**Sentiment analysis bias in 2022 Brazilian election tweets**

**Rafael Santos Gouveia**

**Andrei Fokin Teixeira**

Affiliations: Engineering and Data Science Department - Coimbra University

rgouveia@student.uc.dei.pt

afokin@student.uc.dei.pt

**Abstract**

Recently, democratic elections are meeting the technology of social networks with different consequences for human relationships. One of them is the feeling of the platform users, who are also voters, getting more prominent. This happens because more people are actively participating in political processes, such as the number of data and information about them. Often they are just supporting or opposing candidates, but with polarization observed in different places and moments, identifying hate speeches, fake news or simply the impact of confirmation bias become interesting objectives for the data scientists to support society with good practices. The present study proposes to understand, through a machine learning model, the feelings contained in comments made on Twitter during the 2022 Brazilian elections in relation to the two main candidates in the dispute, Luis Inácio Lula da Silva and Jair Bolsonaro.

**1. Introduction**

Technology impacts all aspects of human life, from infrastructure to communication, from health to entertainment. Its presence is such that it reduces costs, distances and increases people's access and participation to goods and services. Information is one of these goods mainly after the advent of the Internet.

Social networks are examples of open spaces inside the Internet, some used for study, some for photos, some for videos and so on. Sometimes they have many objectives, but they end up specializing in some way. “Many companies and people use them to spread products and services and publish opinions, facts that have turned the social networks into powerful sources of information on various topics” Cristiani et al (2020). Twitter has specialized in being an open, low-censorship wall, where everyone can express free opinions, some of them about policy.

On the readers' side, policy is an aspect of life that gains strength in a digital arena, where one can express more radical opinions with less social punishment, especially in anonymous profiles. On the side of the candidates, in turn, policy is a competitive game, whose objective is the conquest of power and, in democratic societies, convincing the average voter that their campaign is a good option requires sophisticated tools.

When voters are increasingly connected to new spaces for discussion, the migration of parties and candidates to social networks is expected. At the limit, the candidate submits its opinion to the masses opinion and voters support things that make sense to them, in a loop known as “confirmation bias”. Thus, there is a mutual exchange that strengthens a set of ideas that make sense for a particular social group, even after the election, when a government and its opposition are transformed. Therefore, the Internet becomes a space for governance itself, where networks such as Twitter are the communication channel and users have the feeling of greater participation.

In Brazil, where society is heterogeneous and where social networks are a novelty for political purposes, the electoral campaign of 2022 has the potential for aggression in the real world. Therefore, the main motivation of this work is to understand how a machine learning model can be used to classify and identify comments with political content. Insights can be used inside companies not to spread hate.

To this end, two are the main hypotheses raised: (i) the ranking of a good estimator should take into account the initial proportion of sentiments, for example, if there are loads of positive content rows, is expected to have a good-feeling estimator; and (ii) there should be no ranking bias when changing proper names from Lula to Bolsonaro or vice versa.

**2. Related Work**

One of the uses of social networks data is to create security systems against fake news, malicious robots or any abuse of democratic or moral rules. Even when “the real-time sentiment classification on general topics on a real time basis is particularly challenging for data dynamicity and lack of labeled data” (Guerra *et al*, 2011), there are opportunities for the data scientist to improve content classification models. A phenomenon of communication that may help this complex analysis is that “while new terms may arise and old terms may change their meaning, user bias tends to be more consistent and robust over time as a basic property of human behavior” (Guerra *et al*, 2011). Technology must advance faster than human behaviour, so the greater the technical capacity to detect patterns, the greater the chance of good tools for protecting data and/or people, according to the interests of society itself.

Cristiani et al (2020) worked in a sentiment analysis of the 2018 election in Brazil. The goal was to relate people's opinions to traditional street opinion polls: “the results show that Twitter is a great source of information, especially as a source of research on the opinion of its users. The numbers show that the amount of positive messages have a strong relationship with the polls (...). Although the studies did not show the relationship between the tweets classified as neutral and negative, this information can be useful to assist in various types of decision making”.

