FAKE NEWS DETECTION USING MACHINE LEARNING METHOD

Anisha Soman

Post Graduate Student, Computer Science Department, IES college of Engineering, Kerala, India

ABSTRACT

In the modern time the internet is ubiquitous, everyone relies on various online resources for news. Along with the increase in the use of social media platforms like Facebook, Twitter, etc. news spread rapidly among millions of users within a very short span of time. With the rising popularity of social media, people have become more aware of current events and important news, often through sources such as Twitter. One issue with these sources of news is the prevalence of false information, or fake news. Even as some social media platforms take initiative with labels or warnings, fake news continues to have dangerous consequences beyond misinformation. Fake news detection is an emerging research area which is gaining big interest. The goal of this research is to implement a highly effective method of identifying fake news spread on Twitter through the use of Machine Learning method Support Vector Machine.

In this work, we propose a system for Fake news detection that uses machine learning techniques. We used term frequency inverse document frequency (TF-IDF) of bag of words and n-grams as feature extraction technique, and Support Vector Machine (SVM) as a classifier. Using tweets as Twitter dataset, the experiments are considering n-grams as features and passing their TF-IDF values to machine learning model. After tuning the model giving the best results. Also create a module which serves as an intermediate between user and Twitter.

Keyword : - *Fake news, Social media, Machine Learning, Support Vector Machine, TF-IDF, Twitter, Tweet.*

.

1. INTRODUCTION

The rise of technology, social media has become increasingly popular in the personal daily life. Innovations allow people to absorb vast amounts of information on a daily basis. Social media provides its users with a platform to voice their thoughts and connects people around the world.In the past 10 years, we have seen an exponential growth in the number of people using online forums and social networks. Every 60 seconds, there are 510,000 comments generated on Facebook and around 350,000 tweets generated on Twitter. The people interacting on these forums or social networks come from different cultures and educational backgrounds.

However, a significant down side to these advances in technology is the increasing prevalence of false information. Fake news is defined as articles that misrepresent information to deceive and manipulate their audience. They are 70% more likely to be retweeted on Twitter than true ones. On Twitter is a popular social media platform where users can easily share links to articles regardless of validity. As a result, fake news is rampant. Current solutions to combating fake news are often heavily reliant on the initiative of readers. Social media users are encouraged to be vigilant regarding the news they see to avoid being manipulated.

In this paper, we present a novel method and tool for detecting fake news that uses:
• Text preprocessing: consisting of steaming and analyzing the text by removing stop words and special characters.
• Encoding of the text: using bag of words and N-gram then TF-IDF.
• Extraction of the characteristics: this allows a precise

identification of false information. We use the source of a news, its author, the date and the feeling given by the text as features of a news.
• Support vector machine: a supervised machine learning algorithm that allows the classification of new information.

2. RELATED WORKS

Hadeer Ahmed et. al. [1] propose a fake news detection model that uses n-gram analysis and machine learning techniques by comparing two different feature extraction techniques and six different classification techniques. The experiments carried out show that the best performances are obtained by using the so-called features extraction method (TF-IDF). The used the Linear Support Vector Machine (LSVM) classifier that gives an accuracy of 92%. This model uses LSVM that is limited to treat only the case of two linearly separated classes.

Mykhailo Granik et. al. [3] present a simple approach to fake news detection using a naive Bayesian classifier. This approach is tested on a set of data extracted from Facebook news posts. They claim to be able to achieve an accuracy of 74%. The rate of this model is good but not the best, as many other works have achieved a better rate using other classifiers. We discuss these works in the following.

Junaed Younus Khan et. al. [4], the authors present an overall performance analysis of different approaches on three different datasets. This work focused on the text of the information and the feeling given by it, and ignores some features like the source, the author or the the date of the publication that can have a dramatic impact on the result. Besides, in our work, we will show that the integration of the feeling in the detection process does not bring any valuable information.

Vincent Claveau et. al. [5] propose several strategies and types of indices relating to different modalities (text, image, social information). They also explore the value of combining and merging these approaches to assess and verify shared information.

In his paper Florian Sauvageau et. al. [6] describe how users of social networks can ensure the truth of information. They also describe the mechanisms that allow their validation and the role of journalists or what to expect from researchers and official institutions. This work helps people see a little bit of the truth behind the news on social media and not believe anything.

Rupanjal Daigupta et. al. [7] created a new public dataset of valid new articles and proposed a text-processing based machine learning approach for automatic identification of Fake News with 87% accuracy. It appears that this work focuses on the emerging feelings from the text
and not on the contain of the text in it self.

Mykhailo Granik et. al. in their paper [9] shows a simple approach for fake news detection using naive Bayes classifier. This approach was implemented as a software system and tested against a data set of Facebook news posts. They were collected from three large Facebook pages each from the right and from the left, as well as three large mainstream political news pages (Politico, CNN, ABC News). They achieved classification accuracy of approximately 74%. Classification accuracy for fake news is slightly worse. This may be caused by the skewness of the dataset: only 4.9% of it is fake news.

