OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

FAKE NEWS DETECTION USING MACHINE LEARNING APPROACH

Khan Mohammed Salim
PG Student, Computer Science & Engineering Department, EESGOI, Aurangabad

---

Abstract: The role of social media in our day to day life has increased rapidly in recent years. It is now used not only for social interaction, but also as an important platform for exchanging information and news. Twitter, Facebook a micro-blogging service, connects millions of users around the world and allows for the real-time propagation of information and news . Fake news has become a major problem in these social networks . Fake news has become a major problem in these social networks Fake news has vast impact in our modern society. Detecting Fake news is an important step. This work purposes the use of machine learning techniques to detect Fake news ,using Support Vector Machine (SVM)algorithm. The normalization method is important step for cleansing data before using the machine learning method to classify data. This model count the credibility of content and user reputation .This method develops a method for automating fake news detection on social media by learning to predict accuracy assessments in credibility-focused Twitter dataset.

Keywords: Twitter social network; reputation; credibility; machine learning.

---∴∴∴---

I INTRODUCTION

The role of social media in our day to day life has increased rapidly in recent years. Online social media is a popular platform where millions of people can communicate with each other in real time. These are the dynamic data sources where the users can create their own profiles and communicate with each other independent of geographical location. It provides communication platform with large scale and large extent. Furthermore these tools are beyond the boundaries of the physical world in studying human relationship and behaviours. As these social Medias are becoming more popular, cybercriminals have utilize these as a new platform for communicating different types of cybercrimes. Twitter, a microblogging service, Facebook connects millions of users around the world and allows for the real-time propagation of information and news. These factors have resulted in Twitter playing a critical role in world events, especially emergency events, where it has been useful in emergency response and recovery. Nowadays, different cybercrimes are happening such as phishing, spamming, spread of malware and fake news is considered as a major problem along with the recent development of social media. It is a technique by which users get harass from other individual user of the group of user. Online social media such as Facebook, twitter have become integral component of a users life. Because of this, these websites have become the most common platform for spred the fake news. Fake News is an inaccurate, sometimes sensationalistic report that is created to gain attention, mislead, deceive or damage a reputation. Unlike misinformation, which is inaccurate because a reporter has confused facts, fake news is created with the intent to manipulate to the user. Fake news can spread quickly when it provides disinformation that is aligned with the audience's point of view because such content is not likely to be questioned or discounted. Twitter has, however, not only been used for the spread of valid news, and fake news. This fake news can come in the form of spam, AstroTurf is a technique used in political movements to fake support numbers, by making a message appear to have popular origins when in reality it originated from one person or organization, clickbait and more. The increase in the volume of fake news has level controlled to our current times being labeled the age of misinformation and therefore stresses the importance of assessing the credibility of tweets. Hence, we are aimed to utilized useful information in tweets to detect fake news. Selecting the twitter dataset with streaming API and search API tweets is a complex task that requires considerable efforts in building the machine learning model.

There for to develop a fake news detection method by identifying which is the fake news or real news that can be used in machine learning techniques. In particular use many features of twitter such as Structural feature, Content features

and User features these features use for the user reputation and Credibility of content. User and content features use for the calculate the user reputation and Credibility of content. Using these features to train our fake news detection model and improve its performance. Apply machine learning techniques to detect Fake news by using the support vector machine(SVM) algorithm for the classify the tweet with fake or real.

MOTIVATION

In our daily life Fake news is an important issue on social media. Using fake news the more criminal activity are happing in the world it causes defect on human life to avoid and stop criminal activities by using these techniques Our work considers crowd signals for detecting fake news and is motivated by tools recently introduced
by Facebook that enable users to flag fake news.

