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

Karoline Pereira Branco [1]
Gabriel Rodrigo Gomes Pessanha [2]
Eleanderson Campos Eugênio Filho[3]

▪ Received: 10/06/2021 ▪▪ Approved: 07/09/2023 ▪▪▪ Second Approved Version: 11/22/2023

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 Pandemic. Stock Market. Forecast.

1 INTRODUCTION

In January 2020, the world received with apprehension the news that China had communicated to the World Health Organization (WHO) the occurrence, in the city of Wuhan, of a previously unknown disease. What was initially believed to

[1]Msc in applied statistics and biometrics, 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.

[2] PhD in Administration, 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.

[3] PhD in statistics, 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.

be a type of severe pneumonia was later identified as the SARS CoV-2 virus, characterized as a new type of Coronavirus (Estevão, 2020). Named as COVID-19, since the first cases occurred in late 2019, the disease has spread rapidly across the world. Thus, on March 11, 2020, the WHO announced the status of a pandemic (Netto & Corrêa, 2020).

In Brazil, the first confirmed case dates from February 26, 2020. One year and four months after the start of the pandemic, more than 19 million Brazilians fell ill and 542,000 died in the country (Ministry of Health, 2021). As a result of the numerous social and economic impacts, in addition to the generalized crisis observed as a result of the advancement of the disease, the vaccine against Covid-19 is presented across the planet as a hope for a return to "normality" (Meena & Bai V, 2020).

Increasingly, social networks are characterized as an environment that people use to express their feelings and opinions about different entities, products, other people and situations. Thus, the opinion expressed in these media often reflects the public opinion of an entire locality. This makes digital platforms a space for massive and constant data generation. The generated data allow the extraction of information that allows the understanding of various aspects of society (Bollen, Mao & Zeng, 2011; Shayaa et al. , 2017).

Twitter is a social network that works like a microblogging service. On this platform, people can share their latest stories, ideas, opinions and news in real time through messages of up to 280 characters, called tweets (Twitter, 2020). Worldwide, the platform has more than 300 million daily active users (Dubey, 2021). In Brazil, this platform is recognized for its importance in actions to engage the television audience or political demonstrations. In the country, 15.5% of the population are users of the platform (Depexe, 2022), which places the country as the 4th place in the world that most accesses it (Kemp, 2022).

Faced with the moment of a pandemic that the planet is experiencing, the theme "vaccine" is highlighted in virtual discussions and debates. Users for and against vaccination make heavy use of social networks to participate in conversations and debates about vaccines and disseminate information, or disinformation, in order to defend their position (Milani, Weitkamp & Webb, 2020; Rahin & Rafie, 2020; Piedrahita -Valdés et al., 2021).

According to Carosia, Coelho and Silva (2019), sentiment analysis (SA) emerges as a technique that can be applied to social media texts. This technique aims to identify the type of feeling that predominantly emanates from them. The objective of using this tool is to verify whether the feelings regarding a given topic are mostly positive or negative. According to Neuenschwander et al. (2014), although the application of this technique is not trivial, the results it brings reflect a great potential for applications.

A potential application still little explored in Brazil is the use of sentiment analysis as a tool to predict stock market movements (Assis et al., 2018). As assumed in the field of behavioral finance, financial decisions are largely related to the emotions and mood of their agents. Thus, it is plausible to assume that public sentiment can influence stock market values (Bollen, Mao & Zeng, 2011; Nofsinger, 2005). There are studies that report the existence of a relationship between the

population's sentiment and oscillations in the financial market of a country, showing that its knowledge can help investors to make better deals (Carosia, Coelho & Silva, 2019).

Thus, this research aimed to use the sentiment analysis technique to capture the sentiment of Brazilian Twitter users regarding the vaccine against Covid-19 and verify the effect of this sentiment on the movement of the country's stock exchange. The results obtained have two main purposes. The first is to provide information that may be useful for Brazilian public policy makers in designing more efficient vaccination campaign strategies. This is because the results will allow knowing what part of the population thinks about the subject. The second objective is to verify the impact of daily sentiment on oscillations in the Ibovespa index, which is a metric of movement in the Brazilian stock market. This assessment could provide a possible support tool to optimize investors' results.

This work is organized as follows: the next section presents empirical works that carried out sentiment analysis on the topic of vaccination; the later section presents some studies that used sentiment analysis as an auxiliary tool in forecasting the financial market; then, the sentiment analysis technique is presented, then the methodological procedures used in the research are presented, followed by the results and, finally, the final considerations.

2 BACKGROUND

2.1 Immunization and Sentiment Analysis - Related Studies

According to Neuenschwander et al. (2011), the increased use of internet applications for consuming information, sharing opinions and ideas on various subjects has made social networks efficient data collection platforms. Among other possibilities, the data collected allows identifying the feelings and mood of users. Bollen, Mao and Zeng (2011) point out that the texts shared by Twitter users can accurately represent the mood and general feeling of an entire population in real time.

In this context, discussions and debates regarding the vaccine against Covid-19 were no different. Social media was flooded with talk of vaccines. Internet users with positions for and against vaccination acted in the dissemination of information and/or misinformation about vaccines to defend their positions (Milani, Weitkamp & Webb, 2020). Therefore, according to Raghupathi, Ren and Raghupathi (2020), text analysis has been used by several scholars to research attitudes towards vaccination and vaccines for various diseases and in various countries. The objective is to assess the degree of acceptance or hesitation of the population in relation to this type of immunization, as well as the causes that lead to this posture.

Dubey (2021), researched Indian sentiment regarding two Covid-19 vaccines applied in the country's vaccination campaign. Through sentiment analysis of Twitter posts, using the lexical method, the author found that the population's position on both vaccines was overwhelmingly positive. The positivity for the first vaccine was 69% and for the second it was 71%. In addition, through a study that detailed the emotions attributed to each of them, the researcher

observed that those with the highest recurrence were trust and anxiety for both vaccines.

