# Profiling Fake News Spreaders on Twitter
## Notebook for PAN at CLEF 2020

Usman Saeed, Hammad Fahim, and Dr. Farid Shirazi

Data Science Lab (DSL), Ryerson University, Canada
Center for Computing Research (CIC), Instituto Politécnico Nacional (IPN)
usman.saeed, f2shiraz@ryerson.ca, hammad.fahim57@gmail.com

**Abstract** The article presents a model for fake news profiling task on social media data. Fake news poses a great threat to our society and evaluating author plays a critical role in detecting fake news patterns. The article describes machine learning and deep learning algorithm analyses to the binary classification problem for PAN 2020 challenge. All experiments were conducted on the English data set and the results for discriminating fake news spreaders from real news authors were shown. Our final model submitted to TIRA, Bi-LSTM with attention on the training set achieved 70% accuracy.

## 1 Introduction

Due to the growth of social media and technological advancement people tend to spend more time on smart devices and are more inclined to get news and updates through social media, that makes people more exposed to fake news and wide scale misinformation. Fake news has been spreading in the form of reviews[13], advertisements[22], political agendas, news articles, rumours and satires[4] through both social[11] and mainstream electronic mediums. Its extensive use for misleading information, false persuasion and confusion makes it a major threat for public trust on online activities i.e. social community activities, online shopping and positive media reinforcements. Due to the dynamic and heterogeneous nature of the task natural language processing researchers have contributed to multiple solutions to the problem. Variety, Velocity, Volume and Time Latency of the fake news articles are the four fundamental problems [25] encountered by the current academic researchers.

It has been established through previous researches that linguistic-based features from the news are insufficient[25], meanwhile, auxiliary features such as past credibility of the author and spread pattern play a vital role in the detection of fake news. We participated in the PAN 2020 author profiling task[19][17], the challenge was to determine if the authors from Twitter are keen to spread fake news. The best obtained results of the English data set is provided, where several machine learning techniques were applied(Logistic Regression, SVM, Decision tree, Multi-Layer Perceptron, KNN) and deep learning(LSTM, Bi-LSTM with and without attention) to determine the best possible results on the validation set. Best model was submitted (Bi-LSTM with attention) where 70% accuracy was obtained on the test set.

In the following section, the existing work in the research community is described. In section 3, corpora provided by the PAN organizers and task description are presented. In section 4, details of our purposed approach and experiments to evaluate the system are described. In section 5, results and analysis are stated. Section 6 concludes the paper.

## 2 Related Work

Fake news on social media can either be created by bots[12] or real humans. Many bots are created solely to spread misinformation, rumors, spam and can be easily confused with human behaviour. One of the key aspects used as an indicator is social context[25] and distribution pattern of real and fake news. The online users work as a social community to dominate the speed of fake news. Broadly speaking fake news have been studied based on multiple theoretical perspectives including style of the content [26], propagation of fake news [2] and role , engagement and attributes of user in creating fake news [15] [1] [7] [14]. We also see that language variety and cultural idiosyncrasies[18] can influence tasks like author profiling in establishing discourse. Some of the approaches[25] to tackle these issues have been fact checking with the aid of experts, machine learning algorithm, information comparisons etc. Unique emotional patterns[9] and signals[10] between fake news and real news is also studies as manipulation requires emotional language cues[8]. Researcher have used supervised techniques[23][5] like Support Vector Machine(SVM) and Decision Trees repeatedly to detect deception and fraud in text. Supervised deep learning models[21][6][20] like Gated Recurrent Unit(GRU), Bidirectional LSTM and Recurrent Neural networks(RNN) have also shown substantial results by having the ability to capture contextual information in news. It has also been discussed that instead of emphasising on claims, the news sources[25] can provide more valuable insights on fake news. The context between creators analysis and content analyses of the news give rise to the authorship profiling task of fake news.

## 3 Dataset

PAN-2020 provided 30,000 labeled tweets of English language to train and develop the systems. Training corpus consisted of 24,000 labeled tweets, and 6,000 labeled tweets for the development phase (according to the PAN's suggested split of 70 percent for training and 20 percent for testing the models). Evaluation test set had 200 files with 100 tweets per file. Different annotators manually labeled the corpora. More details can be found in overview papers.

## 4 Methodology

We conducted experiments using several machine learning and deep learning methods. For all machine learning methods(Logistic Regression, Support Vector Machine, Decision Tree, Multilayer Perceptron and K-Nearest Neighbor), we used CountVectorizer and Tf-Idf transformer to create feature input vectors. In our deep learning methods, we used 50D Twitter GloVe embeddings[16]. Below, details and features of our final deep learning model are mentioned that was submitted in the PAN 2020 task.

### 4.1 Pre-Processing

As mentioned in the literature review, fake news needs to preserve stylistic features in the tweet to be able to record information that can add value in evaluation. Hence, we used Ekphrasis [3] for the converting all the stylistic information into unique tags before passing it into our deep learning model. We converted number, URL, email address, currency, username, time, date, hashtag, elongated text, all capital text and repeated and emphasised texts special tags to preserve meaningful information.

