 de março de 2020 (Netto & Corrêa, 2020).

No Brasil, o primeiro caso confirmado data de 26 de fevereiro de 2020. Após um ano e 4 meses da pandemia ter chegado no país, observou-se um total acumulado de mais de 19 milhões de doentes e a morte de mais de 542.000 brasileiros (Ministério da Saúde, 2021). Dessa forma, dados os inúmeros impactos sociais e econômicos, além da crise generalizada observada em decorrência do avanço da doença pelo mundo, a vacina contra a Covid-19 se apresenta ao redor do planeta como uma esperança de volta para a "normalidade" (Meena & Bai V, 2020).

Neste contexto, sabe-se que com cada vez mais frequência as mídias sociais são caracterizadas como um ambiente o qual as pessoas usam para

expressar seus sentimentos e opiniões acerca de diversas entidades, produtos, outras pessoas e situações. Assim, tem-se que a opinião expressa nestas mídias muitas vezes reflete a opinião pública de toda uma localidade. Isso faz com que as plataformas digitais se constituam como um espaço de geração massiva e constante de dados, o que possibilita a extração de informações que podem auxiliar a compreensão de vários aspectos da sociedade (Bollen, Mao & Zeng, 2011; Shayaa et al., 2017).

O Twitter é uma rede social que atua com um serviço de microblog. Nesta plataforma as pessoas podem compartilhar suas últimas histórias, ideias, opiniões e notícias em tempo real, escolhendo entre mensagens públicas ou um grupo selecionado em mensagens de até 280 caracteres, chamadas de *tweets* (Twitter, 2020). No mundo todo, a plataforma conta com mais 300 milhões de usuários ativos diariamente (Dubey, 2021). No Brasil, essa plataforma tem sua importância reconhecida em ações de engajamento da audiência televisiva ou de manifestações políticas, por exemplo, sendo que 15,5% dos brasileiros são usuários da plataforma (Depexe, 2022), o que coloca o país como o 4º lugar no mundo que mais a acessa (Kemp, 2022).

Destarte, dado o momento atípico de pandemia que o planeta atravessa, o tema "vacina" entra em pauta e ganha destaque nas discussões e debates virtuais. Usuários pró e contra a vacinação usam as redes sociais de forma intensa para participar de conversas e debates sobre vacinas e para disseminar informações (ou desinformação) em relação a este tipo de imunização, com o intuito de defender seu posicionamento (Milani, Weitkamp & Webb, 2020; Rahin & Rafie, 2020; Piedrahita-Valdés et al., 2021).

Segundo Carosia, Coelho e Silva (2019), a análise de sentimentos (AS) aparece como uma técnica que pode ser aplicada nos textos provindos das redes sociais e tem como objetivo a identificação do tipo de sentimento que emana de forma preponderante destes. O emprego desta ferramenta intenta verificar se os sentimentos em relação a determinada temática são majoritariamente positivos ou negativos. Segundo Neuenschwander et al. (2014), apesar de a aplicação dessa técnica não ser trivial, os resultados trazidos por ela refletem um grande potencial de aplicações.

Um potencial de aplicação ainda pouco explorado no Brasil é o emprego da análise de sentimentos como ferramenta para antever movimentos do mercado acionário (Assis et al., 2018). Conforme suposto no ramo das finanças comportamentais, as decisões financeiras são amplamente relacionadas às emoções e ao humor dos seus agentes, sendo então plausível supor que o sentimento público pode influenciar os valores do mercado de ações (Bollen, Mao & Zeng, 2011; Nofsinger, 2005). Relacionado a isto, existem pesquisas que noticiaram a existência de relação entre o sentimento da população e as oscilações no mercado financeiro de um país, mostrando que o seu conhecimento pode auxiliar investidores a desempenharem melhores negócios (Carosia, Coelho & Silva, 2019).

Destarte, esta pesquisa teve como objetivo o uso da técnica de análise de sentimentos para compreender qual é o sentimento dos usuários do Twitter a respeito da vacina contra a Covid-19 no Brasil e o efeito deste sentimento no movimento do mercado financeiro do país. Dessa forma, os resultados obtidos

aqui buscam atender a duas principais finalidades. A primeira é o fornecimento de informações que podem ser úteis aos formuladores de políticas públicas na elaboração de estratégias de campanha de vacinação mais eficientes, tendo em vista o conhecimento de como um grupo de sua população pensa a respeito da temática. A segunda finalidade se pauta na verificação de qual o impacto do sentimento diário observado nas oscilações do índice Ibovespa, que é uma métrica do movimento do mercado acionário brasileiro, visando trazer aos investidores uma possível ferramenta de apoio para a otimização de seus resultados.

Assim, este trabalho é organizado da seguinte forma: a próxima seção apresenta trabalhos empíricos que também realizaram a análise de sentimentos à respeito da temática da vacinação; a seção posterior apresenta algumas pesquisas que igualmente usaram a análise de sentimentos como um instrumento auxiliar na previsão do mercado financeiro; em seguida é apresentada a técnica de análise de sentimentos e, posteriormente, são apresentados os procedimentos metodológicos empregados na pesquisa, os resultados e, por último, são feitas as considerações finais.

## **2 REFERENCIAL TEÓRICO**

### **2.1 IMUNIZAÇÃO E ANÁLISE DE SENTIMENTOS - ESTUDOS RELACIONADOS**

Conforme Neuenschwander et al. (2011), a elevação do uso de aplicativos de internet pelas pessoas com o intuito de consumir informações e compartilhar opiniões e ideias sobre diversos assuntos fez com que as redes sociais se tornassem eficientes plataformas de coleta de dados para a identificação dos sentimentos e do humor de seus usuários. Nesse sentido, Bollen, Mao e Zeng (2011) apontam que os posts feitos pelos usuários da rede Twitter podem representar de forma precisa o humor e o sentimento geral de uma população em tempo real.

Nesse contexto, com as discussões e debates a respeito da vacina da Covid-19 não foi diferente. As redes sociais foram inundadas por conversas sobre vacinas, nas quais internautas com posicionamentos pró e contra a vacinação, atuaram na disseminação de informações e/ou desinformações a respeito da imunização, defendendo os seus posicionamentos (Milani, Weitkamp & Webb, 2020). Assim sendo, de acordo com Raghupathi, Ren e Raghupathi (2020), a análise de texto tem sido usada por diversos estudiosos para pesquisar atitudes em relação à vacinação e às vacinas de diversas doenças e em inúmeros países, com o intuito de avaliar o nível de aceitação ou hesitação da população a respeito deste tipo de imunização, bem como as causas que levam a determinados posicionamentos.

Dubey (2021), pesquisou o sentimento dos indianos em relação a duas vacinas contra Covid-19 aplicadas na campanha de vacinação do país. Por meio da análise de sentimentos de posts do Twitter, pelo método léxico, o autor verificou que o posicionamento da população sobre ambas as vacinas era preponderantemente positivo. A positividade em relação à primeira vacina era de 69% e em relação à segunda era de 71%. Além disso, por meio de um estudo que detalhou as emoções atribuídas a cada uma delas, o pesquisador observou que aquelas de maior recorrência eram confiança e ansiedade para ambas as vacinas.

Já Meena e Bai V (2020) conduziram um estudo semelhante, no qual os autores buscaram analisar o sentimento dos indivíduos em relação à divulgação da descoberta da vacina da Covid-19 pela Rússia. Os dados para a análise foram coletados no dia em que a notícia sobre a descoberta foi divulgada. Pela utilização do método léxico para a análise de sentimentos, os autores verificaram que o sentimento contido em 90% das publicações era positivo e que apenas 10% dos textos apresentaram sentimentos negativos, mostrando que houve euforia e esperança com relação ao anúncio.

Rahim e Rafie (2020) apontaram em seus estudos na Malásia uma crescente na hesitação em vacinar-se e em vacinar os filhos. Dessa forma, em sua pesquisa tentaram avaliar esta hesitação através da análise de sentimentos de publicações do Twitter. Por meio da utilização do método SVM para classificar sentimentos, os autores constataram que 41% dos *tweets* analisados foram positivos em relação à vacinação, 39% foram neutros e 20% dos usuários da plataforma se mostraram negativos quanto a esta forma de imunização. O medo de efeitos colaterais e questões religiosas foram apontadas pelos autores como fatores criadores de insegurança em relação a vacinar-se.

