

Survey on Automated System for Fake News Detection using NLP & Machine Learning Approach

Subhadra Gurav¹, Swati Sase², Supriya Shinde³, Prachi Wabale⁴, Sumit Hirve⁵

^{1,2,3,4,5}BE(Computer Engineering), Modern Education Society's College of Engineering, Pune, Maharashtra, India.

Abstract - The large use of social media has tremendous impact on our society, culture, business with potentially positive and negative effects. Now-a-days, due to the increase in use of online social networks, the fake news for various commercial and political purposes has been emerging in large numbers and widely spread in the online world. The existing systems are not efficient in giving a precise statistical rating for any given news. Also, the restrictions on input and category of news make it less varied. This paper develops a method for automating fake news detection for various events. We are building a classifier that can predict whether a piece of news is fake based on data sources, thereby approaching the problem from a purely NLP perspective.

Key Words: Natural Language Processing (NLP), Machine Learning, Naïve Bayes, Fake News.

1. INTRODUCTION

Fake news detection topic has gained a great deal of interest from researchers around the world. When some event has occurred, many people discuss it on the web through the social networking. They search or retrieve and discuss the news events as the routine of daily life. Some type of news such as various bad events from natural phenomenal or climate are unpredictable. When the unexpected events happen there are also fake news that are broadcasted that creates confusion due to the nature of the events. Very few people know the real fact of the event while the most people believe the forwarded news from their credible friends or relatives. These are difficult to detect whether to believe or not when they receive the news information. So, there is a need of an automated system to analyze truthfulness of the news.

During the 2016 US president election, various kinds of fake news about the candidates widely spread in the online social networks, which may have a significant effect on the election results. According to a post-election statistical report [4], online social networks account for more than 41.8% of the fake news data traffic in the election, which is much greater than the data traffic shares of both traditional TV/radio/print medium and online search engines respectively. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news timely.

This paper provides an insight into the procedure of detecting fake news. In order to reach a conclusion on the authenticity of the news article, we first take the news event, analyze related data from data sources and then use various classification algorithms to classify the news as legitimate or fake.

Section II describes the work done by various authors in the field of fake news detection. Section III describes the related method and structure of our task. Section IV presents the conclusion of project and segment V describes the various references utilized in our task.

2. LITERATURE SURVEY

In [1], Shloka Gilda presented concept approximately how NLP is relevant to stumble on fake information. They have used time period frequency-inverse record frequency (TF-IDF) of bi-grams and probabilistic context free grammar (PCFG) detection. They have examined their dataset over more than one class algorithms to find out the great model. They locate that TF-IDF of bi-grams fed right into a Stochastic Gradient Descent model identifies non-credible resources with an accuracy of seventy seven.2%.

In [2], Mykhailo Granik proposed simple technique for fake news detection the usage of naive Bayes classifier. They used BuzzFeed news for getting to know and trying out the Naive Bayes classifier. The dataset is taken from facebook news publish and completed accuracy upto seventy four% on test set.

In [3], Cody Buntain advanced a method for automating fake news detection on Twitter. They applied this method to Twitter content sourced from BuzzFeed's fake news dataset. Furthermore, leveraging non-professional, crowdsourced people instead of journalists presents a beneficial and much less costly way to classify proper and fake memories on Twitter rapidly.

In [4], Marco L. Della offered a paper which allows us to recognize how social networks and gadget studying (ML) strategies may be used for faux news detection. They have used novel ML fake news detection method and carried out this approach inside a Facebook Messenger chatbot and established it with a actual-world application, acquiring a fake information detection accuracy of eighty one.7%.

In [5], Rishabh Kaushal carried out 3 getting to know algorithms specifically Naive Bayes, Clustering and Decision

bushes on some of features such as tweet-degree and consumer-level like Followers/Followees, URLs, SpamWords, Replies and HashTags. Improvement of unsolicited mail detection is measured on the premise of general Accuracy, Spammers Detection Accuracy and Non-Spammers Detection Accuracy.

