Fake News Detection

Pooja Chadha, P Aishwarya Naidu, Satvik Vats

Department of Information Science and Engineering, BMS College of Engineering, Bangalore

ABSTRACT

The extensive spread of fake news can have a serious negative impact on individuals and society. It has brought down the authenticity of news ecosystem as it is even more widely spread on social media than most popular authentic news. Fake news has become such a big problem because it possesses the ability to influence and eventually change opinions of the readers and viewers and also impacts the way in which people respond to an authenticated news article. It also has political influence, can encourage mistrust in legitimate media outlet, influence financial markets and damage an individual's reputation.

We aim to develop a model using machine learning and Natural Language Processing (NLP) techniques to determine whether a news is fake or real. This report examines existing 'Fake News' detection systems and highlights the drawbacks of these systems, particularly in relation to the use of classification algorithm that should be chosen and how they all should be combined to get the maximum accuracy. Furthermore, proposed implementations of 'Fake News' detection system using machine learning technology has been researched in order to identify how Term Frequency-Inverse Document Frequency (TF-IDF), Parts of Speech (PoS) and semantic analysis can be combined to attain maximum accuracy for the classification model.

OBJECTIVES

The major objectives of the project are threefold, as follows:

- Using linguistic cues to develop a machine learning based model for accurately determining authenticity of the given news.
- Demonstrate successful use of classification algorithms to classify the feature vector generated from dataset into real or fake.
- To get high accuracy to determine whether a news is fake or true.

INTRODUCTION

The issue of how to tackle the disruption done by 'Fake News' has now become a very crucial issue for both the governments around the world and the big technology companies. , we can see the emergence of the new sources of news for the people and certainly all these sources are being operated by different people, which makes the system more vulnerable to spread of unverified news.

The onset or beginning of social media enabled people to control the propagation of information for the first time, as now the news came out of the less interactive news rooms on their television sets. We cannot leave the responsibility to tackle the problem of 'Fake News' on people only. Thus, a need of a modern-day tool to help people in deciding authenticity of the information they are sharing is felt, something this project aims at doing.

PROPOSED SYSTEM/MODEL/APPLICATION
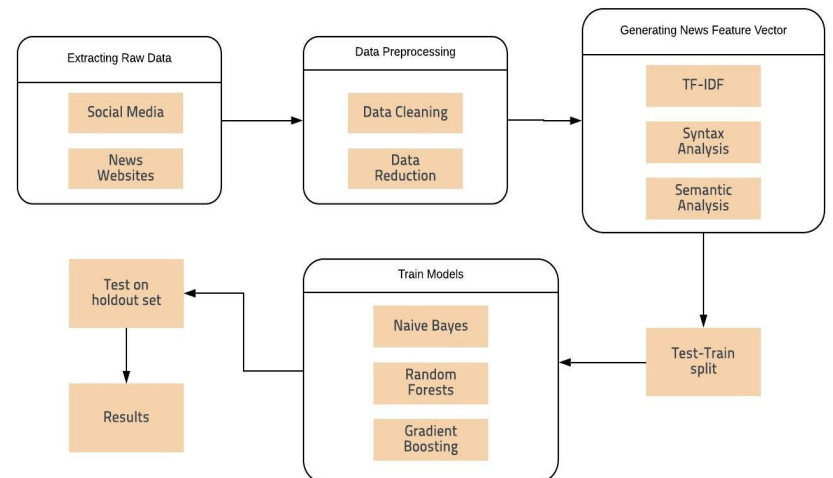
CONCLUSIONS

The developed model demonstrated the different classification algorithms namely, Naive Bayes Classifier, Random Forest Classifier and the Gradient Boost classification algorithm. How with same algorithms used in combination with different linguistic cues acquired different accuracy rates, tells as why using different algorithms for developing the model was necessary. The use of three different classifiers with each one having a different classifying mechanism, allowed to develop a scalable, verifiable and transparent model. In order to get the highest possible accuracy in the presented classification model the three different linguistic cues used with three different classification algorithms were combined and that were used only after giving them appropriate weightage points. This enabled development of a model that is not biased or misguided on the basis of any one of the traits that are being examined.

REFERENCES

[1] Pérez-Rosas, Verónica, et al. "Automatic detection of fake news." arXiv preprint arXiv:1708.07104 (2017).

[2] Gilda, Shlok. "Evaluating machine learning algorithms for fake news detection." 2017 IEEE 15th Student Conference on Research and Development (SCOReD). IEEE, 2017.

[3] Shu, Kai, et al. "Fake news detection on social media: A data mining perspective." ACM SIGKDD Explorations Newsletter 19.1 (2017): 22-36.

[4] Feng, Song, Ritwik Banerjee, and Yejin Choi. "Syntactic stylometry for deception detection." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012.