Nesta mesma linha, Raghupathi, Ren e Raghupathi (2020) analisaram a percepção dos americanos a respeito de vacinas em geral entre janeiro e abril de 2019, período no qual o país passou por um surto de casos de sarampo. Também utilizando das publicações no Twitter como fonte de dados para verificar as emoções das pessoas a respeito do tema, os autores constataram haver um equilíbrio das opiniões dos usuários em relação a eficácia das vacinas, sendo que 43,3% dos *tweets* foram classificados como negativos, 40,4% foram classificados como positivos e 16,3% foram classificados como neutros. Os autores atribuíram o elevado número de sentimentos negativos à recente ligação criada entre vacina e autismo (ainda sem comprovação científica no momento em que a pesquisa foi feita), enquanto que os sentimentos positivos se relacionam a existência de vacina contra o sarampo e às provas de que a vacina efetivamente tem sido eficiente em salvar vidas. A Tabela 1 apresenta alguns outros estudos relacionados à temática.

**Tabela 1**

Pesquisas empíricas que aplicaram a técnica análise de sentimentos em relação à temática "vacina"

Autor(es) /Ano	Fonte de dados	Objetivos	Principais Achados
Milani, Weitkam p & Webb (2020)	Twitter	Analisar o debate pró e contra a vacinação por meio da análise de sentimento de imagens compartilhadas nas redes sociais que se relacionam à imunização por vacina.	Os pesquisadores constataram que imagens que transmitem sentimentos negativos em relação à imunização pela vacina são predominantes na plataforma, em decorrência de ativistas anti-vacina serem mais articulados nas mídias sociais do que as comunidades pró-vacina.

Piedrahit a-Valdés et al. (2021)	Twitter	Analisar em diversos países qual o sentimento preponderante em relação à temática "vacina", entre os anos 2011 e 2019.	Os autores verificaram que os tweets com sentimentos positivos eram sempre preponderantes em relação aos negativos, possuindo também maior engajamento.
Wiysonge et al. (2021)	Twitter	Verificar se o aumento do uso das mídias sociais, com foco no Twitter, e a propagação de campanhas de desinformação advindas do exterior influenciam os níveis de hesitação à vacina de determinada população.	Os pesquisadores constataram que existe significativa ligação entre o aumento do uso das mídias sociais e a elevação na crença de que as vacinas não são seguras, bem como também concluíram que campanhas estrangeiras de propagação de desinformação ocasionam declínio nas taxas médias de vacinação de um país.
Das, Singh & Sharma (2021)	Notícias online	Realizar análise de sentimentos em notícias publicadas online pela mídia indiana e que tivessem vacina como tema.	Os autores encontraram que, apesar de a maior parte das notícias transmitir sentimentos positivos em relação às vacinas, uma porcentagem relevante delas transmite sentimentos de negatividade, indicando que o poder público deve monitorar as redes para que a aceitação das vacinas pela população não seja prejudicado.
Yin et al. (2021)	Weibo	Analisar a opinião pública dos chineses em relação às vacinas e à vacinação contra a Covid-19.	Em relação ao sentimento da população a respeito do contexto em que a vacina está inserida, os pesquisadores constataram que os chineses se mostraram positivos sobre o preço e ocorrência de efeitos colaterais da vacina, bem como também se mostraram positivos em relação às vacinas elaboradas por meio de vírus inativado.

Fonte: Elaborada pelos autores.

Abordados alguns estudos que empregaram a metodologia de análise de sentimento nas redes sociais para verificar a opinião pública a respeito da temática da vacina, na próxima seção serão apresentados alguns trabalhos que aplicaram o método da análise de sentimentos como ferramenta de auxílio para a previsão do movimento do mercado acionário.

## 2.2 ANÁLISE DE SENTIMENTOS E MERCADO FINANCEIRO - ESTUDOS EMPÍRICOS

A possibilidade de antever o comportamento e as oscilações do mercado financeiro atrai tanto a atenção dos acadêmicos quanto a dos investidores, sendo esta uma área que concentra o interesse de muitos pesquisadores que buscam a utilização de variados métodos e técnicas para alcançar este fim (Bollen, Mao & Zeng, 2011). Por conseguinte, Carosia, Coelho e Silva (2019) apontam que, via de regra, os trabalhos que aplicam análise de sentimento ao mercado financeiro analisam computacionalmente textos extraídos de notícias ou publicações nas redes sociais e atribuem a eles, de forma automatizada, uma classificação polarizada entre positiva e negativa. Posteriormente, é aplicada alguma técnica que associa o sentimento momentâneo identificado e a direção tomada pelo mercado financeiro.

Bollen, Mao e Zeng (2011) e Mittal e Goel (2012) desenvolveram trabalhos nos quais buscaram prever o movimento do Dow Jones Industrial Average (DJIA), principal índice acionário da bolsa de valores dos Estados Unidos, usando como variável preditora os resultados da análise de sentimento dos *tweets* observados nas datas analisadas. O primeiro trabalho utilizou a Causalidade de Granger e a rede neural Fuzzy (SOFNN) para relacionar a oscilação do humor público com a oscilação do mercado financeiro para o período de 28/02/2008 a 19/12/2008, conseguindo prever as oscilações diárias do DJIA com 87.6% de precisão. Já o segundo trabalho utilizou as metodologias de regressão linear, regressão logística, SVM (Support Vector Machine) e também a SOFNN para verificar a relação entre o sentimento público encontrado e as oscilações do mercado acionário para o período compreendido entre junho e dezembro de 2009. Os resultados apresentaram acurácia de 75.56% quando da utilização do sentimento público para prever o “sentimento do mercado”.

Na pesquisa de Bharathi e Geetha (2017), os autores buscaram verificar se a inclusão do sentimento como variável preditora geraria um modelo de previsão do comportamento de um ativo com acurácia maior do que um modelo que usasse apenas dados históricos da companhia na previsão. Para tal, os estudiosos realizaram a análise de sentimento em notícias relacionadas ao mercado acionário indiano no mês de abril de 2006 e usaram a série histórica de cotação de uma empresa do país para realizar a pesquisa. O método para a análise de sentimentos foi o léxico e o método para a verificação do impacto da inclusão da variável sentimento nas oscilações no mercado foi feito pela comparação entre as médias móveis obtidas pelos dois modelos, com os valores observados. Como resultado os pesquisadores encontraram acurácia de 64,32% para o modelo ajustado sem o sentimento, enquanto que o modelo com o sentimento apresentou acurácia de 78,75%, demonstrando que o sentimento é uma variável preditora útil para antever o comportamento das ações da empresa analisada.

Lima et al. (2016) aplicaram a análise de sentimento para tentar prever oscilações dos ativos do mercado financeiro brasileiro. Eles empregaram as ações da Petrobras como foco dos estudos. A análise de sentimentos, feita pelo método léxico, teve o intuito de identificar o humor coletivo em relação à Petrobras usando, posteriormente, o humor encontrado para tentar prever o comportamento das ações da companhia por meio do método SVM. Como resultado, os autores conseguiram uma precisão de 82,93% na previsão das oscilações das ações da Petrobras.

Carosia, Coelho e Silva (2019) avaliaram o impacto de notícias e *tweets* relacionados ao mercado financeiro brasileiro por meio da análise de sentimento realizada por meio do método de rede neural artificial do tipo MLP. Posteriormente, os pesquisadores compararam qual das fontes de dados tinha maior poder preditivo sobre o índice Ibovespa. O método SVM foi utilizado para realizar as previsões. Como resultado, constatou-se que o sentimento geral da população brasileira em relação às notícias do dia  $t$  tem maior influência no valor de abertura e no volume negociado da bolsa no dia  $t + 1$ , com precisão de 60,6% para ambos. No entanto, quando se trata do valor de fechamento da troca, verificou-se que os sentimentos contidos nos *tweets* apresentam maior impacto, demonstrando uma precisão de 54%.

Ademais, o trabalho publicado por Araújo e Marinho (2018) também teve o mercado acionário brasileiro como campo de pesquisa. O objetivo deste trabalho foi investigar o impacto da divulgação de notícias online sobre o tema “economia” na performance da BM&FBovespa (B3). Por meio de um classificador de sentimentos, os textos de cada notícia utilizada foram classificados quanto a sua polaridade (positivo ou negativo). Como resultado da pesquisa, os autores descreveram que o número de notícias disponibilizadas a cada 15 minutos influencia a movimentação do mercado financeiro no Brasil, sendo que para a publicação de 19 notícias ou mais nesse período de tempo, pode-se perceber que a qualidade da previsão do mercado financeiro gira em torno de 70%. Além disso, os autores observaram que a repercussão de notícias negativas é maior do que a de notícias positivas e que, portanto, aquelas que expressam sentimentos negativos tendem a apresentar maior impacto na performance da B3 do que as notícias positivas. A Tabela 2 apresenta alguns outros estudos relacionados à temática.