### 4.2 Setup and Evaluation

Our final model with highest accuracy was Bi-LSTM with attention [24]. The Bi-LSTM approach handles the local context from both end to beginning and beginning to end, and attention puts more focus to the information directed by the hidden layer of Bi-LSTM. The model was trained till 8 epochs, the hidden layers were set to size 50, dropout was set to 0.2 and "AdaGradTrainer" was selected. The evaluation score was obtained as described in the task by computing individual accuracy of the users for binary classification. The individual accuracy of all users were than averaged to achieve the final score.

## 5 Results

We present the results on evaluation set for all the used algorithms. Table 1 shows that the highest accuracy was 79.7% which was achieved by B-LSTM with attention. Highest accuracy among the machine learning algorithm was 78.2% through logistic regression which is very close to our best model. Both decision tree and logistic regression achieved better scores than LSTM and Bi-LSTM respectively.

| **Model** | **Accuracy (Validation set)** |
|---|---|
| Logistic Regression | 78.2 |
| SVM | 72.2 |
| Decision Tree | 76.7 |
| Multilayer perceptron | 68.0 |
| KNN | 74.7 |
| LSTM | 74.2 |
| Bi-LSTM | 76.1 |
| Bi-LSTM with Self-Attention | 79.7 |

**Table 1.** Accuracy achieved on validation set

## 6 Conclusion

In this paper, we analysed multiple machine learning and deep learning algorithms to obtain the highest accuracy for detecting fake news patterns among authors. For our final model, we first preserved the stylistic information in the tweets through Ekphrasis tagging of features, then we created 50D GloVe embeddings and trained it on Bi-LSTM with attention. In the end evaluation testing set on TIRA showed 70% accuracy on our highest achieving model. The evaluation phase showed many machine learning algorithms worked well when trained with Tf-Idf and count vectorizers. In future, we would like to experiment with transformer methods and more diverse features.

## References


1. Ashforth, B.E., Mael, F.: Social identity theory and the organization. Academy of management review **14**(1), 20–39 (1989)
2. Bálint, P., Bálint, G.: The semmelweis-reflex. Orvosi hetilap **150**(30), 1430–1430 (2009)
3. Baziotis, C., Pelekis, N., Doulkeridis, C.: Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 747–754. Association for Computational Linguistics, Vancouver, Canada (August 2017)
4. Berkowitz, D., Schwartz, D.A.: Miley, cnn and the onion: When fake news becomes realer than real. Journalism Practice **10**(1), 1–17 (2016)
5. Biyani, P., Tsioutsiouliklis, K., Blackmer, J.: " 8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. In: Thirtieth AAAI Conference on Artificial Intelligence (2016)
6. Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)
7. Deutsch, M., Gerard, H.B.: A study of normative and informational social influences upon individual judgment. The journal of abnormal and social psychology **51**(3), 629 (1955)
8. Ghanem, B., Rosso, P., Rangel, F.: An Emotional Analysis of False Information in Social Media and News Articles. ACM Transactions on Internet Technology (TOIT) **20**(2), 1–18 (2020)
9. Ghanem, B., Rosso, P., Rangel, F.: An emotional analysis of false information in social media and news articles. ACM Trans. Internet Technol. **20**(2) (Apr 2020)
10. Giachanou, A., Rosso, P., Crestani, F.: Leveraging emotional signals for credibility detection. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 877–880. SIGIR'19, Association for Computing Machinery, New York, NY, USA (2019)
11. Guess, A., Nagler, J., Tucker, J.: Less than you think: Prevalence and predictors of fake news dissemination on facebook. Science Advances **5**(1) (2019)
12. Hall, A., Terveen, L., Halfaker, A.: Bot detection in wikidata using behavioral and other informal cues. Proc. ACM Hum.-Comput. Interact. **2**(CSCW) (Nov 2018)
13. Li, J., Ott, M., Cardie, C., Hovy, E.: Towards a general rule for identifying deceptive opinion spam. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1566–1576 (2014)
14. Loewenstein, G.: The psychology of curiosity: A review and reinterpretation. Psychological bulletin **116**(1), 75 (1994)
15. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology **2**(2), 175–220 (1998)

16. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543 (2014)
17. 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)
18. Rangel, F., Franco-Salvador, M., Rosso, P.: A Low Dimensionality Representation for Language Variety Identification. In: International Conference on Intelligent Text Processing and Computational Linguistics. pp. 156–169. Springer (2016)
19. 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)
20. Roy, A., Basak, K., Ekbal, A., Bhattacharyya, P.: A deep ensemble framework for fake news detection and classification. arXiv preprint arXiv:1811.04670 (2018)
21. Socher, R., Huval, B., Manning, C.D., Ng, A.Y.: Semantic compositionality through recursive matrix-vector spaces. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. pp. 1201–1211 (2012)
22. Willmore, A.: This analysis shows how viral fake election news stories outperformed real news on facebook (2016)
23. Wu, K., Yang, S., Zhu, K.Q.: False rumors detection on sina weibo by propagation structures. In: 2015 IEEE 31st international conference on data engineering. pp. 651–662. IEEE (2015)
24. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). pp. 207–212. Association for Computational Linguistics, Berlin, Germany (Aug 2016)
25. Zhou, X., Zafarani, R.: Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 (2018)
26. Zuckerman, M., DePaulo, B.M., Rosenthal, R.: Verbal and nonverbal communication of deception. In: Advances in experimental social psychology, vol. 14, pp. 1–59. Elsevier (1981)