An Efficient Framework For Spammer Detection And Fake User Identificatin on Social Networks

Sivakumar M[1], Deebika B[2], Dharani N [3], Divyha S G[4], Harini S[5]
[1] Assistant Professor, Dept of CSE
[2, 3, 4, 5]Dept of CSE
[1, 2, 3, 4, 5] Sri Eshwar College of Engineering Coimbatore, Tamilnadu, India

Abstract- *Open Social network (OSN)connects millions of users worldwide in which Twitter and Facebook, have a tremendous impact on the daily life and sometimes negative consequences. Online social network sites have become a popular platform for spammers to spread false data or information. Twitter has become the most popular OSN site ever which has an unreasonable amount of spam due to fake users through unwanted tweets.The probability of spreading invalid information through fake identities to consumers has increased the results in the unrolling of harmful content. Recently, spammer detection and recognition of fake users on Twitter has become a popular research area within contemporary online social networks (OSNs). In this paper, an efficient framework is proposed to detect spammers and identify fake users. The proposed work has the following modules which include spammer word training, classification, clustering, and feature extraction. The proposed techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features*

Keywords- Spammer, Open social Network(OSN), fake user.

I. INTRODUCTION

It has gotten very simple to get any sort of data from any source over the world by utilizing the Web. The expanded interest of social destinations grants clients to gather bountiful measure of data and information about clients. Colossal volumes of information accessible on these locales likewise draw the consideration of phony clients. Twitter has quickly become an online hotspot for procuring constant data about clients. Twitter is an online Social Network (OSN) where clients can share everything without exception, for example: news, suppositions; what's more even their temperaments. A few contentions can be held over various themes, for example; governmental issues, current undertakings, and significant occasions. At the point when a client tweets something, it is quickly passed on to his/her devotees, permitting them to extend the got data at a lot more extensive level. With the development of OSNs, the need to examine and investigate client's practices in online social stages has increased. Numerous individuals who do not have a lot of data with respect to the OSNs can undoubtedly be deceived by the fraudsters. There is likewise an interest to battle furthermore, place a control on the individuals who use OSNs just for ads and in this way spam others records. As of late, the recognition of spam in informal communication locales pulled in the consideration of scientists. Spam recognition is a difficult task in maintaining the security of social networks. It is essential to recognizespams in the OSN sites to save users from various kinds of malicious attacks and to preserve their security and privacy. These hazardous manoeuvre adopted by spammers cause massive destruction of the community in the real world. Twitter spammers have various objectives, such as spreading invalid information, fake news, rumours and spontaneous messages. Spammers achieve their malicious objectives through advertisements and several other means where they support different mailing lists and subsequently dispatch spam messages randomly to broadcast their interests. These activities cause disturbance to the original users who are known as non-spammers. In addition, it also decreases the repute of the OSN platforms. Therefore, it is essential to design a scheme to spot spammers so that corrective efforts can be taken their malicious activities. Several research works have been carried out in the domain of Twitter spam detection. To encompass the existing state-of-the-art, a few surveys have also been carried out on fake user identification from Twitter. The aim of this paper is to identify different approaches of spam detection on Twitter and to present taxonomy by classifying these approaches into several categories like fake content, URL based spam detection, detecting spam in trending topics and fake user identification.

II. EXISTING METHODOLOGY

In existing framework, spammer has been examined dependent on content-based client profile. It doesn't uphold the URL based and profile Based client profile. Machine getting the hang of sifting based calculation executed for identification of spammers in the framework. Administrated learning is the most incessant AI approach for performing survey spam identification however acquiring named reviews for preparing is troublesome and manual ID of fake reviews