**Tabela 2**

Pesquisas empíricas que aplicaram a análise de sentimentos na previsão do mercado acionário

Autor(es)/ Ano	Objetivos	Principal Resultado
Valle-Cruz et al. (2020)	Verificar se os sentimentos contidos nas postagens do Twitter advindas de perfis com importância global teriam influência no comportamento de índices financeiros de variados países e setores da economia em épocas de pandemia, usando períodos de surto da H1N1 e da Covid-19 .	Foi descoberto que no período da pandemia de Covid-19 os sentimentos influenciaram a movimentação de mercado de 6 a 13 dias após a postagem, sendo esse prazo de 1 a 2 dias na temporada de H1N1.
Yadav et al. (2019)	Verificar o efeito dos sentimentos da população indiana sobre a movimentação do mercado acionário do país.	O sentimento da população indiana, medido por três métodos diferentes, foi usado na previsão do mercado acionário do país resultando em 3



		previsões que retornaram acurácias que variaram de 75% a 79%
Shah, Isah & Zulkernine (2018)	Avaliar o quanto os sentimentos contidos nas notícias em relação ao setor farmacêutico indiano influenciam no movimento das ações das empresas que compõem o índice Nifty Pharma daquele país.	Os autores verificaram que o modelo de análise de sentimentos proposto teve poder de previsão de 70,59% para antever as tendências de preços das ações no curto prazo.
Ren, Wu & Liu (2018)	Treinar um modelo de previsão das tendências de comportamento de mercado do índice SSE 50 com base no sentimento chinês dominante.	Os pesquisadores encontraram que a previsão das tendências da bolsa é aprimorada em 18,6% quando da inclusão da variável sentimento como preditora dessas oscilações.
Igarashi, Valdevieso & Igarashi (2021)	Verificar a correlação entre os preços das ações e a polaridade das notícias relativas ao mercado acionário brasileiro, usando a ação preferencial da empresa Petrobrás (PETR4) como base para os estudos.	Verificou-se que a polaridade observada nas notícias financeiras teve correlação moderada com as oscilações do ativo da Petrobras.

Fonte: Elaborada pelos autores

A próxima seção apresenta a ferramenta de análise de sentimentos como empregada no presente estudo.

### 2.3 ANÁLISE DE SENTIMENTOS

A atual admiração e aceitação das mídias sociais tem gerado grande volume de informações, possibilitando a realização e incentivando o desenvolvimento de novas técnicas de análise para dados não estruturados, como imagens, vídeos e textos (Santos, Silveira & Lechugo, 2017). Nesse contexto, a análise de sentimentos surge como uma técnica de mineração de dados. Este tipo de técnica visa extrair informações úteis dos dados que sem análise, manipulação e interpretação não possuem teor informativo e não servem de embasamento para processos decisórios (Castro, 2016).

A análise de sentimentos é empregada em diversos tipos de dados para gerar informação útil e conhecimento. Sua principal finalidade é a identificação da opinião pública a respeito de determinado produto, pessoa, evento, organização ou tópico. O objetivo é identificar se o sentimento geral a respeito da temática em estudo é positiva ou negativa. As informações obtidas podem ser usadas com inúmeras finalidades como, por exemplo, montar estratégias de marketing ou mesmo de investimentos (Liu & Zhang, 2012)

Existem múltiplas técnicas para a realização da análise de sentimentos, sendo que estas variam desde métodos mais simples como a abordagem léxica,

até técnicas mais elaboradas, como as executadas por meio de algoritmos de aprendizagem de máquina (machine learning). Um algoritmo de machine learning que pode ser utilizado para este tipo de tarefa é o chamado *Naive Bayes*. Este é um algoritmo de aprendizagem supervisionada comumente utilizado com a função de classificação de sentimentos em textos e que tem apresentado resultados de precisão satisfatórios (Wang & Manning, 2012).

Métodos supervisionados necessitam de um conjunto de dados pré-rotulados para treino do algoritmo de aprendizagem. O algoritmo *Naive Bayes* tem embasamento no Teorema de Bayes, que trata de probabilidades condicionais, e funciona por meio da geração de uma tabela de probabilidades. Nesta tabela é verificada a probabilidade de cada palavra da frase pertencer à classe positiva e negativa. É atribuída como polaridade da frase a classe que possuir a maior probabilidade de ocorrência. A Equação 1, adaptada de Oliveira (2016), ilustra como é feito o cálculo da probabilidade de uma frase ser positiva pelo método *Naive Bayes*. A probabilidade de uma frase pertencer à classe negativa segue a mesma proposta, com a mudança de “positiva” para “negativa” na equação:

$$P(\text{frase}) = P(\text{positiva}) \cdot \prod_{j=1}^n P(V_j|\text{positiva}), \quad (1),$$

em que  $P(\text{frase})$  é a probabilidade de que uma frase qualquer, esta seja classificado como “positiva”;  $P(V_j|\text{positiva})$  é a probabilidade de que a palavra  $V_j$  esteja em uma frase positiva;  $P(\text{positiva})$  representa a probabilidade de uma frase ser classificada como positiva e  $n$  representa a quantidade de palavras da frase.

A próxima seção apresenta o método usado para a realização das previsões das oscilações do mercado acionário.

### 3 PROCEDIMENTOS METODOLÓGICOS

#### 3.1 Método para previsão das oscilações do mercado acionário

Para a verificação da influência do sentimento dos brasileiros usuários do Twitter nas oscilações do mercado acionário nacional, foram realizadas duas previsões. Uma utilizou somente os dados históricos relacionados ao índice Ibovespa, que é a métrica do mercado acionário brasileiro. A outra utilizou além dos dados históricos o sentimento preponderante diário como variável previsor.

O método empregado para a realização das previsões de oscilação do mercado acionário brasileiro foi o Support Vector Machine (SVM). Este método de machine learning é reconhecido por apresentar significativa eficiência tanto em modelos de regressão, quanto em modelos de classificação (Apostolidis-Afentoulis & Lioufi, 2015). Em síntese, o SVM se enquadra na categoria dos métodos supervisionados. Sua ideia central é a criação de um modelo que consiga gerar um hiperplano capaz de separar duas classes, mantendo entre elas a maior distância possível (Pang, Lee & Vaithianathan, 2002).

Em decorrência da pequena quantidade de dados obtidos para a realização da pesquisa, aplicou-se a técnica de validação cruzada para séries temporais. Nesta técnica, foi empregado um algoritmo conhecido como k-fold, no qual um conjunto de  $n$  dados é dividido em  $k$  subconjuntos de aproximadamente mesmo tamanho. Depois, são feitos  $k$  treinamentos. Em cada treinamento um subconjunto é usado como base de teste para a computação

dos erros de classificação, enquanto que os outros subconjuntos são usados para o treinamento do modelo (Shrivastava, 2020). O resultado da acurácia da validação cruzada refere-se à média aritmética dos resultados encontrados em cada um dos  $n$  treinamentos. Essa técnica evita que a avaliação da assertividade do modelo seja comprometida, o que pode ocorrer caso seja realizada apenas uma separação de dados de teste e treino.

Como o algoritmo SVM é da classe supervisionada, para que fosse possível realizar o treinamento do modelo de previsão da tendência da movimentação do mercado acionário, foi necessário o fornecimento de dados para a aprendizagem. Nesse contexto, a primeira etapa da previsão consistiu na geração de atributos para o treinamento do modelo. Os atributos relacionados à série histórica para foram calculados com base no trabalho de Santos (2020) e estão apresentados nas equações (2), (3), (4), (5) e (6):

Índice de força relativa (RSI):

$$RSI = 100 - \frac{100}{1+RS'} \quad (2),$$

$$RS = \frac{\text{Ganho médio nos últimos 5 dias}}{\text{Perda média nos últimos 5 dias}'} \quad (3),$$

em que o ganho médio é calculado pela divisão da quantidade de pontos ganhos pelo índice nos últimos 5 dias por 5, e a perda média é calculada pela divisão da quantidade de pontos perdidos nos últimos 5 dias por 5, a partir dos valores de fechamento diário do índice.

Oscilador estocástico (%K):

$$\%K = 100 * \frac{C - P5}{A5 - P5}, \quad (4),$$

em que C é a quantidade atual de pontos, P5 é a menor quantidade observada de pontos do índice Ibovespa dos últimos 5 dias e A5 é a maior quantidade de pontos observada nos últimos 5 dias. Os dados foram obtidos a partir dos valores de fechamento diário do índice.