II LITERATURE SURVEY

1. Evaluating Machine Learning Algorithms for Fake News Detection. Shlok Gilda et al.(2017) proposed Machine Learning Algorithms for Fake News Detection using a dataset obtained from Signal Media and a list of sources from Open Sources.co, Apply term frequency-inverse document frequency (TF-IDF) of bigrams and probabilistic context free grammar (PCFG) detection to a corpus of about 11,000 articles. compare the performance of models using three distinct feature sets to understand what factors are most
predictive of fake news: TF-IDF using bi-gram frequency, syntactical structure frequency (probabilistic context free grammars, or PCFGs), and a combined feature union using machine classification for identification. This indicates that PCFGs are good for a Fake-News Filter type implementation versus, say, training fake news sites for review. [1]

2. Fake News Detection Using Naive Bayes Classifier. Mykhailo Granik, Volodymyr Mesyura (2017) proposed Fake News Detection Using Naive Bayes Classifier In this paper Dataset, collected by BuzzFeed News, was used for learning and testing the naive Bayes classifier. This paper describes a simple fake news detection method based on one of the artificial intelligence algorithms nave Bayes classifier. Naive Bayes classifiers are a popular statistical technique of email filtering. Naive Bayes typically use bag of words features to identify spam e-mail, an approach commonly used in text classification. and examine how this particular method works for this particular problem given a manually labeled news dataset and to support the idea of using artificial intelligence for fake news detection. [2]

3. A Credibility Analysis System for Assessing Information on Twitter. Majed Alrubaian (2016) Proposed A Credibility Analysis System for Assessing Information on Twitter, In this paper Tweets are collected using two different Twitter application programming interfaces (APIs) propose a novel credibility assessment system that maintains complete entity-awareness in success a detailed information credibility judgment. This model comprises four integrated components, namely, a reputation based model, a feature ranking algorithm, a credibility assessment classifiers engine, and a user expertise model. All of these components operate in an algorithmic form to analyze and assess the credibility of the tweets on Twitter. The reputation-based technique helps to filter neglected information before starting the assessment process. The classifier engine component distinguishes between credible and non-credible content. [3]

4. Detecting Fake News with Machine Learning Method.
Supanya Aphiwongsophon and Prabhas Chongstitvatana Proposed Detecting Fake News with Machine Learning Method, In this paper Fake news can be accurately identified using machine learning methods. In this paper selected data collected from Twitter are summarized with twenty two attributes. From this information, all the machine learning methods: Naive Bayes, Neural Network, Support vector machine, are very good at detecting Fake news There are
classified to two classes with believable and unbelievable. The results from the classification are precision, recall, F-Measure, and accuracy. Neural network and Support Vector Machine are equivalently results with recall, F-measure, and accuracy are 99.90. [4]

5. Automatically Identifying Fake News in Popular Twitter Threads.Cody Buntain, Jennifer Golbeck (2017) Proposed Automatically Identifying Fake News in Popular Twitter Threads. In This paper select the dataset form Twitter by CREDBANK crowdsourced dataset ,PHEME journalist-labeled dataset.They predict accurancy assessments in two credibility focused twitter. PHEME is a curated data set of conversation threads about rumors in Twitter complete with journalist annotations for truth, and CREDBANK is a large-scale set of Twitter conversations about events and corresponding crowdsourced accuracy assessments for each event. While both datasets are built on accuracy assessments, we theorize this question captures two separate qualities for PHEMEs journalists, accuracy is objective or factual truth, whereas CREDBANKs crowdsourced workers equate accuracy with credibility, or how believable the story.better accurancy result for CREADBANK dataset. [5]

III SYSTEM ARCHITECTURE

Features Selection: In this report, use by the twitter dataset for feature selection. The initial motivation for feature selection is that the social data often contain many different features that are difficult to deal with this feature, and most of the features are terminated except for specific tasks. To deal with this problem, Apply feature extraction methods. Feature selection is often preferred over extraction, because the selected features have more understandable and usefulthey

select the three main feature Structural Features Structural features capture Twitter-specific properties of the tweet stream, including tweet volume and activity distributions. User features capture properties of tweet authors, such as interactions, account ages, friend/follower counts, and Twitter verified status. Content features measure textual aspects of tweets, like polarity, subjectivity, and agreement.