has helpless precision. This has led to many experiments using synthetic or small datasets. A final concern related to quality of data is the presence of noise, could be used to evaluate the impact of noise on performance and how its effects many reduced. Further work needs to be conducted to establish how many features are required and what types of features are the most beneficial. Feature selection should not be considered optional when training a classifier in a big data domain with potential for high feature dimensionality. A possibility of less labour intensive means of generating labelled training data is to find and label duplicate reviews as multiple studies have shown duplication or near duplication of review content is a strong indicator of review spam. There are many truthful than fake reviews online, this can be identified through data sampling techniques. Many experiments have avoided the issue by extracting only a small number of features, avoiding the use of n-grams, or by limiting the number of features, through alternative means such as using term frequencies to determine what n-grams are included as features. One of the most notable observations of current research is that experiments should use real world data if possible. As it is difficult to procure accurately labelled real-world datasets, unsupervised and semi-supervised methods are of interest. Additionally, we could find no studies that incorporated distributed or streaming implementations for learning form big data into their spam detection frameworks.In recent years, review spam detection has received significant attention in both business and academia due to the potential impact fake reviews can have on customer behaviour and purchasing decisions. Supervised learning is the most frequent machine learning approach for performing review spam detection. One of the most notable observations of current research is that experiments should use real world data if possible. Tingmin et al. provide a survey of new methods and techniques to identify Twitter spam detection. This survey presents a comparative study of the current approaches. Despite all existing studies, there is still a gap in the existing literature. Therefore, to bridge the gap, we review state-of-the-art in the spammer detection and fake User identification on Twitter.

Drawbacks:

- Less number dataset can be checked
- No real time data are used
- Spammer gets into the system
- Requires some investment
- Low exactness
- More complex

III. PROBLEM STATEMENT

The profile data in social networks consist of static and dynamic parts. Former is the information of the user and latter is the observed results of the user. The static data typically includes user demographics and interests and dynamic data relates to user activities and position in the social networks, where it has merely a smaller number of visible static profiles and no dynamic profile details to the public. Therefore in this research or goal is to identify an approach to determine the spammers and fake profiles in social networks.

IV. PROPOSED SYSTEM

In this undertaking spam identification conspire actualized in light of various sort of highlights. Various sorts of highlights is executed for recognize the spammers in the framework URL based, content based, client movement log, client profiles based plan executed for distinguish the spammers in the framework. AI calculation is prepared dependent on the highlights. We can distinguish the phony clients, spammers precisely. The user can login to their profile after registration and can works according to their wills by accepting the friend request and chat privately; also they can search their needed product for day-to-day life and can review the products. If the user found that the product site is inconvenience they can report and that site is considered as spammer.

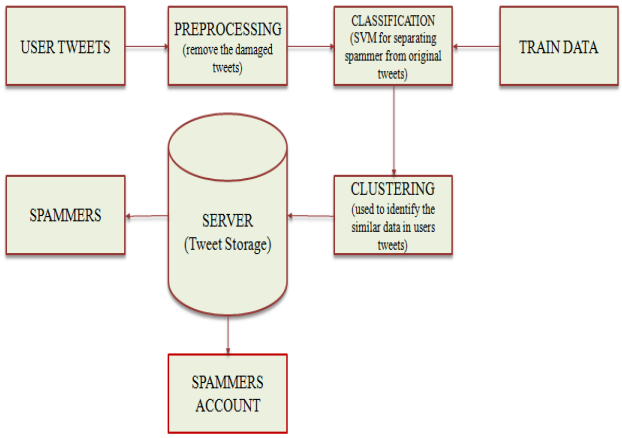


Fig.1: Architecture

The above figure is a proposed platform for efficient identification of fake users and spammer detection. Moreover, the analysis also shows that several machine learning-based techniques can be effective for identifying spams on Twitter. However, the selection of the most feasible techniques and methods is highly dependent on the available data.This study includes the comparison of various previous methodologies

proposed using different datasets and with different characteristics and accomplishments. It is tested with real time data.

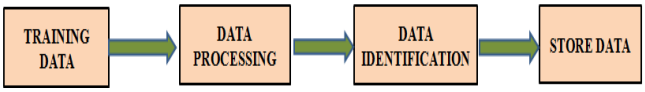
Inputs: The login details of the users are Inputs. The login details are stored in the database to create a unique profile.

Processing: The database is created and updated using MYSQL where the user information are stored

Output: The admin experience negative surveys and report spammers.