Willians (%R) :

$$\%R = 100 * \frac{A5 - C}{A5 - P5}, \quad (5),$$

em que C é a cotação atual, P5 é a menor quantidade observada de pontos do índice Ibovespa dos últimos 5 dias e A5 é a maior quantidade de pontos observada nos últimos 5 dias, sendo os dados obtidos a partir dos valores de fechamento diário do índice.

Taxa de mudança dos preços (PROC (t)):

$$PROC(t) = \frac{C(t) - C(t-5)}{C(t-5)}, \quad (6),$$

em que  $PROC(t)$  representa a taxa de mudança no tempo  $t$ ,  $C(t)$  representa a quantidade de pontos no tempo  $t$  e  $C(t - 5)$ , representa a quantidade de pontos no fechamento do índice no tempo  $t - 5$  dias.

Além disso, para obter a variável referente ao sentimento diário predominante utilizado no segundo processo de previsão, foi utilizada a mesma equação utilizada em Carosia, Coelho e Silva (2019), conforme apresentado na equação (7):

$$\text{Sentimento preponderante diário} = \frac{(nPos - nNeg)}{(nPos + nNeg)}, \quad (7),$$

onde  $nPos$  representa o número de *tweets* classificados como positivos e  $nNeg$  representa o número de *tweets* classificados como negativos no dia analisado. Resultados positivos indicam sentimentos predominantemente positivos e resultados negativos indicam sentimentos predominantemente negativos no dia avaliado. Por fim, a avaliação de desempenho dos dois modelos gerados foi baseada na métrica de acurácia, que mede o número de valores classificados corretamente entre todos os valores classificados pelo modelo. A equação (8) apresenta a fórmula da acurácia:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN}, \quad (8),$$

em que VP/VN representa o número de previsões de alta/baixa no índice Ibovespa classificados corretamente e FP/FN é o número de previsões classificadas incorretamente. A próxima seção apresenta as etapas da pesquisa.

### 3.2 DESENVOLVIMENTO DA PESQUISA

As etapas empregadas comumente na elaboração de uma análise de sentimento são: 1) Coleta dos dados; 2) Pré-processamento dos dados; 3) Treinamento do modelo; 4) Classificação quanto à polaridade; e 5) Análise dos resultados, sendo que esta pesquisa terá também uma sexta etapa que é: 6) A verificação da relação do sentimento diário dos usuários do Twitter em determinado dia com a oscilação do mercado financeiro nacional para este mesmo dia. Abaixo, cada uma das etapas é descrita de forma detalhada.

**1) Coleta dos dados de texto** - Por meio de um API no *software Python 3.9* foi realizada a coleta de 100.000 *tweets* realizados entre os dias 01/12/2020 e 13/07/2021. Os *tweets* coletados não foram focados em usuários específicos. Os textos captados se referem a contas de diversos tipos de usuários brasileiros. Os *tweets* coletados foram divididos em dois grupos. Um grupo foi empregado no treinamento e validação do modelo de classificação de polaridade das mensagens, a exemplo do que foi feito no trabalho de Hasan, Rundensteiner e Agu (2014). O outro grupo foi aplicado ao modelo já treinado, com o intuito de receber rótulos em relação à sua polaridade.

Dos 100.000 *tweets* da base de dados, metade estava relacionada a sentimentos positivos e a outra metade estava relacionada a sentimentos negativos. Para a especificação do grupo com frases de sentimentos positivos foram filtrados os *tweets* que continham palavras relacionadas a emoções positivas, sendo estes captados pelas hashtags: #felicidade, #amor, #gratidão, #alegria, #cura, #esperança e #emoção; já o grupo que continha frases negativas foi especificado por meio do filtro a textos que continham emoções negativas,

sendo captados pelas hashtags: #medo, #tristeza, #vergonha, #raiva, #revolta, #doença, #dor, #doente, #desgosto e #infeliz. Ademais, para a realização na análise de sentimentos proposta neste trabalho, foi possível coletar, para as mesmas datas, 68.408 dados relacionados às hashtags “vacina” e “vacinação”, que posteriormente foram rotulados pelo modelo treinado.

**2) Pré-processamento dos dados** - Foi feita a limpeza dos grupos de dados empregados para treino e validação do modelo - retirada de pontuação, símbolos, links, números, *stopwords*, dados duplicados e dados em língua estrangeira. A limpeza foi realizada para retirar dos textos os caracteres não representativos de sentimentos ou aqueles que não fossem palavras, pois estes não são relevantes no modelo proposto. Ao final deste processo, restaram 79.928 observações para treino e validação do modelo, sendo elas divididas igualmente entre as classes positivas e negativas. Em relação aos dados das hashtags #vacina e #vacinação 55.998 observações permaneceram após a etapa de pré-processamento.

**3) Treinamento do modelo** - Treinamento do modelo - O treino do de classificação de polaridade dos dados seguiu o algoritmo *Naive Bayes*. Para esta tarefa foi empregado o pacote *e1071* no software R, versão 4.0.3. Nesta fase utilizaram-se somente os dados coletados para o treinamento do modelo, que já estavam rotulados quanto a sua polaridade, conforme explicado na etapa 1. Das 79.928 observações separou-se de forma aleatória 75% para o treinamento do modelo, e 25% para a sua validação. Tanto no grupo de treino quanto no grupo usado na validação houve equilíbrio percentual das classes “positivo” e “negativo”.

**4) Classificação dos dados:** O conjunto de *tweets* relacionados às #vacina e #vacinação, foram aplicados ao modelo já treinado. Cada *tweet* foi classificado de acordo com a classe que apresentava a maior probabilidade segundo o modelo *Naive Bayes*. Posteriormente, foi realizada a classificação manual de alguns *tweets* rotulados pelo modelo proposto para verificar sua acurácia.

**5) Análise dos dados** - Por meio da classificação dos dados foi possível desenvolver a análise de sentimentos dos usuários brasileiros do Twitter a respeito do tema “vacina” nas datas avaliadas.

**6) Relação com o movimento de mercado** - Aqui buscou-se verificar, por meio do treinamento de um modelo *SVM* por validação cruzada de séries temporais, se o sentimento preponderante observado em determinado dia ajudaria na previsão da oscilação da bolsa de valores deste mesmo dia. Acredita-se por hipótese que o resultado da inclusão da variável sentimento proporcionará um modelo predictor com resultados mais acurados. Para isso, considerou-se movimentos de queda entre dois dias consecutivos com o valor -1, movimentos de alta com o valor 1 e tendência de estabilidade com o valor 0. Assim, a verificação da influência do sentimento a respeito da vacina nas oscilações do mercado acionário brasileiro foi feita pela comparação da acurácia obtida pelo modelo de previsão que usou somente os dados relacionados à série histórica do índice Ibovespa como variável preditora, com a acurácia do modelo de previsão que utilizou, além dos dados da série histórica, os valores de sentimento preponderante diário.

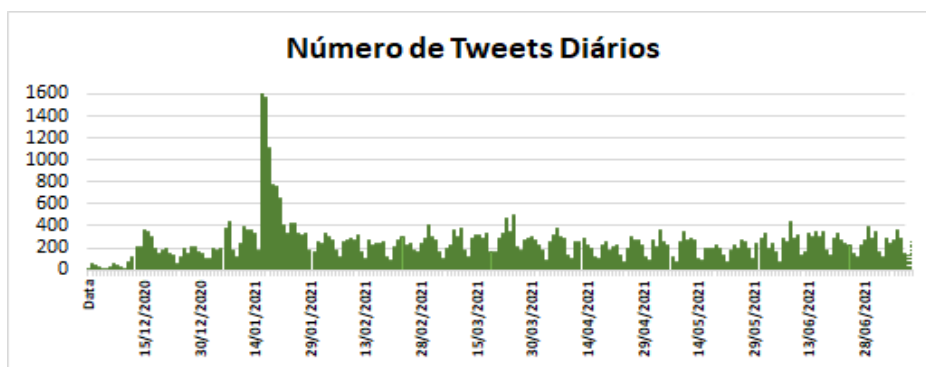
É importante ressaltar que o índice Ibovespa só possui dados disponíveis em dias úteis. Portanto, a base de dados foi ajustada para que contivesse o sentimento predominante apenas para as datas em que a cotação do índice Ibovespa estava disponível. Ao final desse processo, o banco de dados continha informações de 144 dias. A próxima seção expõe os resultados encontrados.