Meena and Bai V (2020) conducted a similar study, in which the authors analyzed the feelings of individuals regarding the discovery of the Covid-19 vaccine by Russia. The data for the analysis were collected on the day the news about the discovery was released. By using the lexical method for sentiment analysis, the authors found that the sentiment contained in 90% of the publications was positive and that only 10% were negative, showing that there was euphoria and hope in relation to the advertisement.

Rahim and Rafie (2020) pointed out in their studies in Malaysia an increase in the population's hesitation in vaccinating themselves and vaccinating their children. Thus, in their research they tried to assess this hesitation through sentiment analysis of Twitter posts. Using the SVM method to classify sentiments, the authors found that 41% of the analyzed tweets were positive in relation to vaccination, 39% were neutral and 20% of platform users were negative in relation to this form of immunization. Fear of side effects and religious issues were pointed out by the authors as factors that create insecurity in relation to getting vaccinated.

Raghupathi, Ren and Raghupathi (2020) analyzed Americans' perception of vaccines in general between January and April 2019. During this period, the country experienced an outbreak of measles cases. Still using publications on Twitter as a data source, the authors found that there was a balance in the opinions of users regarding the effectiveness of vaccines, with 43.3%, 40.4% and 16.3% of tweets being classified as negative, positive and negative. neutral, respectively. The authors attributed the high number of negative feelings to the recent link created between the vaccine and autism, which still lacked scientific evidence at the time of writing the research. Positive feelings were related to the existence of a measles vaccine and evidence that the vaccine has been effective in saving lives. Table 1 presents some other studies related to the vaccine and sentiment analysis.

Table 1
Empirical research that applied the sentiment analysis technique in relation to the theme "vaccine"

Author(s) / Year	Data source	Goals	Main Findings
Milani, Weitkamp & Webb (2020)	Twitter	Understand the debate for and against vaccination through sentiment analysis of images shared on social networks related to vaccine immunization.	It was found that images that show negative feelings about vaccine immunization prevail on the platform, as anti-vaccine activists are more articulated in social networks in pro-vaccine communities.
Piedrahita-Valdés et al. (2021)	Twitter	To analyze in several countries the predominant feeling in relation to the theme	It was found that tweets with positive sentiments were always more prevalent

		"vaccine", between the years 2011 and 2019.	than negative ones, also showing greater engagement.
Wiysonge et al. (2021)	Twitter	To verify whether the increased use of social networks and the dissemination of disinformation campaigns from abroad influence the levels of vaccine hesitancy in a population.	It was found that there is a significant link between increased use of social media and increased belief that vaccines are unsafe. Furthermore, foreign campaigns that spread misinformation have been found to lead to a decline in a country's average vaccination rates.
Das, Singh & Sharma (2021)	Online news	Perform sentiment analysis on news published online by the Indian media that had vaccine as a theme.	The authors found that, although most news transmits positive feelings about vaccines, a relevant percentage of them had negative feelings. This indicates that the government must supervise the networks so that the acceptance of vaccines by the population will not be impaired.
Yin et al. (2021)	Weibo	To analyze Chinese public opinion on vaccines and vaccination for Covid-19.	It was found that the Chinese are positive about the price and occurrence of side effects of the vaccine, as well as being positive about vaccines made with inactivated virus.

Source: Created by the authors.

After presenting some studies that used the sentiment analysis methodology on social networks to verify public opinion on the subject of the vaccine, the next section will present some works that applied the sentiment analysis method as a tool to help predict movement of the stock market.

2.2 Sentiment analysis and financial market - Empirical studies

The possibility of predicting financial market behavior and oscillations attracts both academic and investor attention, and many researchers employ various methods and techniques to achieve this end (Bollen, Mao & Zeng, 2011). Therefore, Carosia, Coelho and Silva (2019) point out that studies that apply sentiment analysis to the financial market first computationally analyze texts extracted from news or publications on social networks and then automatically assign a polarity to them. Subsequently, some technique is applied that associates the identified momentary sentiment and the direction of the financial market.

Bollen, Mao and Zeng (2011) and Mittal and Goel (2012) developed studies in which they sought to predict the movement of the Dow Jones Industrial Average (DJIA), the main stock index of the US stock exchange, using sentiment analysis.

The first work used Granger Causality and the Fuzzy Neural Network (SOFNN) to relate the oscillation of the public's mood with the oscillation of the financial market for the period from 02/28/2008 to 12/19/2008. By this strategy the researchers were able to predict the DIJA oscillations with 87.6% accuracy. The second research applied the methodologies of linear regression, logistic regression, SVM (Support Vector Machine) and also SOFNN to verify the relationship between public sentiment and stock market oscillations in the months of June and December 2009. The results showed an accuracy of 75.56%.

Bharathi and Geetha (2017) sought to verify whether the inclusion of sentiment as a predictor variable would generate an asset forecast model with greater accuracy than a model that used only historical data in the forecast. To this end, sentiment analysis was carried out on news related to the Indian stock market in April 2006 and the historical series of quotations of a company in the country was used. The method for sentiment analysis was the lexicon. To verify the impact of including the sentiment variable on market oscillations, the accuracy of the model that used only the historical series was compared with the model that, in addition to the historical series, included the sentiment variable in the forecast. As a result, the researchers found an accuracy of 64.32% for the adjusted model without sentiment, while the model with sentiment presented an accuracy of 78.75%. This demonstrated that sentiment is a useful predictor variable to predict the behavior of the shares of the analyzed company.

