DOI: 10.36558/rsc.v12i3.7903

Brazilian discussion about COVID-19 lockdown policies on Twitter

Fernando Xavier
University of São Paulo (USP)
Av. Prof. Luciano Gualberto, 380
São Paulo – SP – Brazil
+551130910665
fernando.xavier@gmail.com

Gustavo Rick Amaral
Pontifical Catholic University of São Paulo (PUC-SP)
Rua Caio Prado, 102 - Consolação
São Paulo – SP – Brazil
+551131247216
gustrick@gmail.com

Antonio Mauro Saraiva
University of São Paulo (USP)
Av. Prof. Luciano Gualberto, 380
São Paulo – SP – Brazil
+551130910665
saraiva@usp.br

ABSTRACT
The COVID-19 pandemic affected all countries worldwide, causing big changes in people's routines due to public policies for disease spreading control. Among the most impacting measures were social distancing policies and lockdown, leading to an intense discussion by the population. To describe this discussion in Brazil, this research applied data science and natural language methods to analyze posts on Twitter. It processed more than 12.9 million tweets between 2020 and 2021, and the results highlighted the main topics discussed by Brazilian Twitter users, such as the ideological-political component. The approach employed in this research proved to help extract valuable information in massive data mass.

CCS Concepts
• Computing methodologies→Artificial intelligence • Natural language processing→Information extraction.

Keywords
Natural language processing; covid-19; twitter; topic modelling.

1. INTRODUCTION
In late 2019, the world noticed the initial outbreak of a viral infectious disease in China named COVID-19. This disease, caused by the coronavirus SARS-COV-2, rapidly spread around the world and, in March 2020, was declared a pandemic by the World Health Organization (WHO) [1]. Until October 2022, there are 630,498,015 million cases reported, causing 6,590,207 deaths globally [2].

In the early stages of the pandemic, while a vaccine was not developed and the researchers did not sufficiently know the disease, the first policies used to control the disease spreading were related to containing the direct transmission between people. Because COVID-19 had similarities with other coronavirus diseases [3], the main initial measures adopted were mask obligatory use, social distancing, limiting people’s movement and orientation on disinfection of surfaces, and body cleaning. Due to its high impact on people's routines, measures to restrict the movement were the theme of several discussions. The most restrictive level policy (lockdown) was unpopular and provoked many protests on streets around the world.

People also used social media platforms to express their opinions about lockdown policies. For public health decision-making by health authorities, the information from social media can help assess and measure the impact of the policies on popular opinion and improve communication strategies. Moreover, extracting information from these data can support campaigns against misinformation, which is a big problem for public health due to its risks to population well-being.

This research aims to analyze the Brazilian's opinions on Twitter regarding lockdown policies during the COVID-19 pandemic. More than 12.9 million tweets citing lockdown policies were collected for this research. Text analysis was conducted to analyze discourse about lockdown policies and extract the main topics considering specific subjects and temporal aspects.

It is expected that this work can contribute to identifying the main topics of discussion of the Brazilian population on one of the most important measures to combat infectious diseases. In addition, it is expected that health management departments in decision-making processes can incorporate the method adopted in this research.

This article is organized as follows: the next section contains an analysis of research related to text mining in the health area, while the third section describes the five-steps method adopted. The results of each step are shown and discussed in Section Four, and the final considerations, including main findings, limitations, and future research, are described in Section Five.

2. CONTEXT
The COVID-19 pandemic caused unprecedented changes in people's lifestyle to control the disease dissemination. While vaccines were being developed, several countries defined social distancing policies at different levels. India adopted a national lockdown between March and April 2020 [4], while the first COVID-19 epicenter, Wuhan (China), also adopted a lockdown at the outbreak beginning [5]. Italy was the first COVID-19 epicenter in Europe, imposing a lockdown in March 2020; findings suggest that these measures successfully controlled the disease transmission [6].

In Brazil, there was not a national lockdown policy, but some locations adopted similar measures to promote some level of social distancing. For example, the São Paulo and Rio de Janeiro states ordered a partial lockdown, mainly in the pandemic beginning, with schools and universities closed, no public events, and restrictions on some economic activities [7][8]. Besides slowing down the disease spread, these locations experienced other benefits like improving air quality.

Despite benefits and the need to control de COVID-19 spread, the isolation policies were not unanimous worldwide. These policies impacted the economy of the whole world. A study about India's economy estimates a possible loss of around 10–31% of its Gross Domestic Product due to the pandemic [9]. There were protests against lockdown in many locations worldwide, such as the United States, Serbia, and Germany [10].

Although many people went to the streets to express their opinions about the lockdown and other isolation policies, social media platforms were heavily used to discuss COVID-19, including lockdown measures. A study aimed to analyze the Indians' feelings on Twitter during the government's second and third national lockdown. They found changes from a positive view in the second to a negative one in the third lockdown [11].

3. METHODOLOGY

This research adopted a data science cycle, with well-defined activities from the study planning to text analysis, as shown in Figure 1. The first activity consists of defining the research objectives and which analysis will be necessary to answer the research questions. However, other analyses can be specified during the study conduction according to the preliminary results obtained.

In the first step of this cycle, the study is planned, including the necessary data sources, objectives, methods, and what products must be generated. Products can be many forms of data visualization, such as charts, maps, or tables. In natural language processing projects, the generation of cloud tags summarizes the main topics in a set of documents. After data collection, an exploratory data analysis can be conducted to extract information about the dataset, such as stats related to the records. This step can help identify requirements for the next phases, such as preprocessing techniques and the computing infrastructure needed.