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

### 4.1 ANÁLISE DO SENTIMENTO DOS BRASILEIROS EM RELAÇÃO À VACINA DA COVID-19

Como é possível observar na Figura 1, o número diário de *tweets* publicados com as hashtags #vacina e #vacinação varia comumente entre 100 e 400 postagens. Contudo, a disposição dos dados evidencia que entre os dias 17 e 22 de janeiro houve grande aumento na quantidade de *tweets*. Neste período, o número diário de *tweets* relacionado à temática alcançou a marca de 550 a 1600 publicações em um dia.

Partindo disso, foi realizado um breve estudo de eventos para a melhor compreensão do período em questão. Os resultados deste estudo sugeriram que o aumento no número das publicações relacionadas às hashtags #vacina e #vacinação nestes dias pode estar relacionado à liberação do uso emergencial de duas vacinas contra a Covid-19 pela Agência Nacional de Vigilância Sanitária (Anvisa). Além disso, neste período está englobada a data na qual a primeira pessoa foi vacinada no país (Sabóia, 2021). Além disso, neste período também estava ocorrendo a crise do abastecimento de oxigênio na cidade de Manaus que acabou por tirar a vida várias pessoas que estavam internadas com Covid-19 e necessitavam deste insumo para sobreviver (Lupion, 2021). Este fato pode ter contribuído para o aumento da pauta #vacina entre os internautas, visto que somente a imunização em massa em relação à doença poderia evitar que o sistema de saúde brasileiro observasse mais um colapso (Costa, 2021).



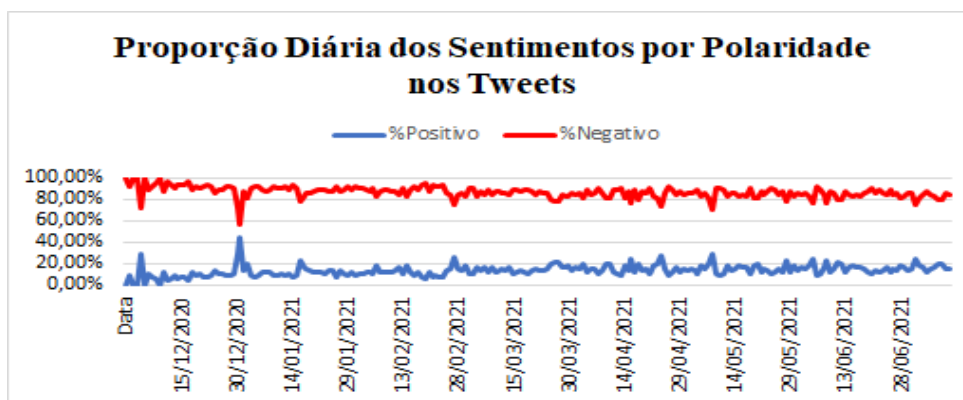
**Figura 1** – Número diário de *Tweets* em relação às #vacina e #vacinação entre os dias 01/12/2020 e 13/07/2021.

Fonte: Elaborada pelos autores.

Destarte, realizou-se o treinamento do modelo conforme apresentado no passo 3, da seção 3.2. A realização da validação cruzada no modelo de classificação proposto tornou possível verificar que este apresentou acurácia de 96,90%, visto que das 19.982 mensagens usadas na validação do modelo, 19.363 foram classificadas com a polaridade correta. Dado o satisfatório nível de acurácia do modelo, aplicou-se nele os dados de texto dos *tweets* que continham as hashtags #vacina e #vacinação, para que fossem rotulados quanto a sua polaridade de forma automatizada.

A análise da polaridade dos sentimentos contidos nas mensagens coletadas permitiu a verificação do tipo de sentimento preponderante nos textos. O resultado demonstrou grande prevalência de mensagens negativas a respeito

da temática. Verificou-se que 86,08% das mensagens rotuladas pelo modelo foram classificadas como negativas e apenas 13,92% tiveram classificação positiva. Quando realizada a validação dos dados por rotulação manual, verificou-se que esta classificação obteve 68% de acurácia. Na Figura 2 é ilustrada a proporção de Tweets positivos e negativos rotulados pelo modelo proposto.

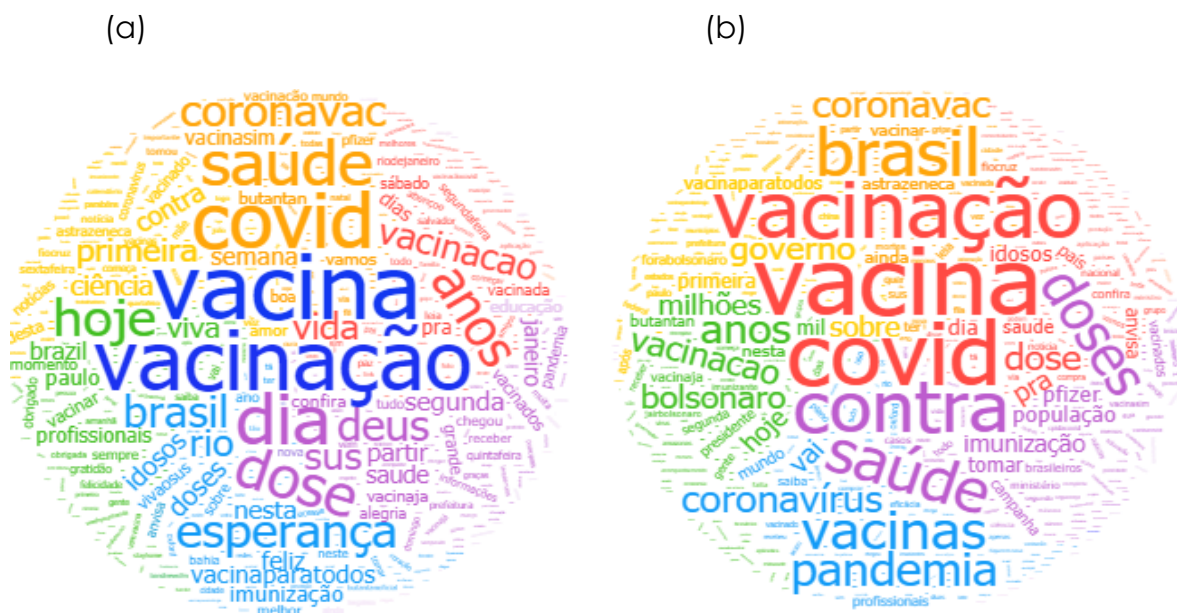


**Figura 2** – Proporção diária dos sentimentos por polaridade nos tweets entre os dias 01/12/2021 e 13/07/2021.

Fonte: Elaborada pelos autores

Assim, constata-se a clara dominância de sentimentos negativos em torno da temática da vacina/vacinação no Brasil, mesma conclusão obtida na pesquisa realizada por Milani, Weitkamp e Webb (2020). Uma explicação possível para a ocorrência deste resultado é que no período em que os tweets foram coletados o Brasil vivenciava a escassez de vacinas. Além disso, em meados de março e abril o país passou pela segunda onda da pandemia. Isto ocasionou inúmeras e levou o Brasil ao topo do mundo no que diz respeito ao número de mortes diárias (Galzo, 2021).

Logo, tem-se que a dominância dos sentimentos negativos pode ser consequência do medo das pessoas que têm vivenciado diretamente os efeitos da pandemia. Também pode surgir da falta de expectativa em relação a agilidade das campanhas de vacinação no país, o que leva a sentimentos de raiva, medo, frustração e preocupação. Sentimentos negativos em relação à vacinação também podem ser expressos por pessoas que não confiam neste método de imunização e que usam as redes para expressar-se ou pela divulgação e espalhamento de fake news a respeito da temática (Massarani, 2021). É necessário um estudo mais aprofundado para afirmar de forma contundente os motivos que levam as hashtags #vacina e #vacinação nas publicações de perfis brasileiros serem expressivamente negativas. Para um grau maior de compreensão sobre os sentimentos em relação à temática foram construídas nuvens de palavras com os termos mais frequentes de cada classe e que estão apresentadas na Figura 3.



**Figura 3** – Nuvem de palavras dos tweets classificados pelo modelo Naïve Bayes como sentimentos positivos (a) e sentimentos negativos (b)  
 Fonte: Elaborada pelos autores

A observação superficial das duas nuvens remete para palavras chaves bem parecidas em ambas. Contudo, quando observa-se a totalidade das palavras de cada uma verifica-se que a principal diferença entre elas é que a positiva coloca sentimentos positivos, como “esperança”, “alegria” e “feliz”, enquanto que a nuvem de palavras negativas apresenta palavras que se relacionam com a política brasileira e o reforço da situação de “pandemia”. Ademais, é importante destacar que, apesar de aparecerem palavras/termos parecidos em ambas as nuvens, essas palavras foram usadas em contextos diferentes. As frases positivas foram assim classificadas por estarem alinhadas com frases de cunho positivo presentes no conjunto de treino do modelo, sendo que o mesmo processo ocorreu com as frases classificadas como negativas.