Lima et al. (2016) applied sentiment analysis to predict oscillations in Brazilian financial market assets. They used Petrobras shares as the focus of their studies. The sentiment analysis, carried out using the lexical method, aimed to identify the collective mood in relation to Petrobras. Subsequently, the humor found was used in an attempt to predict the behavior of the company's shares through the SVM method. As a result, the authors achieved an accuracy of 82.93% in forecasting the oscillations in Petrobras shares.

Carosia, Coelho and Silva (2019) evaluated the impact of news and tweets related to the Brazilian financial market through sentiment analysis performed by MLP-type artificial neural network method. Subsequently, the researchers compared which of the data sources had the greatest predictive power on the Ibovespa index. The SVM method was used to perform the predictions. As a result, it was found that the general sentiment of the Brazilian population in relation to the news on day t has the greatest influence on the opening value and on the trading volume of the stock exchange on day t + 1, with an accuracy of 60.6% for both. However, when it comes to the closing value of the exchange, it was found that the sentiments contained in the tweets have the greatest impact, demonstrating an accuracy of 54%.

The work published by Araújo and Marinho (2018) also had the Brazilian stock market as a research field. The objective of this work was to investigate the impact of the dissemination of online news on the theme “economy” on the performance of BM&FBovespa (B3). Using a sentiment classifier, the texts were classified according to their polarity, positive or negative. As a result of the research, the authors described that the number of news available every 15 minutes influences the movement of the financial market in Brazil. For 19 news or more publications in this period of time, it can be seen that the quality of the

financial market forecast is around 70%. In addition, the authors observed that the repercussion of negative news is greater than that of positive news. Therefore, those that express negative feelings tend to have a greater impact on B3's performance than positive news. Table 2 presents some other studies related.

Table 2
Empirical research that applied sentiment analysis in stock market forecasting

Autor(es)/ Ano	Objetivos	Principal Resultado
Valle-Cruz et al. (2020)	To verify whether the sentiments of tweets from profiles with global importance would influence the behavior of financial indices in different countries and sectors of the economy in times of a pandemic. The periods evaluated were the H1N1 and Covid-19 outbreaks.	It was discovered that in the period of the Covid-19 pandemic, sentiments influenced the market movement from 6 to 13 days after posting, with this period being 1 to 2 days in the H1N1 season.
Yadav et al. (2019)	To verify the effect of the Indian population's sentiments on the movement of the country's stock market.	Indian population sentiment, measured by three different methods, was used in forecasting the country's stock market. The results showed 3 predictions with accuracies ranging from 75% to 79%.
Shah, Isah & Zulkernine (2018)	To evaluate how much the feelings contained in the news regarding the Indian pharmaceutical sector influence the movement of the shares of the companies that make up the Nifty Pharma index in that country.	It was found that the proposed sentiment analysis model had a forecasting power of 70.59% to anticipate stock price trends in the short term.
Ren, Wu & Liu (2018)	To train a model to predict market behavior of the SSE 50 index based on mainstream Chinese sentiment.	The researchers found that the prediction of stock market trends is improved by 18.6% when the variable sentiment is included as a predictor of these oscillations.
Igarashi, Valdevieso & Igarashi (2021)	To verify the correlation between stock prices and the polarity of news related to the Brazilian stock market, using Petrobras preferred stock (PETR4) as the basis for the studies.	It was found that the polarity observed in the financial news had a moderate correlation with the oscillations in Petrobras' assets.

Source: Created by the authors.

The next section presents the sentiment analysis procedure as employed in the present study.

2.3 Sentiment Analysis

The current admiration and acceptance of social media has generated a large volume of information, enabling and encouraging the development of new analysis techniques for unstructured data, such as images, videos and texts

(Santos, Silveira & Lechugo, 2017). In this context, sentiment analysis emerges as a data mining technique. This type of technique aims to extract useful information from data that, without analysis, manipulation and interpretation, do not have informative content and do not serve as a basis for decision-making processes (Castro, 2016).

Sentiment analysis is employed on many types of data to generate useful information and knowledge. Its main purpose is to identify public opinion about a particular product, person, event, organization or topic. The objective is to identify whether the general feeling regarding the subject under study is positive or negative. The information obtained can be used for numerous purposes, such as, for example, setting up marketing or even investment strategies (Liu & Zhang, 2012).

There are multiple techniques for performing sentiment analysis, ranging from simple methods such as the lexical approach, to more elaborate techniques, such as those performed through machine learning algorithms. A machine learning algorithm that can be used for this type of task is called Naive Bayes. This is a commonly used supervised learning algorithm with the function of classifying feelings in texts and which has shown satisfactory accuracy results (Wang & Manning, 2012).

Supervised methods need a pre-labeled dataset for training the learning algorithm. The Naive Bayes algorithm is based on Bayes' Theorem, which deals with conditional probabilities, and works by generating a probability table. In this table, the probability of each word in the sentence belonging to the positive and negative classes is verified. The class that has the highest probability of occurrence is attributed as sentence polarity. Equation 1, adapted from Oliveira (2016), illustrates how the probability of a sentence being positive is calculated using the Naive Bayes method. The probability of a sentence belonging to the negative class follows the same proposal, with the change from "positive" to "negative" in the equation:

$$P(sentence) = P(positive).\prod_{j=1}^{n} \quad P(V_j|positive), \qquad (1),$$

where $P(sentence)$ is the probability that any given sentence is classified as "positive" ; $P(V_j|positive)$ is the probability that the word V_j is in a positive sentence; $P(positive)$ represents the probability of a sentence being classified as positive and n represents the number of words in the sentence.

The next section presents the method used to carry out forecasts of stock market oscillations.

