# Identifying Fake News Spreaders in Social Media

## Notebook for PAN at CLEF 2020

Nikhil Pinnaparaju, Vijaysaradhi Indurthi, and Vasudeva Varma

IIIT, Hyderabad
{nikhil.pinnaparaju, vijaya.saradhi}@research.iiit.ac.in,
vv@iiit.ac.in

**Abstract** With the rise of social networking platforms, everyone now has free access to information from around the work. Anyone from anywhere can now share context with the entire world. This allows for more connectivity around the world and more transparency. However, this also allows for the spread of misinformation and fake news often resulting in undesired and extremely impactful political, economic, social, psychological and criminal consequences. Identifying the fake news spreaders is as important as identifying the fake news itself. We put forward a method to utilize content analysis and more user modelling to capture who is more likely to share fake news. We use TF-IDF as our text transformation method coupled with algorithms simple classification algorithm Logistic Regression and achieve an accuracy of 71.5% and 70% in identifying fake news spreaders in both the English as well as Spanish test set respectively.



## 1 Introduction

Recently we have seen the rise of many social platforms like Facebook, Twitter, Reddit, Snapchat and so many more. These platforms serve as great ways for any and everyone to share content, information and so much more. With this power, comes with bad actors that misuse it to spread disinformation, fake news and rumors. It is important that we identify these bad actors and are able to contain the impact they make on the platform. The task proposed by Rangel et al. [10] allows us to detect these bad actors in both English and Spanish.

For this task we experiment with various machine learning techniques and compare their performance on the task. We use models like Logistic Regression[7], Random Forest[1], Support Vector Machines[5] and XGBoost[4] because of their smaller size in terms of the number of parameters and show they perform well. Another reason for utilizing simpler model architectures is due to the amount of data we have accessible and how data-hungry deep neural architectures can get. All submissions are made through the Tira system[9].

## 2 Background

Traditional methods of fake news detection rely primarily on two techniques either

- Content Based
- User Based

In content based techniques, models are used to try to capture whether a piece of text is fake or not.[3][8] Most work tries to detect fake news based on linguistic features from the text or otherwise capture the style of the text.

The only method direction of work is user based detection, in which they try to assign a credibility to users and detect based on that.[6]

The differentiating aspect of this work is that we are trying to identify fake news spreaders based on the content they share and not using features like follower count, tweet count, etc.

## 3 Identifying Fake News Spreaders

In this task of identifying fake news spreaders, 300 author's tweets have been provided for English and Spanish respectively. For each author 100 tweets are available. The task is to build computational models to identify whether a given author is a fake news spreader or not. The official metric of evaluation is the combined accuracy of both the languages.

## 4 System overview

We chose to participate in both the language tracks, English and Spanish. We formulate the problem of identifying fake news spreaders by treating it as a document classification problem. We concatenate all the tweets of a given author and consider it as a single big document corresponding to the author. With this approach, each author is represented by the collection of all his tweets concatenated together.

Empirical observations showed that it is the terms of the tweets which are significant in identifying if the author is a spreader of fake news or other wise. Since the presence of specific terms is key to this task, we use a very simple transformation - TF-IDF algorithm to transform the training data into numeric vector representations for training as TF-IDF is sequence invariant i.e the sequence of the terms do not matter. We could have used some recent embedding models like Word2vec or GloVe but as the document size is large and consists of around 100 tweets, the average embedding technique dilutes the word embeddings and the resulting transformation would not hold the semantic representation of all the tweets of that author. Hence we did not delve in word embeddings.

The following pre-processing is done before the training data is transformed with TF-IDF. For each tweet, we remove all the occurrences of retweets ('RT'), mentions of user ('#user#'), mentions of hashtags ('#hashtag#') and mentions of urls ('#url#'). In addition all the text is lowercased.

The transformed representations are then fed into a simple classification algorithm like Logistic Regression. The advantage with the logistic regression is that the resulting model can be interpreted.

## 5 Experimental setup

In this shared task, the training dataset consisted of tweets tweeted by 300 authors. For each author, 100 tweets tweeted by him are available for training. In our experimental setup, we used 5-fold cross validation. For each fold, we trained on the 80% of the authors and evaluated on the remaining 20% of the authors. We keep the experimental same for both the languages.

