

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 6, June 2019

## Fake News Detection Using Machine Learning

Muhammed Murshid.T<sup>1</sup>, Nikhil P P<sup>2</sup>, Ranjith.EP<sup>3</sup>, Jerrin Joe Francis<sup>4</sup>

Dept. of Computer Science and Engineering, Eranad Knowledge City Technical Campus, Campus

Malappuram, India<sup>1,2,3</sup>

Asst. Professor, Dept. of Computer Science and Engineering, Eranad Knowledge City Technical Malappuram, India<sup>4</sup>

**ABSTRACT:** Consuming news from social media is becoming increasingly popular nowadays. Social media brings benefits to users due to the inherent nature of fast dissemination, cheap cost, and easy access. However, the quality of news is considered lower than traditional news outlets, resulting in large amounts of fake news. Detecting fake news becomes very important and is attracting increasing attention due to the detrimental effects on individuals and the society. The performance of detecting fake news only from content is generally not satisfactory, and it is suggested to incorporate user social engagements as auxiliary information to improve fake news detection. Thus it necessitates an in- depth understanding of the correlation between user profiles on social media and fake news. In this paper, we construct real-world datasets measuring users trust level on fake news and select representative groups of both “experienced” users who are able to recognize fake news items as false and “naive” users who are more likely to believe fake news. We perform a comparative analysis over explicit and implicit profile features between these user groups, which reveals their potential to differentiate fake news.

### I. INTRODUCTION

These days’ fake news is creating different issues from sarcastic articles to a fabricated news and plan government propaganda in some outlets. Fake news and lack of trust in the media are growing problems with huge ramifications in our society. Obviously, a purposely misleading story is “fake news “ but lately blathering social media’s discourse is changing its definition. Some of them now use the term to dismiss the facts counter to their preferred viewpoints.

The importance of disinformation within American political discourse was the subject of weighty attention, particularly following the American president election. The term 'fake news' became common parlance for the issue, particularly to describe factually incorrect and misleading articles published mostly for the purpose of making money through page views. In this paper, we are trying to produce a model that can accurately predict the likelihood that a given article is fake news.

Facebook has been at the epicenter of much critique following media attention. They have already implemented a feature to flag fake news on the site when a user sees it; they have also said publicly they are working on to distinguish these articles in an automated way. Certainly, it is not an easy task. A given algorithm must be politically unbiased – since fake news exists on both ends of the spectrum – and also give equal balance to legitimate news sources on either end of the spectrum. In addition, the question of legitimacy is a difficult one. However, in order to solve this problem, it is necessary to have an understanding on what Fake News is. Later, it is needed to look into how the techniques in the fields of machine learning, natural language processing help us to detect fakenews.

### II. RELATED WORK

There exists a large body of research on the topic of machine learning methods for deception detection, most of it has been focusing on classifying online reviews and publicly available social media posts. Particularly since late 2016 during the American Presidential election, the question of determining 'fake news' has also been the subject of particular attention within the literature. Conroy, Rubin, and Chen [1] outlines several approaches that seem promising towards the aim of perfectly classify the misleading articles. They note that simple content-related n- grams and shallow parts-of-speech (POS) tagging have proven insufficient for the classification task, often failing to account

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 6, June 2019

for important context information. Rather, these methods have been shown useful only in tandem with more complex methods of analysis. Deep Syntax analysis using Probabilistic Context Free Grammars (PCFG) have been shown to be particularly valuable in combination with n-gram methods. Feng, Banerjee, and Choi [2] are able to achieve 85%-91% accuracy in deception related classification tasks using online review corpora.