Cody Buntain et. al. [11] develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibilityfocused Twitter datasets: CREDBANK, a crowd sourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumours in Twitter and journalistic assessments of their accuracies. They apply this method to Twitter content sourced from BuzzFeed‟s fake news dataset. A feature analysis identifies features that are most predictive for crowd sourced and journalistic accuracy assessments, results of which are consistent with prior work. They rely on identifying highly retweeted threads of conversation and use the features of these threads to classify stories, limiting this work‘s applicability only to the set of popular tweets. Since the majority of tweets are rarely retweeted, this method therefore is only usable on a minority of Twitter conversation threads.

In his paper, Shivam B. Parikh et. al. [12] aims to present an insight of characterization of news story in the modern diaspora combined with the differential content types of news story and its impact on readers. Subsequently, we dive

into existing fake news detection approaches that are heavily based on textbased analysis, and also describe popular fake news datasets. We conclude the paper by identifying 4 key open research challenges that can guide future research. It is a theoretical Approach which gives Illustrations of fake news detection by analysing the psychological factors.

3. FAKE NEWS DETECTION SYSTEM

The proposed fake news detection system is based on machine learning approaches. The architecture and the important steps in building the proposed system is described here.

3.1 Architecture

Figure-1 shows an overall work flow or architecture of the proposed fake news detection system. The main processes in the proposed system are Dataset Collection, Data Pre-processing, Feature extraction and Model Training.

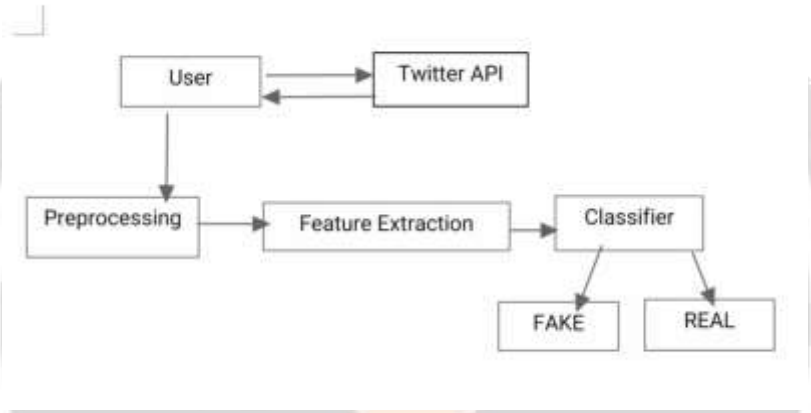

Fig -1: System Architecture

The dataset chosen here is the publically available tweets. The first task is to pre-process the selected dataset. The pre-processed dataset is then used in the feature extraction phase, where various features are extracted from it. The extracted features are used to train the system. The trained system will be able to classify whether the tweet is fake or real.

3.2 Dataset Collection

The dataset used here is publicly available twitter dataset. It contain 31,925 observations. The dataset contain id, label and tweet fields.

3.3 Data Pre-Processing

In the data preprocessing stage, the system perform the removal of unnecessary columns from the datasets and enumerating the classes. For dataset, the system retrieve the tweets corresponding to the tweet-ID present in the dataset and use Twitter API for this purpose. Then convert the tweets to lowercase and remove the unnecessary space pattern, URLs, twitter mentions, retweet symbols, stopwords. The system use the Porter Stemmer algorithm to reduce the inflectional forms of the words. After combining the dataset in proper format, it randomly shuffle and split the dataset into two parts train dataset containing 70 percentage of the samples and test dataset containing 30 percentage of the samples.

3.4 Feature Extraction

The system extract the n-gram features from the tweets and weight them according to their TFIDF values. The goal of using TF-IDF is to reduce the effect of less informative tokens that appear very frequently in the data corpus. Experiments are performed on values of n ranging from one to three. The system consider unigram, bigram and trigram features. The formula that is used to compute the TFIDF of term t present in document d is: tf-idf(t, d) = tf(t) * idf(t, d).

3.5 Classification

The system consider prominent machine learning algorithm used for text classification is Support Vector Machines. Then system train each model on training dataset by performing grid search for all the combinations of feature parameters and perform 10-fold cross-validation. The performance of each algorithm is analyzed based on the average score of the cross-validation for each combination of feature parameters.

Further, the hyperparameters of algorithm giving the result that tuned for their respective feature parameters, which gives the best result. Again, 10-fold cross validation is performed to measure the results for each combination of hyperparameters for that model. The model giving the highest crossvalidation accuracy is evaluated against the test data.The system have used scikit-learn in Python for the purpose of implementation.

