Automatic detection of fake tweets about COVID-19 Vaccine in Portuguese

Rafael Geurgas

geurgas@ifi.unicampp.br

State University of Campinas

Leandro Tessler

State University of Campinas

Research Article

Keywords: Disinformation, COVID-19, Neural Networks, Automatic Classification

Posted Date: October 4th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3393186/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Social Network Analysis and Mining on March 8th, 2024. See the published version at https://doi.org/10.1007/s13278-024-01216-x.

Automatic detection of fake tweets about the COVID-19 Vaccine in Portuguese

Rafael Geurgas[1*] and Leandro R. Tessler[1]

[1]IFGW, Unicamp, 13083-970 Campinas, SP, Brazil.

*Corresponding author(s). E-mail(s): geurgas@ifi.unicamp.br;
Contributing authors: tessler@unicamp.br;

Abstract

The COVID-19 pandemic induced an unprecedented wave of of disinformation in social media in Brazil. In particular Twitter (currently X) was used to spread fake news about COVID-19 vaccines that helped to induce vaccine hesitation. This article presents a BERT-based neural network for automatic detection of fake tweets. The optimized architecture relies upon BERTimbau, a BERT implementation pre-trained in Brazilian Portuguese, fine-tuned using three fully connected layers. All 2,857,908 tweets in Portuguese containing the word *vacina* (vaccine in Portuguese) were collected over a 7-month period. A subset of 14,400 tweets was manually classified as real or fake. The network was fine-tuned using the 1,144 curated fake tweets and a random sample of 2000 real tweets. Optimal results were achieved melting the last four layers of the BERTimbau. The best results obtained were 77.1% f1-score and 76.9% accuracy. These results are already acceptable for practical applications. They can be improved increasing the size of the training dataset. State of the art performance was obtained training the same neural network architecture with a larger curated balanced English language training dataset.

Keywords: Disinformation, COVID-19, Neural Networks, Automatic Classification

1 Introduction

For more than ten years, the main sources of information for Brazilians have been online. Online news sources comprise both traditional organizations and

small independent information channels, as well as social networks. Since 2020, social networks have surpassed traditional television networks Newman et al (2023).

While this shift enables rapid and extensive news coverage, it also provides a fertile environment for the spread of misinformation.

The political polarization that took place in Brazil has given rise to a significant group of individuals who exhibit radicalized tendencies Layton et al (2021). This group exhibits a strong tendency towards consuming news that reinforce their preconceived opinions, rather than seeking a deeper, balanced perspective. It is also more susceptible to conspiracy theories Uscinski et al (2022) and disinformation Ecker et al (2022), often favoring independent news sources aligned with their political beliefs while nurturing mistrust towards traditional and reliable news outlets. They prefer to rely on their own judgment to assess and filter information. This causes the widespread acceptance and propagation of false narratives. This phenomenon not only amplifies belief in conspiracy theories but also facilitates the rapid dissemination of fake news.

In contrast to traditional news outlets, social media platforms allow users to act as commentators and even news sources. This unique ecosystem tends to reinforce existing biases and strongly enhance the rapid propagation of disinformation. Empirical studies have revealed that false news spread faster and more widely than accurate information on platforms like Twitter, now knows as X Vosoughi et al (2018).

The implications of disseminating health-related disinformation are far-reaching and potentially catastrophic. During the COVID-19 pandemic, society had to respond promptly and effectively to contain the spread of the virus. When the pandemic began, the idea of having an effective vaccine in a short time seemed very far from reality. The Brazilian government claimed that social distancing policies would cause irreparable damage to the economy for a long time. President Bolsonaro then encouraged the adoption of ineffective "early treatment", based on drugs promoted by the US president Donald Trump. Since then, the official Brazilian government COVID-19 police was never science-based Ricard and Medeiros (2020). Government officials downplayed the importance of social distancing and mask usage, and later, when vaccines were available, cast doubt on their safety and efficacy Galhardi et al (2020).

