**Spammer Detection and Fake User Identification on Social Networks**

**Mr. P. Ramesh1, Asst Prof CSE:BVCE, 542ram@gmail.com**

**K.N.V.P.S.B. Ramesh2, Asst Prof CSE:BVCE, ramesh.kb17@gmail.com**

**Chinta Revanth3, CSE:BVCE**

**Chodapaneedi Venkata Sandhyarani4, CSE:BVCE**

**Dirisala Naga Sai Manikanta5, CSE:BVCE**

**Lavudu Ramesh6, CSE:BVCE**

|  |  |
| --- | --- |
| ***Article Info***  ***Page Number:*** ***5197 - 5212***  ***Publication Issue:***  ***Vol 71 No. 4 (2022)***  ***Article History***  ***Article******Received****: 25 March 2022*  ***Revised****: 30 April 2022*  ***Accepted****: 15 June 2022*  ***Publication****: 19 August 2022* | **Abstract**  Millions of people from all over the world use social networking sites. The things that people do on social networking sites like Twitter and Facebook have a big effect on their daily lives, sometimes in a bad way. Spammers use popular social networking sites to spread a lot of useless and harmful information. For example, Twitter has become one of the most overused platforms of all time. Because of this, it lets a lot of spam through. Fake users send unwanted tweets to other users to advertise services or websites, which not only bothers real users but also slows down the use of resources. Also, fake identities have made it easier to spread false information to users, which makes it easier for harmful content to get out. Recently, finding spammers and fake users on Twitter has become a common topic of research in today's online social networks (OSNs). In this paper, we look at some of the ways that spammers on Twitter can be found. Also, a taxonomy of Twitter spam detection methods is given, which groups the methods by how well they can find: I fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The techniques that are shown are also compared based on different characteristics, such as user characteristics, content characteristics, graph characteristics, structure characteristics, and time characteristics. We hope that the study we are presenting will be a good way for researchers to find the most important new developments in Twitter spam detection in one place.  Spammer Detection, Fake User, Social Networks (OSNs), and Twitter Spam are some of the key words. |

INTRODUCTION:

By using the Internet, it has become very easy to get any kind of information from any source in the world. Users can find out a lot about each other on social sites because they are becoming more popular. Fake users are also interested in these sites because they have so much information [1]. Twitter has quickly become an online place to find out about users in real time. Twitter is an Online Social Network (OSN) where people can share news, opinions, and even how they are feeling. There are many things that can be argued about, such as politics, current events, and important events. When a user tweets something, it is immediately sent to his or her followers, who can then share the information with a much larger group of people [2]. As OSNs have changed, the need to study and understand how people use online social platforms has grown. Fraudsters can easily trick many people who don't know much about OSNs. People who use OSNs only to advertise and spam other people's accounts are also being asked to stop and be controlled.

Researchers have recently become interested in how spam is found on social networking sites. Spam detection is a hard part of making sure that social networks are safe. It is important to know how to spot spam on OSN sites to protect users from different types of attacks and to keep their security and privacy. Spammers use these dangerous strategies, which cause a lot of damage to the community in the real world. Spammers on Twitter have different goals, like spreading wrong information, fake news, rumours, and random messages. Spammers reach their bad goals by putting up ads and using other methods, such as joining different mailing lists and sending spam messages at random to promote their own interests. These things are annoying to the real users, who are called "non-spammers." It also brings down the reputation of the OSN platforms. So, it is important to come up with a way to find spammers so that steps can be taken to stop them from doing bad things [3].

Several studies have been done on how to tell if something on Twitter is spam. A few surveys have also been done on fake Twitter user IDs to get a full picture of the current state of the art. Tingmin et al. [4] give an overview of new techniques and methods for detecting Twitter spam. The above survey shows a comparison of how things are done now. On the other hand, the authors of [5] did a survey about the different ways spammers act on the social network Twitter. The study also has a review of the literature that shows spammers do exist on the Twitter social network. Even with all the studies that have been done, there is still a hole in the literature. So, to close the gap, we look at the current state of the art in identifying spammers and fake users on Twitter. This survey also gives a taxonomy of the ways Twitter spam can be found and tries to give a detailed description of recent changes in the field.

The goal of this paper is to find different ways to find spam on Twitter and to present a taxonomy by putting these different ways into different groups. For sorting, we have found four ways to report spammers that can help us figure out who is using a fake identity. Spammers can be found by: I looking for fake content, (ii) looking for spam in trending topics, (iii) looking for spam in URLs, and (iv) looking for fake user identification. Table 1 compares the techniques that are already being used and helps users understand the importance and usefulness of the proposed methods. It also compares their goals and results. Table 2 shows how different Twitter features can be used to spot spam. We hope that this survey will help people find a lot of different information about how to find spammers in one place.