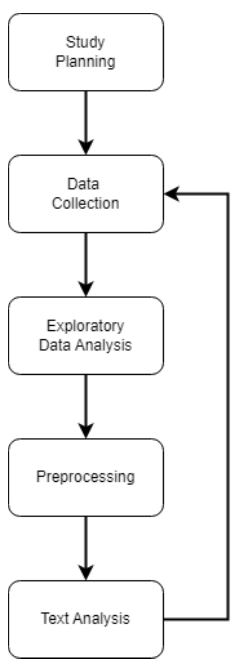

Figure 1. Research steps

Once the data sources are defined, the next step is to collect the data, sometimes in an automated way, such as a script. When the data sources are social media platforms, it is common to collect data using an application programming interface (API) or web scraping. Although not mandatory, this data can be stored in plain files or a database for further analysis. It is important to note that different studies can use the same database and keeping data can be a better project decision to guarantee research traceability. Another factor in favor of the data storage strategy is that some API platforms can return different results if the request is made in different periods.

After data collection, the next step contains activities to prepare data for the analysis step. Depending on the research objective, this preprocessing step includes tasks to select necessary features, remove noise and define strategies for the missing data problem. Considering data related to social media, it can be necessary to remove certain kinds of data, such as images or links. However, not all this data can be classified as noise because this is highly dependent on the research objectives. Even in natural language processing projects, some of this data can represent valuable information for analysis. An example is emojis, a class of images that can be used to define a mood about a text, especially in social media, where informal language and text length platforms' limitations are very common.

Different methods, such as traditional statistical analysis and machine learning, can be applied for the text analysis step. The choice of methods depends on many factors, such as problem type and data characteristics. It is also common to use combined methods, using the best of each. In text analysis research, it is common to use natural language processing combined with machine learning algorithms. This research adopted five strategies: word clouds, analysis using the most common hashtags, bigrams extraction, topic modeling, and sentiment analysis. For all strategies, results were compared over time.

4. RESULTS AND DISCUSSION

4.1 Step 1: Study planning

The main objective of this research is to analyze the Brazilian' discussion about lockdown policies on the social media platform Twitter. To develop this study, the main objective was divided into two research questions:

- Question 1: What are the main general topics around the lockdown discussion?

- Question 2: Are there temporal variations in the discussion topics?

Because the two research questions need the same data source, the activities in the data collection step were executed once. Python and Linux scripts were used for this step to collect and select the required data.

4.2 Step 2: Data Collection

Python scripts were developed to collect data through Twitter's API using search keywords related to COVID-19 between April 2020 and May 2021. Because the objective is only to analyze opinions in the Portuguese language and some words are also used in other languages, the language parameter was included in the API requests. Table 1 contains the original keywords used in API requests and their English translation.

Table 1. Keywords used to search tweets

Type	Original keyword	Keyword in English

COVID-10 names	Covid coronavirus corona	Covid coronavirus corona
Policies	Lockdown isolamento quarentena	Lockdown isolation quarantine

In addition to the tweet text, Twitter's API returns all information about a tweet, such as date, language, and user information. For this research, it was decided not to collect retweets without quotes, and only two fields (data of tweet and text) were stored in CSV files for further analysis.

Table 2. Keywords used to filter tweets

Original keyword	Keyword in English
lockdown	lockdown
quarentena	quarantine
isolamento	isolation
isolada	isolated
isolado	isolated

The last activity in the data collection step was the selection of only tweets related to the lockdown policies. This step was necessary because some search keywords do not necessarily return tweets about the lockdown but other COVID-19 subjects. Then, a Linux script was used to search only lines containing the words related to the lockdown policies in each CSV file, as described in Table 2.

4.3 Step 3: Exploratory data analysis

Using the parameters described previously, 12,497,029 tweets were selected between April 2020 and April 2021. As shown in Figure 2, it is important to note that the discussion around isolation policies decreased over time. This fact can be related to many reasons. At the pandemic's beginning, there was little information about COVID-19 because the research about it was relatively new. Moreover, the only known measures were those related to the control of the dissemination of other coronavirus diseases, such as isolation policies, mask-wearing, and cleaning.

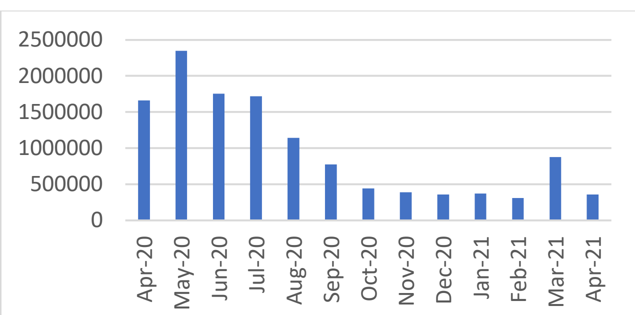

Figure 2. Tweets count

Although there was no national isolation policy, many Brazilian cities adopted measures to restrict the people's movement. However, due to the economic impact of these measures, many cities started to ease the movement restrictions when the total number of cases decreased in Brazil. With cases rising after the Christmas season, new isolation measures were adopted in many locations in Brazil. The Sao Paulo state, for example, decreed a set of restrictions on movement on March 2021. This may be a reason for the rise in this period, as shown in Figure 2.

Additionally, the top 10 hashtags used in all the periods were selected. As described in Table 3, the main hashtags were related to the pandemic and entertainment. At the beginning of the pandemic in Brazil, a large part of the population practiced isolation measures, and promoting online events, such as artist lives, was common. Although it did not appear among the top 10 hashtags, the political aspect was evident. Among the top 100 hashtags used in the period, there were 21,439 mentions of the president of Brazil in tweets related to isolation policies.