3 METHODOLOGICAL PROCEDURES

3.1 Method used for predicting stock market oscillations

In order to verify the influence of the sentiment of Brazilian Twitter users on the oscillations of the national stock market, two forecasts were made. One used only historical data related to the Ibovespa index, which is the Brazilian stock market metric. The other used, in addition to historical data, the prevailing daily sentiment as a predictive variable.

The method used to carry out forecasts of oscillations in the Brazilian stock market was the Support Vector Machine (SVM). This machine learning method is

recognized for its significant efficiency both in regression models and in classification models (Apostolidis-Afentoulis & Lioufi, 2015). In summary, the SVM fits into the category of supervised methods. Its central idea is the creation of a model that manages to generate a hyperplane capable of separating two classes, keeping the greatest possible distance between them (Pang, Lee & Vaithianathan, 2002).

Due to the small amount of data obtained for the research, the cross-validation technique for time series was applied. In this technique, an algorithm known as k-fold was employed, in which a set of n data is divided into k subsets of approximately the same size. Then, k training sessions are performed. In each training, a subset is used as a test base for computing misclassifications, while the other subsets are used for model training (Shrivastava, 2020). The result of the cross-validation accuracy refers to the arithmetic mean of the results found in each of the n training sessions. This technique prevents the assessment of the model's assertiveness from being mistaken, which can occur if only a separation of test and training data is performed.

As the SVM algorithm is of the supervised class, in order to be able to carry out the training of the forecast model of the stock market movement trend, it was necessary to provide data for learning. In this context, the first stage of the forecast consisted of generating attributes for model training. Attributes related to the historical series were calculated based on the work of Santos (2020) and are presented in equations (2), (3), (4), (5) and (6):

Relative strength index (RSI):

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

$$RS = \frac{Mean\ gain\ in\ the\ last\ 5\ days}{Mean\ loss\ in\ the\ last\ 5\ days}, \quad (3),$$

where the mean gain is calculated by dividing the number of points gained by the index in the last 5 days by 5, and the mean loss is calculated by dividing the number of points lost in the last 5 days by 5. Data were obtained from the index's daily closing values.

Stochastic Oscillator (%K):

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

where C is the current number of points; P5 is the lowest number of points observed in the Ibovespa index in the last 5 days and A5 is the highest number of points observed in the last 5 days. Data were obtained from the index's daily closing values.

Willians (%R) :

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

where C is the current number of points; P5 is the lowest number of points observed in the Ibovespa index in the last 5 days and A5 is the highest number of points observed in the last 5 days. Data were obtained from the index's daily closing values.

Price change rate (PROC (t)):

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

where PROC(t) represents the rate of change at time t, C(t) represents the number of points at time t and C(t - 5), represents the number of points at the closing of the index at time t - 5 days.

In addition, to obtain the variable referring to the predominant daily sentiment used in the second forecasting process, the same equation used in Carosia, Coelho and Silva (2019) was used, as shown in equation (7):

$$Predominant\ daily\ sentiment = \frac{(nPos - nNeg)}{(nPos + nNeg)}, \quad (7),$$

where *nPos* represents the number of tweets classified as positive and *nNeg* represents the number of Tweets classified as negative on the analyzed day. Positive results indicate predominantly positive feelings and negative results indicate predominantly negative feelings on the evaluated day. Finally, the performance evaluation of the two generated models was based on the accuracy metric, which measures the number of values correctly classified among all values classified by the model. Equation (8) presents the accuracy formula:

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

where VP/VN represents the number of correctly classified up/down forecasts in the Ibovespa index and FP/FN is the number of incorrectly classified forecasts. The next section presents the research steps.

3.2 Research steps

The steps commonly used in the elaboration of a sentiment analysis are: 1) Data collection; 2) Data pre-processing; 3) Model training; 4) Classification according to polarity; and 5) Analysis of the results, and this research will also have a sixth step which is: 6) Verification of the relationship between the daily sentiment

of Twitter users on a given day with the oscillation of the national financial market for that same day. Below, each of the steps is described in detail.

1) Collecting text data - Through an API in the Python 3.9 software, 100,000 tweets were collected between 12/01/2020 and 07/13/2021. The collected tweets were not targeted to specific users. The captured texts refer to accounts of different types of Brazilian users. The collected tweets were divided into two groups. A group was employed in the training and validation of the message polarity classification model, as was done in the work of Hasan, Rundensteiner and Agu (2014). The other group was applied to the already trained model, in order to receive labels about its polarity.

Of the 100,000 tweets in the database, half were related to positive sentiments and the other half were related to negative sentiments. To specify the group with phrases of positive feelings, tweets containing words related to positive emotions were filtered, which were captured by the hashtags: #happiness, #love, #gratitude, #joy, #cure, #hope and #emotion; the group that contained negative phrases was specified as the texts that contained negative emotions, being captured by the hashtags: #fear, #tristeza, #shame, #rage, #revolt, #disease, #pain, #sick, # heartbreak and #unhappy. Furthermore, in order to carry out the sentiment analysis proposed in this work, it was possible to collect, for the same dates, 68,408 data related to the hashtags "vaccine" and "vaccination", which were later labeled by the trained model.

2) Data pre-processing - The groups of data used for training and validation of the model were cleaned - removal of punctuation, symbols, links, numbers, stopwords, duplicate data and data in a foreign language. The cleaning was carried out to remove from the text characters that do not represent feelings or those that are not words, as these are not relevant in the proposed model. At the end of this process, 79,928 observations remained for training and validation of the model, which were divided equally between the positive and negative classes. Regarding the data from the hashtags #vacina and #vacinação, 55,998 observations remained after the preprocessing stage.

