II Workshop de Matemática, Estatística e Computação Aplicadas à Indústria

Automatic Identification of Fake News in Portuguese

Luiz Giordani,[1] Gilsiley Darú,[2] Rhenan Queiroz, Vitor Buzinaro, Davi K. Neiva, Daniel Camilo Fuentes Guzman,[3] Marcos Jardel Henriques, Francisco Louzada Neto
ICMC-USP

1 Introduction

With the intensification of the use of online communication tools in the last decade, especially social networks, the amount of manipulated or even totally false news has increased significantly. Their impact on political and social issues has become increasingly evident [2]. As a result, this topic has become the focus of work of many researchers around the world, encouraging the development of statistical and computational techniques aimed at controlling and reducing the spread of fake news (See for example [2], [5], [3], [1] and [4]).

Public opinion on various relevant issues of society has been negatively influenced by the widespread dissemination of fake news. In Brazil, for example, it is observed that manipulated information is scatter through the various social networking platforms created with the aim of allowing people to express themselves freely [1]. To try to get around this problem, there are fact and news verification agencies as well as several professionals who make use of searches and surveys with the intention of being as complete as possible, researching the origin of the information transmitted, validating these together with the opinions of experts on the subject in specific for checking the information [4].

In this work, natural language processing methodologies were applied, together with statistical modeling in the development of a tool capable of assisting in the classification of potentially fake news (Web Platform for automatic verification of fake news). In addition, with a theoretical basis, this tool provides statistics that point out the most common characteristics of these types of news. Furthermore, the platform uses statistical models and techniques such as Word2vec, Web Scraping, Boosting, among others, on real data, evaluating them based on new news. The combination of these methodologies makes the platform provide the probabilities of the verified news being, or not, false; with an accuracy of approximately 96%.

During the development of this work, it was possible to evaluate/identify which are the most relevant technologies to deal with the classification of fake news, bring about a technical debate

[1] luiz.giordani@usp.br
[2] ghdaru@usp.br
[3] camiguz89@usp.br

Tabela 1: Model results in current news

			True percentages		
Sources	Material	N	TF-IDF	CBOW	Skip-Gram
Jornal cidade online	Doubtful	26.239	14,04%	10,72%	10,69%
Senso incomum	Doubtful	2.889	38,80%	23,23%	28,35%
Conexão política	Doubtful	15.867	26,41%	25,25%	22,60%
Brasil de fato	Credulous	25.412	80,01%	80,68%	80,86%
G1	Credulous	11.907	82,20%	84,20%	81,18%
El país	Credulous	16.809	90,78%	87,54%	86,66%

*Source: Prepared by the authors

about how they work and which are the most efficient ways for them to be applied to textual news in Portuguese, aiming to reach the highest possible level of precision in the identification of *fake news*. In addition, it is a work with a social nature of extreme relevance. Because, with it, it became possible to build a web platform that is easy to access and use, so that the general population can use it. In other words, the purpose of the platform is to make it another tool to combat fake news, where each user can access a page on the internet, paste the news and check if it is something false or true.

2 Results

Different models of machine learning were tested, configuring the hyperparameters using the *GridSearch* technique and comparatively measuring their performance when applied to the Fake.Br corpus.

The LightGBM TF-IDF model presented the best result, reaching 96.17% of accuracy and F1-Score of 96.16%. An increase of approximately 8% in the F1-Score presented by the model *Bag of Words* of [3], and an increase of approximately 3% in the F1-Score in relation to the model *Bag of Words* exposed by [5]. Subsequently, the LightGBM CBOW model has an accuracy of 95.67% and an F1-Score of 95.65%. In third place, the LightGBM Skip-gram model with 95.53% accuracy and F1-Score 95.49%.

Even with the adoption of different *features* for each model, there is a uniformity in the computed results, with the exception of the SVM model, which was not efficient in the classification of *fake news* from the Fake.Br corpus.

