Detection of Fake Twitter Accounts with Machine Learning Algorithms

Gayatri Nair[1], Vaibhav Davande[2], Kajal Dewade[3], Shraddha Gupta[4], Prof.P.R. Kulkarni[5]
[1,2,3,4,5]*SIEM, Computer Engineering Dept., Nashik*

Abstract **- Social networks have become a part of human life in many areas today. Many activities such as communication, promotion, advertisement, news and agenda have been started to be carried out over social networks. Some malicious accounts on Twitter are used for purposes such as creating false information and agenda. This is one of the main problems in social networks. Therefore, it is important to detect malicious accounts. In this study, machine learning-based methods were used to detect fake accounts that could mislead people. The dataset created for this purpose was pre-processed and fake accounts were detected by machine learning algorithms. Decision tree, logistic regression and support vector machine algorithms are used to detect fake accounts. The classification success of these methods has been compared and it has been proven that logistic regression gives more successful results.**

Index Terms **- Social Networks, Twitter, Bot Detection, Advanced Machine Learning, Classification.**

I.INTRODUCTION

LOGIN

The extensive use of social media has made it easier for these environments to be ignored by malicious people. Twitter social media application, known for its corporate identity, is used by all segments of the society. According to Alexa, Twitter is among the most visited sites in the world.

Despite coming in the ranks, abuse has also increased in this social platform [1]. It is possible to easily access Twitter's application from different platforms [2]. This attracts malicious users more. Often malicious Twitter accounts have goals such as gaining more followers, affecting a certain community, making people into their own organizations, manipulating people for the stock market, spreading fake news, and using private information to blackmail people.

In social networks, bots are computer software that display automatic reactions that imitate human behavior [3]. Today, it is thought that there are 15% bots among active Twitter users [4]. Cyborg accounts are similar to bots. These accounts are half bot half human characteristics. It is opened by people, but their next actions are similar to bot accounts [5].

II. EXISTING SYSTEM

In paper [1], a review of the number of data mining approaches used to detect anomalies. An uncommon direction is made to the investigation of informal community driven irregularity identification systems which are comprehensively named execution based, structure based and phantom based. Every last one of this gathering further fuses a number of procedures which are talked in the paper. The paper has been closed with various future headings and territories of research that could be tended to and worked upon.

In paper [2], it is said that the Twitter trends are secure from the manipulation of hateful users. We gather in excess of 69 million tweets from 5 million records. Utilizing the gathered tweets, we first direct an information investigation and find proof of Twitter pattern control. At that point, we learn at the subject level and derive the key factors that can decide if a theme starts slanting because of its prevalence, inclusion, transmission, potential inclusion, or notoriety. What we find is that with the exception of transmission, all of the elements above are firmly identified with inclining. At long last, we further research the inclining control from the viewpoint of traded off and phony records and talk about countermeasures.

In paper [3], social network users build trust relationships with the accounts they follow. This conviction can create for an assortment of reasons. For instance, the client may know the proprietor of the believed record face to face or the record may be worked by an element regularly considered as dependable, for example, a general news office.

Improperly, should the power over a record fall under the control of a digital lawbreaker, he can trust without much of a stretch adventure this trust to facilitate his very own pernicious. We show how we can utilize comparable methods to distinguish bargains of individual high-profile accounts. High-profile accounts much of the time have one trademark that makes this location solid; they show reliable conduct after some time. We demonstrate that our framework, were it sent, would have had the option to distinguish and counteract three true assaults against well-known organizations and news exercises. Moreover, our framework, as opposed to well-known media, would not have fallen for an organized trade off affected by a US café network for attention reasons.

In paper [4], they proposed a new credibility analysis system for assessing information authority on Twitter to prevent the proliferation of fake or malicious information. The proposed framework comprises four incorporated segments: a notoriety-based segment, a validity classifier motor, a user experience component, and a feature ranking algorithm. The segments operate together in an algorithmic structure to study and access the reliability of Twitter tweets and users. They also tested the performance of the system based on two different datasets from 489,330 premium twitter accounts.

In paper [5], they forward the research discussed in this paper and apply these same engineering features to a set of fake human accounts in the hope of advancing the successful detection of fake identities created by humans on SMPs. Diverse way to deal with contemplating validity on Twitter: they looked to demonstrate how name worth inclination influences the decisions of micro blog creators. In this investigation, the creator demonstrated the relationship between name worth inclination and the quantity of adherents. Accordingly, it is hard to gauge the validity of a client in these systems and to check his/her posts. As online interpersonal organizations have turned out to be progressively valuable for dispersing data to more extensive crowds, tending to the previously mentioned difficulties to decide the validity of clients in OSNs requires the advancement of hearty methods for estimating client and substance believability.