3) Model training - The data polarity classification training followed the Naive Bayes algorithm. For this task, the e1071 package was used in the R software, version 4.0.3. In this phase, only the data collected for model training were used, which were already labeled according to their polarity, as explained in step 1. Of the 79,928 observations, 75% were randomly separated for model training, and 25% for its validation. Both in the training group and in the group used in the validation, there was a percentage balance of the "positive" and "negative" classes.

4) Data classification: The set of Tweets related to #vaccine and #vaccination were applied to the already trained model. Each tweet was classified according to the class that had the highest probability according to the Naive Bayes model. Subsequently, a manual classification of some tweets labeled by the proposed model was carried out to verify its accuracy.

5) Data analysis - By classifying the data, it was possible to develop an analysis of the feelings of Brazilian Twitter users regarding the topic "vaccine" on the evaluated dates.

6) Relation with market movement - Here, we sought to verify, through the training of an SVM model by cross-validation of time series, whether the prevailing sentiment observed on a given day would help in predicting the oscillations of the stock exchange on that same day. It is hypothesized that the result of including the sentiment variable will provide a predictive model with more accurate results. For this, we considered downward movements between two consecutive days with the value -1, upward movements with the value 1 and stability trend with the value 0. The verification of the influence of sentiment about the vaccine on oscillations in the Brazilian stock market was carried out by comparing the accuracy obtained by the forecast model that used only data related to the historical series of the Ibovespa index as a predictor variable, with the accuracy of the forecast model which used, in addition to data from the historical series, the values of the prevailing daily sentiment.

It is important to point out that the Ibovespa index only has data available on working days. Therefore, the database was adjusted to contain the prevailing sentiment only for the dates on which the Ibovespa index quotation was available. At the end of this process, the database contained information for 144 days. The next section exposes the results found.

4 ANALYSIS AND DISCUSSION OF RESULTS

4.1 Analysis of Brazilian sentiments regarding the Covid-19 vaccine

As shown in Figure 1, the daily number of tweets published with the hashtags #vacina and #vacinação commonly varies between 100 and 400 daily posts. However, the arrangement of the data shows that between the 17th and the 22nd of January there was a large increase in the number of tweets. During this period, the daily number of tweets related to the theme reached the mark of 550 to 1600 publications in one day.

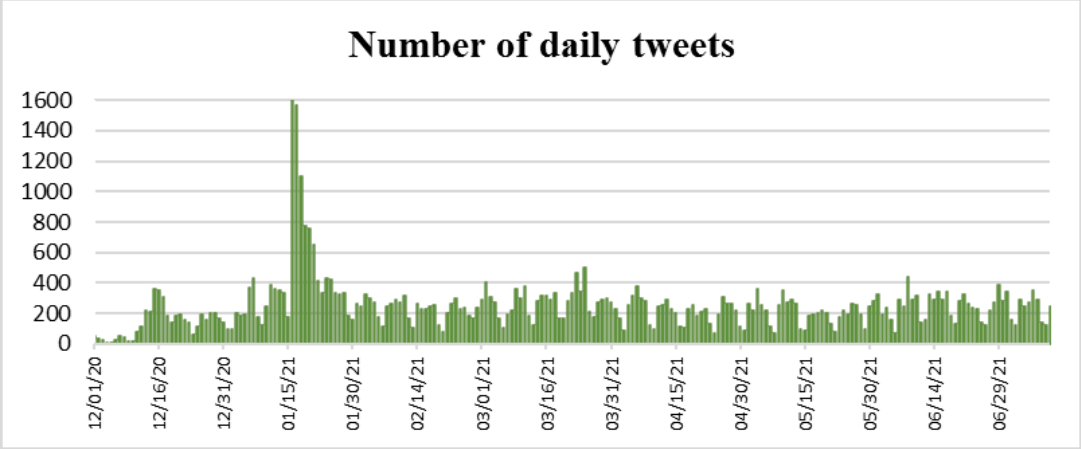

Figure 1 – Daily number of Tweets regarding #vaccine and #vaccination between 12/01/2021 and 07/13/2021.
Source: Created by the authors.

A brief study of events was carried out for a better understanding of this period. The results of this study suggested that the increase in the number of publications related to the hashtags #vacina and #vacinação between those

days may be related to the permission for the emergency use of two vaccines against Covid-19 by the National Health Surveillance Agency (Anvisa). In addition, this period includes the date on which the first person was vaccinated in the country (Sabóia, 2021). Furthermore, during this period there was also an oxygen supply crisis in the city of Manaus, which claimed the lives of several people who were hospitalized with Covid-19 and needed it to survive (Lupion, 2021). This fact may have contributed to the rise of the topic #vaccine among Internet users, since only mass immunization against the disease could prevent the Brazilian health system from experiencing another collapse (Costa, 2021).

Thus, the model was trained as shown in step 3 of section 3.2. Cross-validation of the proposed classification model made it possible to verify that it had an accuracy of 96.90%. This is because of the 19,982 messages used in model validation, 19,363 were classified with the correct polarity. Given the satisfactory level of accuracy of the model, tweets containing the hashtags #vacina and #vacinação were applied to it. Thus the messages were automatically labeled according to their polarity.

The analysis of the polarity of the feelings contained in the collected messages allowed the verification of the predominant type of feeling in the texts. The result showed a high prevalence of negative messages regarding the theme. It was found that 86.08% of the messages labeled by the model were classified as negative and only 13.92% were classified as positive. When the data were validated by manual labeling, it was found that this classification obtained 68% accuracy. Figure 2 illustrates the proportion of positive and negative Tweets labeled by the proposed model.