In [6], Saranya Krishnan used superior framework to identify faux information contents. Initially, they've extracted content material capabilities and consumer functions via Twitter API. Then functions together with statistical analysis of twitter user accounts, reverse picture searching, verification of fake news assets are used by facts mining algorithms for class and analysis.

3. METHODOLOGY

The basic idea of our project is to build a model that can predict the credibility of real time news events.

As shown in Fig. 1, the proposed framework consists of four major steps: Data collection, Data preprocessing, Classification and Analysis of results.

We first take key phrases of the news event as an input that the individual need to authenticate. After that live data is collected from Twitter Streaming API. The filtered data is stored in the database (Mongo DB). The data preprocessing unit is responsible for preparing a data for further processing. Classification will be based on various news features, twitter reviews like Sentiment Score, Number of Tweets, Number of followers, Number of hashtags, is verified User, Number of retweets and NLP techniques.

We are going to describe fake news detection method based on one artificial intelligence algorithm – Naïve Bayes Classifier. Sentiment Score will be calculated using Text Vectorization algorithm and NLTK(Natural Language Toolkit).

By doing the evaluation of effects acquired from classification and analysis, we are able to decide the share of news being fake or real.

4. CONCLUSION

Many people consume news from social media instead of traditional news media. However, social media has also been used to spread fake news, which has negative impacts on individual people and society. In this paper, an innovative model for fake news detection using machine learning algorithms has been presented. This model takes news events as an input and based on twitter reviews and classification algorithms it predicts the percentage of news being fake or real.

REFERENCES

- [1] Shloka Gilda, "Evaluating Machine Learning Algorithms for Fake News Detection", 2017 IEEE 15th Student Conference on Research and Development (SCORED).
- [2] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).
- [3] Cody Buntain, Jennifer Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads", 2017 IEEE International Conference on Smart Cloud.
- [4] Marco L. Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, Luca de Alfaro, "Automatic Online Fake News Detection Combining Content and Social Signals", ISSN 2305-7254, 2017.
- [5] Arushi Gupta, Rishabh Kaushal, "Improving Spam Detection in Online Social Networks", 978-1-4799-7171-8/15/\$31.00 ©2015 IEEE.
- [6] Saranya Krishnan, Min Chen, "Identifying Tweets with Fake News", 2018 IEEE International Conference on Information Reuse and Integration for Data Science.
- [7] Arushi Gupta, Rishabh Kaushal, "Improving Spam Detection in Online Social Networks", 978-1-4799-7171-8/15/\$31.00 ©2015 IEEE.
- [8] Conroy, N., Rubin, V. and Chen, Y. (2015). "Automatic deception detection: Methods for finding fake news.", Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.
- [9] S. Maheshwari, "How fake news goes viral: A case study", Nov. 2016. [Online]. Available: <https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html> (visited on 11/08/2017).
- [10] Nikita Munot, Sharvari S. Govilkar, "Comparative Study of Text Summarization Methods", International Journal of Computer Applications (0975 – 8887) Volume 102–No.12, September 2014.

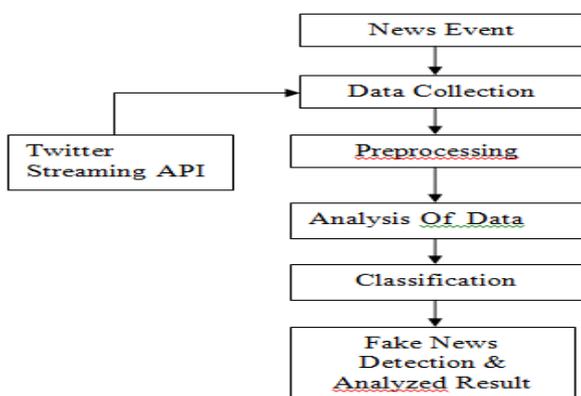


Fig -1: Block Diagram