ForceNLP at FakeDeS 2021: Analysis of Text Features Applied to Fake News Detection in Spanish

Jorge Reyes-Magaña[1,2][0000−0002−8296−1344] and Luis Enrique Argota Vega[1][0000−0003−2988−9563]

[1] Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, México
luiso91@comunidad.unam.mx
[2] Universidad Autónoma de Yucatán, Mérida, Yucatán, México
jorge.reyes@correo.uady.mx

Abstract. This paper presents our approach to the Task "Fake News Detection", which aims to decide if a news item is fake or real by analyzing its textual representation. The corpus consists of news compiled mainly from Mexican web sources: established newspaper websites, media companies websites, special websites dedicated to validating fake news, and websites designated by different journalists as sites that regularly publish fake news. Our approach is based mostly on different types of n-grams. For the task we use the classifiers: Logistic Regression, Support Vector Machines and Multinomial Naive-Bayes. Our approach achieved an average F1-score with respect to the other teams in the competition.

Keywords: Fake news · Machine learning · Text features

1 Introduction

The new era of spreading information is here, the transmission speed of all kinds of news is vertiginous. The use of social networks has encouraged and provoke not only to be more informed, yet misinforming about the reality of the world. Facebook presents the 50% of the total traffic to fake news sites and 20% total traffic to reputable websites (9). The impact of this kind of fake news is difficult to measure, some of the possible affected areas are economics, politics, security, health, among others. For this reason the detection of this kind of untrue statements turns to be essential in most of the automatics systems, in order to keep the facts veracity that will lead people to make decisions according to true facts.

Besides, in most cases, the fake information turns to be more striking, and when the users see this kind of news, they feel the duty of sharing because the information seems to be very important and should be passed on, provoking the fast-spreading and making in some cases, viral information. If we can contribute in some way to stop this kind of behavior from the beginning, the benefit will be for all. An example of the kind of damage that false information could cause is about the supposed effects of vaccines in general, which could influence people for instance not to take the COVID-19 vaccine, as we all know, the pandemic has paralyzed the world and even now is causing so much pain, affecting the daily life all over the world.

The system developed for filtering fake news is based on annotated corpora, the organizers provided us a set of truthful and fraudulent previously reviewed news (8), the testing corpus contained information associated with COVID-19, although the corpus used in the 2019 edition was given as a training set with other information topics. The 2021 task edition (4) has as purpose to measure the quality of the methods when the corpora have different topics during all the competition phases. Posing in this way, a more challenging competition. The FakeDeS is a task to be presented during the IBERLEF 2021 (6) (Iberian Languages Evaluation Forum) .

The rest of the paper is organized as follows: Section 2 presents some related work regarding fake news detection, the description of the corpora used during all the competition is described in Section 3, our methodology is presented in Section 4, containing some preliminary results using the corpus available in the development phase, guiding us to the improvement of each approach. The final results of all systems with the evaluation corpus is reported in Section 5. The paper ends with some conclusions in Section 6.

2   Related work

Shu et al.(11) give us a formal definition of fake news, as follows: **Fake news** is a news article that is intentionally and verifiably false. During the study, they focused on two principal branches referent to the features that characterize better fake information, the first is based on traditional news media and they claimed that this approach mainly relies on news content, while in social media, extra social context auxiliary information can be used to as additional information to help detect fake news.

Pérez-Rosas et al. (9) present two fake news datasets, the former is based on information of different domains via crowdsourcing, and the latter was gathered through the Web. Authors developed classification models with a linear SVM classifier and five-fold cross-validation. They combined a series of features, like lexical, syntactic, and semantic information, including some properties that represent text readability.

Additionally, *Reis et al.* (10) present a large study of the most important features to consider in fake news classification, they grouped into different elements that include, a) Textual features (syntax, lexical, psycholinguistic, semantic and

subjectivity), b) News Source Features (bias, credibility and trustworthiness and domain location), and c) Environment Features (engagement and temporal patterns). They found that the prediction performance of the features combined with existing classifiers like k-Nearest Neighbors, Naive Bayes, Random Forests, Support Vector Machine with RBF kernel, and XGBoost, have a useful degree of discriminative power for detecting fake news.

