Media Manipulation Unmasking Using Mincing Bayes Classifier

Manjushree T L

M.Tech Student

Department of Computer Science and Engineering

India

Abstract— Accounting to the expeditious digitization across all channels and mediums, the menace of fake news has been burgeoning at a colossal scale. Majority of the countries all across the world are trying to combat this challenge. This paper explores the application of Natural Language Processing and Machine Learning techniques to identify fake news accurately. Pre-processing tools are used to clean the data and apply feature extraction on them. Then a fake news detection model is built using four different techniques. Finally, the paper investigates and compares the accuracy of techniques which are Naive Bayes, Support Vector Machine (SVM), neural network and long short-term memory (LSTM) to find the best fit for the model.

Keywords: Machine Learning, Fake News Classification, Probability

I. INTRODUCTION

Over the last decade, there have encounters of flux in misinformation that spread like wildfires. The surge in fake news was noticed during the 2016 presidential elections that happened in the US that determined the fate of these elections. In many cases, it is seen that the sharing of hoax news has been more than that of accurate news. In a massive market like India, the scope of fake news propaganda has been artfully misused by many groups.

Researches indicate that Facebook and WhatsApp are the platforms that are utilized for spreading fake news. An approximate of one in two Indians have agreed to have received fake news during the 2019 Lok Sabha elections.

According to research at Ohio University by M Laeeq Khan, the majority of the population do not verify the authenticity and veracity of a piece of information before forwarding it on social media. Khan conducted his work in the US and later extended it to Indonesia to test his hypothesis about spreading misinformation on a medley of online platforms. "Fake news and spreading of it could be rightly termed as the major issues of our time. No other study can explain the vital role of people in halting the spread of misinformation", says Laeeq. It was found that gender, age, and social class do not play a pivotal role in spreading misinformation, but rather the literacy regarding the internet and social media that mattered the most. The research which comprised of 396 participants stated that the information verification attitude of the people was taken for granted. It was found that the simple habit of googling the veracity of information before sharing it could prove crucial in fighting against the issue of fake news.

It is also seen that spam messages and fake news have striking similarities. They use manipulative ways to win over the reader's opinions. Most of them also have grammatical mistakes and they also use a similar restricted set of words among them. Since both the media share such similar properties, we can use similar approaches to detect fake news accurately. One way to tackle fake news is to manually classify news as real or fake. Even though that seems like the simplest solution it is not practical with the jillions of news that get produced to manually label it. Hence, there is a need to look for a pragmatic technical solution to do the same. The proposed method in this research is to exploit the advancement in machine learning. To do the same, the classification model has been trained with various machine learning algorithms to label the data. The results from the study indicate neural networks to be the case in which news highest accuracy is achieved.

Here the data is divided into test dataset and train dataset and the train dataset is divided into groups of similar information. Test data is later matched with these groups and accuracy is found using Naive Bayes classifier. It helps in knowing whether a given news is fake or real. It provides maximum accuracy and helps to determine the fake news.

II. LITERATURE SURVEY

In the paper by Shu A et al., they investigated how news can be classified as true or not by focusing on a few attributes that are repeatedly encountered in fake news. In their opinion, these characteristics were based on "psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets". This paper also analyses the different challenges one will encounter while studying this topic. [1]

The paper is written by Rubin et al. deals with the domain of fake news which is composed of satirical news. Satire news intentionally provides hints revealing its own deception. While fake news wants the readers to believe a false fact, satire news must eventually be understood as a jest. This paper provides an in-depth view of the features of humor and satire news along with the style of the authors reporting. The paper has considered the news articles from twelve contemporary news topics in four different domains which are civics, science, business, and soft news. The paper proposes a Support Vector Machine based algorithm which can detect satire news based on features like Absurdity, Humor, Grammar, and Punctuation. The models achieved an accuracy of 90% and a recall of 84%. The aim is to reduce the negative impact of satire news on readers. [2]

In the paper by Kelly Stahl et al., they have considered past and current techniques for fake news identification in text formats while elucidating how and why news fake exists in any case. This paper incorporates a discussion on how the writing style of a paper can also impact on its classification. They had implemented their project using Naïve Bayes Classifier and Support Vector Machines methods. They had looked into the semantic analysis of the text for classification. [3]

In this paper by Marco L. Delia Vedov et al., they say that "we propose a novel ML fake news detection method which, by combining news content and social context features, outperforms existing methods in the

literature, increasing their already high accuracy by up to 4.8%". The proposed model was then tested on a real-time application and they achieved high accuracy by testing it on a FB messenger chatbot. The accuracy achieved by them is close to 82%. [4]

In the research paper written by Mykhailo Granik et al., they have explored a simple approach to detect fake news using Naïve Bayes classifier. Their approach was conducted on Facebook news posts rather than news articles on the internet. The paper achieved an accuracy of 74% in the classification process. It also went on to state that simple machine learning models like the Naïve Bayes can achieve a moderate accuracy and in future, more artificial intelligence techniques could be used to tackle the menace of fake news. [5]