Within this context, vaccine disinformation has been disseminated not only by government officials but also by their followers and influencers, who acted as sources of fake news and amplifiers. Detecting and preventing the rapid spread of disinformation is of utmost importance to mitigate its dire effects Bin Naeem and Kamel Boulos (2021). One possible approach is to use independent, trained professionals to meticulously analyze and verify the legitimacy of news content. However, this traditional method is both expensive and time-consuming, especially in an era characterized by instantaneous information. Alternatively, artificial intelligence offers a promising solution, capable of efficiently processing and classifying multiple texts within seconds. An artificial

intelligence layer can at least filter potentially fake news to be curated by humans.

The objective of this study is to develop an operational model that can effectively discern Fake from Real tweets about COVID-19 vaccines in Portuguese. By leveraging the power of artificial intelligence, we aim to contribute to the identification and mitigation of vaccine misinformation on social media platforms.

2 Background

Transformer based algorithms have been become the standard approach to Natural Language Processing (NLP) Vaswani et al (2017). They rely on the so called self-attention mechanism, that allows for processing whole sentences at once resulting in highly efficient and reliable outputs. A very convenient, widely used architecture for language representation is the Bidirectional Encoder Representations from Transformer (BERT) Devlin et al (2018). The BERT architecture consists of an input (or embedding) layer, a stack of transformer encoders and a classification layer. Text must be transformed in tokens before being processed by BERT. The tokenizer attributes numerical values to each semantic element of a sentence. BERT is available in two sizes: BERT-base with 12 encoder layers, 768 neurons per layer resulting in 110 million parameters; and BERT-large with 24 encoder layers, 1024 neurons per layer, with 340 million parameters. At the time of writing, the most recent version of BERT was trained using the entire English Wikipedia (2,500 million words) and the BookCorpus (800 million words) Zhu et al (2015). The training process took 4 days using Google Cloud TPUs Muller (2022). Training the whole BERT for specific applications is costly and impractical. The best strategy is to use fine-tuning: adding supplemental layers of neurons to the BERT output and train only those layers with a specific dataset Sun et al (2019). Fine-tuning can be improved by unfreezing the top layers of BERT in the optimization process, allowing small adjustments in already optimized parameters, a process that requires much less computational power than optimizing from scratch Lee et al (2019).

Fine-tuned BERT was successfully used to automatically detect fake news about COVID-19 in English in Twitter Chakraborty et al (2021).

3 Methods

To design a practical detector of fake tweets about vaccines in Portuguese, we opted for BERT-base because it is reliable enough and requires less computational resources than BERT-large. Although a multilingual version of BERT-base is available Devlin (2019), the best approach is to use BERTimbau Souza et al (2020), a BERT model trained specifically with a Web corpus of Brazilian Portuguese.

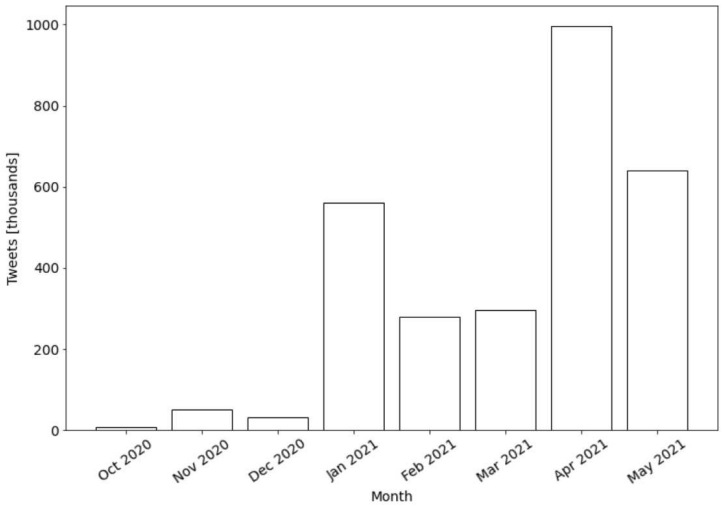

Fig. 1 Monthly distribution of tweets in Portuguese containing the word *vacina*.

3.1 Database

The database consisted of all 2.857.908 tweets in Portuguese containing the word *vacina*, vaccine in Portuguese, collected using the Twitter API between October 30, 2020 and May 25, 2021. This corresponds to a period of intense activity about vaccines in social networks in Brazil. Vaccination for COVID-19 started in December 2020 in the UK. Brazil had the first vaccinated person on January 8, 2021. Figure 1 shows the monthly distribution of tweets in the database. Notice the significant increase after vaccination started in Brazil.