Table 3. Main hashtags between April 2020 and April 2021

Hashtag	Count
quarentena	102575
livelocalmariliamendonca	53406
covid	51527
lockdown	25581
coronavirus	24096
bbb	22074
fiqueemcasa	19259
melim	16700
ficaemcasa	13561
pandemia	13109

It was verified in the tweets that they contained references to other users, which on Twitter may indicate an interaction/conversation with other users or a quote to a user, as in a reference in a news. According to data in Table 4, an increase in the percentage of tweets containing references to other users is observed over time. The same trend can be observed in relation to Tweets containing links to external pages, in which there was a higher percentage of posts containing links in April 2021 than in April 2020. This could characterize an increase in posts sharing information (such as news) and a decrease in Personal Stories posts.

Table 4. Percentage of tweets with references and links

Period	Tweets with reference to other users (%)	Tweets with links (%)
April/2020	27.23	11.58
October/2020	32.33	14.94
April/2021	43.45	17.30

Regarding the published text, measures were calculated regarding the average length of posts and the average number of words in each period. According to Table 5, a change in the average length of posts can be observed over time, with an increase of about 37.6% between April 2020 and April 2021. The analysis of the average number of words also indicates an increase over time, around 34.3% between the months of April 2020 and 2021.

Table 5. Average length and word count by tweet

Period	Average length	Average word count by tweet
April/2020	109.70	18.16
October/2020	126.78	20.98
April/2021	150.99	24.39

Another characteristic that may indicate the classification of a text in terms of formality is the use of abbreviated words [12]. Using a Brazilian corpus of abbreviated words [13], posts from the three periods were analyzed, with results presented in Table 6. It is possible to observe that in the third period (April 2021) there was a decrease in the percentage of tweets that used at least one abbreviated word, which could indicate an increase in posts with a greater degree of formality compared to other months.

Table 6. Abbreviated words stats

Period	Tweets with abbreviated words (%)
April/2020	38.65
October/2020	39.24
April/2021	35.56

4.4 Step 4: Preprocessing

As the objective was to analyze only textual components, other elements of the tweets were removed, such as links, emojis, and numbers. As tweets may contain mentions of other users, these elements were also removed from the text. For some of the analyzes described below, part-of-speech tagging was performed to identify and select the words needed for each analysis.

4.5 Step 5: Text analysis

In this step, some techniques were adopted to describe the content of the tweets, such as bigram extraction, hashtags, and topic modeling. To analyze the complete content, tweets from the entire period were selected. However, given the dynamics of the pandemic in Brazil, specific periods (semesters or months) were selected for temporal evaluation of the discussion on social isolation measures.

4.5.1 Cloud Tags

In order to summarize the topics discussed on Twitter, three tag clouds were created with the main words discussed in the analyzed periods (April/2020, October/2020, and April/2021). For the development of these tag clouds, the SpaCy package was initially used for the part-of-speech tagging task. Then, the words of three classes were extracted (proper name, nouns, and adjectives) and with them, the visualizations of the most frequent words of each analyzed period were constructed.

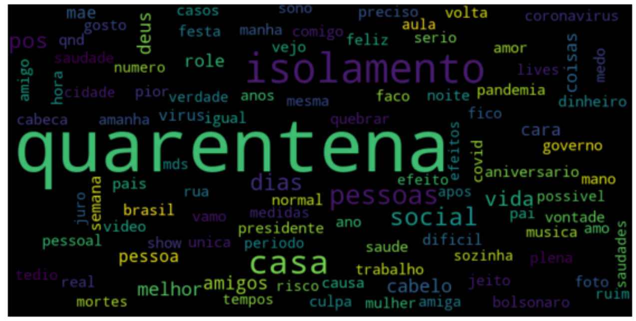

Figure 3. Word cloud for April 2020

In the three tag clouds (Figures 3, 4, 5) it is possible to notice the presence of the word *quarantine* (“quarentena” in Portuguese) more prominently, especially in the period of April/2020. However, it should be noted that this word was used as a filter in the selection of tweets, as described in Table 7. It is possible to observe in the three tag clouds that the weight of the word quarantine has been decreasing over time, unlike the other two words used in the filtering of tweets: *isolation* (“isolamento” in Portuguese) and lockdown, which gained weight over time. This data is indicative of the perception of the Brazilian people in relation to social isolation measures, although a lockdown has not been adopted in the country as was applied in other countries, such as China and Italy.

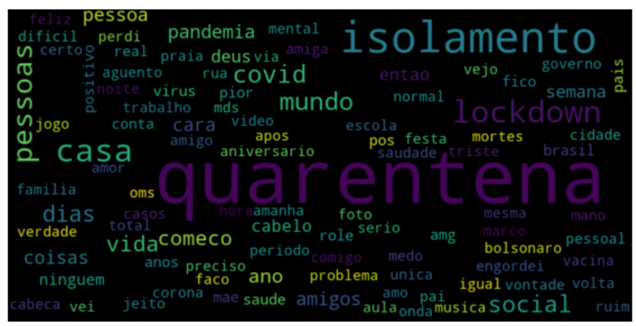

Figure 4. Word cloud for October 2020

Another difference observed over time was in relation to the formality of the speech. It is noticed, especially in the two analyzed periods of 2020 (April and October) the greater presence of words more present in informal speeches, especially those abbreviated as: "mds" (abbreviation of "Meu Deus" - *My God* in English), "vei" (*dude* in English), "qnd" (abbreviation for "when" - *when* in English). In the cloud for April 2021, words such as "pandemia" (*pandemic* in English), "Bolsonaro", "mortes" (*deaths* in English) are more prominent. One hypothesis may be related to the types of users who were discussing isolation measures in the three periods.