Chaudhry et al (2020) analyzed tweets of the 2020 US elections and compared Trump's performance in his two electionss. A time series analysis was used to also bring dynamics through the year of 2020 in different states and for different subjects. One conclusion is that “for all states where sentiment results did not corroborate with election results, long-term trends before and after the election reveal that there was an increase in the positive sentiment of the winning candidate”. The current paper will not take into account, but it would be interesting for future works to have insights on how former-president Lula da Silva returned to power after 12 years and how to use macro data to predict future election scenarios.

Ricci (2020) brings some insights about preprocessing data. Regarding the use of emojis, by removing then "the Random Forest algorithm had the greatest gain of all", precisely the algorithm used in this paper. "In the case of stopwords removal, the Logistic Regression and SVM algorithms had a decrease in accuracy, while Naïve Bayes and Random Forest algorithms had a gain. This reveals that also the type of specific treatment affects different algorithms in different ways".

**3. Data & Approach**

This work will be carried out in the sequence of actions: (i) obtaining data from Twitter; (ii) preprocessing data by removing accentuation, obvious writing errors and orthographic mistakes; (iii) train a random dataset with two different models (Random Forest and Decision Trees) to help checking accuracy; and (iv) after a good accuracy, do a simple test write some random sentence and changing the name of the two main candidates to see if there is some bias related to the persona name.

**3.1 Inicial data information**

This project has the Twitter Application Programming Interface (API) as its source of data. Talking about the tweets themselves, 413 comments were collected at all and divided as follows: 20% for training and 80% for testing. From the whole universe, 236 were taken about Lula, 213 about Bolsonaro. There is some intersection between them, because in some tweets, both are mentioned. Figure 1 shows the groups.

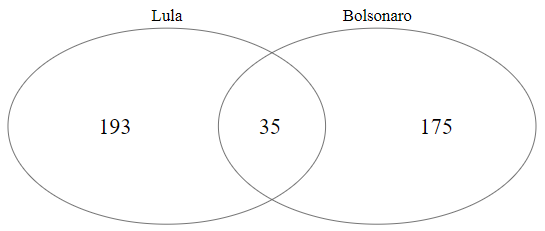


Figure 1: Number of tweets for each candidates

Initially, we have only used "lula", "bolsonaro" and “eleicoes”, but we have just created a list of new words to better characterize each group and collect more information, without context loss:

|  |  |
| --- | --- |
| Candidates | Key words |
| Lula da Silva | “lula”, “lulalivre”, “lulaladrao”, “molusco”, “lulapresidente”, “lula13” |
| Jair Bolsonaro | “bolsonaro”, “bozo”, “bozonaro”, “jair”, “elenao”, “elesim”, “mito”, “messias”, “bolsonaro22” |
| Eleições 2022 | “eleicao”, “eleicao22”, “eleicoes2022”, “eleicoes22” |

Table 1: Dictionary of words used to group candidates

The period is about October, after the result of the 1st round, where there was no decision and when the campaign continues until the end of the month.

Tweets whose location is identified and within Brazil are sought, as we want to see opinions from people who are experiencing and being impacted by the process.

Thus, we collect tweets outside the official accounts of the two candidates, no matter if they were original or retweeted and being indifferent to photos, videos, emojis and any other non-textual artifice.

**3.2 Programming approach**

After the data is defined, the tweets are extracted directly from Twitter's API with the help of Python's Tweepy library, which will link the Python code to the network's social media information. This tool allows us to collect tweets from the day the query was performed or up to a certain date. We use some filters for the queries, one of these filters is the language of the tweets in 'pt-br'. The result of this query is a json document that has all the data and metadata of a tweet.

To clean up the data we used a few steps, the first was to check what data we were going to use that was useful for sentiment analysis. After this step necessary treatments were made on the tweets' texts, which were to remove user names, urls, accents and special characters. Also tweets were normalized to take data as string type on utf-8 decode scheme. Finally, “nltk” and “stop words' libraries were used to pre-process data.