The tweets originate form 918,368 unique users.

The language used on Twitter is different from formal language and even from the web language used to train BERTimbau. The use of slang, abbreviations and especially emojis is very common in Twitter. Since emojis are associated with semantic meaning in tweets, they were translated into common words. A dictionary containing 1,068 emojis and their description in Portuguese was created. The descriptive text was adapted to Portuguese from Emojipedia[1]. Although stopword removal from the dataset is not necessary when using Transformers, we removed URL links because they are not semantic elements.

The tweets about vaccines contains frequent words that either do not appear or are very uncommon in the corpus used to train BERTimbau. We therefore added custom tokens of significant words to the tokenizer: *vacina* (vaccine), *vacinação* (vaccination), Pfizer, *cloroquina* (chloroquine), *ivermectina* (ivermectin), *Brasil* (Brazil), *Covid* (COVID-19), *Bolsonaro* (the Brazilian president), *Jair* (his firs name) and *Doria* (the governor of the state of São Paulo).

[1] https://emojipedia.org/

To create a training data subset, 16.731 tweets were randomly selected and manually labeled by the authors in three classes: not COVID-19 related, Fake and Real. To make the labeling process was as homogeneous as possible, the authors established rigid criteria. In the beginning of the process they labeled the same messages and checked for consistency.

In the training data subset, 2,309 messages turned out to be about vaccines not related to COVID-19[2]. These tweets were discarded from the training dataset.

Automatic classification models struggle to detect irony and sarcasm Potamias et al (2020). For this reason, the 436 tweets that contained irony were also excluded from the training dataset. Some prominent examples are tweets mocking about 'turning gay after receiving the vaccine' and 'turning into an alligator after receiving the vaccine'. Both make fun of president Bolsonaro statements in public speeches at the time. Following fake statements spread on Twitter[3], president Bolsonaro more than once associated the COVID-19 vaccine with a supposed increase in HIV/AIDS infections, associated by his followers with homosexual behavior. President Bolsonaro often made public statements casting doubts about the effectiveness and safety of the vaccines. He once said that the contract with Pfizer had a clause exempting the company from responsibility over any side effects, and ironically added that a vaccinated individual could not complain if he or she turned into an alligator, a reference to another belief of his followers, that the vaccine could damage the DNA of human cells.

The effective training dataset consisted of 14,000 tweets. Some examples of classified tweets are in Table 1. Several Fake messages are written in non-standard Portuguese. This may affect the quality of the training process.

Figure 2 shows wordclouds of the most frequent words in the Fake and Real tweets subsets. Common stop words were disregarded to build the the wordclouds. In the Fake Tweets subset, the word *chinesa* (Chinese) is the 9th most common word. This word is often used with a negative connotation in relation to Coronavac (21st most common word), the vaccine developed by China (18th most common word). Coronavac development was partly financed by the State of São Paulo as determined by governor Doria (7th most common word). Doria evolved from political ally to fierce opponent of president Bolsonaro. *Tratamento* (treatment), the 13th most common word, and *precoce* (early), the 24th most common word, also appear in the Fake News subset referring to different forms of early treatment with ivermectin and *cloroquina* (chloroquine), the 15th most common word. Although early treatment was widely supported and recommended by Bolsonaro government officials and by the president himself, there is no scientific evidence of efficacy of any of these treatments.

[2]During the collection period there were news about a cancer vaccine being developed at the University of Oxford McAuliffe et al (2021) and consequently a wave of tweets concerning this subject have circulated.

[3]Reuters Fact Check: COVID-19 vaccines do not cause HIV or AIDS. https://www.reuters.com/article/factcheck-vaccines-hiv-idUSL1N2UW10H.

Table 1 Some examples of sentences from the training dataset with the assigned classification and sub-classification. Many messages are written in non-standard Portuguese.

text	translation	classification	sub
por um Brasil com mais vacina e nenhuma chacina	for a Brazil with more vaccines and no slaughter	Real	-
astrazenica/oxford não é uma vacina, é um atropelamento de um caminhão	astrazeneca/oxford is not a vaccine, it is like being hit by a truck	Fake	Irony
vacinar a população com uma vacina sem comprovação científica, típico de uma ditadura, o povo que sofre as consequências da irresponsabilidade	vaccinate the population with a vaccine without scientific proof, typical of a dictatorship, the people who suffer the consequences of irresponsibility	Fake	Public Health
tomei a vacina da gripe, sou a pessoa mais imunizada do mundo agora	I just got the flu jab, I am the most immunized person in the world now	Not COVID-19	-

Fig. 2 Left: Word cloud for the Fake Tweets subset. **Right:** Word cloud for the Real Tweets subset.