Dessa forma as hashtags pró-vacina que ganharam espaço no Brasil, por exemplo: #vacinaparatodos, #vacinajá e #vacinasim, aparecem nas duas nuvens, sendo que na nuvem dos tweets positivos elas estão inseridas em um contexto de incentivo e esperança em relação à imunização e na nuvem dos tweets negativos elas aparecem com um sentido de insatisfação relacionada aos problemas da campanha de vacinação no Brasil.

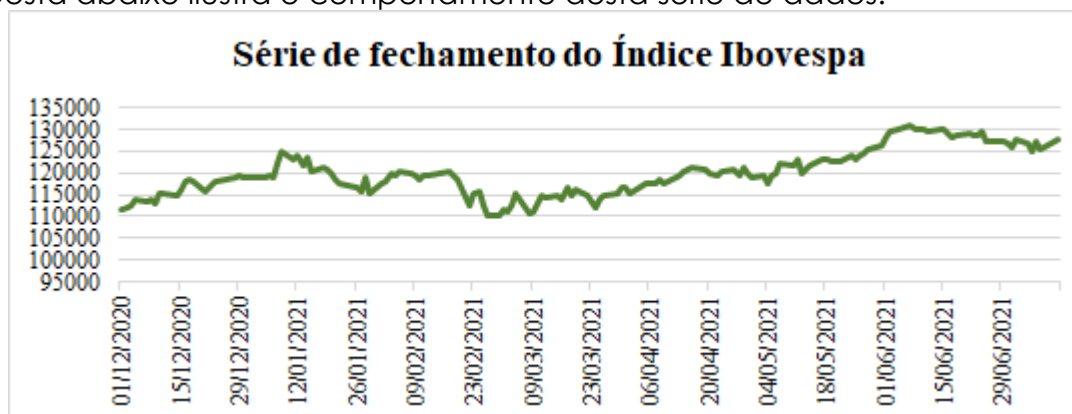
Como conclusão desta seção reafirma-se que o grupo de brasileiros analisados nesta pesquisa tem sentimentos predominantemente negativos em torno da temática da vacina e da vacinação contra a Covid-19. Assim como apontado por Das, Singh e Sharma (2021), nas mídias online existe a divulgação de razoável número de notícias que transmitem sentimentos negativos sobre a imunização por vacina. Wiysonge et al. (2021) aponta que isto pode gerar inseguranças e receios nos usuários das mídias sociais a seu respeito.



O conhecimento de quais sentimentos são majoritários na população a respeito da temática e alguns dos alvos que desencadeiam essas emoções, tornam possível um diálogo mais efetivo com o povo. Isso possibilita com que formuladores de políticas públicas possam elaborar estratégias de publicidade governamentais que tratem das inseguranças e receios da população a respeito da imunização, possibilitando campanhas de vacinação mais eficientes. Além disso, para que as campanhas de vacinação tenham sucesso deve-se aumentar a credibilidade e confiança da população nos órgãos públicos em relação a este aspecto. Para verificar se além da hesitação vacinal, o sentimento negativo a respeito da vacina tem outras implicações, a próxima seção apresenta a previsão das oscilações do mercado financeiro com e sem a variável “sentimento” como previsora.

#### 4.2 USANDO O SENTIMENTO DOS BRASILEIROS USUÁRIOS DO TWITTER COMO VARIÁVEL PREDITORA DAS OSCILAÇÕES DO MERCADO FINANCEIRO

Para avaliar a hipótese proposta nesta pesquisa de que a inclusão do sentimento dos brasileiros usuários do Twitter em relação à vacina/vacinação contra a Covid-19 no Brasil como variável preditora é capaz de tornar as previsões a respeito do mercado acionário mais assertivas, o primeiro passo foi coletar informações sobre o valor diário de fechamento do índice Ibovespa. A Figura 4 disposta abaixo ilustra o comportamento desta série de dados.



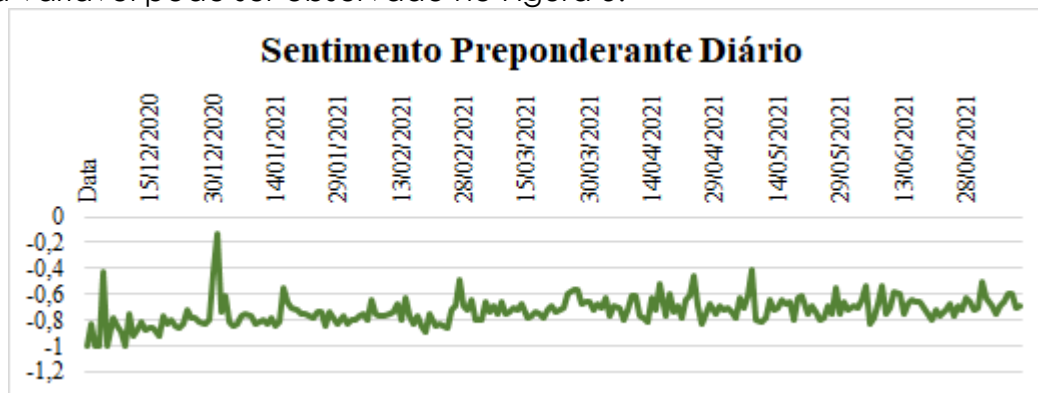
**Figura 4** – Série com valores de fechamento do índice Ibovespa entre os dias 01/12/2020 e 29/06/2021, com dados do site Infomoney.

Fonte: Elaborada pelos autores

A visualização da Figura 4 mostra que as oscilações dos valores de fechamento do índice analisado entre 12/2020 e 07/2021, foram bastante elevadas. Neste período o índice variou entre o mínimo de 110.035 pontos e o máximo de 130.776 pontos. Dessa forma, tem-se que a grande volatilidade dos dados torna a modelagem e previsão desta série uma tarefa bastante complexa (Santos, 2020). Como explicado na metodologia deste trabalho, buscou-se prever não os valores da série, mas sim a sua tendência de comportamento. A primeira projeção dos valores do Ibovespa foi realizada por validação cruzada com o método SVM e utilizou as variáveis *RSI*, *%K*, *%R* e *PROC(t)*, calculadas a partir da série histórica do índice, como atributos para o treinamento do modelo.

Para ajustar o modelo de validação cruzada para séries temporais, usou-se como padrão o kernel “radial basis”, com valor de custo igual a 1 e k igual a 5

separações. O resultado da validação cruzada obteve acurácia média de 92,27% no modelo de previsão proposto somente com as variáveis relacionadas à série histórica do índice Ibovespa, com desvio padrão igual a 8,32%. Posteriormente, o segundo modelo de previsão foi ajustado com o acréscimo da variável “sentimento preponderante diário” dos usuários brasileiros do Twitter em relação à vacina/vacinação da Covid-19 como previsor. O comportamento desta variável pode ser observado no Figura 5.



**Figura 5** – Comportamento da série de dados do sentimento preponderante diário entre 01/12/2021 e 13/07/2021.

Fonte: Elaborada pelos autores

Como pode ser observado na Figura 5, a série de dados relacionada ao sentimento preponderante diário é bastante volátil e varia entre dias com maior e menor negatividade a respeito do tema da vacina/vacinação. Sobre a correlação entre os valores de fechamento do índice Ibovespa com o sentimento preponderante diário verificou-se o valor de apenas 0,3677, que pode ser considerada fraca, sendo este um indicativo de que os movimentos dessas variáveis não estão muito alinhados.

Assim, o novo modelo SVM por validação cruzada, também ajustado pelo kernel “radial basis”, com valor de custo 1 e k igual a 5 separações, obteve acurácia de 92,05%, com desvio padrão igual a 7,67%. Portanto, considerando o desvio padrão, dentre as possibilidades de variações de resultados afirma-se que ambos os modelos são igualmente assertivos. Na Tabela 3 é apresentada uma síntese dos resultados encontrados em ambos os modelos de previsão propostos.

