

## **Fake news Detection on Social Media**

**Dr. Soumitra Das, Sachin Vasle, Aditi Paikrao,  
Ayesha Shaikh, Anshumati Londhe**

Department Computer Engineering, SPPU,Pune

***Abstract** - Fake news may be intentionally created to promote economic, political and social interests and can lead to negative impacts on humans beliefs and decisions. Hence, detection of fake news is an emerging problem that has become extremely prevalent during the last few years. Most existing works on this topic focus on manual feature extraction and supervised classification models leveraging a large number of labeled (fake or real) articles. In contrast, we focus on content-based detection of fake news articles, while assuming that we have a small amount of labels, made available by manual fact-checkers or automated sources. We argue this is a more realistic setting in the presence of massive amounts of content, most of which cannot be easily fact checked. So, we represent collections of news articles as multi-dimensional tensors, leverage tensor decomposition to derive concise article embeddings that capture spatial/contextual information about each news article. Results on real-world data sets show that our method performs on par or better than existing fully supervised models, in that we achieve better detection accuracy using fewer labels. In our proposed system we perform fake news detection using a ensemble learning. Firstly we perform pre-processing on extract data which is extracted from news. After that the classifier predicted news as a fake or real. In particular, our proposed method achieves 92.84 percent accuracy.*

**Key Words:** Internet of Thing (IoT), fake, real, news, accuracy

### **1. INTRODUCTION**

Fake news is any form of false story or content spread on the internet to influence people's view to gain inimical benefits. Detecting fake news in the digital world is a significant challenge in overcoming the wide spread dissemination of rumors and biases. Although there has been significant progress in fake news detection, a concrete set of solutions is yet to be established as the standard. Companies such as Face book, Twitter and Google are facing challenges in tackling this problem to ensure a platform where people can trust the news feed content. Fake news affects human judgment and behavior. In the spring of 2018, there was an article spreading the news that "Cadbury chocolate is infected with HIV-Positive Blood" with a video of boxes of the chocolate being burnt. This post gained traction on Face book, especially in South-Asian countries such as India. Rumors began to spread, damaging

Cadbury's reputation, and even people who did check other sources to make sure that the news was false became hesitant about buying those chocolates. Thus, it is imperative to detect and limit the spread of fake news. This project built a classification model using machine learning to detect fake news and implemented the machine learning model using a web-based application. In this project, the fake news detection is a binary classification problem-news is either fake or reliable. A user-friendly web interface was built to enable users to easily query an news source using a URL and determine if the news is fake.

## **2. LITERATURE SURVEY**

### **2.1 Naive Bayes classification Fake news detection:**

Information preciseness on Internet, especially on social media, is an increasingly important concern, but web-scale data hampers, ability to identify, evaluate and correct such data, or so called "fake news," present in these platforms. In this paper, we propose a method for "fake news" detection and ways to apply it on Face book, one of the most popular online social media platforms. This method uses Naive Bayes classification model to predict whether a post on Face book will be labeled as REAL or FAKE. The results may be improved by applying web Scrapping method. Web

Scrapping is a technique employed to extract large amounts of data from different websites and to store as desired. Received results suggest, that fake news detection problem can be addressed with machine learning methods. This method is easy to implement and requires a small amount of training data to estimate the parameters.

### **2.2 Linear SVM Classifier Automatic Detection of Fake News:**

The proliferation of misleading information in everyday access media outlets such as social media feeds, news blogs, and online newspapers have made it challenging to identify trustworthy news sources, thus increasing the need for computational tools able to provide insights into the reliability of online content. In this paper, we focus on the automatic identification of fake content in online news. Our contribution is twofold. First, we introduce two novel datasets for the task of fake news detection, covering seven different news domains. We describe the collection, annotation, and validation process in detail and present several exploratory analyses on the identification.