While in the first two periods, the discussion seems to be more focused on the reports and opinions of citizens about the isolation measures, in the last period the discourse seems to be conducted in a slightly less formal way, which could indicate a greater concentration of profiles related to publications. news outlets, such as portals and journalists. Another evidence supporting this hypothesis is the relative increase in posts made by verified accounts in relation to isolation measures, as detailed in Table 7.

While in the first period only 0.59% of posts were made by verified accounts, a year later this number increased to 1.48%. Institutional accounts (governments, newspapers) and journalists usually have the "verified" seal on Twitter and this increase in the participation of these accounts in the discussion about isolation measures may explain the change in the degree of formality between the three periods over time.

Table 7. Keywords used to filter tweets

Period	Tweets from Verified Accounts (%)
April/2020	0.59
October/2020	0.95
April/2021	1.48

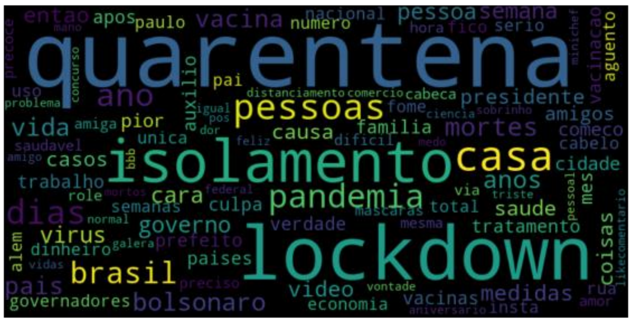

Figure 5. Word cloud for April 2021

Other hypotheses for the differences between the periods refer to the moment of the pandemic in which the country was. In April 2021, Brazil had the highest number of deaths, as shown in Figure 6. In addition, in this period, the country had already started vaccination against COVID-19, which would explain the presence of the word "vacinas" (*vaccines* in English) in the tag cloud of the period. The presence of the words "Bolsonaro", "president" (*president* in Portuguese), and "governadores" (*governors* in English) may be indicative of a discussion more related to political issues in the period.

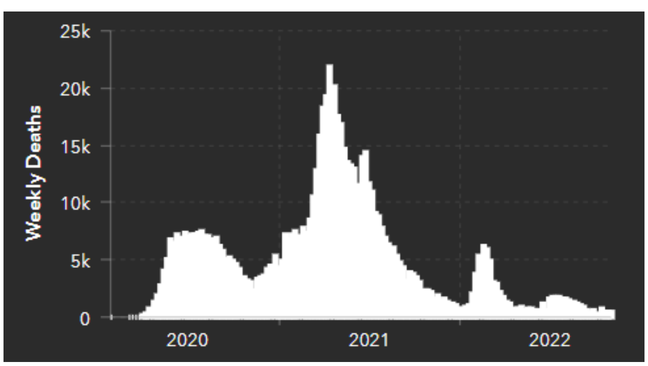

Figure 6. Deaths by COVID-19 in Brazil. Source: COVID-19 Dashboard [2]

4.5.2 Hashtags

To identify the main topics and to compare possible temporal changes in the discussion, the most frequent hashtags were identified for three months: April and October 2020 and April 2021. Table 8 lists the top 5 hashtags by month, and how each semester differs from others in pandemic history is clear.

Some aspects can be observed by analyzing the top 5 listed in Table 8. First, the list of the five most cited hashtags contains three groups of subjects: pandemic, quarantine, and entertainment. In all three months, there were mentions of Brazilian TV shows in tweets about the lockdown. Especially in the first months of the pandemic, online events were common, and many promoted isolation policies with the hashtag #ficaemcasa (#stayathome). In addition, an increase in mentions of the word lockdown can be observed, which did not appear in the top 5 of the first month and started to appear more in the following months.

The first months of 2020 were the pandemic beginning, which can explain why most top 5 hashtags were related to the coronavirus disease or the lockdown subject. The hashtag *quarantine* ("quarentena" in Portuguese) had almost three times more mentions than the covid hashtag when considering only tweets with words used as filters. The other two hashtags were related to the campaign 'stay-at-home' promoted by governments, artists, and other influencers.

The data about the three periods can illustrate, in some way, how the lockdown measures were adopted in Brazil during the pandemic. While many locations adopted isolation measures in the pandemic beginning, the lower cases and pressure made by economic sectors left many governments to ease the lockdown measures in the second semester of 2020. With the growth in cases in the first months of 2021, some locations returned to lockdown measures adoption.

Table 8. Main hashtags per period

Apr-2020	Oct-2020	Apr-2021
#quarentena (20314)	#quarentena (2639)	#bbb (2564)
#covid (7980)	#afazenda (2532)	#lockdown (2440)
#bbb (7135)	#covid (1934)	#covid (2085)
#coronavirus (4830)	#lockdown (847)	#quarentena (1709)
#ficaemcasa (4182)	#coronavirus (808)	#coronavirus (730)

4.5.3 Bigrams

Sometimes, the frequency of words alone may not describe a discussion on an issue enough. Thus, using bigrams or trigrams can be helpful in identifying the context in which a given word is used. Then, using the package NLKT, the main bigrams per semester were extracted, and the five most frequent were presented in Table 9.