In the table (1), we compare the results of the three models that performed better when applying the classifier to the set of news collected via *Web Scraping*.

The classification of a text as true was considered when the probability indicated by the model to fit this classification was greater than 0.5. In general, the three models assertively differentiate between dubious and legitimate materials, but it is worth noting that the models that used mapping through *Word2Vec* presented a rate of classification as true lower for sources of dubious origin; which is an indication that these models (*CBOW* and *Skip-gram*) have a greater sensitivity to point

to materials of such origin.

3 Final Comments

It is important to point out some limitations that the platform still has, and that will possibly be improved/overcome over time. Some of them are about the type of text that the platform is prepared to work with. So far it has been implemented to verify the veracity of news texts only. Texts outside the journalistic nature, such as direct questions, texts from Instagram, Twitter, Facebook, Whatsapp, messages via cell phones, among others, do not carry enough textual structures for the platform to be able to analyze them.

4 Conclusions

The results indicate that training models with more recent vocabularies, added through *Word2Vec*, provided an assertiveness very close to what had been previously obtained in the classification of *fake news*. There is still a discussion about which model would be chosen, since methodologies that create less complex models, such as logistic regression, allow a better understanding of which components contribute most to the classification, while boosting models, such as LightGBM, present greater complexity and, consequently, less interpretability. Furthermore, with the application of these models in more recent news, it was observed that the predictions are in line with what was expected, presenting predictions with a higher rate of classification as *fake news* for news of dubious origin, while the texts from more prestigious sources had a much lower rate of classification as *fake news*.

A web interface was created to make available the *fake news* classification models as well as to make it possible to obtain the probability of a new news being fake, allowing the user to choose which model(s) will be applied.

Finally, the increase in the performance of these classification models are essential to solve the problem of spreading fake news, a problem that is increasingly present in society. In addition, these fake news are increasingly elaborate, highlighting the need to develop more refined modeling techniques, as well as the constant updating (retraining) of these models to be able to identify more recent patterns among fake news.

Acknowledgments

The authors were supported by CNPq, FAPESP, FAPEMIG and CAPES of Brazil.

Appendix A: Platform for Automatic Identification of Fake News

The following is a step-by-step description of how to properly use the platform to detect fake news. The figure (1) presents the initial screen of the web platform that performs automatic detection of fake news developed by the authors of this work. This one was developed so that anyone

Figura 1: Web Platform for Automatic Fake News Detection

can use it easily. That is, at any time an individual can put into practice the act of verifying the veracity of certain news that he has in his hands.

To use the platform access the address http://www.fakenewsbr.com and paste the text of the news to be verificate in the box located in the central region of the platform. After pasting the news in the field click on the *Analyze* button. This button is just below the box where the news was pasted. After a few seconds the results are displayed on the screen. With this, the person who is carrying out the verification will have results that will help him in making a decision to conclude on the veracity of the news he has just read.

Referências

[1] Furnival, Ariadne Chloe Mary, and Tábita Santos. *Desinformação e as fake news: apontamentos sobre seu surgimento, detecção e formas de combate*. Conexão-Comunicação e Cultura (2019).

[2] Haridas, Nandhini. *Detecting the spread of online fake news using natural language processing and boosting technique*. Diss. Dublin, National College of Ireland, 2019.

[3] Monteiro, Rafael A., et al. *Contributions to the study of fake news in portuguese: New corpus and automatic detection results*. International Conference on Computational Processing of the Portuguese Language. Springer, Cham, 2018.

[4] Parra, Francisco de Borja Martínez, and Susana Torrado Morales. *Fake news y redes sociales: análisis del fact-checker Newtral durante las elecciones al Parlamento de Andalucía de 2018*. Revista de estilos de aprendizaje 13.26 (2020): 106-117.

[5] Silva, Renato M., et al. *Towards automatically filtering fake news in Portuguese*. Expert Systems with Applications 146 (2020): 113199.