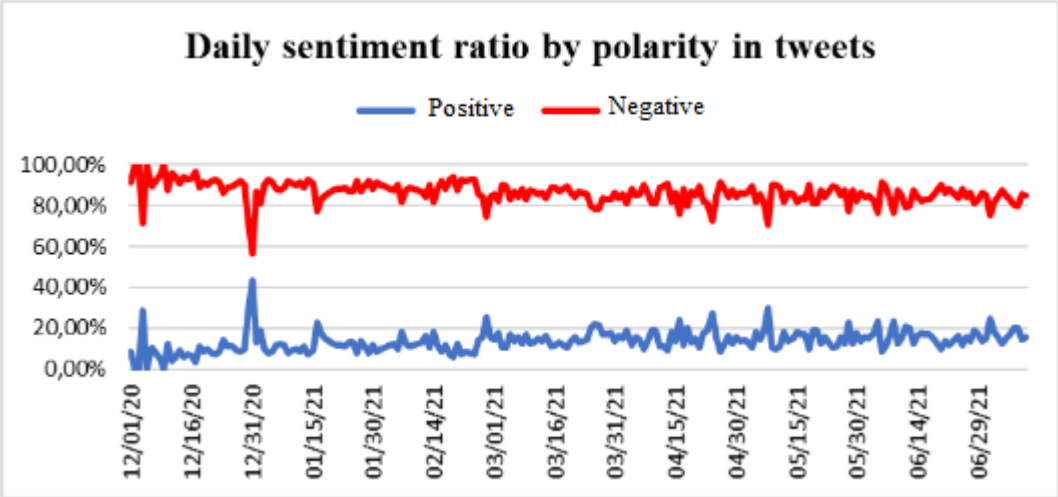

Figure 2 – Daily proportion of feelings by polarity in Tweets between 12/01/2021 and 07/13/2021. Source: Created by the authors.

Thus, there is a clear dominance of negative feelings around the theme of vaccine/vaccination in Brazil, the same conclusion obtained in the research carried out by Milani, Weitkamp and Webb (2020). A possible explanation for the occurrence of this result is that, in the period in which the tweets were collected, Brazil was experiencing a shortage of vaccines. In addition, in mid-March and April, the country experienced the second wave of the pandemic. This caused countless deaths and took Brazil to the top of the world with regard to the number of daily deaths (Galzo, 2021).

Therefore, the dominance of negative feelings may be a consequence of the fear of people who have directly experienced the effects of the pandemic. It can also arise from the lack of expectations regarding the speed of vaccination campaigns in the country, which leads to feelings of anger, fear, frustration and concern. Negative feelings regarding vaccination can also be expressed by people who do not trust this method of immunization and who use the networks to express themselves or by disseminating and spreading fake news on the subject (Massarani, 2021). A more in-depth study is needed to bluntly affirm the reasons why the hashtags #vacina and #vacinação in publications of Brazilian profiles are expressively negative. For a greater degree of understanding about feelings regarding the theme, word clouds were built with the most frequent terms of each class and are shown in Figure 3.

(a) (b)

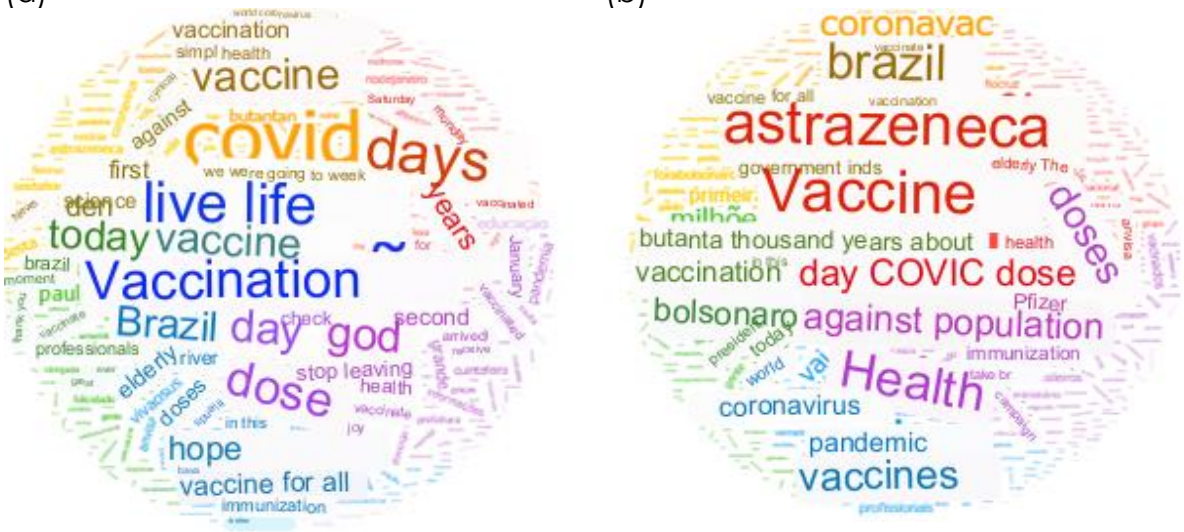

Figure 3 – Word cloud of tweets classified by the Naive Bayes model as positive sentiments (a) and negative sentiments (b).
Source: Created by the authors.

A superficial observation of the two clouds leads to very similar keywords in both. However, when the totality of the words of each one is observed, it is verified that the main difference between them is that the positive one puts positive feelings, such as "hope", "to live" and "life", while the cloud of negative words presents words that relate to Brazilian politics and the reinforcement of the "pandemic" situation. Furthermore, it is important to highlight that, although similar words/terms appear in both clouds, these words were used in different contexts. The positive phrases were classified as such because they were aligned with positive phrases present in the model's training set, and the same process occurred with the phrases classified as negative.

Thus, the pro-vaccine hashtag that gained space in Brazil as "#vaccineforall" appear in both clouds. In the cloud of positive tweets, they are inserted in a context of encouragement and hope in relation to immunization. In the cloud of negative tweets they appear with a sense of dissatisfaction related to the problems of the vaccination campaign in Brazil.