1.1 The Effects of Spam on OSNs

Since the number of malicious user accounts on social networks has grown a lot, so have the effects of malicious activities. Based on the report shared by Nexgate in 2013, spam distribution has increased by up to 35% in the first half of 2013. And the report talks about a few things, such as: At least 5% of all social structure applications are used to send spam. 2. Malicious user accounts post a lot more and faster than real user accounts on social networks. 3. A spammer sends out bad information on at least 23 social networks. 4. For every seven social media accounts, there are five spammers. 5. 15% of all social spam messages include a spam-spreading URL. Reviews show that the number of identity fraud cases has reached 13 million per year over the past six years, and that social spammers cost trust, productivity, and profit $200 million per year. As bad things happen online more and more, it is important to get rid of fake user accounts that could harm real users.

1.2 Identifying spammers and catching fake users

1) Fake content: If an account has a small number of followers compared to the number of people who follow it, it's less likely to be trustworthy and more likely to be spam. In the same way, content-based features include the reputation of tweets, HTTP links, mentions and replies, and "trending" topics. For the time feature, an account is a spam account if it sends a lot of tweets in a certain amount of time.

2) Spam URL Detection: The user-based features are found by looking at things like how old the account is and how many favourites, lists, and tweets the user has. The JSON structure is parsed to find the user-based features that have been found. On the other hand, tweet-based features include the number of I retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs. We will use an algorithm for machine learning called Nave Bayes to check if a spam URL is in a tweet.

3) Finding Spam in Trending Topics: In this method, the content of tweets will be sorted by the Naive Bayes algorithm to see if they contain spam words or not. This algorithm will look for spam URLs, words with adult content, and tweets that are the same. If Nave Bayes thinks a tweet is SPAM, it will return 1, and if it doesn't think it's SPAM, it will return 0.

4) Fake User ID: These include the number of followers and people who follow the account, the age of the account, etc. On the other hand, content features are linked to the tweets that users post. Spam bots post a lot of duplicate content, while people who don't spam don't post duplicate tweets. In this method, features (following, followers, and tweet content to find spam or non-spam content using Naive Bayes Algorithm) will be taken from tweets and then put into spam or non-spam categories using Naive Bayes Algorithm. Later, a random forest algorithm will be used to train this feature to figure out if an account is fake or not. The features.txt file is where all of the extracted features will be saved. Naïve Bayes classifier saved inside 'model' folder.

1.3 Motivation: Researchers are now trying to figure out how to spot spam on social networking sites. Keeping social networks safe is hard because it is hard to find spam. It is important to be able to spot spam on OSN sites so that users are safe from all kinds of dangerous attacks and can keep their security and privacy. Spammers' risky actions cause a lot of damage to communities in the real world. Spammers on Twitter have many different goals, such as spreading false information, fake news, rumours, and comments that they just made up on the spot. Spammers reach their destructive goals in a number of ways, such as by supporting multiple mailing lists and sending spam messages at random to let people know what they are interested in. The original users, who are called "non-spammers," find these things annoying. It also hurts the reputation of the OSN platforms. So, it's important to come up with a plan for finding spammers so that they can be stopped from doing their bad things.

1.4 Problem Statement:

Thanks to the Internet, it is now pretty easy to get any kind of information from any source in the world. People can gather a lot of information about other people on social media sites, which are becoming more and more popular. Because these sites offer a lot of information, fake users are drawn to them. Twitter has become very popular as a way to find out about people in real time. With the Internet, it is now very easy to get any kind of information from any source in any part of the world. Social media platforms are becoming more and more popular, which means that users can find out a lot about other users. Because these platforms offer a lot of information, fake users are drawn to them. Because there is so much information on these sites, fake users are drawn to them.

1.5 The goal is to find spam tweets and fake accounts on Twitter, an online social network.

1.6 Work Scope:

Social networking services are used by millions of people, and they all have a big effect on daily life, with some bad results. Spammers have turned popular social networking sites into places where they can spread a lot of useless and harmful content. letting an excessive amount of spam through. Fake users send unwanted tweets to other users to promote services or websites. This hurts real users and wastes time and money. Also, fake identities have made it easier to spread false information to users, which has led to the spread of dangerous materials.

1.7 Applications: The main goal of the applications is to find spammers and fake users. Online social networking sites are a good example.

2.2: Our research contribution is that we can tell if a tweet is a normal message or spam. By finding and getting rid of these spam messages, you can help social networks get a good name in the market. If social networks didn't get rid of spam messages, they would lose users. Now, most people rely on social networks to find out about news, business, and family. Protecting social networks from spammers can help them gain a good reputation.