In the Real Tweets dataset the words *sorridente* (smiling), 28th most common word, and *coração* (heart), 22nd most common word. Although *sorridente* (smiling) and *coração* (heart) may seem out of context, they appear as emojis expressing happiness and love for receiving the vaccine dose.

3.1.1 Fake Tweets Subset

All tweets containing false or misleading information were classified as fake. They were checked against news published in prestigious traditional Brazilian newspapers and news outlets as Folha de São Paulo[4], Estadão[5] and G1[6]. All tweets that contained statements involving treatments for COVID-19, facts

[4] https://www.folha.uol.com.br/
[5] https://www.estadao.com.br/
[6] https://g1.globo.com/

about the vaccine or government statements that could not be verified as Real were classified as Fake.

3.1.2 Real Tweets Subset

All tweets that could not be classified as Fake were considered Real. They include expressions of opinion as diverse as people happy to receive the vaccine, people commenting that family members received the vaccine and news about the vaccine. This subset is highly heterogeneous, with different forms of writing and a wide collection of topics.

3.1.3 Dataset Statistics

The effective labeled dataset is composed of 14,000 tweets, of which 1,144 (8.17%) were classified as Fake and 12,856 (91.83%) as Real. Class imbalance can be detrimental to the training of a neural network Mirus et al (2020); Hensman and Masko (2015). To minimize the imbalance effects, the subset was reduced by randomly suppressing Real Tweets to the ratio of 64% Real Tweets and 36% Fake Tweets. Bringing the balance closer to 50% Real and 50% Fake Tweets decreases the accuracy of the training because of the reduction of absolute size of the training dataset. The randomization was stratified to ensure that each subset had the same Real/Fake ratio Geron (2018) The labeled dataset was randomly distributed in 70% for training, 15% for validation and 15% for testing.

BERT algorithms require that all sentence inputs have the same number of tokens. One important BERT fine-tuning hyperparameter is the maximum sentence size. A very small maximum sentence size will cause truncation of longer sentences. A very large maximum sentence size will cause padding with null tokens. The best results should be obtained minimizing both truncation and padding, but truncation leads to worse performance because it suppresses information. The distribution of sentence lengths (in tokens) in the labeled dataset is represented in Figure 3. Most sentences are shorter than 60 tokens, but some are as long as 100 tokens.

4 Results

The performance of neural networks for classification is normally evaluated by the accuracy and f1-score, a metric that combines precision and recall. Because only a subset of the Real Tweets is used for training, the accuracy of the trained neural network may depend on the particular subset chosen. To check this dependence we trained the network with 50 different randomly chosen Real Tweets subsets. We adjusted a Gaussian curve to the results and obtained a weighted f1-score average $\mu = 0.75$ with a standard deviation $\sigma = 0.02$. The same happens to the Fake Tweets f1-score, with a slightly higher $\sigma = 0.03$. This shows that indeed the performance of the network depends on the particular chosen Real Tweets subset. We also observed that even for the same Real Tweets subset the f1-score may vary depending on the seed used for random

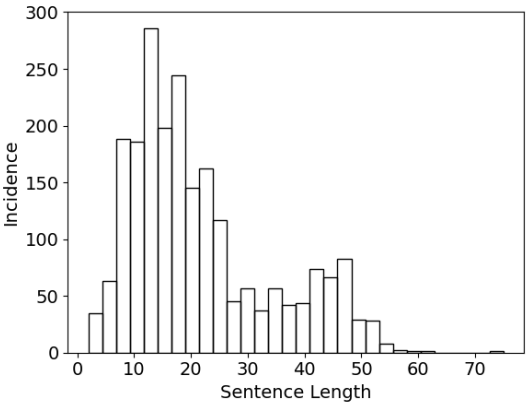

Fig. 3 Distribution of sentence lengths (in tokens) in the training subset.