## **3. PROBLEM DEFINATION**

As an increasing amount of our lives is spent interacting online over the internet, more and more people tend to seek and consume news from social media, news agency homepages, search engines. On the other hand, it enables the proliferation of fake news", i.e., low quality news with intentionally false information. Popular social media platforms such as Face book, twitter have proven to be an

effective means of channels for spreading these false news due to their wide reach and the speed in which information is spread.

#### 4. PROPOSED SYSTEM

News content is archived from news link. Whole news content as text data given to pre-processing. In pre-processing feature extraction and tokenization of news text data is done. Pre-processing is required to be done for further modeling. These tokens then transfer to process model in form of vectors. Ensemble learning is used to predict result. In ensemble learning naive bayes classifier, logistic regression, SVM classifier and SGD classifier and random forest classifier is used. All the different algorithms have different working, so they will produce different outcomes. All results from all of classifier are then combined to get final result or the outcome. From all the classifiers, the best performing classifier is used to predict outcome. The fig.1 shows the architecture diagram of the proposed system.

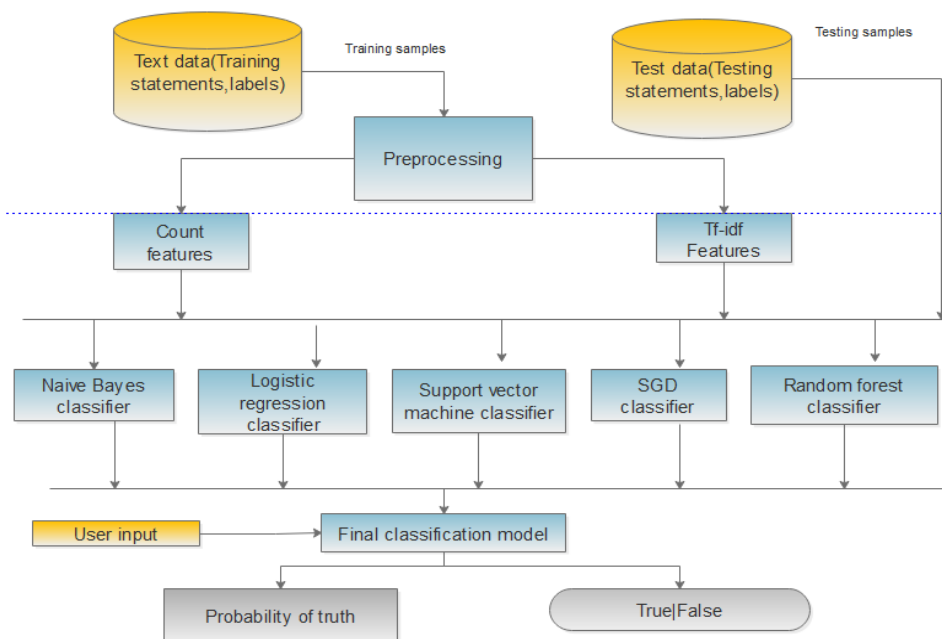


Fig.1: Architecture Diagram

#### 5. CONCLUSIONS

Due to digital revolution and increase in internet users, there are almost 600 million internet users in India alone. Users are mostly naive and new, so spreading and believing in fake news is increased, so classification of news is very important. Our system is developed in order to curb this spread of fake news. Our proposed system has achieved an accuracy of 92.84 percent which is greater than existing system proposed in our base paper.

## REFERENCES

- [1]. KaiShu, SuhangWang, "News Detection on Social Media: A Data Mining Perspective", *Computer Science Engineering*, Arizona State University, Tempe,AZ,USA.
- [2]. Akshay Jain, Amey Kasbe, "Fake News Detection", *2018 IEEE International Students Conference on Electrical, Electronics and Computer Sciences*, Bhopal,India.
- [3]. Amitabha Dey, Rafsan Zain Ra, "Fake News Pattern Recognition using Linguistic Analysis", *Joint 7th International Conference on Informatics, Electronics Vision (ICIEV)*, Kitakyushu, Japan.
- [4]. NamwonKim, DeokjinSeo, "Fake News Pattern Recognition using Linguistic Analysis", *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China.