The paper by Namwon Kim et al. detects fake news based on unified key sentence information. The model extracts the key sentences of the article to the question from the article and then later combining the word vector for each key sentence. They make use of a Korean dataset for their analysis. It performs an efficient matching operation for the word vectors obtained using bidirectional LSTM. They have achieved an accuracy ranging from 64 to 69%. Their future plan is to develop a more advanced model which applies the model independently to each key sentence. [6]

In this paper by Hadeer Ahmed et al., in this project an elaborate study was done using n-gram while being compared with different feature extraction methods. Many feature extraction techniques and quite a few machine learning techniques were utilized in this study. The paper reflected the highest output while implementing unigram as the extraction method and SVM as the classification method. [7]

In this paper by Saranya Krishnan et al., they have proposed a generalized framework to predict tweet credibility. First, they have extracted the essential features and user features through the Twitter API. If a tweet has an image or image URL, the reverse image search is performed to check whether the same image has been tagged with different information in the past. In addition, if any URL is present in any tweet, it will be cross-checked against the fake news sources to see whether it is a part of a fake news websites dataset. All these features are then used by the data mining algorithms to classify tweets as fake or real. [8]

In this paper by Stefan Helmstetter et al., tweets were assimilated in real time. On arrival, these tweets were background checked to verify if they were classified properly. Then the incoming tweets were tested against a classification algorithm. Despite having a noisy dataset a decent result was obtained. [9]

In this paper by Akshay Jain, he suggested a classification methodology that could be used in real time data that gets generated on Facebook. It could also be used in other social media giants like Twitter and WhatsApp. He has utilized a simple Naive Bayes classification to classify the incoming data as trustworthy or non-trustworthy. [10]

III. EXISTING SYSTEM AND PROPOSED SYSTEM

A. Existing System

There are projects which are being done in this domain which use algorithms such as Hybrid Cloud approach to detect the fake news. But such algorithms have very less accuracy and take more storage. This algorithm uses human as input sometimes, so the risk that the data given by a single human is very high which hinders the accuracy of the fake news detection so an algorithm with a efficiency greater than the current algorithm is needed. Hybrid approach based models need lager data sets to train the data and this method also does not sometimes classify the data so there is a higher risk of matching with the unrelated data which in turn will affect the accuracy of the news.

B. Proposed system

The proposed system uses naïve Bayes algorithm for detecting the fake news, it has a higher accuracy level. The data is divided here into two parts (test and train). The train data is trained and is classified into groups with similar datasets. After the data is trained the test data is assig ned to the group which has similar characteristics with the group. Now, the naïve Bayes algorithm is used to detect the accuracy of the fake news with which spams can be stopped. Also here weights are given to each and every individual word, the least important word is given less weight and most important word is given most weight. Also the tfidf vector is used to count the number of word and also the number of unique words is counted and at the same time the weights are allotted to each and every word. In this way the unimportant words are not taken into consideration and the accuracy of only important words are matched and detected from the dataset, this saves the time to detect the accuracy of the news.

IV. MINCING BAYES CLASSIFIER WORKING AND ITS MODULE FLOW

In machine learning, naive Bayes classifiers are the part of simple machine learning. Naive Bayes is popular algorithm which is used to find the accuracy of the news whether its real or fake using multinomial NB and pipelining concepts. There are number of algorithms that focus on common principle, so it is not the only algorithm for training such classifiers. To check if the news is fake or real naïve Bayes can be used.It is a kind of algorithm is used in text classification. The use of token is correlated with the news that may be fake or not fake in naïve Bayes classifier and then the accuracy of the news is calculated by using Bayes theorem.
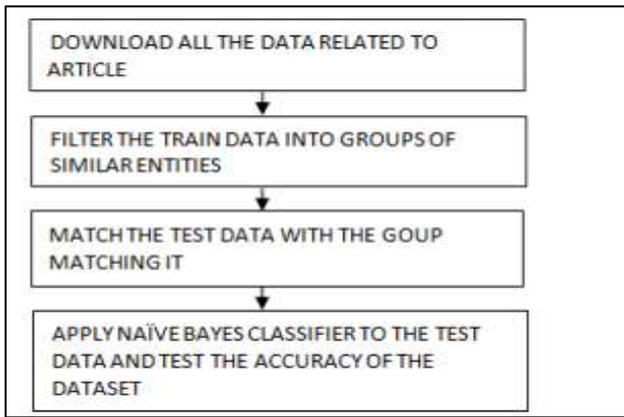
Fig. 1: module flow

First extract all the data which is to be checked or download it if available and then divide the data into test and train, train the data and then apply Bayes theorem in this way Naïve Bayes theorem is applied.

A. Data Pre-processing

This contains all the data which must be checked thoroughly and preprocessed. First, we go through the train, test and validation data files then performed some preprocessing like tokenizing, stemming etc. Here the data is checked thoroughly if it has missing values.

