ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Machine Learning Framework For Identification Of Fake User On Social Networks

Mr. B. Naga Raju,[1]N. Sri Hasa[2], N. Apurupa[3], P. Tejaswini, N. Swarnanjali[5], Sk. Sameer[6]

[1]Asst. Professor, Department of Computer Science and engineering

[2,3,4,5,6]Student, Department of Computer Science and engineering

[1,2,3,4,5,6]QIS College of Engineering and Technology, Vengamukkapalem, Ongole-523272

Abstract-Uncovering spammers has become one of the most challenging problems in OSNs. It is essential to recognize fake accounts to preserve their security and privacy. spammers have various objectives, such as spreading invalid information, fake news, rumors, and spontaneous messages. These activities cause disturbance to the original users and it 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 to counter their malicious activities. The current project involves the development of Describing concept to detect spam tweets and fake user account from online social network called twitter. To perform detection, we use twitter dataset and 4 different techniques called Fake Content, Spam URL Detection, Spam Trending Topic and Fake User Identification. Using above 4 techniques we can identify whether tweet is normal or spam.

Keywords— Social network, Spammer, Machine learning, Extreme learning machine

I. INTRODUCTION

It has become quite unpretentious to obtain any kind of information from any source across the world by using the Internet. The increased demand of social sites permits users to collect abundant amount of information and data about users. Huge volumes of data available on these sites also draw the attention of fake users [1]. Twitter has rapidly become an online source for acquiring real-time information about users. Twitter is an Online Social Network (OSN) where users can share anything and everything, such as news, opinions, and even their moods. Several arguments can be held over different topics, such as politics, current affairs, and important events. When a user tweets something, it is instantly conveyed to his/her followers, allowing them to outspread the received information at a much broader level [2]. With the evolution of OSNs, the need to study and analyze users‘ behaviors in online social platforms has intensified. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people‘s accounts.

Recently, the detection of spam in social networking sites attracted the attention of researchers. Spam detection is a difficult task in maintaining the security of social networks. It is essential to recognize spams in the OSN sites to save users from various kinds of malicious attacks and to preserve their security and privacy. These hazardous maneuvers 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, rumors, 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 to counter their malicious activities [3].

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

Several research works have been carried out in the domain of Twitter spam detection. To encompass the existing state-ofthe-art, a few surveys have also been carried out on fake user identification from Twitter. Tingmin et al. [4] provide a survey of new methods and techniques to identify Twitter spam detection. The above survey presents a comparative study of the current approaches. On the other hand, the authors in [5] conducted a survey on different behaviors exhibited by spammers on Twitter social network. The study also provides a literature review that recognizes the existence of spammers on Twitter social network. Despite all the 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. Moreover, this survey presents a taxonomy of the Twitter spam detection approaches and attempts to offer a detailed description of recent developments in the domain.

The aim of this paper is to identify different approaches of spam detection on Twitter and to present a taxonomy by classifying these approaches into several categories. For classification, we have identified four means of reporting spammers that can be helpful in identifying fake identities of users. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Table 1 provides a comparison of existing techniques and helps users to recognize the significance and effectiveness of the proposed methodologies in addition to providing a comparison of their goals and results. Table 2 compares different features that are used for identifying spam on Twitter. We anticipate that this survey will help readers find diverse information on spammer detection techniques at a single point.

we introduce SIGPID, a malware detection system based on permission usage analysis to cope with the rapid increase in the number of Android malware. Instead of extracting and analyzing all Android permissions, we develop 3-levels of pruning by mining the permission data to identify the most significant permissions that can be effective in distinguishing between benign and malicious apps. SIGPID then utilizes machine-learning based classification methods to classify different families of malware and benign apps. Our evaluation finds that only 22 permissions are significant. We then compare the performance of our approach, using only 22 permissions, against a baseline approach that analyzes all permissions. The results indicate that when Support Vector Machine (SVM) is used as the classifier, we can achieve over 90% of precision, recall, accuracy, and F-measure, which are about the same as those produced by the baseline approach while incurring the analysis times that are 4 to 32 times less than those of using all permissions. Compared against other state-of-the-art approaches, SIGPID is more effective by detecting 93.62% of malware in the data set, and 91.4% unknown/new malware samples.

