Fake News Detection

Dr. Deepika Ajalkar , Tushar Nimbalkar, Krushnakant Hangarge, Sandesh Nigade, Ninad Jagtap

INOFORMATION TECHNOLOGY

G H RAISONI COLLEGE OF ENGINEERING AND MANAGEMENT WAGHOLI, PUNE

Abstract :

Politics suffered from a great set back due to fake news. Fake news is intentionally written to mislead the audience to believe the false propaganda, which makes it difficult to detect based on news content. The fake news has disturbed the mindset of the common people. Due to this widespread of the fake news online it is the need of the hour to check the authenticity of the news. The spread of fake news has the potential for extremely negative impact on society. The proposed approach is to use machine learning to detect fake news. Using vectorisation of the news title and then analysing the tokens of words with our dataset. The dataset we are using is a predefined curated list of news with their property of being a fake news or not. Our goal is to develop a model that classifies a given article as either true or fake.

IndexTerms :

Fake News, Self Learning, Pattern Matching, Response Generation, Artificial Intelligence, Natural Language Processing, Context Free Grammar, Term Frequency Inverse Document Frequency, Stochastic Gradient Decent, Word2Vec.

I. INTRODUCTION

Fake news definition is made of two parts: authenticity and intent. Authenticity means that fake news content false information that can be verified as such, which means that conspiracy theory is not included in fake news as there are difficult to be proven true or false in most cases. The second part, intent, means that the false information has been written with the goal of misleading the reader.
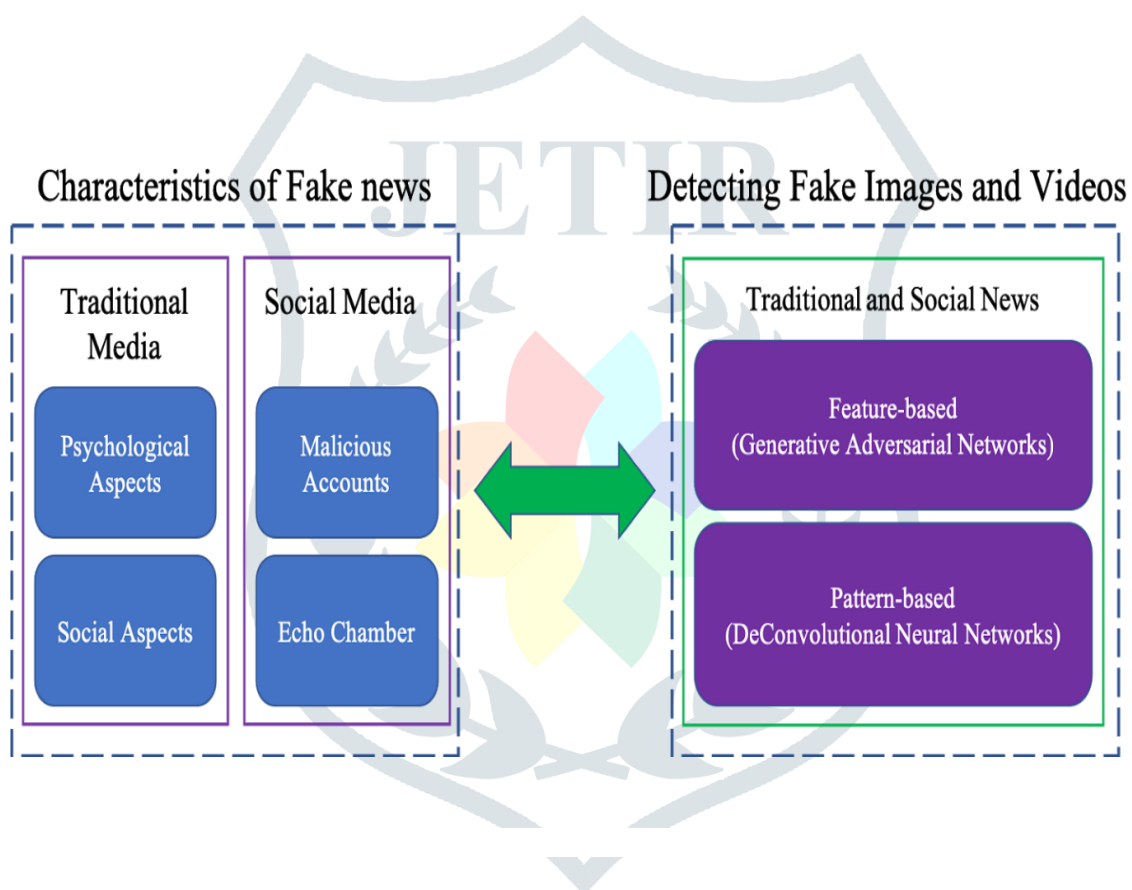
METHODOLOGY