After cleaning up, Scikit-learn algorithm Count Vectorizer (CV) are used to transform the text into a vector based on the frequency of each word that occurs in the text and TfidfVectorier measure how relvant a word is to a document and then creates a matrix with this values, and to compare how it will affect the prediciotn of the model. Then, another two Scikit-learn algorithms Random Forest Classifier (RFC), Decision Tree Classifier (DTC) and Support Vector Classification(SVC) are used to train and test vectors and transform them into sentiment classificators.

The three methods we used are supervised machine learning algorithms used for classification modeling. Their classifiers create a set of decision trees from a randomly selected subset of the training set. “The main distinction between them is that DTC are graphs that illustrate all possible outcomes of a decision using a branching approach. In contrast, RFC outputs are a set of decision trees that work according to the output” (Talari, 2022). And SVC “works mapping data points to a high-dimensional space and then finding the optimal hyperplane that divides the data into two classes.” (Kumar, 2022).

Last but not least, after training each model, some metrics are used to see their efficiencies, such as:

* accuracy: the number of hits in our model divided by the total sample
* f1-score: the harmonic mean between precision and recall, a global metric for the model.

**4. Experimentation**

After extraction, the data for Lula, Bolsonaro and the election went through the following pre-processing:

1. removing whitespace before and after sentences;
2. removing accentuation and cedillas;
3. removing all special characters, leaving just letters and numbers;
4. removing "@" from users;
5. removing "#" from hashtags; and
6. removing stop words.
7. is normalized the words from the dataset using lemmatize, here the lemma of the word is found so that the context of the sentence is taken into account.

The separate data for each candidate and ranking was as it follows:

|  |  |  |
| --- | --- | --- |
| Candidates | Classification | Quantity |
| Lula da Silva | 1 | 109 |
| 0 | 25 |
| -1 | 102 |
| Jair Bolsonaro | 1 | 60 |
| 0 | 28 |
| -1 | 125 |

Table 2: Real values by candidate and emotion

Many tweets have obvious classification, but facing doubtful scenarios must be dangerous for the estimator. It’s mandatory to the following some criteria, so here it follow some adopted rules:

* having a journalistic format or a simple report of some event: neutral;
* when the comment is clearly positive for one and clearly negative for another: positive;
* critical but thoughtfully written: negative;
* presence of adversative adverbs for positive and negative content: neutral.

The final database was passed through RFC., DTC and SVC, whose rating targets is the following list:



Figure 2: Decision list

Below, the test result:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | AccuracyCV | AccuracyTfidf | F1-Score CV | F1-Score Tfidf |
| RF | 0,67 | 0,67 | 0,66 | 0,66 |
| DT | 0,57 | 0,51 | 0,57 | 0,49 |
| SVC | 0,67 | 0,68 | 0,62 | 0,67 |

Table 3: Test results by model

Table 3 shows us that even when using two vectorization methods, CountVectorizer and TfidfVectorizer, the results of Accuracy and F1\_score are very close when comparing the same models, and when comparing the results of the training algorithms we see that DCT falls short of RFC and SVC, and the last two fall short of the same results.

Once the way the estimator works (average) was known, an experiment was conducted to understand how accurate it is with respect to the candidates' names (bias). In this way, it is possible to understand not only the efficiency of the estimator but also its consistency.

A test dataframe was created with new phrases to swap the names Lula and Bolsonaro in each one. Below in Table 5 are some of the phrases used, where X is the dedicated space to fill with the name:

|  |  |
| --- | --- |
| # | Random created sentence |
| 1 | Gosto muito do presidente X. |
| 2 | X foi um ladrão e merece ir para a cadeia. |
| 3 | X foi um dos piores presidentes do Brasil, roubou e fez o que queria para se manter no poder |
| 4 | X foi muito bom para o povo brasileiro, deu oportunidades e emprego |