Besides, the research done by *Karimi et al.* (5) studied the degree of false news. They proposed a coherent and interpretable framework, that involves automated feature extraction, multi-source fusion and fakeness discrimination, showing that that their model can effectively distinguish different degrees of the fakeness of news.

3   Corpus

The training corpus consists of news compiled mainly from a diversity of Mexican web sources and covers the following 9 topics: Science, Sport, Economy, Education, Entertainment, Politics, Health, Security, and Society. The data were gathered from January to July of 2018. The principal sources used to collect the information were established newspaper websites, media companies websites, special websites dedicated to validating fake news, websites designated by different journalists as sites that regularly publish fake news. The corpus has 971 news, 480 were labeled as Fake and the remaining as True, all the news followed a manual labeling process:

- A news article is **true** if there is evidence that it has been published on reliable sites.
- A news article is **fake** if there is news from reliable sites or specialized websites in the detection of deceptive content that contradicts it or no other evidence was found about the news besides the source.

Organizers collected the true-fake news pair of an event so there is a correlation of news in the corpus.

The distributed corpus during the development phase contained the following information:

- **Topic**: Science/ Sport/ Economy/ Education/ Entertainment/ Politics, Health/ Security/ Society
- **Category**: Fake/ True
- **Source**: The name of the source media.
- **Headline**: The title of the news.
- **Text**: The complete text of the news.
- **Link**: The URL where the news was published.

For the systems evaluation, they provided a new testing corpus containing 572 elements, that were news related to COVID-19 and news from other Ibero-American countries. This variation in the testing corpus produces that the system should be prepared to dodge thematic and language variation. Besides, the test data only includes Id, Headline and Text columns.

4 Methodology

This section presents the process we employed to prepare texts for further classification. When we deal with text information having the idea of discovering knowledge, we face the problem about lack of structure. This absence is just apparent, the text itself presents a kind of structure but so much complex and hard to work computationally. Depending on the operations used in this stage of pre-processing, these will be the kind of patterns to discover in the collection. Before the feature extraction, we performed the pre-processing steps, described in 4.1, to improve the n-grams representation.

Additionally, there are several methods for increasing the characteristics of the system, in order to feed the classifier and have more elements to discriminate the data.

4.1 Pre-processing

- All texts were standardized to lowercase, avoiding the repetition of the same words.
- Stopwords were removed.
- We deleted the numbers that appear in text.
- We deleted punctuation, since it does not add any additional information when processing text data.
- The sequences of several blank spaces, tabs and line breaks were standardized to a single blank space.

Due to the differences in both corpora, development and testing, we decide to apply the pre-processing only in the main text of the news.

4.2 Features

We took into account several n-grams features for the representation of texts:

- **Character.**
- **Word.**
- **POS tags.** Are sequences of continuous part-of-speech (POS) tags. They capture syntactic information and are useful. For this feature we used the Spacy tagger.
- **Skipgrams.** We capture groups of 2 words with skips of 1 to 3 words.
- **Function words.** The frequency of this words is one of the best characteristics to detect hate speech and aggressiveness (1), so in this case we want to see if this can help us to discriminate fake news. We built function words n-grams from 2 to 4 tokens using the spanish stopwords list from NLTK (3).
- **Punctuation symbols.** With this feature we want to tackle the coherence and cohesion to the written text. Prior to the corpus pre-processing, we built n-grams of 2 to 4 punctuation symbols.

We use two variations of features as seen in Table 1. The columns associated with the approaches represent the lengths of n-grams that were applied using all the features when the tested classifiers were executed; meaning that approach-1 contained 17 features and approach-2, 15. We select the feature combinations due to the performance showed during the phases.

Table 1. Features applied to the models.