## III. PROPOSED SYSTEM

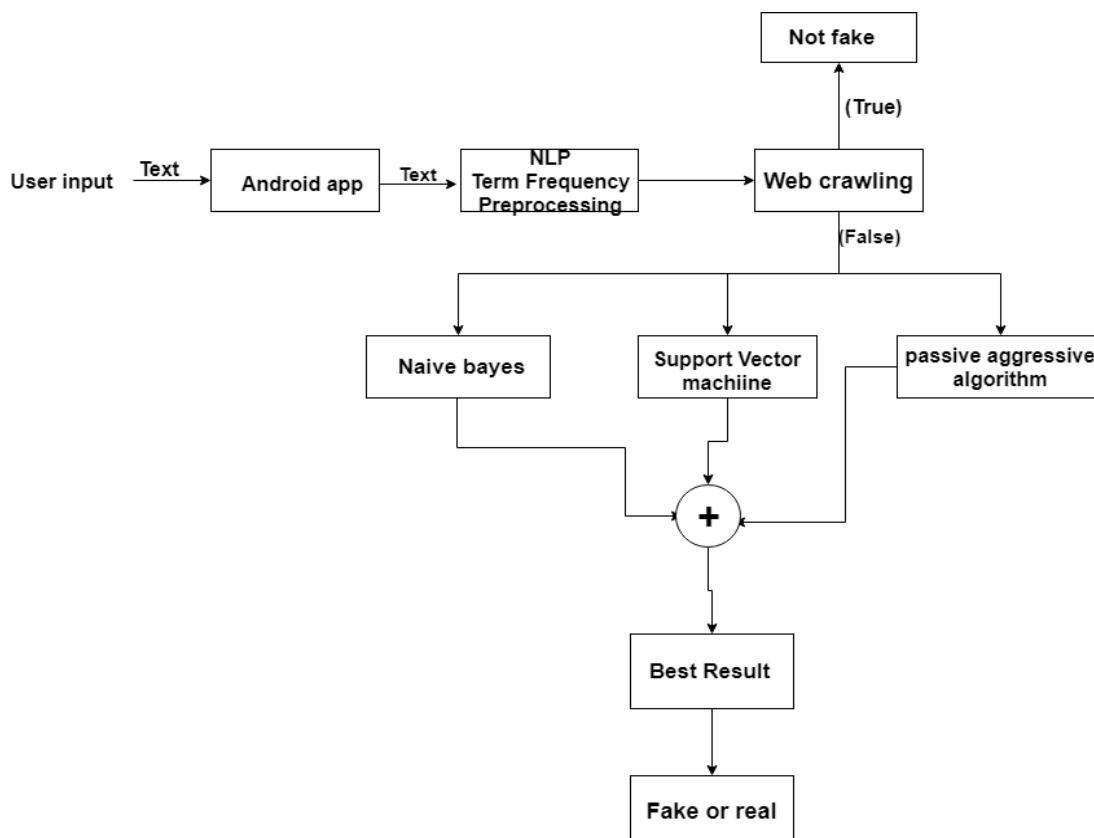


Fig:1 -: machine learning technique

Mainly There Is Two steps first we take input from the user and searching if it is available in trusted web sites. And the second stage is if the user input not found in the trusted sites then go to the machine learning technique. The above figure shows the basic procedure of our project , detection of fake news is done by combining web crawling and machine learning process, first user input is taken and it processed by NLP, Then web crawling is used to find whether it is in our trusted domain. If not a model is build based on the count vectorizer or a tf-idf matrix ( i.e )( word tallies relatives to how often they are used in other articles in our dataset)canhelp.Sincethisproblemisakindoftext classification, Implementing Combination of Naive Bayes classifier, support vector machine and passive aggressive algorithm will be best as this is standard for text-based processing. The actual goal is in developing a model which was the text transformation (count vectorizervstfidfvectorizer) and choosing which type of text to use (headlines vs

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 8, Issue 6, June 2019

full text). Now the next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given textdataset.

## IV. RESULT

For testing the performance the Sci-kit Learn's GridSearch functionality is utilized to efficiently execute this task. After execution the system will classify the user input by the help of the trusted web sites and the trained data. If we insert a news or text then it will show an out as “the given news is fake or real”. The data is classified by a result percentage of 80. Implementing Combination of Naive Bayes classifier, support vector machine and passive aggressive algorithm will be best as this is standard for text-based processing.

## V. CONCLUSIONS

The research showed, that quite simple machine learning algorithms (naïve base classifier, support vector machine and passive aggressive) and web crawling is best for classification of fake news. This automated system detecting fake news in popular social media. Furthermore crowd sourced workers provides a useful and less expensive means to classify true and false stories on social medias. Such a system could be valuable to social media users by augmenting and supporting their own credibility judgments. The social-based method offer good performances, these performances worse for news items that collect only few social interactions and these approaches can increase the overall detection accuracy.

## REFERENCES

- [1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media: A data mining perspective,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.
- [2] H. Allcott and M. Gentzkow, “Social Media and Fake News in the 2016 Election,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, May 2017.
- [3] H. Ahmed, I. Traore, and S. Saad, “Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques,” in Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, ser. Lecture Notes in Computer Science. Springer, Cham, Oct. 2017, pp. 127–138.
- [4] K. Shu, S. Wang, and H. Liu, “Exploiting Tri- Relationship for Fake News Detection,” arXiv:1712.07709 [cs], Dec. 2017.
- [5] L. Sydell, “We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned,” nov 2016. [Online].
- [6] C. Silverman, L. Strapagiel, H. Shaban, E. Hall, and J. Singer-Vine, “Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate,” oct 2016. [Online]. Available: <https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis>
- [7] Z. Jin, J. Cao, Y. Zhang, and J. Luo, “News Verification by Exploiting Conflicting Social Viewpoints in Microblogs,” in AAAI, 2016, pp. 2972–2978.