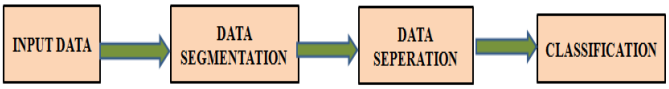
The following are the modules which are implemented:

1. ***Spammer Words Training:*** In this module, the spammer word training is used. The trained dataset is uploaded to the module to identify and separate the spammers from the other tweet messages. The spam identifying is divided into three modules:
    - Classification
    - Clustering
    - Feature extraction



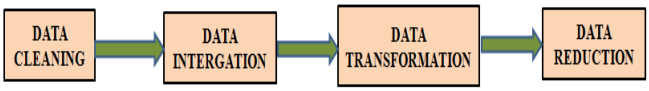
The data that are to be trained should be in the input format. The input data is processed by the training data and it is identified by data identification and finally the identified data is saved and shown as stored data which is the output.

2. ***Feature extraction:***The process of feature extraction is useful when we need to reduce the number of resources needed for processing without losing important or relevant information. It can also reduce the amount of redundant data for a given analysis. Feature extraction involves reducing the number of resources required to describe a large set of data. The general way to label an e-mail as spam or non-spam is to set up a finite set of discriminative features and use a classifier for the detection. Two different methods are proposed to select the most discriminative features among a set of reasonably arbitrary features for spam e-mail detection. Text feature extraction is the process of taking out a list of words from the text data and then transforming them into a feature set which is usable by a classifier. This work emphasizes on the review of available feature extraction methods. It starts from an initial set of measured data and builds derived values.



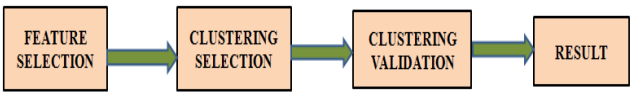
When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features. Determining a subset of the initial features is said to be feature Extraction.

3. ***Classification:*** The classification process is used to identify the category of the data's. It is used to identify impossible data combinations, missing data's, out of range value, etc., It is used to remove the damaged data's and the empty data's in the overall dataset. The two common approaches used for filtering spam mails are knowledge engineering and machine learning. Emails are classified as either spam or ham using a set of rules in knowledge engineering. A particular machine learning algorithm is then used to learn the classification rules from these email messages.



Naïve Bayes classifiers are a popular statistical technique of e-mail filtering. Here, bag of words features are used to identify spam e-mail, an approach commonly used in text classification. It is one of the oldest ways of doing spam filtering, with roots in the 1990s.

4. ***Clustering:*** Clustering is the task of grouping a set of objects in such a way that objects in the same group is more similar to each other than to those in other groups. First it, select the data in the input then it validate the input data for the analysing process. While analysing it forms a group which is of similar types and then it validate to give the output. The formation of object should be of similar to each other to form a cluster.



Advantages of proposed system:

- High accuracy
- Privacy improved
- Detect and protect the spammers

V. SYSTEM ANALYSIS

Admin Module and User Module:

In Admin module, the admin has to login by using valid username and password. After login successful he can do some operations such as adding Categories, Adding Products for that Categories, Viewing and authorizing users, View Spam accounts details, viewing friend request & response, all recommended posts, all posts with all reviews, all positive and negative reviews, removing products, viewing all purchased products, viewing positive and negative reviews chart on products. The admin adds the category details such as category name. These details will be stored into the database. He adds Product posts for categories which include details such as, product image, product name, cost, description and uses of that product. These details will be stored into the database. These details will be further searched and accessed by the users in order to recommend to their friends and to buy products. In the users module, the admin can view the list of users who all registered. In this, the admin can view the users' details such as, user name, email, address, phone number and authorize the users. The admin can view all the friend requests and responses. Here all the requests and responses will be displayed with their tags such as Id, requested user image, requested user name, username request to, status and time & date. If the user accepts the request then the status will be changed to accept or else the status will remain as waiting. The admin can view all the recommended products. If any recommendations happened for particular products, those details will be shown along with products. Details include product name, recommended user name, user recommended to name and the date.

In this, the admin can view all posts with their Positive and Negative Comments posted by users based on their opinions.

Positive: If the user comment contains at least one of the words which are listed in positive words, then that comment will be treated as a positive comment.