In the general analysis of the bigrams for the word "isolation" ("isolamento", in Portuguese) and "quarantine" ("quarentena", in Portuguese), it was perceived that the occurrences of the expression "this isolation" and "this quarantine" – used to refer to isolation or quarantine as a specific event in time – diminished immensely over time. In the first semester of 2020, there were

5,374 occurrences of the bigram "this isolation" and 387,643 occurrences of the bigram "this quarantine". In the second semester of the same year, the occurrences of "this quarantine" dropped to 287,344 (diminished by 25,87%) and the occurrences of “this isolation" to 4,939 (decreased by 8,09%). In the first semester of 2021, there were 34,457 occurrences of "this quarantine" (diminished by 88,01%) and 1,926 of "this isolation" (reduced by 61%). In the whole period (2020.1 to 2021.1), the occurrences of "this quarantine" diminished by 91,11% and "this isolation" by 64,11%.

Table 9. Main bigrams per semester

1s2020	2s2020	1s2021
Nessa quarentena (387643)	Nessa quarentena (287344)	Isolamento social (65942)
Quarentena acabar (149665)	Quarentena pra (110285)	Todo mundo (39721)
Quarentena pra (128913)	Todo mundo (99801)	Nessa quarentena (34457)
Isolamento social (127233)	Isolamento social (94645)	Contra lockdown (24846)
Essa quarentena (113562)	Antes quarentena (76115)	Ano passado (22938)

The interpretation is that these numbers represent an important change in the discourses and debates on Twitter about the restrictive measures for Covid-19 in Brazil. The heated political debate about this topic did not seem to cease for the whole period. So, the hypothesis is that, over the three semesters, a terminological substitution took place. The center of gravity of the discussions has shifted. At the end of 2020, when Brazil was heading to the "third covid wave", the debate began to gravitate around a new term: "lockdown". The analysis of the bigrams for the word "lockdown" indicates the increase in the occurrences of the bigrams that refer to "lockdown" as an event or that refer to the position concerning the topic (the restrictive measures). For example, the expression "[to] make [the] lockdown" ("fazer [o] lockdown", in Portuguese) ou "against [the] lockdown" ("contra [o] lockdown", in Portuguese) has increased from 3,654 occurrences in 2020.2 to 16,012 in 2021.1 in the first case and from 3,182 to 24,846 in the second case. It was an increment of 338.20% in the occurrences of the expression "[to] make [the] lockdown" and an increment of 680.83% in the occurrences of the expression "against [the] lockdown".

So, the hypothesis is that, over the three semesters (2020.1 – 2021.1), the more generic words, "isolation" and "quarantine", seem to have been gradually replaced by the more technical word, "lockdown". However, the term "lockdown" was only superficially technical. The debate about the types of isolation (the variations and gradation of the Covid restrictive measures) was reproduced with the new term (although the use of the word "lockdown" to refer to "types of isolation" is inconsistent with its technical definition).

The analysis of the bigrams for the word "isolation" indicates that, in the first semester of 2020, this term was used to discuss types of isolation. In this period, the bigram "horizontal isolation" ("isolamento horizontal") occurred 3,891 times, "total isolation" ("isolamento total") occurred 5,469 times, and vertical isolation" ("isolamento vertical") occurred 7,204 times. These bigrams make up 5,84% of the total of isolation bigrams. In the second semester of 2020, the occurrence of this "type of isolation" category dropped to 1,27%. The "vertical isolation" bigram, for example, occurred only 2,967 times.

In comparison with the isolation bigrams, the analysis of the bigrams for the word "lockdown" indicates that, in the first semester of 2021, there were more "types of lockdown" being discussed: "total lockdown" ("lockdown total"), 7,193 times; "real lockdown" (lockdown [de] verdade), 5,068 times; "full lockdown" (“pleno lockdown”), 2,277 times; "severe lockdown" ("lockdown severo"), 2,077; "general lockdown" ("lockdown geral"), 1,362 times; "rigorous lockdown" ("lockdown rigoroso"), 1,182 times.

4.5.4 Topic analysis

4.5.4.1 Topic modelling

Since subjects can vary significantly over a semester, three specific months were selected for topic modeling: April and October 2020, in addition to April 2021. The topic modeling method adopted in this study was the Latent Dirichlet Allocation (LDA), a probabilistic model of how a document could be related to the topics that has been employed in other studies with Twitter data [14].

After extracting the 50 most important words of each topic in each period, the topics were classified into the following categories: 1) "report about personal matters (in the context of the pandemic)", 2) "general commentaries about the pandemic", 3) "political debate (about the pandemic)" and 4) "entertainment".

The first category – "report about personal matters" – is composed of words related to primary groups (family, friends, work colleagues) and connected to personal activities of daily living (dressing, feeding, mobility, and other associated tasks). In this category, people tweeted about how they have suffered during the pandemic with the loss of socialization, the new routine, etc. For example, people tweeted about the creative ways they found to cut their hair.

While the first category comprises commentaries about personal matters related to the pandemic, the second category is "general commentaries about the pandemic". The second category is composed of words that express feelings and opinions on the pandemic and associated themes.

The political opinions and views were separated into a third category in which the names of politicians appeared more frequently. The main words of the third category refer to political agents and positions as well as to central terms of the political debate like "economy", "death", "masks", "lockdown", "vaccine", "treatment", "isolation", etc.

Commentaries about TV shows, YouTube concerts, and artists compose the fourth category – "entertainment". In April/2021, an interesting phenomenon was observed in the topic categorized under the label "entertainment": political thematization interfered with entertainment themes. Some of the political and ideological discussion on lockdown and other restrictive measures appeared mingled with commentaries about the TV show "Big Brother Brazil" because one of the participants declared political support to one of the main political parties in Brazil and expressed opinions about the restrictive measures adopted by some states' governors.