Features (n-grams)	Approach **1**	Approach **2**
Characters	[3,4,5]	[3,4,5]
Words	[2,3,4]	[2,3]
Skipgrams	[2,3]	[1,2]
PosTags	[2,3,4]	[2,3]
stop words	[2,3,4]	[2,3,4]
Punctuation symbols	[2,3,4]	[2,3,4]

4.3 Classifier

We used three different well -known classifiers, all of them described in (2; 7). The selected models are: Multinomial Naive Bayes (MNB), Logistic Regression (LR), and Support Vector Machines (SVM). We also select CountVectorizer with a threshold of 3. We tested all the models during the competition phases in conjunction with the approaches described in Table 1.

5 Results

During the development phase the best result we got is using the Multinomial Naive Bayes classifier, having an F1-score of 0.7576, this model let us rank in position 7 of this phase. The feature approach used in this case was number 1. The models using Logistic Regression and Support Vector Machines applied on both approaches didn't overcome the results obtained with the Multinomial classifier.

On the contrary, during the evaluation phase, we had different results, our best model turns to be the Logistic Regression applied with the feature approach 1 and the worst was Multinomial Naive Bayes, as seen in Table 2. The F1-score of our best model ranked us in number 8 of the competition. We weren't able to try more model combinations with the approaches, due to the rules of maximum submissions during this phase of the competition.

6 Conclusions

Fake news detection is still an ongoing challenge to be resolved, the wide range of information that can be changed to produce false statements increase the difficulty to have a forthright solution. Hence, the importance of this type of task

Table 2. F1-score in evaluation phase.

Model	Approach	F1-score
LR	1	**0.6925**
SVM	1	0.6722
LR	2	0.6921
MNB	1	0.4928

that will let us see, understand, and improve the methodologies developed all over the world. According to the official results, we can see that our approach stated below the baseline of a SVM featured with character trigrams, having an F1-score of 0.7062. We consider, that the complexity added to our model as having different features didn't worth it, the gain, in this case, is very low. We also believe that the unbalanced corpora regarding different columns in the development phase, provided good means of discriminant information during the classifier training, as well as containing only news in Mexican Spanish. Unfortunately, we didn't have the same elements in the evaluation. We believe this causes an impact on the general results.

Bibliography

[1] Argota Vega, L.E., Reyes-Magaña, J.C., Gómez-Adorno, H., Bel-Enguix, G.: Mineriaunam at semeval-2019 task 5: Detecting hate speech in twitter using multiple features in a combinatorial framework. In: Proceedings of the 13th International Workshop on Semantic Evaluation. pp. 447–452 (2019)

[2] Aurelien, G.: Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. OReilly (2019)

[3] Bird, S.: Nltk: the natural language toolkit. In: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. pp. 69–72 (2006)

[4] Gómez-Adorno, H., Posadas-Durán, J.P., Bel-Enguix, G., Porto, C.: Overview of fakedes task at iberlef 2021: Fake news detection in spanish. Procesamiento del Lenguaje Natural **67**(0) (2021)

[5] Karimi, H., Roy, P., Saba-Sadiya, S., Tang, J.: Multi-source multi-class fake news detection. In: Proceedings of the 27th international conference on computational linguistics. pp. 1546–1557 (2018)

[6] Montes, M., Rosso, P., Gonzalo, J., Aragón, E., Agerri, R., Álvarez Carmona, M., Álvarez Mellado, E., Carrillo-de Albornoz, J., Chiruzzo, L., Freitas, L., Gómez-Adorno, H., Gutiérrez, Y., Jiménez Zafra, S.M., Lima, S., Plaza-de Arco, F.M., Taulé, M.: Ceur workshop proceedings, 2021. In: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) (2021)

[7] Müller, A.C., Guido, S.: Introduction to machine learning with Python: a guide for data scientists. OReilly Media (2018)

[8] Posadas-Durán, J.P., Gómez-Adorno, H., Sidorov, G., Escobar, J.J.M.: Detection of fake news in a new corpus for the spanish language. Journal of Intelligent & Fuzzy Systems **36**(5), 4869–4876 (2019)

[9] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R.: Automatic detection of fake news (2017)

[10] Reis, J.C., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Supervised learning for fake news detection. IEEE Intelligent Systems **34**(2), 76–81 (2019)

[11] Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter **19**(1), 22–36 (2017)