Table 2 Comparing neural network shapes. The **best configuration** is represented in bold text.

Neural Network length	Layer width	f1-score weighted	f1-score Fake Tweets	Accuracy
2	(768; 2)	0.620	0.553	0.612
3	(768; 128; 2)	0.609	0.512	0.602
3	**(768;512;2)**	**0.771**	**0.696**	**0.769**
3	(768; 768; 2)	0.651	0.588	0.644
4	(768; 512; 512; 2)	0.663	0.583	0.656
5	(768; 512; 512; 512; 2)	0.591	0.509	0.583

number generation. The standard deviation is the $\sigma = 0.2$ both for weighted and Fake Tweets f1-score. The obtained standard deviations are used as an estimator for the uncertainty of the results.

Several hyperparameters were optimized to obtain the best possible performance. The process must be done carefully because some hyperparameters are interdependent. We started by optimizing the network shape. The results are shown in Table 2. The network depth (number of layers) was varied from two to five, with the best result obtained with three layers. We also tested different middle layer widths, 128, 512 and 768 neurons. The best result was for 512 neurons.

The output layer used Softmax as activation function. With the neural network shape optimized we proceeded to improve the remainder hyperparameters: activation function in the bottom layers, optimizer, batch sizes, dropout regularization and learning rate. The performance for different hyperparameter combinations are shown in Table 3.

Retraining the last BERT layers during downstream tasks can improve results Lee et al (2019). We tested the effects of melting the top layers while unfreezing between 0 and all 10 BERTimbau layers for different maximum input lengths, between 50 and 120. The results are shown in Figure 4. Unfreezing BERTimbau layers does not cause a clear change in network performance.

Table 3 Optimization of performance.

Hyperparameter	Value	f1-score weighted	f1-score Fake Tweets	Accuracy
Best Configuration		0.771	0.696	0.769
Activation Function	Softmax	0.603	0.546	0.595
Optimizer	SGD	0.622	0.558	0.614
Optmizer	Adam	0.622	0.558	0.614
Batch Size	64	0.620	0.563	0.612
Batch Size	All data	does not converge		
Dropout Rate	0.5	0.646	0.565	0.638

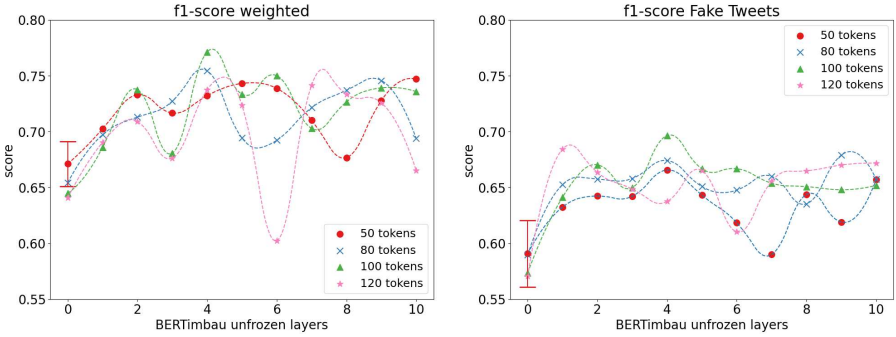

Fig. 4 Left: Overall f1-score versus the number of unfrozen BERTimbau layers. **Right:** Fake Tweets f1-score versus the number of unfrozen BERTimbau layers. The dotted lines are guides for the eyes.

The best result was obtained when keeping the bottom 8 layers frozen and retraining the top 4 layers. This means that to achieve top performance we had to optimize approximately 37 million BERTimbau parameters and 0.4 million fine-tuning parameters. The curves are noisy, probably because of the reduced size of the dataset.

The hyperparameters that yielded the best results are shown in Table 4. The precision, recall and f1 scores for **best configuration** of our system are represented in Table 5. The corresponding confusion matrix is represented in Figure 5.