According to the analysis of domain experts, it was possible to observe that the only topic where there was a clear boundary between the others was the topic classified as "political debate". The algorithm used selected five topics and, after analyzing the most frequent words in each topic and period, there were cases in which a topic could be classified in more than one class. In these cases, the decision about the classification of the topic was made by consensus among the project researchers. Thus, the distribution of posts by class was calculated and presented in Table 10.

Table 10. Tweets percentage by topic

Topic	**1s2020**	**2s2020**	**1s2021**
Personal reports	20.86	46.19	9.63
Commentaries about pandemic	19.39	17.55	55.87
Political debate	20.36	13.57	25.63
Entertainment	39.39	22.69	8.87

With the exception of the entertainment topic, which had a consistent decline in the three periods, in the others there was no trend. However, some explanations can be made in the comparison between the months of April 2020 and 2021. In the topic that contained posts with personal reports about isolation measures, there was a drop of more than 50% between the two months. It can be explained by the fact that in the second year of the pandemic, the isolation measures were not completely new for Brazilians, contrary to what happened in 2020. Regarding the topic of political debate, an increase can be seen between the two periods, possibly related to the fact that, in April 2021, Brazil had a record number of deaths from COVID-19 and possible discussions about the adoption of new isolation measures in the country.

4.5.4.2 Word clouds

Additionally, word clouds were generated for each topic and analyzed period. For this summary of subjects, the same strategy was adopted to generate a tag cloud for all the posts in each period: the SpaCy package was used to select the terms of the proper nouns, nouns and adjectives classes and to view the 100 most frequent terms in the clouds of words.

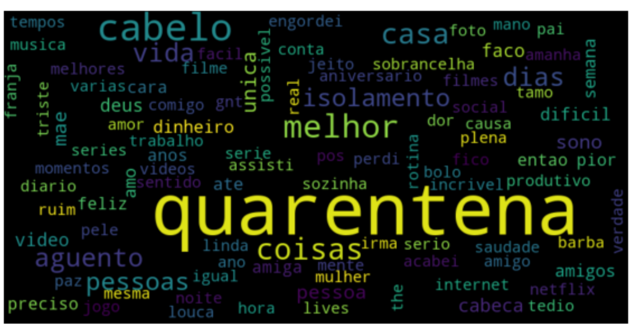

Figure 7. April 2020 – word cloud for topic 0

Figure 8. April 2020 – word cloud for topic 1

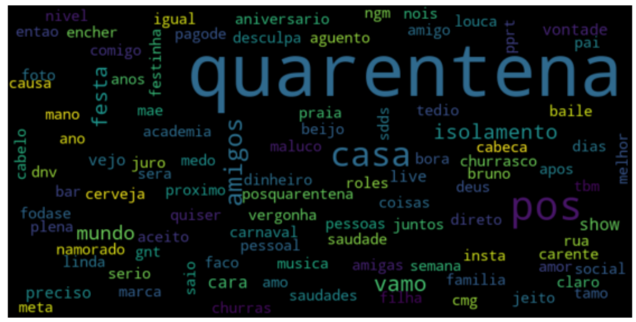

Figure 9. April 2020 – word cloud for topic 2

Regarding the month of April 2020, it is noteworthy that the in only one topic (3 – Figure 10) there is no predominance of the word "quarantine" (”quarentena” in Portuguese) in relation to other words. In this topic, it is possible to observe, unlike the others, that the main word related to isolation measures is "isolation". It is also possible to notice that this topic contains other highlight words, such as: "president" (“*presidente"* in Portuguese), governors (“*governador(es)”* in Portuguese), and names of Brazilian politicians. It may be an indication that the posts on this topic may be more related to political themes and possibly made by journalists or the media's Twitter profiles.

Figure 10. April 2020 – word cloud for topic 3

Figure 11. April 2020 – word cloud for topic 4

In topics 1 (Figure 8), 2 (Figure 9) and 4 (Figure 11) it is possible to notice in the tag cloud the presence of abbreviated words, such as "cmg" (abbreviation for the word "me"), "ngm" (abbreviation for the word "nobody") and "tbm" (short for "also"). These abbreviations may indicate a greater degree of informality in these topics' posts , which may be related to personal reports or opinions about the isolation measures. It is also possible to notice in topic 4 the presence of the name of a famous Brazilian entertainment program ("bbb"), which could help to classify posts in this topic as "entertainment".

In the topics related to the October 2020 posts, it is possible to observe that the word "isolation" is more prominent in topic 4 (Figure 16), where it is predominant, and in topic 0 (Figure 12), being the third most frequent word. In the other topics (1 - Figure 13, 2 - Figure 14 and 3 - Figure 15), the predominance of the word quarantine is observed.

Figure 12. October 2020 – word cloud for topic 0

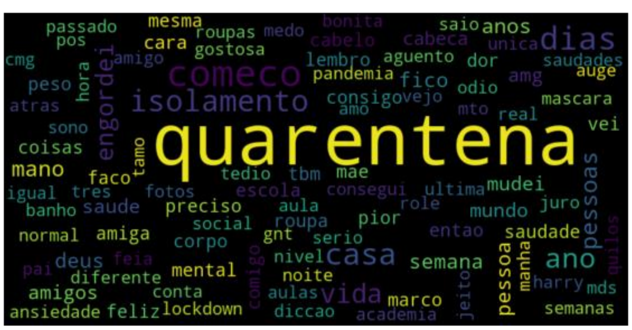

Figure 13. October 2020 – word cloud for topic 1

As in the previous period, one can observe the use of a more informal language in topics where the word "quarantine" has absolute prominence, with the use of abbreviated words. There are also some verbs in the first person, which were possibly classified by the SpaCy package as nouns. The highlighting of these verbs in these topics may indicate that these topics could be related to personal reports or opinions about the pandemic or isolation measures.