**Tabela 3**

Resultados dos dois modelos de previsão ajustados

Modelo	Variáveis Previsoras	Acurácia Média	Desvio Padrão
1	<i>RSI, %K, %R e PROC(t)</i>	92,27%	8,32%
2	<i>RSI, %K, %R, PROC(t) e Sentimento preponderante diário</i>	92,05%	7,67%

Fonte: Elaborada pelos autores

A conclusão do experimento é que, diferente do que foi apresentado em pesquisas semelhantes em países como Estados Unidos (Bollen, Mao & Zeng, 2011; Mittal & Goel, 2012), China (Ren, Wu & Liu, 2018) e Índia (Yadav et al., 2019; Shah, Isah & Zulkernine, 2018), no modelo de classificação e de previsão proposto a inclusão da variável de sentimento relacionada à vacina/vacinação em um dia

não melhorou o processo de previsão das oscilações do mercado acionário do país no dia seguinte. Este resultado se diferencia também do encontrado na pesquisa realizada por Carosia, Coelho e Silva (2019) que conseguiram acurácia na utilização do sentimento geral diário para prever os valores de abertura, fechamento e o volume de negociações no mercado brasileiro.

Os autores Araújo e Marinho (2018) e Lima et al. (2016) que também conseguiram relacionar sentimentos diários à movimentação acionária no Brasil, adotaram procedimentos metodológicos diferentes dos adotados nesta pesquisa. Isto pode ter contribuído para os resultados mais assertivos de suas pesquisas. O primeiro trabalho realizou a análise dos sentimentos e variação de mercado a cada 15 minutos e não a cada dia. O menor período de tempo pode ter contribuído para que o modelo captasse melhor as variações de humor durante o dia e sua relação com o índice Ibovespa.

Já o segundo trabalho segmentou mais o seu campo de pesquisa, avaliando se os sentimentos dos internautas especificamente relacionados a uma empresa, afetariam seu desempenho na bolsa. Esta especificação maior pareceu trazer maiores ganhos no poder de previsão, do que usar o mercado em geral. Esta constatação também pode ser feita verificando que o trabalho de Igarashi, Valdevieso e Igarashi (2021), no qual os autores também consideraram apenas uma empresa na análise, conseguiram ao menos uma correlação moderada entre os sentimentos e a movimentação de mercado, diferentemente do que aconteceu nesta pesquisa, na qual a correlação encontrada foi baixa.

A pesquisa aqui proposta se assemelha às pesquisas anteriores dentro da sua proposta de avaliar o impacto de uma variável não convencional nas movimentações do mercado acionário. Apesar de não terem sido confirmados os resultados esperados como hipótese inicial, acredita-se que o diferencial desta pesquisa está em apresentar uma metodologia inovadora de previsão de mercado que já tem sido usada fora do Brasil e também servir como embasamento para a construção de novas ideias relacionadas a esta tarefa. Como contribuição adicional, esta pesquisa trouxe sugestões metodológicas desempenhadas por outros pesquisadores que podem auxiliar pesquisas futuras a terem maior sucesso. Por fim, na próxima seção estão explicitadas as considerações finais deste trabalho, apresentando suas conclusões, limitações e oferecendo sugestões para pesquisas futuras.

## 5 CONCLUSÕES

Pela análise do sentimento dos usuários brasileiros do Twitter sobre o tema vacinação contra a Covid-19, constatou-se que existem sentimentos predominantemente negativos a esse respeito. Isso demonstra a necessidade de intervenção dos formuladores de políticas públicas, principalmente no que diz respeito às campanhas de vacinação e melhor informação à população sobre esse tipo de imunização. Além disso, deve haver o foco em agilizar o processo de imunização no país, tendo em vista que a demora nesse processo é uma crítica frequentemente mencionada em mensagens negativas. Essas estratégias tenderiam a ajudar a recuperar a credibilidade de entidades públicas que também aparecem com destaque em textos negativos.

Quanto à relação entre os sentimentos sobre a vacina e as oscilações observadas no mercado de ações brasileiro, o método utilizado nesta pesquisa não foi capaz de confirmar a hipótese inicial de que o conhecimento da primeira ajudaria a prever o comportamento da segunda. Isso porque a inclusão da variável preditiva “sentimento” no modelo de classificação não retornou um modelo com melhor acurácia do que aquele gerado apenas com a série histórica do índice Ibovespa.

Nesse contexto, apesar dos resultados, não é possível afirmar de forma generalizada que o sentimento não tem influência sobre os resultados do mercado acionário, pois além de alguns pesquisadores terem encontrado essa relação, pode ser que o sentimento sobre a temática da vacinação realmente não seja importante para a sua previsão, mas o sentimento sobre outras temáticas sejam. Entretanto, mesmo para a temática da vacinação, pode ser que caso a metodologia fosse alterada, por exemplo, caso fosse empregado um outro método para a classificação dos sentimentos ou outra fórmula para a obtenção do sentimento preponderante diário, o resultado encontrado poderia ter sido diferente, sendo essas sugestões para pesquisas futuras. Destarte, sugere-se também a realização de pesquisas que comparem os resultados obtidos por diferentes métodos de classificação textual.

O trabalho apresentou como principal limitante a pequena quantidade de dados. Isto porque a vacina para a Covid-19 e sua necessidade são recentes, então eram relativamente poucos os dias que tratavam do assunto. Além dos poucos dias que continham *tweets* sobre o tópico, ao fazer a compatibilização entre os *tweets* e os dias de funcionamento da bolsa brasileira, os dados de finais de semana e feriados foram perdidos.

Uma saída para este problema que pode ser adotada em trabalhos futuros é a avaliação de frações de tempo menores. Ou seja, avaliar o sentimento e o mercado financeiro em períodos de minutos ou horas, ao invés de períodos diários. Assim a pesquisa terá mais dados para os processos de modelagem. Conclui-se atestando que a principal contribuição desta pesquisa foi a apresentação de um método inovador para a previsão das oscilações do mercado acionário brasileiro, que é a utilização do sentimento público, convertido em séries numéricas como atributo preditivo.

## REFERÊNCIAS

- Araújo, J. G. Jr., & Marinho, L. B. (2018). Using online economic news to predict trends in Brazilian stock market sectors. *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, 1. 37-44. doi: <https://doi.org/10.1145/3243082.3243087>
- Apostolidis-Afentoulis, V., & Lioufi, K. I. (2015). *SVM classification with linear and RBF kernels*. 21, 1-7. doi: <http://doi.org/10.13140/RG.2.1.3351.4083>
- Bharathi, S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, 10(3), 146-154. doi: <http://doi.org/10.22266/ijies2017.0630.16>

- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8. doi: <https://doi.org/10.1016/j.jocs.2010.12.007>
- Castro, L. N. D., & Ferrari, D. G. (2016). *Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações*. Editora Saraiva.
- Carosia, A. E. O., Coelho, G. P., & Silva, A. E. D. (2019). The influence of tweets and news on the brazilian stock market through sentiment analysis. In *Proceedings of the 25th Brazilian Symposium on Multimedia and the Web*. 1, 385-392. doi: <https://doi.org/10.1145/3323503.3349564>
- Costa, A. G. (2021). *Entenda quando um sistema de saúde entra em colapso e como sair da crise*. CNN Brasil, from: <https://www.cnnbrasil.com.br/saude/2021/03/03/entenda-quando-um-sistema-de-saude-entra-em-colapso-e-como-sair-da-crise>.
- Das, M. K., Singh, D., & Sharma, S. (2021). Media news on vaccines and vaccination: The content profile, sentiment and trend of the online mass media during 2015–2020 in India. *Clinical Epidemiology and Global Health*, 10, 1-6. doi: <https://doi.org/10.1016/j.cegh.2020.100691>
- Depexe, S. (2022). NVivo e Twitter: da coleta de dados à exploração do dataset #coronavirusbrasil. *Comunicação & Inovação*, 23(51), 37-55. doi: <https://doi.org/10.13037/ci.vol23n51.8412>
- Dubey, A. D. (2021) *Public Sentiment Analysis of COVID-19 Vaccination Drive in India*. doi: <http://dx.doi.org/10.2139/ssrn.3772401>
- Estevão, A. (2020). COVID-19. *Acta Radiológica Portuguesa*, 32(1), 5-6. doi: <https://doi.org/10.25748/arp.19800>.
- Galzo, W. (2021) *Brasil lidera número de mortes diárias por Covid-19 no mundo em março*. CNN Brasil, from: <https://www.cnnbrasil.com.br/saude/2021/03/30/brasil-e-o-pais-que-mais-registra-mortes-diaras-por-covid-19-em-marco>.
- Hasan, M., Rundensteiner, E., & Agu, E. (2014). *Emotex: Detecting emotions in twitter messages*, from: <http://web.cs.wpi.edu/~emmanuel/publications/PDFs/C30.pdf>
- Igarashi, W., Valdevieso, G. S., & Igarashi, D. C. C. (2021). Análise de sentimentos e indicadores técnicos: uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. *Brazilian Journal of Business*, 3(1), 470-486. doi: <https://doi.org/10.34140/bjbv3n1-029>
- KEMP, S. (2022). *Another year of bumper growth. We are Social*, from: <https://wearesocial.com/uk/blog/2022/01/digital-2022-another-year-ofbumper-growth-2/>.