B. Feature Extraction

In this dataset we have done feature extraction and selection methods from scikit and python. To perform feature selection, we use a method called as tf -idf. We have also used word to vector to extract the features, also pipelining has been used to ease the code.

C. Classification

Here the classification of the data is done in to parts that is test data and train data and the train dataset is classified into groups with similar entities. Later the test data is matched, and the group is assigned to whichever it belongs to and then further the Naïve Bayes classifier is applied and the probability of each and every word is calculated individually. If the word whose probability is to be calculated is not available in the dataset of the train data then the Laplace smoothing is applied here. Finally, the data is determined if it's fake or real.

D. Prediction

Our finally selected and best performing classifier was algorithm which was then saved on disk with name file_modal.sav. Once you close this repository, this model will be copied to user's machine and will be used by predict.py file to classify the fake news with accuracy. It takes a news article as input from user then model is used for final classification output that is shown to user along with probability of truth.

V. System architecture

The first step in the detection of fake news is extracting the training data either by downloading it from a file or from online.
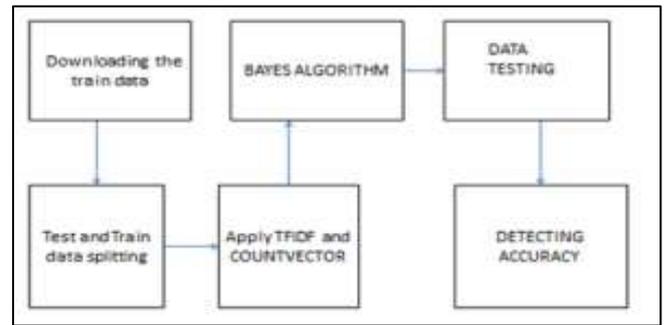

Fig. 2: System architecture

There are two methods to count the words. The Fit method and the Transform method. The fit method is used to give a specific serial number to each and every word and the transform method is used to count the number of times a particular word is occurring in the data set. Instead of using both the methods separately we can use it as a whole single method called fit transform method which helps us in saving both the space and time. Term frequency is required to count number of times a word is occurring, and inverse document frequency is used to give weight to the words. It gives maximum weight to the most important words and minimum weight to the least important words. So, we club both the methods into a single method to save the time and space in the detection called as tfidf which calculates the height of a particular word. Now the dataset is split into two parts that is test and train dataset. Now multinomial Naïve Bayes algorithm is used to classify the train data in groups of similar entities. The test data is no matched with the group of the train data it's matching with. After the data is matched naïve Bayes algorithm is applied to the test dataset and the probability of each and every word is checked and approximate percentage value is calculated and in this way the accuracy of the fake news is determined. Therefore, in this way it is determined whether a given news is fake or real.

VI. Conclusions

Therefore by usingmincing bayes classifiers theorem we can conclude that any news from a large or small dataset can be classified as fake or real news by matching it with the previous dataset values in less time which in turn helps the users to believe in a particular news. In this the deep learning and machine learning concepts are used to detect the fake new using naïve Bayes classifier. The data set is loaded for which the news is to be classified and then the data is to be split as test and train data and pipelining is to be done to detect the accuracy. As the fake news is increasing day by day the people are not believing even if the news is real and this drifts the thoughts of the common people from the real issue

References

[1] Shu K., Sliva A., Wang S., Tang J., Liu H., "Fake News Detection on Social Media: A Data Mining Perspective", ACM SIGKDD Explorations Newsletter, 2017, 19(1), 22-36.

[2] Rubin, V., Conroy N., Chen Y., Cornwell S., "Fake News or Truth? Using Satirical Cues to Detect

Potentially Misleading News", Proceedings of the Second Workshop on Computational Approaches to Deception Detection Association for Computational Linguistics, 2016.

[3] Kelly Stahl., "Fake news detection in social media", B.S. Candidate, Department of Mathematics and Department of Computer Sciences, California State University Stanislaus, 2018.

[4] Marco L. Delia Vedova, Stefano Moret, Eugenio Tacchini, Massimo Di Pierro," Automatic Online Fake News Detection Combining Content and Social Signals", Research Gate May 2018.

[5] Mykhailo Granik, Volodymyr Mesyura, "Fake News Detection Using Naive Bayes Classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON).

[6] Namwon Kim, Deokjin Seo, Chang-Sung Jeong, "FAMOUS: Fake News Detection Model based on Unified Key Sentence Information", 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)

[7] Hadeer Ahmed, Issa Traore, Sherif Saad, "Detection of Online Fake News Using N-Gram Analysis and Machine Learning Techniques", Research Gate May 2017.

[8] Saranya Krishnan, Min Chen.," Identifying tweets with fake news", 2018 IEEE Conference on Information Reuse Integration for Data Science.

[9] Stefan Helmstetter, Heiko Paulhem, "Weakly Supervised Learning for Fake news detection on Twitter", 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[10] Akshay Jain, "Fake News Detection", IEEE 2018 International Students" Conference on Electrical, Electronics and Computer Sciences