To evaluate the performance of our approach it was tested with the publicly available COVID-19 Fake Tweets Dataset. This is a 10,700 tweets dataset about COVID vaccines, manually tagged as Fake or Real Patwa et al (2021). The main difference in relation to the Portuguese dataset is that it is balanced, consisting of 5,600 tweets labeled Real and 5,100 tweets labeled False. Using this dataset to fine tune CT-BERT, a BERT-base based model pre-trained on a large corpus of Twitter messages on the topic of COVID-19, an impressive f1 = 98.42% has been reported for Fake tweets Glaskowa et al (2021). Fine tuning BERT-base the same authors obtained f1 = 96.75 % for Fake tweets. We compare with our results, we fine tuned BERT-Base instead of BERTimbau-base retaining all hyperparameter values that were optimized for the Portuguese dataset. The results are shown in Table 6.

Table 4 Best neural network configuration.

Hyperparameter	Used
Number of layers	3
Layer widths	(765; 512; 2)
Pre-trained BERT model	BERTimbau Base cased
Tokenizer	BERTimbau Base cased
Activation Function	ReLU
Output Activation Function	LogSoftmax
Optimizer	AdamW
Loss Function	Cross entropy
Batch Size	32
Dropout Rate	0.8
Learning Rate	3×10^{-5}
Maximum input length	100
Unfrozen BERTimbau layers	4
Epochs	3

Table 5 Classification report for the best configuration.

	precision	recall	f1-score
Real Tweets	83.8%	79.1%	81.4%
Fake Tweets	66.5%	73.1%	69.6%
accuracy			76.9%
weighted avg			77.1%

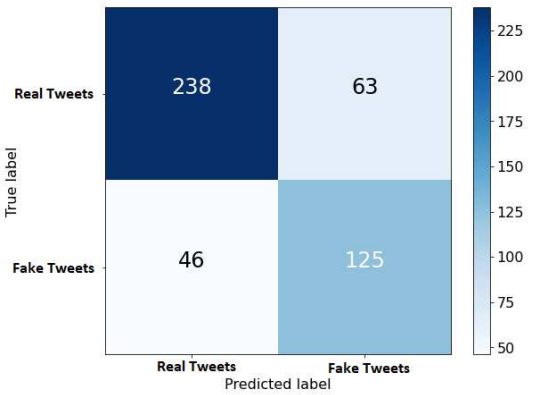

Fig. 5 Confusion matrix.

5 Discussion

We optimized and trained a BERT-based algorithm to detect fake tweets about vaccines in Portuguese. The overall performance is better than the performance for Fake Tweets. This can be due to the larger (64%) Real Tweets proportion in the training dataset. The performance indicators were not as good as what

Table 6 Classification report for BERT-base using the hyperparameters optimized for the Tweets dataset in Portuguese applied to the COVID-19 Fake News Dataset Chakraborty et al (2021).

	precision	recall	F1-score
Real Tweets	95.1%	98.3%	96.3%
Fake Tweets	98.5%	94.7%	96.4%
accuracy			96.3%
weighted avg			96.3%

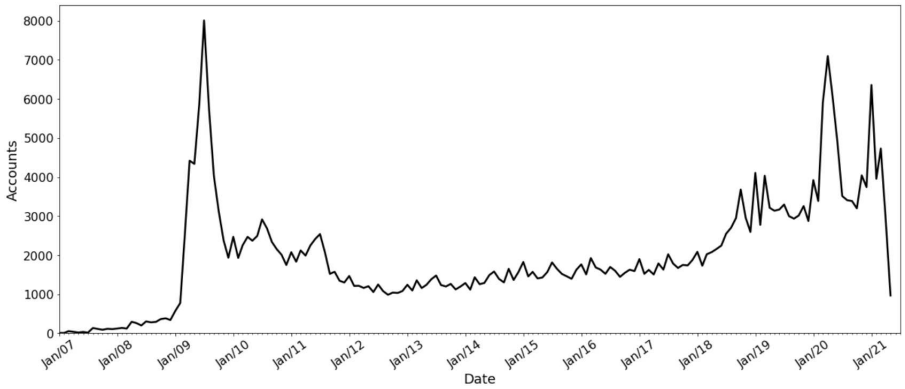

Fig. 6 Number of accounts that spread Fake Tweets versus account creation date.

has been reported for a similar situation using a balanced dataset in English Glaskowa et al (2021). However, if we apply the hyperparameters optimized for the Portuguese dataset to BERT-base in English using the COVID-19 Fake News Dataset we obtain the same performance indicators. This means that the main limitation of the present work is the Portuguese database, with a low incidence of Fake tweets. Furthermore, some peculiarities of the Portuguese dataset may also have contributed. Very often Fake tweets are written in non-standard Portuguese, making them difficult to understand even for a human reader. This is detrimental for the fine tuning process and consequently the performance of the neural network decreases.