II. RELATEDWORKS

Twitter has rapidly become an online source for acquiring real-time information about users. When a user tweets something, it is instantly conveyed to his/her followers, allowing them to outspread the received information at a much broader level. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensied. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people's accounts.In the existing system no accurate spam detection system that why lot of spam account could not be identified in this way lots of carpeted data was coming in to the social network.

C.Chen et.al has proposed Statistical structures built constant identification of drifted Twitter spam-Twitter spam has become a major topic now a days. Late works centered on relating AI methods for Twitter spam location which utilize the measurable features of tweets. Here tweets acts

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

as a data index, be that as it may, we see that the factual belongings of spam tweets vary by certain period, and in this way, the presentation of prevailing AI built classifiers reduces. This problem is alluded to as "Twitter Spam Drift". In order to switch this dispute, , we first do a deep investigation on the measurable features for more than one million spam and non-spam tweets. At this point we suggest a new Lfun conspire. The projected plan is changing spam tweets since unlabelled tweets and consolidates them into classifier's preparation procedure. Numerous tests are made to measure the projected plan. The results show the present Lfun plan can altogether improve the spam discovery exactness in genuine world scenarios.[9]

C. Buntain and J. Golbeck has proposed Automatically recognizing phony news in prevalent Twitter strings Information quality in online life is an undeniably significant issue, however web-scale information impedes specialists' capacity to evaluate and address a significant part of the incorrect substance, or "phony news," current stages in this paper builds up a technique for computerizing counterfeit news location on Twitter by figuring out how to foresee precision evaluations in two validity cantered Twitter datasets: CREDBANK, which supports the exactness for instance in Twitter a publicly supported dataset of exactness appraisals for occasions in Twitter, and PHEME, which contains a set of rumours and nonrumours, We use this to Twitter set content taken from BuzzFeed's fake news dataset and models arranged against freely reinforced experts beat models reliant on journalists' assessment and models arranged on a pooled dataset of both openly upheld workers and authors. All of the three datasets, balanced into a uniform group, is additionally openly accessible. An element examination at that point recognizes features that are generally prescient for publicly supported and journalistic precision evaluations, consequences which can be related with previous results.[10] C. Chen et.al has performed A performance evaluation of machine learning based streaming spam tweets detection-the popularity of twitter Twitter pulls in an ever increasing number of spammers. Spammers send undesirable tweets to Twitter clients to advance sites or administrations, here destructive to typical clients. So as to stop spammers, scientists have proposed various components. The focal point of late workings is based on utilization of AI methods into Twitter spam location. In any case, tweets are recovered in a gushing way, and Twitter gives the Issuing API to designers and analysts to get to open tweets continuously. There come up short on a presentation valuation of present AI created gushing spam recognition techniques. Here we crossed over any barrier via doing a presentation valuation that is since 3 distinctive shares of data, features, and ideal. For constant spam location, here extricated 12 lightweight features for tweet portrayal. Spam location was then changed to a double arrangement issue in the component space and can be explained by regular AI calculations. We assessed the effect of various components to the spam recognition execution that included non-spam to spam proportion, highlight discretization preparing data size, time related data, data testing, and AI calculations. The outcomes show the spilling spam tweet discovery is as yet a major test and a strong location system should consider the three parts of information, include, and model.[11]

F. Fathaliani and M. Bouguessa has proposed A modelbased methodology for recognizing spammers in interpersonal organizations In this paper, we see the errand of distinguishing spammers in informal communities from a blend displaying viewpoint, in view of which we devise a principled unaided way to deal with identify spammers. In our methodology, we initially speak to every client of the informal community with an element vector that mirrors its conduct and connections with different members. Next, in light of the evaluated clients Highlight vectors, we propose a measurable system that uses the Dirichlet circulation so as to distinguish spammers. The proposed methodology can naturally segregate among spammers and genuine clients, while existing solo approaches require human intercession so as to set casual edge parameters to distinguish spammers. Besides, our methodology is general as in it very well may be applied to various online social destinations. To