They have investigated the nature of spam users on Twitter with the goal to improve existing spam detection mechanisms. For detecting Twitter spammers, they have used several new features, which are more effective and robust than existing used features.

III. OBJECTIVES

To implement different algorithms to get better Spam Detection i.e., IP Address, Account used, Negative Word Dictionary using Senti-strength, Ontology. Graphical representation of work. To deal with 6 different types of Spam Reviews. To present Opinion Mining on Spam Filtered Data. To implement Ontology in Spam Detection To present an algorithm that does Opinion Mining with Spam Detection.

IV. SYSTEM ARCHITECTURE

There are various ways to detect Spam Reviews in order to the Opinion mining to be more accurate and useful have been studied. A detailed discussion about the existing techniques, to find out whether the review is spam or not is presented. Other Techniques are incorporated like IP Address Tracking and Ontology to detect Spam Reviews in order to get more accurate results from Opinion mining.

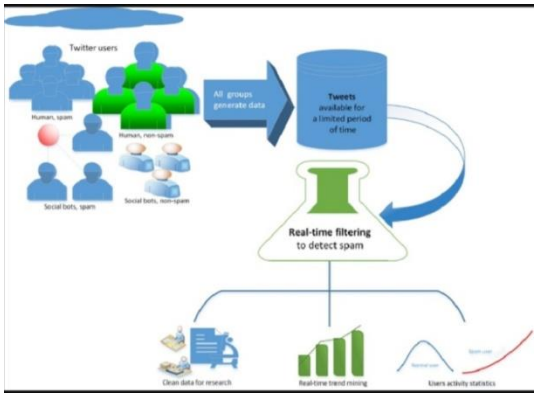

Figure 4.1: System Architecture

For detection of fake online reviews, we start with raw text data. We have used a dataset which was already labeled by the previous researchers. We remove unnecessary texts like articles and prepositions in the data. Then these text data are converted into numeric data for making them suitable for the classier.

Figure 4.2: Data Flow Diagram 0

Important and necessary features are extracted and then classification process took place. As we have

used 'gold standard' dataset prepared by Ott et al. [3], we did not require the steps like handling missing values, removing inconsistency, removing redundancy etc. Instead, we needed to merge the texts, create a dictionary, and map the texts to numeric value as the tasks of preprocessing. We have used word frequency count, sentiment polarity and length of the review as our features. We have taken 2000 words as features. Hence the size of our feature vector is 1602002. We have not taken n-gram or parts of speech as features because these are the derived features from a bag of words and may cause over-fitting. The process of feature extraction is summarized in the figure 1. From the figure 1, we can see that, when we are working with i'th review, it's corresponding features are generated in the following procedure.

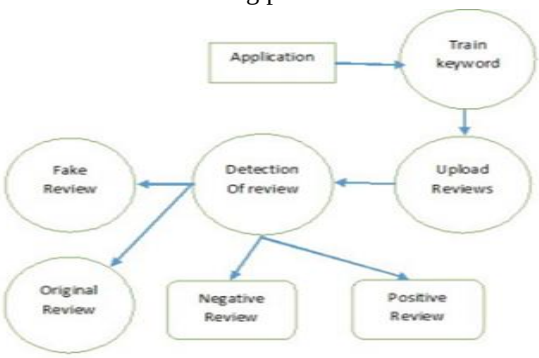

Figure 4.3: Data Flow Diagram 1

Each review goes through the tokenization process rst. Then, unnecessary words are removed, and candidate feature words are generated.

Each candidate feature words are checked against the dictionary and if its entry is available in the dictionary then it's frequency is counted and added to the column in the feature vector that corresponds to the numeric map of the word.

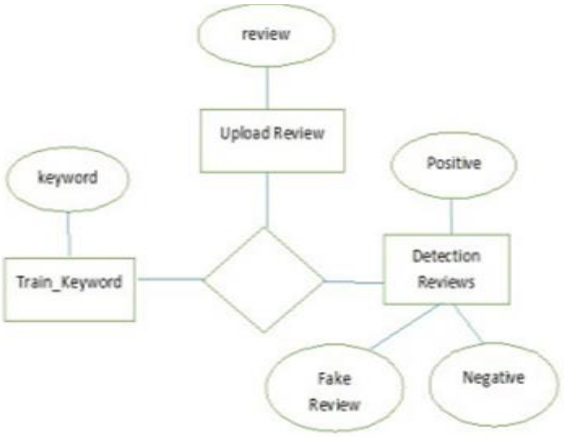

Figure 4.4: Data Flow Diagram 2

Alongside with counting frequency, the length of the review is measured and added to the feature vector.