Table 4: Dictionary of sentences to test bias

Each model result are presented below, where 1 is a positive sentiment and -1 a negative one:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Treated sentences | RF TFIDF | RF CV | DT TFIDF | DT CV | SVC TFIDF | SVC CV |
| gosto presidente lula | -1 | -1 | -1 | -1 | -1 | -1 |
| gosto presidente bolsonaro | -1 | -1 | -1 | -1 | -1 | 1 |
| lula ladrao merece ir cadeia | 1 | 1 | -1 | 1 | 1 | 1 |
| bolsonaro ladrao merece ir cadeia | 1 | 1 | 1 | 1 | 1 | 1 |
| lula piores presidentes brasil roubou fez queria manter poder | 1 | 1 | 1 | 1 | 1 | 1 |
| bolsonaro piores presidentes brasil roubou fez queria manter poder | 1 | 1 | -1 | 1 | 1 | -1 |
| lula bom povo brasileiro deu oportunidades emprego | 1 | 1 | 1 | 1 | 1 | 1 |
| bolsonaro bom povo brasileiro deu oportunidades emprego | -1 | 0 | -1 | 0 | 1 | -1 |

Table 5: Bias experiment results

The bias experiment showed an interesting thing: even with a well-balanced amount of positive and negative tweets from candidate Lula and a much larger unbalanced amount of negative tweets from candidate Bolsonaro, the estimator tied the former to -1 ratings and the latter to 1.

These are unanticipated results for both initial hypotheses: results didn't respect the initial proportion of each sentiment and they treated name candidates in a factitious manner.

**5. Conclusions**

Table 5 shows us that the results of the models and their respective numerical/vectorial representation of the text, TfidfVectorizer and CountVectorizer, we realize that the results tend to classify sentences with 'bolsonaro' in the tweets as positive and that contain 'lula' as negative and neutral. We noticed this both in the different models, this may have as the amount of data is low and some tweets used mention both candidates in the same sentence, we noticed this when we checked the figure where we compared the amount of classification per mention of the candidate, where we noticed that it is above the amount of data present in the study.

Another reason that led to this problem was the difficulty in retrieving this data, first because of the Twitter API that does not allow downloading tweets from periods longer than 7 days from the request date, another is that we have no control of which tweets will come, so there is the possibility of having duplicate tweets, data can come unbalanced etc., ie, we have no control over how the data will come and how the balancing will be, so it took a very long time to get enough data so that there was a generous amount and balanced so that the models get more accuracy in the classifications.

For future work we will use a dataset with a larger amount of tweets to see if this same set of variables, such as two types of tweet vectorization algorithms, three classification algorithms will maintain the biased behavior or if the bias will decrease.

**6. Bibliographic references**

A Cristiani, D Lieira, H Camargo., 2020. *A sentiment analysis of brazilian elections tweets*, in: Anais do VIII Symposium on Knowledge Discovery, Mining and Learning, p. 153-160.

Chaudhry, H.N.; Javed, Y.; Kulsoom, F.; Mehmood, Z.; Khan, Z.I.; Shoaib, U.; Janjua, S.H. Sentiment., 2020. *Analysis of before and after Elections: Twitter Data of U.S. Election*. Electronics 2021, 10, 2082. Available in: <https:// doi.org/10.3390/electronics10172082>.

PH Calais Guerra, A Veloso, W Meira Jr, V Almeida, 2011. *From bias to opinion: a transfer-learning approach to real-time sentiment analysis*, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, p. 150-158.

Ricci, R. D. *Análise de sentimentos no Twitter sobre a Reforma da Previdência no ano de 2019*, 2020, Monografia, Universidade Federal de Uberlândia, Uberlândia, MG, p. 70.

Talari Sakumar, *Random Forest vs Decision Tree: Key Differences*, 2022. Available on: <https://www.kdnugget s.com/2022/02/random-forest-decision-tree-key-differences.html>.

Kumar Ajitesh, *SVM Classifier using Sklearn: Code Examples*, 2022. Available on: < https://vitalflux.com/svm-classifier-scikit-learn-code-examples/ >.