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

exhibit the appropriateness of the proposed technique, we led probes genuine information extricated from Instagram and Twitter.[15]

C. Meda et.al has proposed Spam identification of Twitter traffic: A system dependent on irregular backwoods and non-uniform element inspecting Law Enforcement Agencies spread an essential job in the examination of open information and need powerful strategies to channel problematic data. In a genuine situation, Law Enforcement Agencies break down Social Networks, for example Twitter, observing occasions and profiling accounts. Clients' characterization and spammers' ID is a helpful method for mitigate Twitter traffic by unhelpful substance. Analyses are done on a prominent datasets of Twitter clients. The given Twitter dataset is comprised of clients marked as genuine clients or spammers, portrayed by 54 features. Exploratory results exhibit the viability of improved highlight testing technique.[21]

III. PROPOSED SYSTEM ARCHITECTURE

In this paper, we perform a review of techniques used for detecting spammers on Twitter. Moreover, taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The presented techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features.

Advantages of proposed system

We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform.

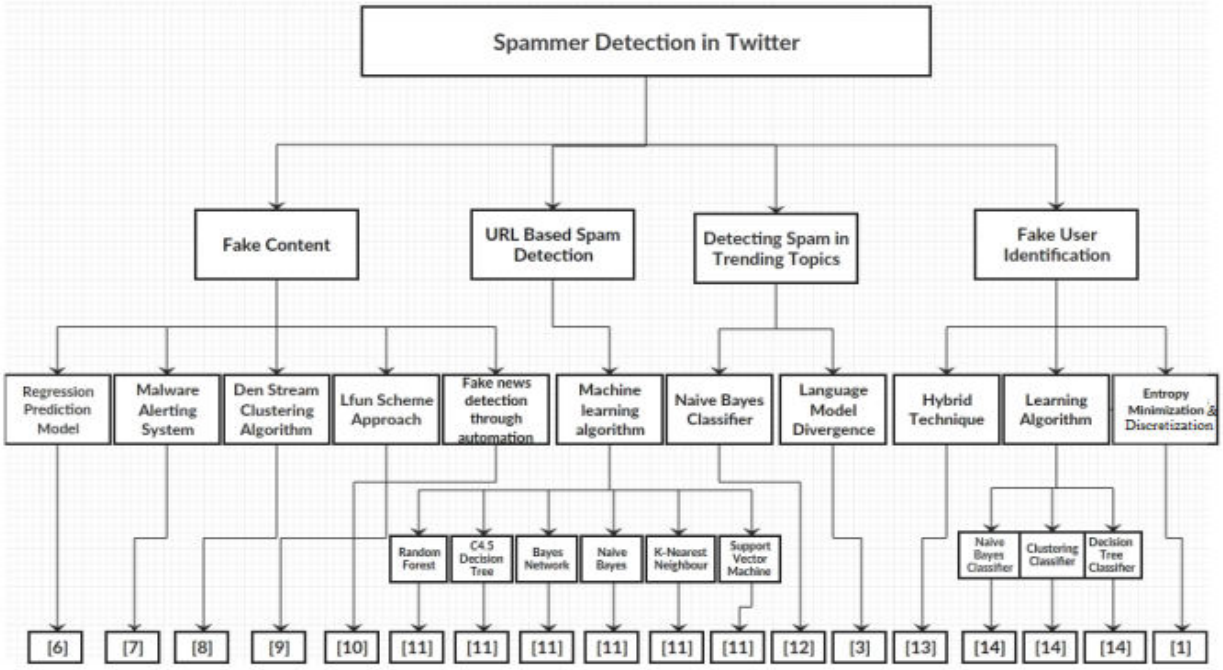

Fig. 1 Proposed system architecture

Description of 4 techniques to detect tweet is spam or normal.

The presented techniques are also compared based on various features, such as user features (retweets, tweets, followers etc.), content features (tweet content messages). 1) Fake Content: If the number of followers is low in comparison with the number of followings, the credibility of an

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

account is low and the possibility that the account is spam is relatively high. Likewise, feature based on content includes tweets reputation, HTTP links, mentions and replies, and trending topics. For the time feature, if many tweets are sent by a user account in a certain time interval, then it is a spam account.