Finally, the sentiment score which is available in the data set is added in the feature vector. We have assigned negative sentiment as zero valued and positive sentiment as some positive valued in the feature vector.

We have implemented both semi-supervised and supervised classifications. For semi-supervised classification of the data set, we have used Expectation-Maximization (EM) algorithm. The Expectation Maximization algorithm, first proposed by Karimpour et al. [9], is designed to label unlabeled data to be used for training. The algorithm operates as follows: A classier is first derived from the labeled dataset. This classier is then used to label the unlabeled dataset. Let this predicted set of labels be PU. Now, another classifier is derived from the combined sets of both labeled and unlabeled datasets and is used to classify then labeled dataset again. This process is repeated until the set PU stabilizes. After a stable PU set is produced, we have trained the classification algorithm with the combined training set of both labeled and unlabeled datasets and deploy it for predicting test dataset [8]. The algorithm is given below.

V. PROPOSED SYSTEM

The drawback in the existing system have been overcome by implementing this Application. Our proposed model is still based on use of People use Twitter to share their feelings, news, events and to post their daily activities such as eating, drinking, travelling and so forth. Therefore, malicious users can check everyone's activities from their timeline and twitter becomes a place for hateful users to commit the frauds. These users which are having hostile intentions create fake accounts and spread various fake news, fake links and photos. Most of the internet users are not aware of these fake accounts; they accepted the requests and suffer in the process. Therefore, detecting fake accounts on twitter is obligatory for everyone who uses it

VI. ADVANTAGES

Due to machine learning techniques, it improves accuracy of fake account detection systems. (The network or computer is constantly monitored for any invasion or attack.

(The major advantage of Support Vector Machines that classify our composite data model.
(Twitter's major advantage is, Twitter has limited message size of 140 characters per tweet, it can include a message or link on your website as it is free and also free for the advertisements, you do not have to face the problem with bunch of posters like the other social networking

VII. CONCLUSION

we have maintained the highest accuracy in detecting fake accounts by different classifying algorithms. The results show the increase of the accuracy results of two of the classification algorithms after using the suggested attributes with their corresponding heaviness. The classification algorithms are proposed to improve detecting fake accounts on social networks, where the SVM trained model.

VII. FUTURE SCOPE

Implement Strong Security and develop for organization. We have to implement a system to Fake reviews have proliferated in every area of ecommerce, from electronics to clothes to books to children's toys. Review tracking website Fake spot, which analyzes reviews from popular e-commerce sites, estimates that a third of online reviews on sites like Walmart.com, Amazon.com, and Sephora.com are fake.

VIII. ACKNOWLEDGEMENT

We are thankful to our guide Prof.P.R. KULKARNI, SIEM, Nashik for the Guidance. We needed his essential guidance and suggestions.

REFERENCES

[1] R.Kaur and S.Singh, "A survey of data mining and social network analysis based anomaly detection techniques", Egyptian informatics diary, vol.17, no.2, pp.1992-216, 2016.
[2] Yubao Zhang, Xin Ruan, Haining Wang, Hui Wang, and Su He "Twitter Trends Manipulation: A First Look Inside the Security of Twitter Trending" IEEE Exchanges on Data Crime scene investigation and Security (Volume: 12, Issue: 1, Jan. 2017).
[3] Manuel Egele, Gianluca Stringhini, Christopher Kruegel, and Giovanni Vigna "Towards Detecting Compromised Accounts on Social Networks", IEEE Exchanges on Reliable and Secure Processing (Volume: 14, Issue: 4, July-Aug. 1 2017).
[4] Majed Alrubaian, Muhammad Al-Qurishi, Mohammad Mehedi Hassan, and Atif Alamri, "A Credibility Analysis System for Assessing Information on Twitter", IEEE Exchanges on Reliable and Secure Processing (Volume: 15, Issue: 4, July-Aug. 1, 2018).

AUTHORS PROFILE

Gayatri Nair
Student, B.E. Computer Dept., Sandip Institute of Engineering & Management, Nashik

Vaibhav Davande
Student, B.E. Computer Dept., Sandip Institute of Engineering & Management, Nashik

Kajal Dewade
Student, B.E. Computer Dept., Sandip Institute of Engineering & Management, Nashik

Shraddha Gupta
Student, B.E. Computer Dept., Sandip Institute of Engineering & Management, Nashik