Because the subject has strong political implications, there is the possibility that accounts propagating Fake tweets have been created with this purpose. Figure 6 shows the number of accounts who disseminated Fake tweets versus the account creation date. Notice that besides a peak in 2009 when Tweeter became popular in Brazil, there are two peaks: one in early 2020 that corresponds to the beginning of the COVID-19 pandemic and one in early 2021 that coincides with the vaccination. These accounts were probably created respectively to comment COVID-19 and vaccination related topics. Because of the political implications of the subject they may include numerous inauthentic bots.

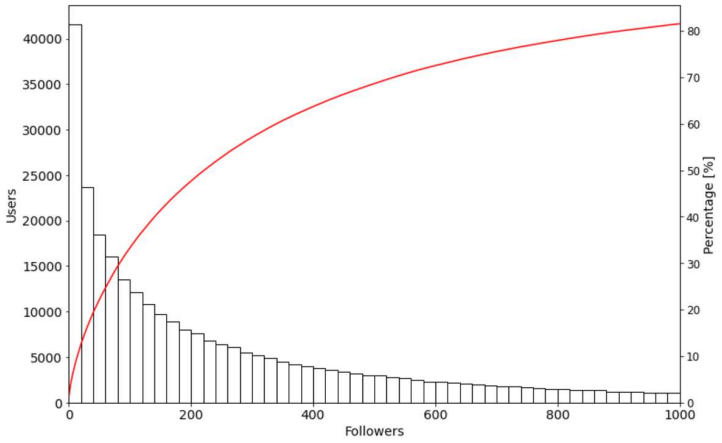

Fig. 7 Histogram of the number of accounts versus number of followers up to 1000 followers. The red line represents the percentage of users with less than the number of followers. 83% of the accounts have less than 1000 followers.

Most users who produced Fake Tweets have an irrelevant number of followers, but a few can have more than 1.5 million followers. In Figure 7 we represent the histogram of number of accounts versus number of followers and the total percentage of users below that number of followers. Notice that 83% of the users have spread fake tweets have less than 1000 followers. There are, however, few accounts with more than 1 million followers that are certainly relevant sources of fake messages that are amplified in the social network.

6 Conclusions

We have collected all 2.857.908 tweets in Portuguese containing the word *vacina* between October 30, 2020 and May 25, 2021. We used this database to fine tune a BERTimbau-base based neural network, and obtained a f1-score of 69.6% for Fake tweets and 77.1% weighted average, Using the hyperparameters optimized for this dataset in an identical architechture with BERT-base and a well balanced dataset of roughly the same size in English we obtained 96.4% F1 for Fake Tweets and 96.3% weighted average. These findings indicate that our current approach holds a very good potential for automatic detection of fake news in Portuguese.

7 Acknowledgements

This article was produced within the aegis of BI0S - Brazilian Institute of Data Science, FAPESP contract 20/09838-0. It was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

Bin Naeem S, Kamel Boulos MN (2021) Covid-19 misinformation online and health literacy: A brief overview. International Journal of Environmental Research and Public Health 18(15). https://doi.org/10.3390/ijerph18158091

Chakraborty T, Shu K, Bernard HR, et al (2021) Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers, vol 1402. Springer Nature, https://link.springer.com/book/10.1007/978-3-030-73696-5

Devlin J (2019) Bert multilingual model. https://github.com/google-research/bert/blob/master/multilingual.md, Accessed: 2023-02-23.