2) Spam URL Detection: The user-based features are identified through various objects such as account age and number of user favourites, lists, and tweets. The identified user-based features are parsed from the JSON structure. On the other hand, the tweet-based features include the number of (i) retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs. Using machine learning algorithm called Naïve Bayes we will check whether tweets contains spam URL or not.

3) Detecting Spam in Trending Topic: In this technique tweets content will be classified using Naïve Bayes algorithm to check whether tweet contains spam or non-spam words. This algorithm will check for spam URL, adult content words and duplicate tweets. If Naïve Bayes detect tweet as SPAM then it will return 1 and if not detected any SPAM content then Naïve Bayes will return 0.

4) Fake User Identification: These attributes include the number of followers and following, account age etc. Alternatively, content features are linked to the tweets that are posted by users as spam bots that post a huge amount of duplicate contents as contrast to non-spammers who do not post duplicate tweets. In this technique features (following, followers, tweet contents to detect spam or non-spam content using Naïve Bayes Algorithm) will be extracted from tweets and then classify those features with Naïve Bayes Algorithm as spam or non-spam. Later this features will be train with random forest algorithm to determine account is fake or non-fake. All extracted features will be saved inside features.txt file. Naïve Bayes classifier saved inside model‘ folder.

Using above techniques we can detect whether tweets contains normal message or spam message. By detecting and removing such spam messages help social networks in gaining good reputation in the market. If social networks did not remove spam messages then its popularity will be decreases. Now a days all users are heavily dependent on social networks to get current news and business and relatives information and thus protecting it from spammer help it to gain reputation.

IV. RESULTS AND DISCUSSION

The results obtained after executing the implementation code is shown from Fig.2 to Fig.8.

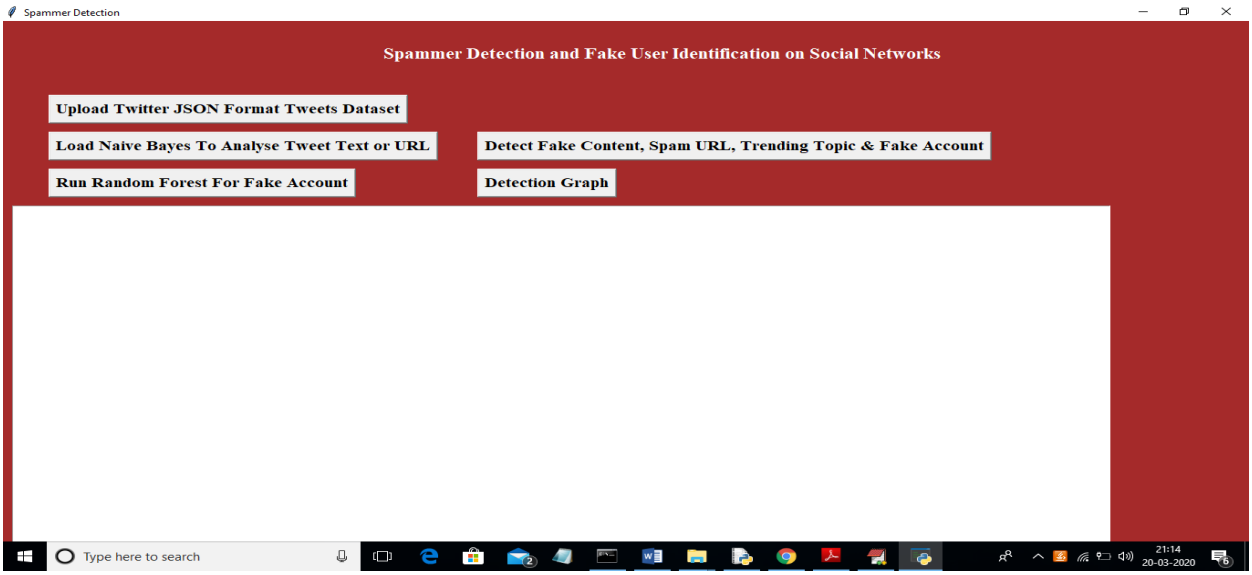

Fig. 2Upload Twitter JSON Format Tweets Dataset

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen click on Upload Twitter JSON Format Tweets Dataset‘ button and upload tweets folder