Figure 14. October 2020 – word cloud for topic 2

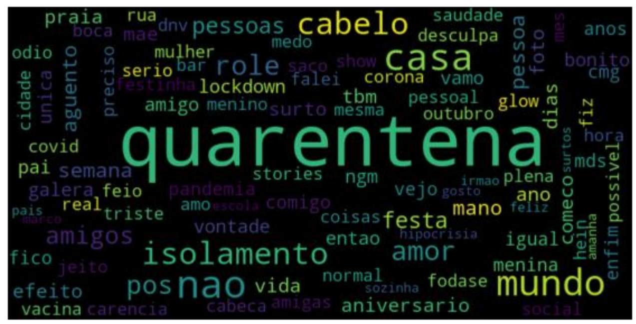

Figure 15. October 2020 – word cloud for topic 3

Figure 16. October 2020 – word cloud for topic 4

In topics 0 and 4, references to Brazilian politicians can also be noted, which could be an indication that the discussion in these topics could have a more political bias. In topic 0 of this period, there are references to other countries, such as Italy and Argentina, which were other countries that adopted isolation measures [15] [16].

In the period of April 2021, there is an increase in the frequency of the words "isolation" and "lockdown", with only one topic where the word quarantine was more prominent in relation to the others (Topic 3 - Figure 20). As in the other analyzed periods, there is a greater use of informal language in this topic than in the other topics, which could suggest that the use of the word "quarantine" as a way of referring to isolation measures would be more common in the population.

Figure 17. April 2021 – word cloud for topic 0

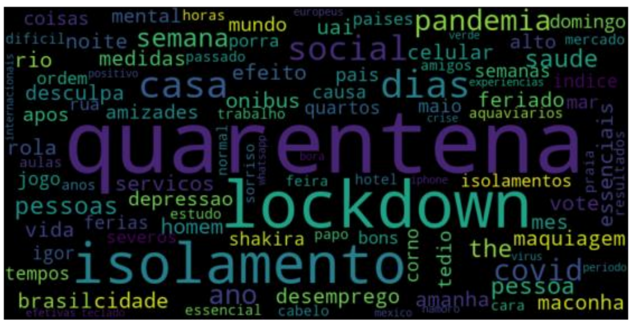

Figure 18. April 2021 – word cloud for topic 1

In Topic 2 (Figure 19), it is noteworthy that the word "lockdown" has greater prominence in relation to the others, something that did not happen in any of the other topics and analyzed periods. Observing a greater degree of formality in the words of this topic, it could be an indication that the profiles with more formal language (eg, journalists or politicians) could be referring to isolation measures such as lockdown. The presence of words related to elective positions (eg mayor and president) and names of politicians can be observed. These indications may suggest that discussions on this topic may be more related to political factors.

Figure 19. April 2021 – word cloud for topic 2

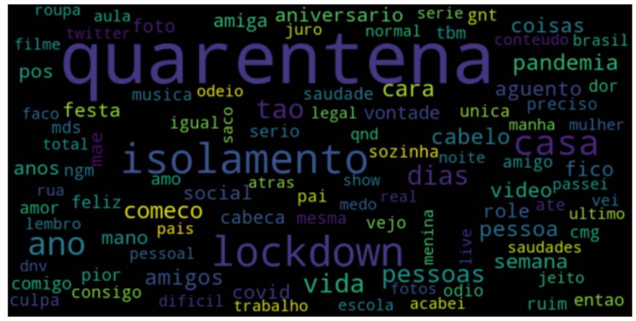

Figure 20. April 2021 – word cloud for topic 3

In Topic 4 (Figure 21) there is a greater predominance of proper names (participants of a Brazilian reality show) and a soccer team, indicating that the posts in this topic are related to entertainment programs. In this topic, the word quarantine is more prominent in relation to the others, although the difference in frequency is smaller than in the other analyzed periods.

Figure 21. April 2021 – word cloud for topic 4

The analysis of word clouds from the three periods can bring significant evidences to help in the classification of topics by domain experts. In addition, one can observe how different audiences refer to isolation measures in specific ways. While in the topics with a higher degree of formality, the measures were mainly referred to with the words "isolation" and "lockdown", in the other topics, with a greater degree of informality, the use of the word quarantine is predominant. This type of information can be very useful to support health departments in analyzing the results of communication campaigns.

4.5.5 *Sentiment analysis*

The last text analysis was the sentiment analysis by calculating the polarity of each period. For this analysis, we used a lexicon with 77,355 terms in Portuguese [17], which contains, for each term, polarity scores regarding positivity/negativity.

Initially, part-of-speech tagging of each tweet was performed in order to apply the analysis only to nouns, adjectives, and adverbs. Thus, only words pertaining to these three classes were selected and, for each word, the polarity of the lexicon used was calculated. Then, for each tweet/period, the total positive and negative scores were calculated, as well as the difference between the two scores (column "diff"). With this data, four metrics were calculated for each period:

- Average grade for each period: (sum of positive scores - sum of negative scores) / total tweets

- % of positive Tweets: (total of tweets with diff > 0) / total tweets

- % of negative Tweets: (total of tweets with diff < 0) / total tweets

- % Neutral Tweets: (total of tweets with diff = 0) / total tweets

Table 11. Polarity metrics

Metric	1s2020	2s2020	1s2021
Score average	0.15	0.11	0.03
Positive tweets (%)	62.13	56.07	45.82
Negative tweets (%)	20.84	25.02	31.04
Neutral tweets (%)	17.03	18.91	23.14

According to the data presented in Table 11, there is a decrease in the average grade over time, which could suggest a worsening in people's view of isolation measures. This hypothesis is supported by the trend identified in the other two metrics: an increase in the percentage of tweets with a negative score and a decrease in the percentage of tweets with a positive score.