Negative: If the user comment contains at least one of the words which is listed in negative words, then that comment will be treated as a negative comment. The comments of all posts will be displayed. Comments include Positive, Negative, Non-Positive and Non-Negative. It includes details such as, commented user name, comment and date, the products which are purchased by users will be displayed. It includes details such as, purchased user name, and purchased products, price of the products and the date of purchase. The number of positive Reviews got by the particular product will be displayed in a chart. The number of negative Reviews got by the particular product will be displayed in a chart. In this module, the products which have got the negative comments from more than five users will be listed and removed by the admin.

In User Module, there are N numbers of users present. Users should register before doing any operations. Once a user registers, their details will be stored to the database. After registration is successful, he has to login by using authorized user name and password. Once login is successful user will do some operations like viewing their profile account details like spam or normal, search users and send friend requests, viewing friend requests, searching posts and recommendations to friends and viewing all product recommendations sent to him by his friends, commenting on posts, purchasing products and viewing their product search history. The user can search the users based on names and the server will give responses to the user like User name, user image, E mail id, phone number and date of birth. If you wish to send a friend request to a particular user then click on the "request" button, then request will be sent to that particular user. In this, the user searches for products based on the products description. The user can recommend searched products to his friends, comment on posts and he can add the products to cart to buy those added products later by using their created account. The user can view the friend requests which are sent by other users. Which includes sending user details with their tags such as user name, user image, date of birth, Email ID, phone number and Address and user can accept the request by clicking on the "waiting" link. The user can view all the products which are recommended by his friends. This includes recommended user name and his image, recommended products details. He can view all the searched products names and categories, the keywords which he used to search the products. This includes details such as, searched product, used keyword and date of search. He can create his bank account by providing details such as, account number, branch, address, email id. Later he can add money to his account and can view his account details.

VI. CONCLUSION

In this paper,a new model is implemented for spammer detection and fake user identification on tweeter where false data and unwanted tweets can be reduced. It also explains the taxonomy of Twitter spam detection with the functions of fake content, spam based on URL, spam in trending topics, fake users. False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level. Another associated topic that is worth

investigating is the identification of rumour sources on social media. Spammer detection in an open social network is a serious issue for researchers to be focussed more in the future to provide 100% percent accuracy in the result.

REFERRENCES

[1] S. Ghosh, G. Korlam, and N. Ganguly, "Spammers' networks within online social networks: A case-study on Twitter," in Proc. 20th Int. Conf.Companion World Wide Web, Mar. 2011, pp. 41_42.

[2] C. Chen, S. Wen, J. Zhang, Y. Xiang, J. Oliver, A. Alelaiwi, and M. M. Hassan, "Investigating the deceptive information in Twitter spam," Future Gener. Comput. Syst., vol. 72, pp. 319_326, Jul. 2017.

[3] I. David, O. S. Siordia, and D. Moctezuma, "Features combination for the detection of malicious Twitter accounts," in Proc. IEEE Int. Autumn Meeting Power, Electron. Comput.(ROPEC), Nov. 2016, pp. 1_6.

[4] M. Babcock, R. A. V. Cox, and S. Kumar, "Diffusion of pro- and anti-false information tweets: The black panther movie case," Comput. Math. Org. Theory, vol. 25, no. 1, pp. 72_84, Mar. 2019.

[5] S. Keretna, A. Hossny, and D. Creighton, "Recognising user identity in Twitter social networks via text mining," in Proc. IEEE Int. Conf. Syst., Man, Cybern., Oct. 2013, pp. 3079_3082.

[6] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, "A machine learning approach for Twitter spammers detection," in Proc. Int. Carnahan Conf. Secur. Technol. (ICCST), Oct. 2014, pp. 1_6.

[7] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Real-time Twitter content polluter detection based on direct features," in Proc. 2nd Int. Conf. Inf. Sci. Secur. (ICISS), Dec. 2015, pp. 1_4.

[8] H. Shen and X. Liu, "Detecting spammers on Twitter based on content and social interaction," in Proc. Int. Conf. Netw. Inf. Syst. Comput., pp. 413_417, Jan. 2015.

[9] G. Jain, M. Sharma, and B. Agarwal, "Spam detection in social media using convolutional and long short term memory neural network," Ann. Math. Artif.Intell., vol. 85, no. 1, pp. 21_44, Jan. 2019. 57