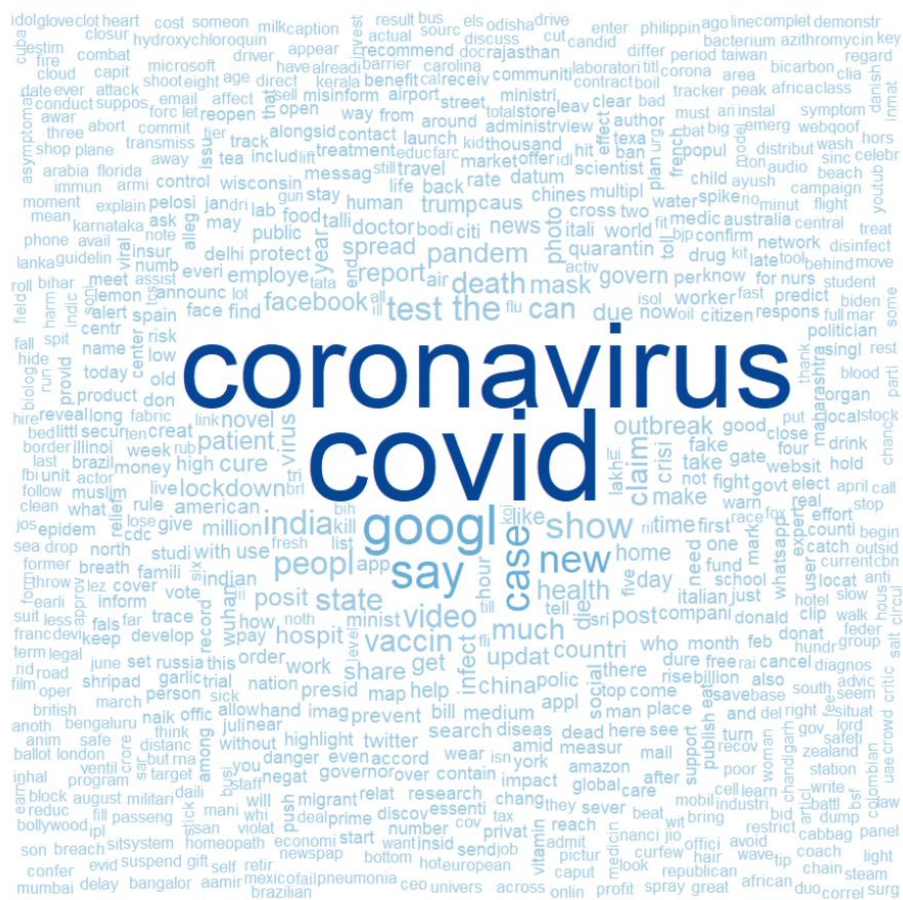# COVID-RELATED FAKE NEWS CLASSIFICATION

NATURAL LANGUAGE PROCESSING

ELIA ALONSO RUEDA
elia.alonsoru@alumnos.upm.es
Intelligent Systems - EIT Health

Contents

Problem to Solve

The spread of SARS-CoV-2 and consequently, the critical situation due to COVID-19 emergence prompted health officials to establish stringent control measures to halt the spread of the disease, leaving the world in a grim situation. However, the number of coronavirus infections has increased again and again in many countries resulting in the most devastating public health crisis seen in the 21st century [1], which is said to have originated from Wuhan, China, and has been in the news for months [2].

Of course, a large amount of news are consumed using social media, however, this is like using a double-edged sword: on the one hand, it is cost-effective and fast but, on the other hand, the information may be fake, and it is important not only to detect them but also stop their spreading. It goes without saying that fake news is not a new phenomenon, but when the world is fighting a situation as severe as the Covid-19 pandemic, the spread of fake news related to that can induce panic in society. This can lead to ineffectiveness in implementing control measures imposed by governments together with dangeorus proceedings resulting from erroneus beliefs of different treatments or vaccines functions, among others [2].

Thus, the foremost objective is focused on creating a tool that classifies social media statements as true or false through Natural Language Processing (NLP) techniques in order to avoid, to the extent possible, the spread of fake news and its possible consequences.