Proposed Framework

I am expanding on the current literature by introducing ensemble techniques with various linguistic feature sets to classify news articles from multiple domains as true or fake. The ensemble techniques along with Linguistic Inquiry and Word Count (LIWC) feature set used in this research are the novelty of our proposed approach.
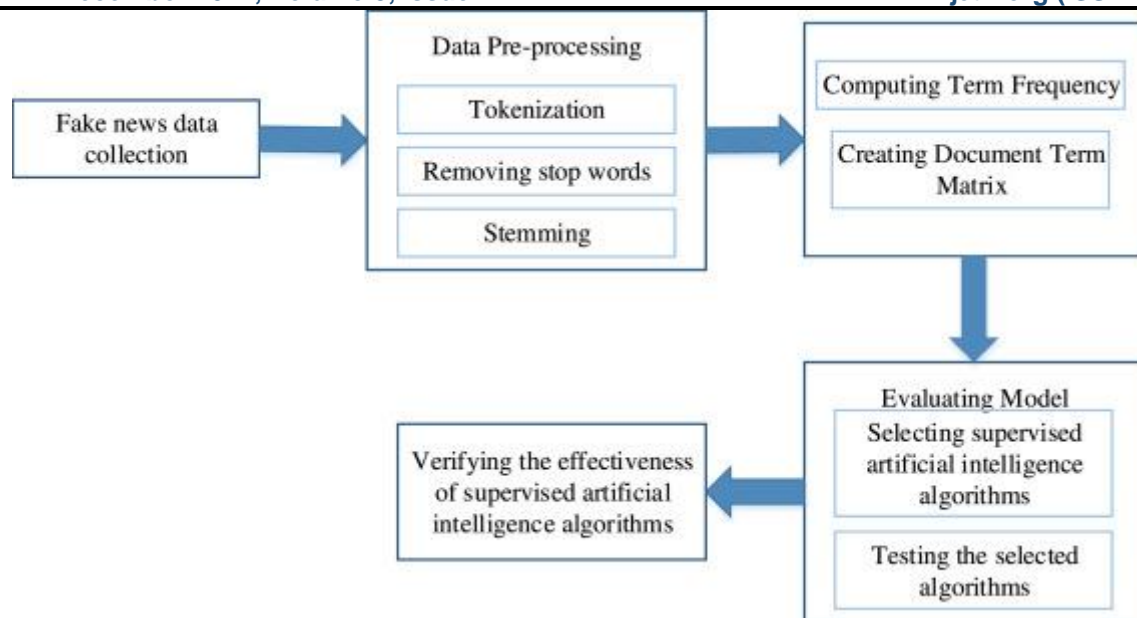
There are numerous reputed websites that post legitimate news contents which are used for fact checking. In addition, there are open repositories which are maintained by researchers to keep an up-to-date list of currently available datasets and hyperlinks to potential fact checking sites that may help in countering false news spread. However, we selected three datasets for our experiments which contain news from multiple domains (such as politics, entertainment, technology, and sports) and contain a mix of both truthful and fake articles, and merged the three datasets into large dataset. The datasets are available online and are extracted from Kaggle.

Tools And Techniques

Softwares:

1. Python
2. Jupyter Notebook

Technologies:

1. Machine Learning
2. Python Libraries

Algorithms

1. 1.Naïve Bayes
2. Logistic Regression
3. Support Vector Machine (SVM)
4. Natural Language Processing

Advantages

1. Low Cost
2. Easy Access
3. Rapid Dissemination

Acknowledgement

Conclusion

Finally, this application is only one that would be necessary in a larger toolbox that could function as a highly accurate fake news classifier. Other tools that would need to be built may include a fact detector and a stance detector. In order to combine all of these "routines," there would need to be some type of model that combines all of the tools and learns how to weight each of them in its final decision.

Result
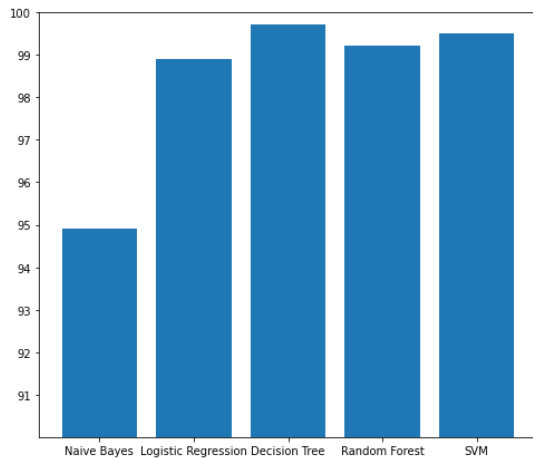
CLASSIFIER	ACCURACY
Naïve Bayes	94.91%
Support Vector Machine	99.52%
Random Forest	99.22%
Logistic Regression	98.91%
Decision Tree	99.91%

References

[1] M. Risdal. (2016, Nov) Getting real about fake news. [Online]. Available: https://www.kaggle.com/

mrisdal/fake-news

[2] J. Soll, T. Rosenstiel, A. D. Miller, R. Sokolsky, and J. Shafer. (2016, Dec) The long and brutal history of fake news. [Online]. Available: https://www.politico.com/magazine/story/2016/12/

fake-news-history-long-violent-214535

[3] C. Wardle. (2017, May) Fake news. it's complicated. [Online]. Available: https://firstdraftnews.com/

fake-news-complicated/

[4] T. Ahmad, H. Akhtar, A. Chopra, and M. Waris Akhtar, "Satire detection from web documents using

machine learning methods," pp. 102–105, 09 2014.

[5] C. Kang and A. Goldman. (2016, Dec) In washington pizzeria attack, fake news brought real guns. [Online]. Available: https://www.nytimes.com/2016/12/05/business/media/

comet-ping-pong-pizza-shooting-fake-news-consequences.html

[6] C. Domonoske. (2016, Nov) Students have 'dismaying' inability to tell fake news from real,

study finds. [Online]. Available: https://www.npr.org/sections/thetwo-way/2016/11/23/503129818/

study-finds-students-have-dismaying-inability-to-tell-fake-news-from-real