As a conclusion of this section, it is reaffirmed that the group of Brazilians analyzed in this research has predominantly negative feelings regarding the vaccine and vaccination against Covid-19. As pointed out by Das, Singh and Sharma (2021), in the online media there is a reasonable number of news reports that convey negative feelings about vaccine immunization. Wiysonge et al. (2021) points out that this can generate insecurities and fears in social media users about them.

The knowledge of which feelings are majority in the population regarding the subject and some of the targets that trigger these emotions, make possible a more effective dialogue with the people. This allows public policy makers to develop government advertising strategies that address the population's insecurities and fears about immunization, enabling more efficient vaccination campaigns. In addition, for vaccination campaigns to be successful, the population's credibility and trust in public agencies in this regard must be increased.

In order to verify whether, in addition to vaccine hesitation, negative sentiment about the vaccine has other implications, the next section presents the forecast of financial market oscillations with and without the variable "sentiment" as a predictor.

4.2 Using the sentiment of Brazilian Twitter users as a predictor of financial market oscillations

To evaluate the hypothesis proposed in this research that the inclusion of the sentiment of Brazilian Twitter users regarding the vaccine/vaccination for Covid-19 in Brazil as a predictor variable is capable of making predictions about the stock market more assertive, the first step was to collect information on the daily closing values of the Ibovespa index. Figure 4 shown below illustrates the behavior of this data series.

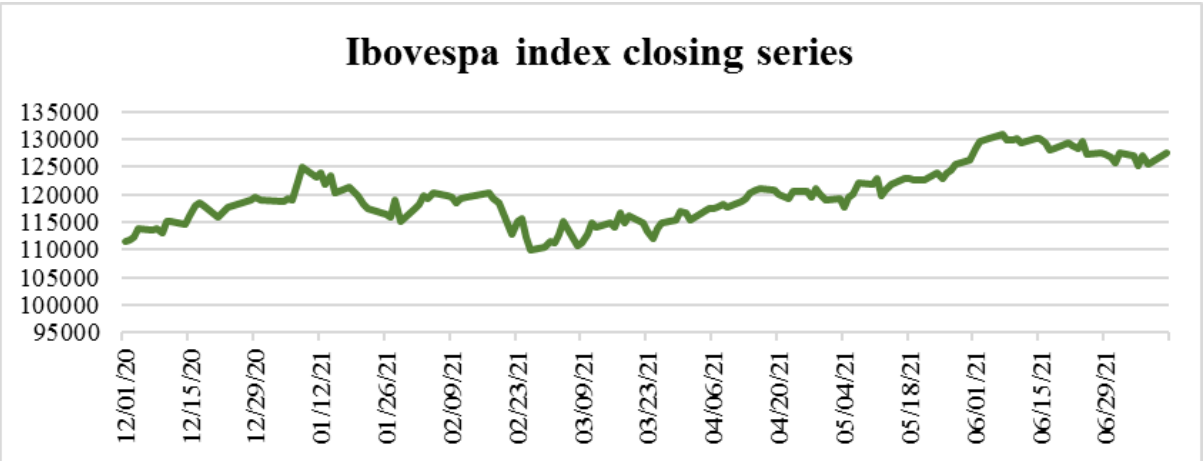

Figure 4 – Series with closing values of the Ibovespa index between 12/01/2021 and 07/13/2021, with data from the Infomoney website.
Source: Created by the authors.

The visualization of Figure 4 shows that the oscillations of the closing values of the analyzed index between 12/2021 and 07/2021 were quite high. In this period, the index varied between a minimum of 110,035 points and a maximum of 130,776

points. Thus, the high oscillations of the data makes modeling and forecasting this series a very complex task (Santos, 2020). As explained in the methodology of this work, we sought to predict not the values of the series, but rather its trend. The first projection of Ibovespa values was carried out by cross-validation with the SVM method and used the variables RSI, %K, %R and PROC(t), calculated from the historical series of the index, as attributes for model training.

To adjust the cross-validation model for time series, the "radial basis" kernel was used as default, with cost value equal to 1 and k equal to 5 separations. The result of the cross-validation obtained an average accuracy of 92.27% in the proposed forecast model only with the variables related to the historical series of the Ibovespa index, with a standard deviation equal to 8.32%. Subsequently, the second prediction model was adjusted by adding the variable "preponderant daily sentiment" of Brazilian Twitter users regarding the Covid-19 vaccine/vaccination as a predictor. The behavior of this variable can be seen in Figure 5.

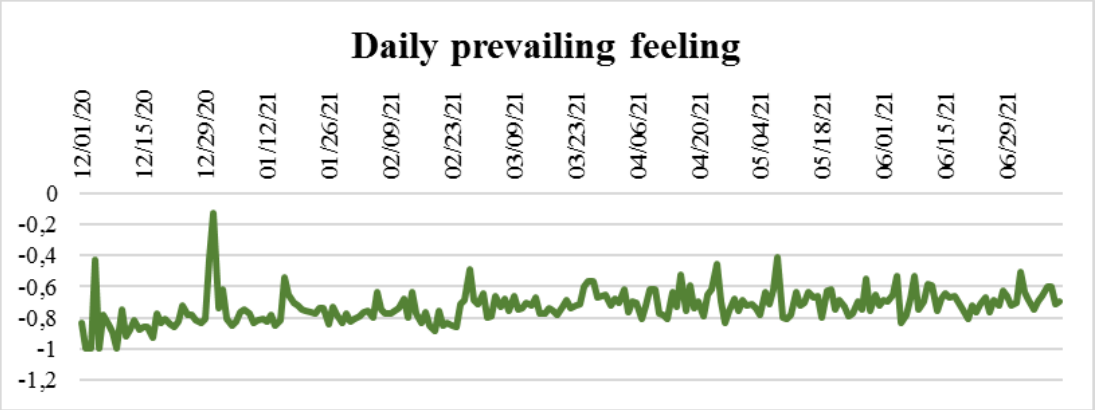

Figure 5 – Behavior of the daily prevailing sentiment data series between 12/01/2021 and 07/13/2021.
Source: Created by the authors.