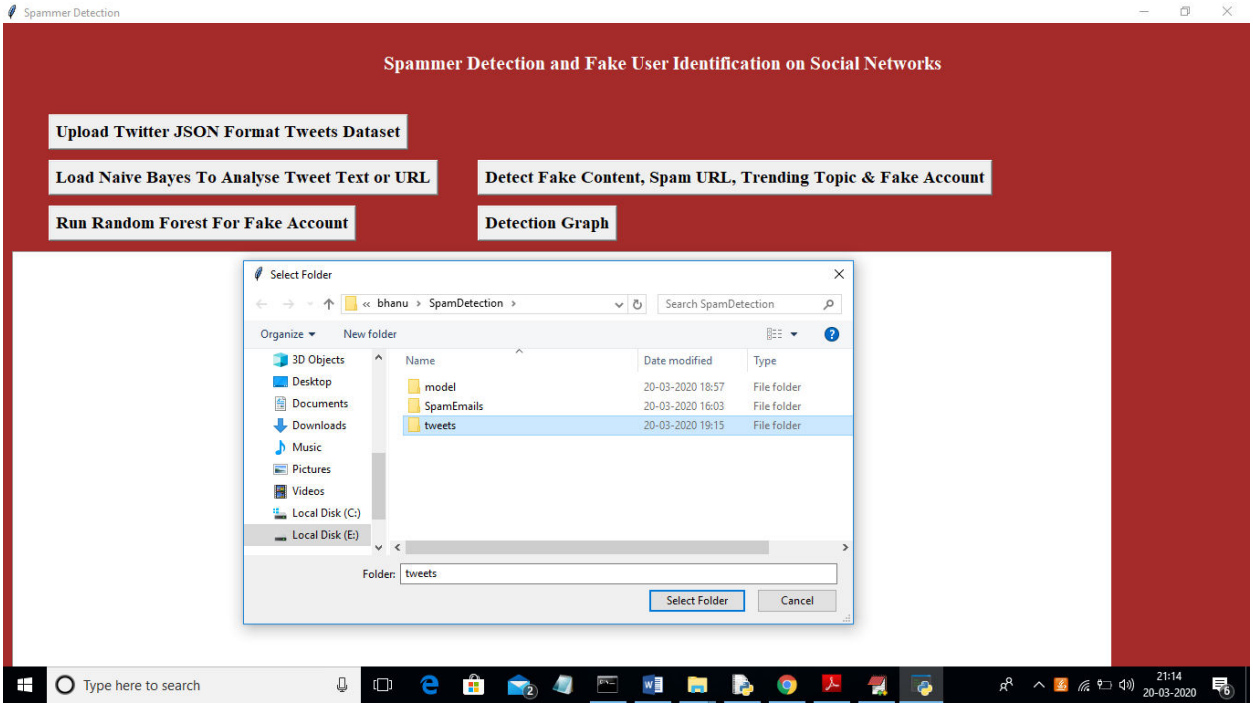

Fig.3Browse Twitter JSON Format Tweets Dataset

In above screen I am uploading tweets‘ folder which contains tweets from various users in JSON format. Now click open button to start reading tweets.

Fig. 4 Load Naive Bayes ToAnalyse Tweet Text or URL

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen we can see all tweets from all users loaded. Now click on Load Naive Bayes ToAnalyse Tweet Text or URL‘ button to load Naïve Bayes classifier

Fig.5 Detect Fake Content, Spam URL, Trending Topic & Fake Account

In above screen naïve bayes classifier loaded and now click on Detect Fake Content, Spam URL, Trending Topic & Fake Account‘ to analyse each tweet for fake content, spam URL and fake account using Naïve Bayes classifier and other above mention technique.