According to the data presented in Table 11, there is a decrease in the average grade over time, which could suggest in general lines, a worsening in people's view of isolation measures. This hypothesis is supported by the trend identified in the other two metrics: an increase in the percentage of tweets with a negative score and a decrease in the percentage of tweets with a positive score. This general modification in people's view of isolation measures can be explained with two auxiliary hypotheses. On the one hand, the decrease in the percentage of tweets with a positive score might have been caused by the fact that, in the later analyzed periods, people were tired of the measures, and some stopped defending it. On the other hand, the increase in the percentage of tweets with a negative score might have been caused by the political polarization that the nation was experiencing (at the time of the pandemic and, to some extent, because of it).

5. FINAL CONSIDERATIONS

This research aimed to analyze Brazilian users' discussion about lockdown policies on Twitter. Three studies were conducted considering the whole period, user types, and the temporal aspect.

Regarding research question 1, the main topics were entertainment (TV shows), political discussion, general commentaries about the pandemic, and personal stories. Results show that the debate on these topics varied over time (research question 2), caused by the events at the moment, such as pandemic status, the advance of vaccination, and the reduction of cases.

Results suggested that specific analysis can be more helpful than considering the whole period and users' data. The use of information extraction techniques in large volumes of text proved to be very useful in identifying the main topics discussed on Twitter. Public Health departments can adopt these techniques to monitor people's opinions. Furthermore, it is important that the political aspect was linked to the discussions on isolation measures. This may have negatively impacted the acceptance of the measures, and these impacts may be the subject of future studies.

Regarding limitations, it is important to know that Twitter users do not represent the population. This platform's users differ from other social media, such as Instagram and Facebook. Another limitation is that Twitter is less used in Brazil compared to countries like England and the United States. Moreover, it is important to analyze the influences caused by high movements not necessarily related to the subject of interest. This is the case of the increased engagement generated by artists or fans.

Further research can analyze the role of other kinds of data in describing the mood of a subject. For example, extracting meaning from emojis can reveal more information about a post. Moreover, additional user information, such as description, age, and location, can be used to extract more insights.

6. ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

7. REFERENCES

[1] Cucinotta, D. and Vanelli, M. 2020. WHO Declares COVID-19 a Pandemic. Acta Biomed, v. 91, p. 157–160

[2] Dong, E. and Du, Ho.. Gardner, L. 2020. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases, v. 20, n. 5, p. 533-534, 2020.

[3] Cevik, M., Tate, M., Lloyd, O., et al. 2021. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. The Lancet Microbe, v. 2, n. 1, p. 13–22.

[4] Mahato, S., Pal, S. and Ghosh, K. G. 2020. Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. Science of The Total Environment, v. 730, p. 139086.

[5] Lau, H., Khosrawipour, V., Kocbach, P., et al. 2020. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. Journal of travel medicine.

[6] Guzzetta, G., Riccardo, F., Marziano, V., et al. 2020. The impact of a nation-wide lockdown on COVID-19 transmissibility in Italy. arXiv preprint arXiv:2004.12338.

[7] Nakada, L. Y. K. and Urban, R. C. 2020. COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil. Science of the Total Environment, v. 730, p. 139087.

[8] Dantas, G., Siciliano, B., França, B. B., Da Silva, C. M. and Arbilla, G. 2020. The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil. Science of the total environment, v. 729, p. 139085.

[9] Kanitkar, T. 2020. The COVID-19 lockdown in India: Impacts on the economy and the power sector. Global transitions, v. 2, p. 150–156.

[10] Jørgensen, F. J., Bor, A., Lindholt, M. F. and Petersen, M. B. 2020. Lockdown evaluations during the first wave of the COVID-19 pandemic. Aarhus: Aarhus University. doi, v. 10.

[11] Chehal, D., Gupta, P. and Gulati, P. 2021. COVID-19 pandemic lockdown: An emotional health perspective of Indians on Twitter. International Journal of Social Psychiatry, v. 67, n. 1, p. 64–72.

[12] Sheikha, F. A. and Inkpen, D. 2010. Automatic classification of documents by formality. In: Proceedings of the 6th international conference on natural language processing and knowledge engineering (nlpke-2010). IEEE, 2010. p. 1-5.

[13] Nascimento, R. S. et al. 2021. Avaliando contribuições na substituição de termos informais em classificação de texto de redes sociais com NetSpeak-BR. In: Anais do X Brazilian Workshop on Social Network Analysis and Mining. SBC, 2021. p. 181-186.

[14] Zou, L. and Song, W. W. 2016. LDA-TM: A two-step approach to Twitter topic data clustering. In 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). . IEEE. http://ieeexplore.ieee.org/document/7529581/.

[15] Cerqueti, Roy et al. 2022. The sooner the better: lives saved by the lockdown during the COVID-19 outbreak. The case of Italy. The Econometrics Journal, v. 25, n. 1, p. 46-70, 2022.

[16] Larrosa, J. MC. 2021. SARS-CoV-2 in Argentina: Lockdown, mobility, and contagion. Journal of Medical Virology, v. 93, n. 4, p. 2252-2261, 2021.

[17] Pedro Thales. 2021. SentiWordNet-PT-BR. (August 2021). Retrieved October 4, 2022 from https://github.com/Pedro-Thales/SentiWordNet-PT-BR