Devlin J, Chang M, Lee K, et al (2018) BERT: pre-training of deep bidirectional transformers for language understanding https://arxiv.org/abs/1810.04805

Ecker UK, Lewandowsky S, Cook J, et al (2022) The psychological drivers of misinformation belief and its resistance to correction. Nature Reviews Psychology 1(1):13–29. https://doi.org/10.1038/s44159-021-00006-y

Galhardi CP, Freire NP, Minayo MCdS, et al (2020) Fact or fake? an analysis of disinformation regarding the covid-19 pandemic in brazil. Ciência & Saúde Coletiva 25:4201–4210. https://doi.org/10.1590/1413-812320202510.2.28922020

Geron A (2018) Hands-on machine learning with scikit-learn and tensor flow. O'Reily Media Inc, Sebastopol, CA

Glaskowa A, Glazkov M, Trifonov T (2021) g2tmn at constraint@aaai2021: Exploiting ct-bert and ensembling learning for covid-19 fake news detections. In: Combating Online Hostile Posts in Regional Languages during Emergency Situation. Springer International Publishing, pp 116–127, https://doi.org/10.1007/978-3-030-73696-5_12

Hensman P, Masko D (2015) The impact of imbalanced training data for convolutional neural networks. Degree Project, KTH Royal Institute of Technology, https://wwwkthse/social/files/588617ebf2765401cfcc478c/PHensmanDMasko_dkand15pdf

Layton ML, Smith AE, Moseley MW, et al (2021) Demographic polarization and the rise of the far right: Brazil's 2018 presidential election. Research & Politics 8(1). https://doi.org/10.1177/2053168021990204

Lee J, Tang R, Lin J (2019) What would elsa do? freezing layers during transformer fine-tuning https://arxiv.org/abs/1911.03090

McAuliffe J, Chan HF, Noblecourt L, et al (2021) Heterologous prime-boost vaccination targeting mage-type antigens promotes tumor t-cell infiltration and improves checkpoint blockade therapy. Journal for ImmunoTherapy of Cancer 9(9). https://doi.org/10.1136/jitc-2021-003218

Mirus F, Stewart TC, Conradt J (2020) The importance of balanced data sets: Analyzing a vehicle trajectory prediction model based on neural networks and distributed representations. In: 2020 International Joint Conference on Neural Networks (IJCNN), pp 1–8, https://doi.org/10.1109/IJCNN48605.2020.9206627

Muller B (2022) Bert 101 state of the art nlp model explained. Hugging face https://huggingface.co/blog/bert-101, Accessed 09/20/2023.

Newman N, Fletcher R, Eddy K, et al (2023) Reuters institute digital news report 2023, http://www.digitalnewsreport.org/2023

Patwa P, Sharma S, Pykl S, et al (2021) Fighting an infodemic: Covid-19 fake news dataset. In: Chakraborty T, Shu K, Bernard HR, et al (eds) Combating Online Hostile Posts in Regional Languages during Emergency Situation. Springer International Publishing, pp 21–29, https://doi.org/10.1007/978-3-030-73696-5_3

Potamias RA, Siolas G, Stafylopatis AG (2020) A transformer-based approach to irony and sarcasm detection. Neural Computing and Applications 32(23):17,309–17,320. https://doi.org/10.1007/s00521-020-05102-3

Ricard J, Medeiros J (2020) Using misinformation as a political weapon: Covid-19 and bolsonaro in brazil. Harvard Kennedy School Misinformation Review 1(3). https://doi.org/10.37016/mr-2020-013

Souza F, Nogueira R, Lotufo R (2020) Bertimbau: Pretrained bert models for brazilian portuguese. In: Cerri R, Prati RC (eds) Intelligent Systems. Springer International Publishing, pp 403–417, https://doi.org/10.1007/978-3-030-61377-8_28

Sun C, Qiu X, Xu Y, et al (2019) How to fine-tune bert for text classification? In: Sun M, Huang X, Ji H, et al (eds) Chinese Computational Linguistics: 18th China National Conference, CCL 2019, Kunming, China, October 18–20, 2019, Proceedings 18. Springer International Publishing, pp 194–206, https://doi.org/doi.org/10.1007/978-3-030-32381-3_16

Uscinski J, Enders A, Diekman A, et al (2022) The psychological and political correlates of conspiracy theory beliefs. Scientific reports 12(1):21,672. https:

//doi.org/doi.org/10.1038/s41598-022-25617-0

Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need abs/1706.03762. https://arxiv.org/abs/1706.03762

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151. https://doi.org/10.1126/science.aap9559

Zhu Y, Kiros R, Zemel R, et al (2015) Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp 19–27, https://doi.org/10.1109/ICCV.2015.11