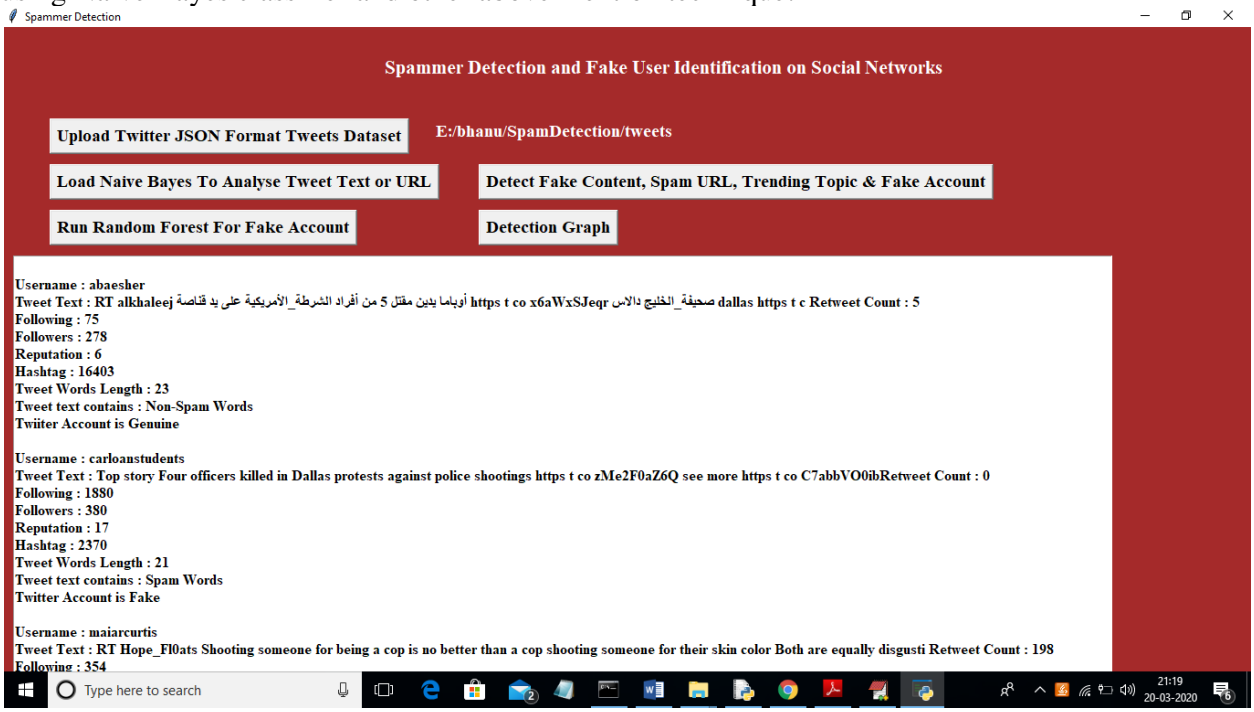

Fig. 6 Spam detection

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above screen all features extracted from tweets dataset and then analyse those features to identify tweets is no spam or spam. In above text area each records values are separated with empty line and each tweet record display values as TWEET TEXT, FOLLOWERS, FOLLOWING etc with account is fake or genuine and tweet text contains spam or non-spam words. Now click on _Run Random Forest Prediction' button to train random forest classifier with extracted tweets features and this random forest classifier model will be used to predict/detect fake or spam account for upcoming future tweets. Scroll down above text area to view details of each tweet.