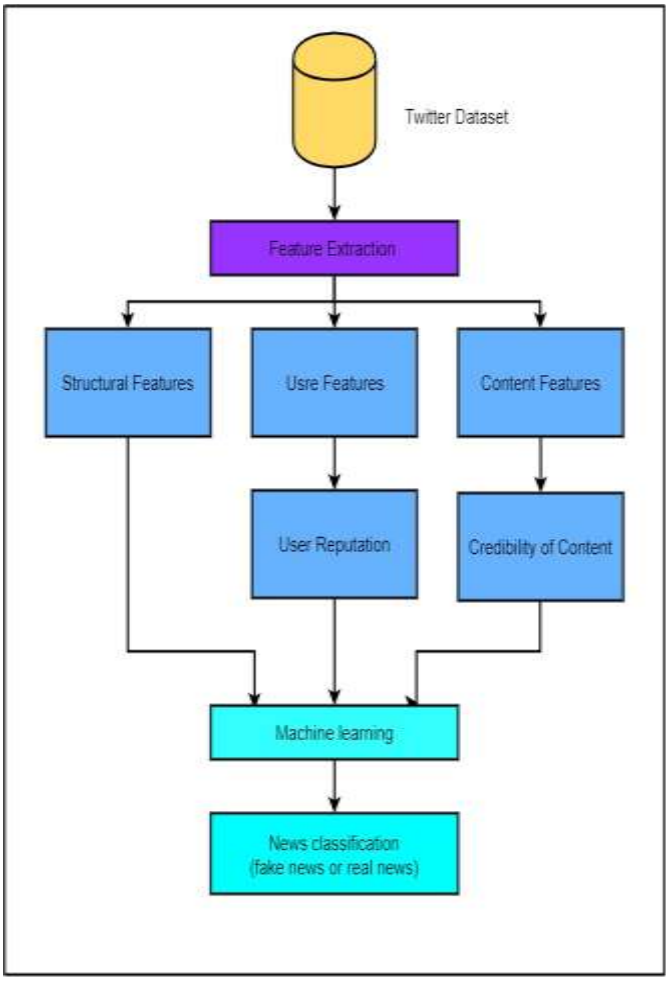

Figure 1: Architecture Diagram

IV RESULTS

Figure 2: Home Page

Figure 3: Load Dataset

Figure 4: BuzzFeed News Fact-Checking Dataset

V CONCLUSIONS

In this project identifying misinformation is authoritative in online social media platforms, because information is circulated easily across the online community by unsupported sources. To be able to automatically detect fake news and stop misinformation circulation, can be useful in analyzing activist movements. User accounts who include many URLs, @username mentions and hash tags in their tweet. This paper also solve the problem of assessing information credibility on Twitter. The issue of information credibility has come under scrutiny, especially in social networks that are now being used actively as first sources of information. structures as well as the users sentiments to identify and evaluate topically relevant and credible sources of information and User reputation.

REFERENCES

[1] Shlok 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] Mohammad Mehedi Hassan, Member, IEEE and Atif Alamri, A Credibility Analysis System for Assessing Information on Twitter Member, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING.

[4] Supanya Aphiwongsophon , Prabhas Chongstitvatana,Detecting Fake News with Machine Learning Method,IEEE.

[5] Cody Buntain ,Jennifer Golbeck,Automatically Identifying Fake News in Popular Twitter Threads. 2017 IEEE International Conference on Smart Cloud.

[6] Juan Cao, Junbo Guo, Xirong Li, Zhiwei Jin, Han Guo, and Jintao Li,Automatic Rumor Detection on Microblogs: A Survey, arXiv:1807.03505v1 [cs.SI] 10 Jul 2018.

[7] Kai Shu,SuhangWang, Huan Liu,Understanding User Profiles on Social Media for Fake News Detection,2018 IEEE Conference on Multimedia Information Processing and Retrieval

[8] Tanushree Mitra and Eric Gilbert,CREDBANK: A Large-Scale Social Media Corpus with Associated Credibility Annotations ,School of Interactive Computing GVU Center Georgia Institute of Technology tmitra3, gilbert@cc.gatech.edu,Proceedings of the Ninth International AAAI Conference on Web and Social Media.

[9] Jacob Ross, Krishnaprasad Thirunarayan,Features for Ranking Tweets Based on Credibility Newsworthiness Kno.e.sis: Ohio Center of Excellence in Knowledge-enabled Computing Department of Computer Science and Engineering Wright State University Dayton, Ohio 45435 ross.138, t.k.prasad@wright.edu,2016 International Conference on Collaboration Technologies and Systems