- Lima, M. L. (2016). *Um modelo para predição de bolsa de valores baseado em mineração de opinião* [Dissertação de mestrado]. Universidade Federal do Maranhão. From: Biblioteca digital de teses e dissertações: <https://tedebc.ufma.br/jspui/handle/tede/tede/297>.
- Liu B., & Zhang L. (2012) A Survey of Opinion Mining and Sentiment Analysis. In: Aggarwal C., Zhai C. (eds) *Mining Text Data*. Springer, Boston, MA. doi: [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)
- Lupion, B. (2021) *A sucessão de erros que levou à crise de oxigênio em Manaus*. DW Brasil, from: <https://www.dw.com/pt-br/a-sucess%C3%A3o-de-erros-que-levou-%C3%A0-crise-de-oxig%C3%AAnio-em-manaus/a-56275139>.
- Massarani, L., Waltz, I., Leal, T., & Modesto, M. (2021). Narrativas sobre vacinação em tempos de fake news: uma análise de conteúdo em redes sociais. *Saúde e Sociedade*, 30(2), 1-16. doi: <https://doi.org/10.1590/S0104-12902021200317>
- Meena, R., & Bai V, T. (2020). *Russia's Covid-19 Vaccine: Social discussion and first emotions*. doi: <https://doi.org/10.21203/rs.3.rs-95570/v1>
- Milani, E., Weitkamp, E., & Webb, P. (2020). The visual vaccine debate on Twitter: A social network analysis. *Media and Communication*, 8(2), 364-375. doi: <https://doi.org/10.17645/mac.v8i2.2847>
- Ministério da Saúde (2021, 19 de julho). *Painel Coronavírus. Coronavírus Brasil*, from Ministério da Saúde: <https://covid.saude.gov.br/>
- Mittal, A., & Goel, A. (2012). *Stock prediction using twitter sentiment analysis*. Stanford University, from: <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>
- Neuenschwander, B., Pereira, A. C., Meira, W. Jr., & Barbosa, D. (2014). Sentiment analysis for streams of web data: A case study of brazilian financial markets. In *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web*, 1, 167-170. doi: <https://doi.org/10.1145/2664551.2664579>
- Netto, R. G. F., & Corrêa, J. W. N. (2020). Epidemiologia do surto de doença por coronavírus (covid-19). *Desafios-Revista Interdisciplinar da Universidade Federal do Tocantins*, 7(3), 18-25. doi: <https://doi.org/10.20873/uftsuple2020-8710>
- Nofsinger, J.R. (2005). Social mood and financial economics, *Journal of Behaviour Finance*, 6(3), 144-160. doi: <https://doi.org/10.1016/j.jocs.2010.12.007>
- Oliveira, A. D. C. M. D. (2016). *Identificando emoções em manchetes de notícias escritas em português do Brasil utilizando Naïve Bayes* [Monografia] Universidade Federal do Ceará. from Biblioteca digital de teses e dissertações: <http://www.repositorio.ufc.br/handle/riufc/24819>

- Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Conference. on Empirical Methods in Natural Language Processing*, 1, 1-9. doi: <https://doi.org/10.3115/1118693.1118704>
- Piedrahita-Valdés, H., Piedrahita-Castillo, D., Bermejo-Higuera, J., Guillem-Saiz, P., Bermejo-Higuera, J. R., Guillem-Saiz, J., Sicilia-Montalvo, J. A., & Machío-Regidor, F. (2021). Vaccine hesitancy on social media: Sentiment analysis from June 2011 to April 2019. *Vaccines*, 9(1), 28. doi: <https://doi.org/10.3390/vaccines9010028>
- Rahim, N. H. A., & Rafie, S. M. (2020). Sentiment analysis of social media data in vaccination. *International Journal*, 8(9), 5259-5264. doi: <https://doi.org/10.30534/ijeter/2020/60892020>
- Raghupathi, V., Ren, J., & Raghupathi, W. (2020). Studying public perception about vaccination: A sentiment analysis of tweets. *International journal of environmental research and public health*, 17(10), 1-23. doi: <https://doi.org/10.3390/ijerph17103464>
- Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. *IEEE Systems Journal*, 13(1), 760-770. doi: <http://doi.org/10.1109/JSYST.2018.2794462>
- Sabóia, G., Mazieiro, G., Andrade, H., & Adorno, L. (2021). Anvisa aprova uso emergencial das vacinas CoronaVac e AstraZeneca no Brasil, Agência Nacional de Vigilância Sanitária, from: <https://noticias.uol.com.br/saude/ultimas-noticias/redacao/2021/01/17/anvisa-aprova-pedido-de-vacina-do-butantan-e-da-fiocruz.htm>
- Santos, F. P., Silveira, I. F., & Lechugo, C. P. (2017). Análise da percepção dos alunos sobre as práticas docentes por meio da mineração de dados educacionais. *Análise*, 38(05), 9-24. doi: <http://www.revistaespacios.com/a17v38n05/a17v38n05p09.pdf>
- Santos, G. C. (2020). *Algoritmos de Machine Learning para previsão da B3* [Dissertação de Mestrado]. Universidade Federal de Uberlândia, from Biblioteca digital de teses e dissertações: <https://repositorio.ufu.br/handle/123456789/29897>
- Shah, D., Isah, H., & Zulkernine, F. (2018). Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data (Big Data)*, 1, 4705-4708. doi: <http://doi.org/10.1109/BigData.2018.8621884>.
- Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., & Al-Garadi, M. A. (2017). Sentiment analysis of big data: Methods, applications, and open challenges. In *IEEE Access*, 6, 37807-37827. doi: <https://doi.org/10.1109/ACCESS.2018.2851311>
- Shrivastava, S. (2020). *Cross Validation in Time Series*, from: <https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4>

Twitter, I. (2020). *Perguntas frequentes de novos usuários*, from: <https://help.twitter.com/pt/new-user-faq>.

Valle-Cruz, D., Fernandez-Cortez, V., López-Chau, A., & Sandoval-Almazán, R. (2021). Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods. *Cognitive computation*, 1, 1-16. doi: <https://doi.org/10.1007/s12559-021-09819-8>

Wang, S. I., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, 2(1), 90-94. doi: <https://aclanthology.org/P12-2018.pdf>

Wiysonge, C. S., Ndwandwe, D., Ryan, J., Jaca, A., Batouré, O., Anya, B. P. M., & Cooper, S. (2021). Vaccine hesitancy in the era of COVID-19: could lessons from the past help in divining the future?. *Human vaccines & immunotherapeutics*, 1-3. doi: <https://doi.org/10.1080/21645515.2021.1893062>

Yadav, A., Jha, C. K., Sharan, A., & Vaish, V. (2019). Sentiment analysis of financial news using unsupervised and supervised approach. In *International Conference on Pattern Recognition and Machine Intelligence*, 1, 311-319. doi: [http://doi.org/10.1007/978-3-030-34872-4\\_35](http://doi.org/10.1007/978-3-030-34872-4_35)

Yin, F., Wu, Z., Xia, X., Ji, M., Wang, Y., & Hu, Z. (2021). Unfolding the determinants of COVID-19 vaccine acceptance in China. *Journal of medical Internet research*, 23(1), e26089. doi: <http://doi.org/10.2196/26089>

<b>Planilha de Contribuição dos Autores</b>			
<b>Contribuição</b>	<b>Karoline Pereira Branco</b>	<b>Gabriel Rodrigo Gomes Pessanha</b>	<b>Eleanderson Campos Eugênio Filho</b>
1. Idealização e concepção do assunto e tema da pesquisa		X	
2. Definição do problema de pesquisa	X	X	X
3. Desenvolvimento da Plataforma Teórica	X	X	X
4. Delineamento da abordagem metodológica da pesquisa	X	X	X
5. Coleta de dados	X		X
6. Análises e interpretações dos dados coletados	X	X	X
7. Conclusões da pesquisa	X	X	X
8. Revisão crítica do manuscrito	X		
9. Redação final do manuscrito, conforme as normas estabelecidas pela Revista.	X		
10. Orientação		X	X