But, what is NLP? It is referred to the Branch of computer science (specifically, artificial intelligence) concerned with giving computers the ability to process, understand and extract information from a given text by combining computation linguistics-rule-based modeling of human language, statistical, machine learning and deep learning models [3].

In this work, in order to tackle the classification task, a Naive Bayes classifier was employed, which is a probabilistic machine learning model widely used for classification tasks. The crux of the classifier is based on the Bayes theorem, through which it is possible to find the probability of A happening, given that B has occurred; in other words, B is the evidence and A is the hypothesis. An important assumption is that the predictors/ are independent (presence of one particular feature does not affect the other), reason why it is called naive [4].

Experiments

All the computational work was performed using RStudio [5]. The specific packages and/or functions will be indicated for each procedure. The data the classification task was carried out with was extracted from the Kaggle repository. It contained a list of 10.201 COVID fake and true statements shared all over the internet, organized in two columns: one of them, the statements as such and the second, the outcome: a binary label where 0 means the headline is fake and 1 means that it is true.

First of all, the original data set [6] was briefly analized and a large unbalance was noticed because the vast majority of the class was "fake". In order to turn it into a balanced data set mantaining the initial one's characteristics as far as possible, the difference between both labels was calculated and, by sampling, the excess of fake statements was removed.

Then, before starting NLP itself: missing values, encoding and normalization basic checks were made through the utf8 package. Everything seemed to be in order; however, as it is usual when extracting text from social media, some non-alphanumeric characters were observed and removed. Moreover, in the same loop, text was lemmatized using the textstem package.

The next step was cleaning the data from a NLP point of view, which means creating the VCorpus (tm package) and performing some transformations such as removing punctuation, stopwords or stemming, among others. Once this was done, it was converted into a quanteda type corpus (quanteda package) and the label (true or fake) was added to the corpus for the further classification procedure. Before said analysis, clean data was displayed and easily visualized through a wordcloud plot.

Train and test subsets were built with dfm_subset function after sampling data and splitting it into 75% and 25%, for train and test, respectively.

Finally, a naive bayes model (quanteda.textmodels) was employed to carry out the fake news classification using a multinomial distribution. The quality of the performance was measured with several calculations such as accuracy, sensitivity or precision provided by the confusionMatrix function.

The code and all the materials required to run it is provided in the following GitHub repository:

https://github.com/elia-alonso-rueda/NLP

Results

Several estimated features scores of the most frequent words in both true and fake statements were provided by the NaiveBayes model, allowing to find which ones are the most frequent or more probable to find in both types of social media statements. It is, the probability of a social media statement being classified as true or false (hypothesis) given a specific word or feature (evidence). Note that these values may slighly change through the different runs.

Related to the model performance, accuracy, precisión or sensitivity reached values over 80%, which can be improved but are still good to tackle fake news classification. Note that it would be worthwhile to try other models such as svm or random forest and evaluate their performance as well to assess which one is the best.

Finally, in order to conclude this work, the wide usefulness of this study must be highlighted given the importance of denying fake news in social media, not only related to covid, but also to the rest of topics in the world.

Bibliography

[1]   S. J. R. da Silva and L. Pena, "Collapse of the public health system and the emergence of new variants during the second wave of the COVID-19 pandemic in Brazil," *One Health*, vol. 13, p. 100287, Dec. 2021, doi: 10.1016/J.ONEHLT.2021.100287.

[2]   A. Raj and M. P. Goswami, "Is fake news spreading more rapidly than COVID-19 in India? A representative study of people's perspective on controlling the spread of fake news on social media," *Journal of Content, Community and Communication*, vol. 10, no. 6, pp. 208–220, Jun. 2020, doi: 10.31620/JCCC.06.20/15.

[3]   "What is Natural Language Processing? | IBM." https://www.ibm.com/cloud/learn/natural-language-processing (accessed Jan. 30, 2022).

[4]   "Naive Bayes Classifier. What is a classifier? | by Rohith Gandhi | Towards Data Science." https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c (accessed Jan. 30, 2022).

[5]   RStudio, "RStudio | Open source & professional software for data science teams - RStudio." https://www.rstudio.com/ (accessed Jan. 08, 2022).

[6]   "COVID Fake News Dataset | Kaggle." https://www.kaggle.com/thesumitbanik/covid-fake-news-dataset (accessed Jan. 30, 2022).