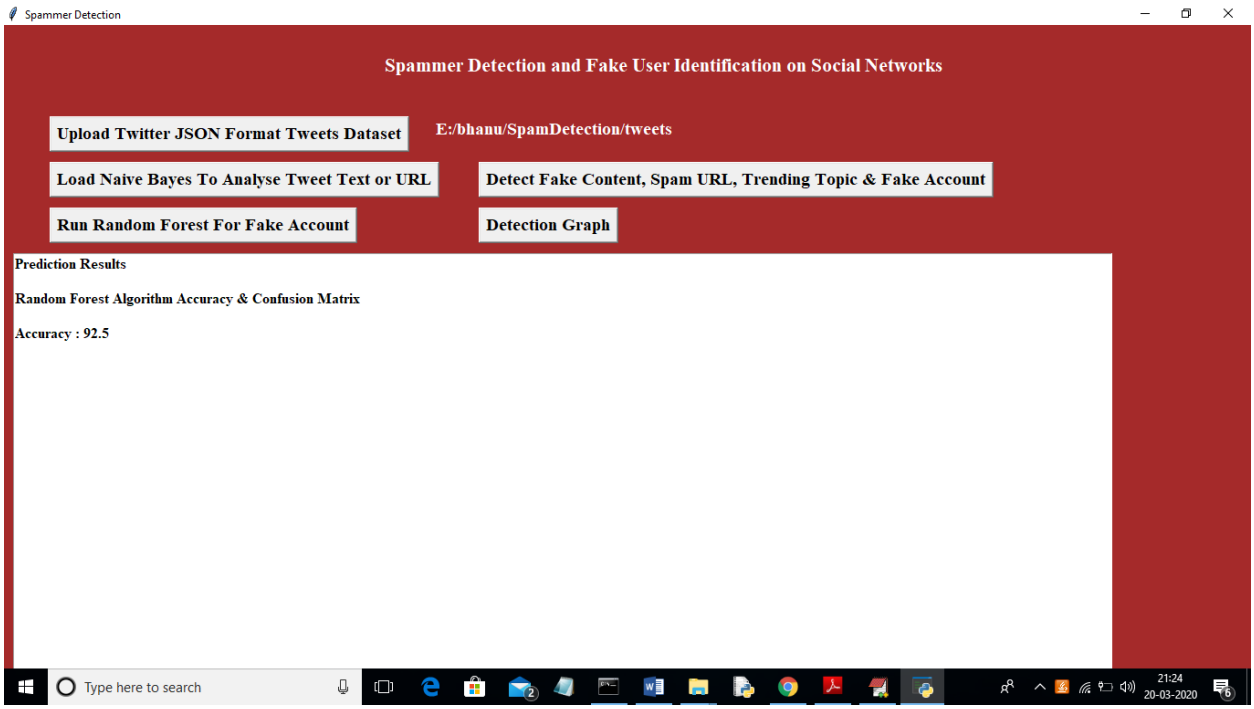

Fig. 7 Accuracy

In above screen we got random forest prediction accuracy as 92%, now click on _Detection Graph' button to know total tweets and spam and fake account graph.