As can be seen in Figure 5, the series of preponderant daily sentiment is quite volatile and varies between days with greater and lesser negativity regarding the topic of vaccine/vaccination. Regarding the correlation between the closing values of the Ibovespa index and the prevailing daily sentiment, a value of only 0.3677 was verified, which can be considered weak, which is an indication that the movements of these variables are not aligned.

Thus, the new cross-validation SVM model, also adjusted by the "radial basis" kernel, with a cost value of 1 and k equal to 5 separations, achieved an accuracy of 92.05%, with a standard deviation of 7.67%. Therefore, considering the standard deviation, among the possibilities of variations in results, it is stated that both models are equally assertive. Table 3 presents a summary of the results found in both proposed forecasting models.

Table 3
Results of the two fitted forecast models

Model	Predictive Variables	Mean Accuracy	Standard deviation
1	*RSI, %K, %R* and *PROC(t)*	92.27%	8.32%

2	*RSI, %K, %R, PROC(t)* and *Daily prevailing feeling*	92.05%	7.67%

Source: Created by the authors.

The conclusion of the experiment is that, unlike what was presented in similar studies in countries such as the United States (Bollen, Mao & Zeng, 2011; Mittal & Goel, 2012), China (Ren, Wu & Liu, 2018) and India (Yadav et al., 2019; Shah, Isah & Zulkernine, 2018), in the proposed classification and forecasting model, the inclusion of the sentiment variable related to the vaccine/vaccination in one day did not improve the process of forecasting oscillations in the country's stock market the next day. This result also differs from that found in the research carried out by Carosia, Coelho and Silva (2019) who achieved accuracy in using the general daily sentiment to predict opening and closing values and trading volume in the Brazilian market..

The authors Araújo and Marinho (2018) and Lima et al. (2016) who were also able to relate daily feelings to stock movements in Brazil, adopted methodological procedures different from those adopted in this research. This may have contributed to the more assertive results of their research. The first work carried out the analysis of sentiment and market variation every 15 minutes and not every day. The shorter period of time may have contributed to the model better capturing changes in mood during the day and their relationship with the Ibovespa index.

The second work segmented more the focus of the research. This evaluated whether Internet users' feelings related specifically to a company would affect its performance on the stock exchange. This larger specification seemed to bring greater gains in predictive power than using the general market. This finding can also be made by verifying that the work by Igarashi, Valdevieso and Igarashi (2021), in which the authors also considered only one company in the analysis, achieved at least a moderate correlation between sentiments and market movement, contrary to what happened in this research, in which the correlation found was low.

The research proposed here is similar to previous research within its proposal to assess the impact of an unconventional variable on stock market movements. Despite the fact that the results expected as an initial hypothesis were not confirmed, it is believed that the differential of this research lies in presenting an innovative market forecasting methodology that has already been used outside Brazil and also serving as a basis for the construction of new ideas related to this task. As an additional contribution, this research brought methodological suggestions performed by other researchers that may help future research to be more successful. Finally, in the next section, the conclusions and limitations of this research are explained, as well as suggestions for future research.

5 CONCLUSION

By analyzing the sentiment of Brazilian Twitter users on the subject of vaccination against Covid-19, it was found that there are predominantly negative feelings in this regard. This demonstrates the need for intervention by public policy makers, especially with regard to vaccination campaigns and better information

to the population about this type of immunization. In addition, there should be a focus on streamlining the immunization process in the country, given that the delay in this process is a criticism often mentioned in negative messages. These strategies would tend to help recover the credibility of public entities that also appear prominently in negative texts.

As for the relationship between sentiments about the vaccine and the oscillations observed in the Brazilian stock market, the method used in this research was not able to confirm the initial hypothesis that knowledge of the first would help to predict the behavior of the second. This is because the inclusion of the predictive variable "sentiment" in the classification model did not return a model with better accuracy than the one generated only with the historical series of the Ibovespa index.

Despite the results, it is not possible to state in general that sentiment does not influence stock market results. This is because some researchers have managed to find a relationship between these variables. So it may be that sentiments about vaccination are not really important to your prediction, but sentiments about other topics are. In addition, even for the topic of vaccination, it may be that if the methodology were changed, for example, if another method was used to classify feelings or another formula to obtain the predominant daily feeling, the result could have been different. These are suggestions for future research. Thus, it is also suggested to carry out research that compares the results obtained by different methods of textual classification.

The work presented the small amount of data as the main limiting factor. This is because the vaccine for Covid-19 and its need are recent, so there were relatively few days dealing with the subject. In addition to the few days that contained tweets on the topic, when matching the tweets with the days on which the Brazilian stock exchange was open, data for weekends and holidays were lost.

A way out of this problem that can be adopted in future work is the evaluation of smaller time fractions. That is, assessing sentiment and the financial market in minute or hourly timeframes, rather than daily timeframes. This way the research will have more data for the modeling processes. It concludes by attesting that the main contribution of this research was the presentation of an innovative method for forecasting oscillations in the Brazilian stock market, which is the use of public sentiment, converted into numerical series as a predictive attribute.

REFERENCES

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 Brazillian 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 #coronavirusnobrasil. *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-diarias-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

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. Standford 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). I*dentificando 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: http://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

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