We use sklearn [2] for all our experiments. We experiment with four classification algorithms - Logistic Regression, Random Forest, SVM and XGBoost. We also use the default hyper parameters provided by the sklearn as we didn't want to overfit to the training dataset. First, we show the 5-fold cross validation performance of these algorithms. Then, we pick the best performing algorithm and train the model again, this time utilising the full training data available and use this model to make predictions on the task's test set which is not publicly available.

## 6 Cross Validation Results

| **Algorithm** | **Accuracy** |
|---|---|
| Logistic Regression | 0.7209 |
| RandomForest | 0.7000 |
| SVM | 0.7330 |
| XGBoost | 0.7000 |

**Table 1.** Cross Validation for Fake news spreaders task for English language using TF-IDF and Logistic Regression, SVM and XGBoost

| **Algorithm** | **Accuracy** |
|---|---|
| Logistic Regression | 0.6866 |
| RandomForest | 0.7230 |
| SVM | 0.7133 |
| XGBoost | 0.6900 |

**Table 2.** Cross Validation for Fake news spreaders task for Spanish language using TF-IDF and Logistic Regression, SVM and XGBoost

| Language | Algorithm | Accuracy |
|---|---|---|
| English | Logistic Regression | 0.7150 |
| Spanish | Logistic Regression | 0.7000 |

**Table 3.** Test scores for Fake news spreaders task for English and Spanish languages using TF-IDF and Logistic Regression

Table 1 and Table 2 show the cross validation scores for both the languages using different classification methods like logistic regression, Random Forest, SVM and XGBoost. These scores are the mean of the 5-fold cross validation scores obtained.

For English language, We see that TF-IDF with SVM has obtained better accuracy than every other method. SVM performed slightly better than logistic regression. XGBoost and Random Forest obtain the same accuracy.

For Spanish language, Random Forest obtained the best accuracy and Logistic Regression the least.

For uniformity and simplicity, we chose to submit the model which uses TF-IDF with Logistic Regression for the final run on the unknown test set.

Table 3 shows the performance of the model on the final unseen test set. We see that our model has obtained an accuracy of 0.7150 and 0.7000 on English and Spanish languages respectively. Since these accuracies are similar to the accuracies we obtained using 5-fold cross validation, we infer that the model is able to generalise to unseen test set and not overfit the training data.

## 7 Conclusion

To conclude, we describe the methods we applied for the task. We show the processing steps involved along with the results achieved by each of the models. Future work would be along attempting to apply and use deep learning and state of the art methods and see their performance on this task.

## References


1. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (Oct 2001). https://doi.org/10.1023/A:1010933404324, https://doi.org/10.1023/A:1010933404324
2. Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., VanderPlas, J., Joly, A., Holt, B., Varoquaux, G.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop: Languages for Data Mining and Machine Learning. pp. 108–122 (2013)
3. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th international conference on World wide web. pp. 675–684 (2011)
4. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. pp. 785–794 (2016)
5. Cortes, C., Vapnik, V.: Support-vector networks. Machine learning **20**(3), 273–297 (1995)

6. Gupta, M., Zhao, P., Han, J.: Evaluating event credibility on twitter. In: Proceedings of the 2012 SIAM International Conference on Data Mining. pp. 153–164. SIAM (2012)
7. Pearl, R., Reed, L.J.: On the rate of growth of the population of the united states since 1790 and its mathematical representation. Proceedings of the National Academy of Sciences of the United States of America **6**(6), 275 (1920)
8. Popat, K., Mukherjee, S., Strötgen, J., Weikum, G.: Credibility assessment of textual claims on the web. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. pp. 2173–2178. ACM (2016)
9. Potthast, M., Gollub, T., Wiegmann, M., Stein, B.: TIRA Integrated Research Architecture. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World. Springer (Sep 2019)
10. Rangel, F., Giachanou, A., Ghanem, B., Rosso, P.: Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. In: Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR-WS.org (Sep 2020)