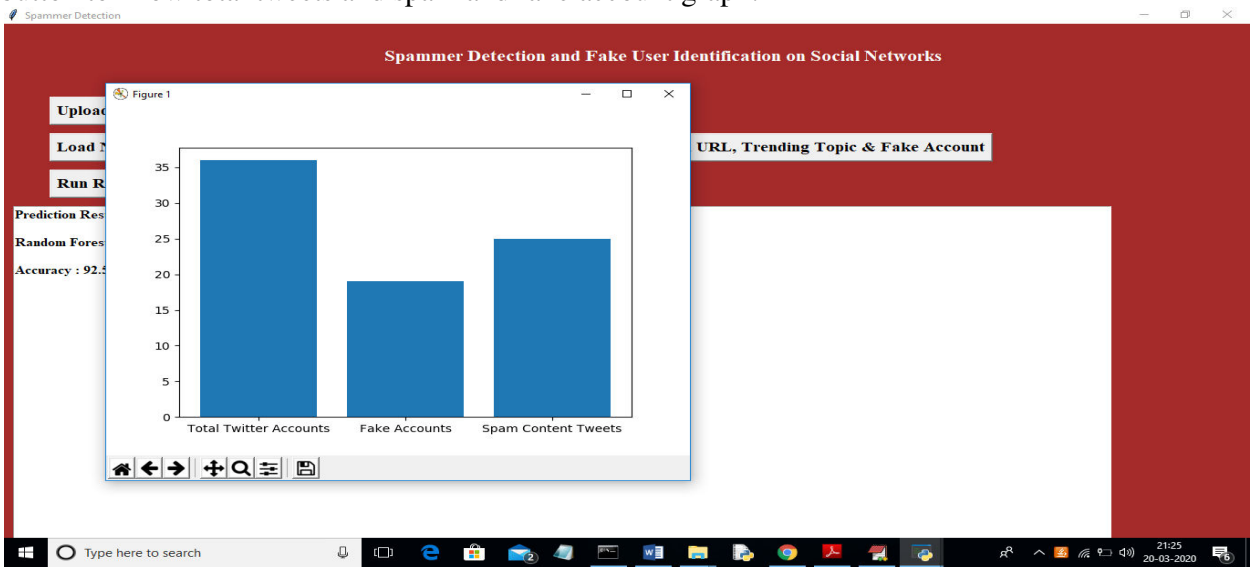

Fig. 8 Accuracy Graph

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

In above graph x-axis represents total tweets, fake account and spam words content tweets and y-axis represents count of them.

V. FUTURE SCOPE AND CONCLUSION

Here the paper is a implementation of analysis method utilized on behalf of distinguishing spammers on Twitter. We additionally exhibited taxonomy of Twitter spam identification method are considered as false contented recognition, URL built spam identification, spam location in inclining points, and phony client recognition strategies. We likewise analysed the introduced strategies dependent on a few features, for example, client features, content features, chart features, structure features, and time features. Besides, the procedures were likewise looked at regarding their predefined objectives and datasets utilized. It is foreseen that the introduced audit will assist scientists with finding the data on best in class Twitter spam discovery procedures in a united structure. Notwithstanding the improvement of proficient and viable methodologies for the spam discovery and phony client distinguishing proof on Twitter, there are as yet certain open zones that need extensive consideration by the analysts.

REFERENCES

[1] S. Y. Bhat, M. Abulaish, "Community-based features for identifying spammers in online social networks," In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013, p. 100-107.

[2] C. Grier, K. Thomas, V. Paxson, et al, "@ spam: the underground on 140 characters or less," In Proceedings of the 17th ACM conference on Computer and communications security, 2010, p. 27-37.

[3] Y. Liu, B. Wu, B. Wang, et al, "SDHM: A hybrid model for spammer detection in Weibo," Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on. 2014, p. 942-947.

[4] H. J. Rong, Y. S. Ong, A. H. Tan, Z. Zhu, et al. "A fast pruned-extreme learning machine for classification problem," Neurocomputi., vol. 72, pp. 359-366, Dec. 2008.

[5] C. W. Hsu, C. J. Lin. "A comparison of methods for multiclass support vector machines," Neural Networks, IEEE Transactions on, vol. 13, pp. 415-425, Mar. 2002.

[6] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," In Neural Networks, 2004 IEEE International Joint Conference on. 2004, vol. 2, p. 985 – 990.

[7] Y. Hirose, K. Yamashita, S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," Neural Networks, vol. 4, pp. 61-66, 1991.

[8] H. Shen, Z. Li, "Leveraging social networks for effective spam filtering," Computers, IEEE Transactions on, vol. 63, pp. 2743-2759, Jul. 2013.

[9] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: theory and applications," Neurocomputing, vol. 70, pp. 489-501, Dec. 2006.

[10] C. R. Rao, S. K. Mitra, Generalized inverse of matrices and its applications. New York: Wiley, 1971, vol. 7.

[11] X. Zheng, N. Chen, Z. Chen, C. Rong, G. Chen, W. Guo, "Mobile Cloud Based Framework for Remote-Resident Multimedia Discovery and Access," Journal of Internet Technology, vol. 15, pp. 1043-1050, Nov. 2014.

[12] G. E. Hinton, "Learning multiple layers of representation," Trends in cognitive sciences, vol. 11, pp. 428-434, Oct. 2007.

ISSN- 2394-5125 VOL 9, ISSUE 04, 2022

[13] Y. Bengio, "Scaling up deep learning," Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. 2014, p.1966-1966.
[14] S. Zhou, Q. Chen, X. Wang, "Active deep learning method for semi-supervised sentiment classification," Neurocomputing, vol. 120, pp. 536-546, Nov. 2013.