3.6 Interfacing with Twitter

The final model is configured to interface with Twitter through the use of Twitter API particularly to collect data tweets via Twitter REST API. In python, the library Tweepy helps add this functionality with simplicity. Twitter APIs, besides basic information such as the tweet text and the author of the tweet, returns data structure contains additional information which can be used to provide further analysis. For each maximum 280 character tweet, API returns a JSON document containing several items of metadata presented as key and value pairs, out of which id and text are most important for the sake of this study.

4. RESULTS

To get the best decision model with highest accuracy, we tuned many parameters. First we tried to get the best parameters from both bag of words and n-gram techniques which give the best recognition rate on rour dataset.The classifier has given about 100% accuracy in classifying the fake news texts. We can see a snapshot of the predicted labels for the news tweets by the system in the below image.

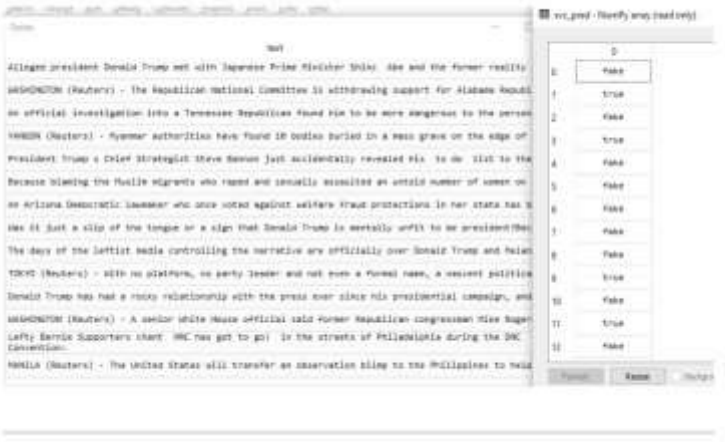

Fig -2: Label prediction for tweets

We will classify the tweet texts using the Support Vector Classifier model and evaluate its performance using evaluation matrices it shown in Fig -3.

```
Accuracy of SVM Classifier: 99.55%

Confusion Matrix of SVM Classifier:

[[4720   20]
 [  20 4220]]

Classification Report of SVM Classifier:

              precision    recall  f1-score   support

        fake       1.00      1.00      1.00      4740
        true       1.00      1.00      1.00      4240

    accuracy                           1.00      8980
   macro avg       1.00      1.00      1.00      8980
weighted avg       1.00      1.00      1.00      8980
```

Fig -3: Performance of system

5. CONCLUSIONS

The proposed system is able to detection fake and real news on Twitter through machine learning using n-gram features weighted with TFIDF values. Upon evaluating the model on test data, we achieved 95.6 percentage accuracy. This paper presents a method of detecting fake news using support vector machine, trying to determine the best features and techniques to detect fake news. We started by studying the field of fake news, its impact and its detection methods. We then designed and implemented a solution that uses a dataset of news preprocessed using cleaning techniques, steaming, Ngram encoding, bag of words and TF-IDF to extract a set of features allowing to detect fake news.

5.1 Future Work

1) Satire Analysis: In the future, it would be beneficial to expand the data set and add satirical news stories to train the LSTM and GRU models. This would allow both models to more accurately classify satire and comedy as fake news.
2) Gradual Classification: Modifying the models to classify articles on a spectrum would be more complex, but could be more helpful for readers rather than simply determining whether it is real or fake. This would show readers how much of the article should be trusted, rather than completely writing off a source if it is only partially fake. Adding this to
the project would entail changing the network from a binary classifier into a multi-classification system.
3) Usage Beyond Twitter: While the models were trained using data from Twitter, verifying the systems' effectiveness on other platforms could further reduce the prevalence of fake news. Facebook and Snapchat, for instance, are popular platforms that feature news pages, and applying these models there could help users differentiate real from fake news.

6. REFERENCES

[1]. Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments, pages 127–138. Springer, 2017
[2]. Niall J Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1):1–4, 2015.
[3]. Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pages 900–903.IEEE, 2017
[4]. Junaed Younus Khan, Md Khondaker, Tawkat Islam, Anindya Iqbal, and Sadia Afroz. A benchmark study on machine learning methods for fake news detection. arXiv preprint arXiv:1905.04749, 2019
[5]. Cédric Maigrot, Ewa Kijak, and Vincent Claveau. Fusion par apprentissage pour la détection de fausses informations dans les réseaux sociaux. Document numerique, 21(3):55–80, 2018
[6]. Florian Sauvageau. Les fausses nouvelles, nouveaux visages, nouveaux défis. Comment déterminer la valeur de l'information dans les sociétés démocratiques? Presses de l'Université Laval, 2018
[7]. DSKR Vivek Singh and Rupanjal Dasgupta. Automated fake news detection using linguistic analysis and machine learning
[8]. William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017
[9]. M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017
[10]. H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in Twitter," 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018
[11]. C. Buntain and J. Golbeck, "Automatically Identifying Fake News in Popular Twitter Threads," 2017 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, 2017
[12]. S. B. Parikh and P. K. Atrey, "Media-Rich Fake News Detection: A Survey," IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, 2018