3 System Planned

The goal of this work is to find different ways to find spam on Twitter and to create a taxonomy that divides these ways into groups. We've come up with four ways to report spammers that can help us figure out if a user is pretending to be someone else. Spammers can be found by using false content, URL-based spam detection, spam detection in popular subject areas, and fake user identification. Existing methods and helps the user understand the importance and usefulness of the proposed method, as well as compare their goals and results. There are many ways to find spam on Twitter. By holding this poll, we hope that people will be able to find a lot of information about how to find spammers in one place.

3.1 Algorithms: Random forest, Naive Bayes, and Extreme Machine Learning are used in this study (EML)

3.1.1 Bayes the Simple:

When compared to more complex algorithms, the Naive Bayes classifier can be very fast. Because the class distributions aren't mixed up with each other, each one can be evaluated on its own as a one-dimensional distribution. In turn, this helps get rid of some of the problems caused by the "dimensionality curse."

The Nave Bayes classifier is a type of probabilistic classifier that is based on Bayes' theorem and assumes that the features are very independent from each other based on the value of the class variable. This method is a group of algorithms for learning with help.

3.1.2 Random forest algorithm: This model has three random parts: picking training data at random when making trees, picking some subsets of features when splitting nodes, and only looking at a subset of all features when splitting each node in each simple decision tree. In a random forest, each tree learns from a random sample of the data points when it is being trained. A random forest model is made up of a large number of decision trees. The model basically takes the average of what the trees predict, which is why it is called a forest. Also, the algorithm includes three random ideas: picking training data at random when making trees, picking some subsets of variables at random when splitting nodes, and deciding that only a subset of all variables should be used to split every node in each basic decision tree. During the training of a random forest, each basic tree learns from a random sample of the dataset.

3.1.3 Machine learning to the extreme (EML)

Extreme machine learning (EML) are feed-forward neural networks for classification, regression, clustering, sparse approximation, compression, and feature learning with a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes (not just the weights connecting inputs to hidden nodes) do not need to be tuned. These hidden nodes can be given at random and never changed (this is called a random projection with nonlinear transforms), or they can get them from their ancestors without changing. Most of the time, learning the output weights of hidden nodes is done in a single step, which is basically the same as learning a linear model. The main person who made these models, Guang-Bin Huang, called them "extreme learning machines" (ELM).

3.1.4 Detection Random Forest & Extreme ML to get accuracy comparison between random forest and extreme machine learning to get total accounts, fake accounts, and real accounts graph

**3.3 Architecture/Framework**:

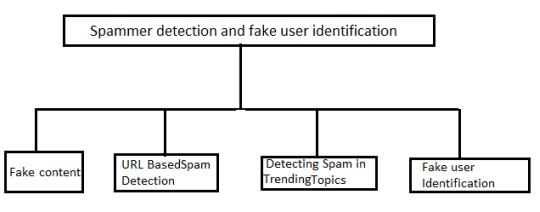


Fig. No 1

**3.4 Algorithm and Process Design:**

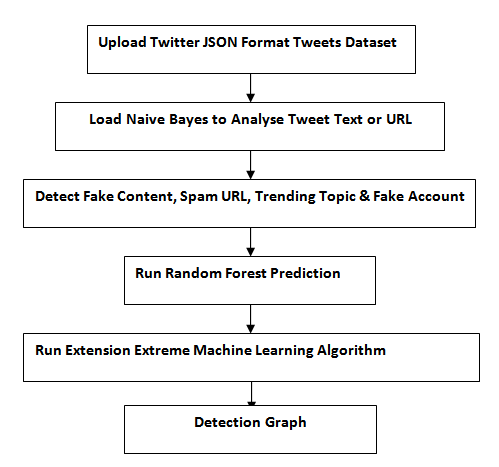
****

Fig. No.2

**4.2 Metrics for Evaluating:**

Load Naive Bayes to analyse Tweet Nave Bayes classifier can be loaded with text or a URL.

To Detect Fake Content, Spam URL, Trending Topic, and Fake Account,' to use Naive Bayes classifier to look at each tweet for fake content, spam URL, and fake accounts.

All features are taken from the tweets dataset and then those features are used to figure out whether a tweet is spam or not. In the text area above, each record's values are separated by a blank line, and each tweet record shows values like TWEET TEXT, FOLLOWERS, FOLLOWING, etc., as well as whether the account is fake or real and whether the tweet text has spam words or not.

Random Forest Prediction' to train random forest classifier with extracted tweet features, and this random forest classifier model will be used to predict/detect fake or spam account for future tweets.

We got 87.5% accuracy with the extreme machine learning algorithm, and it's better at figuring out whether an account is fake or real than random forest.

**4.3 Outcome:**

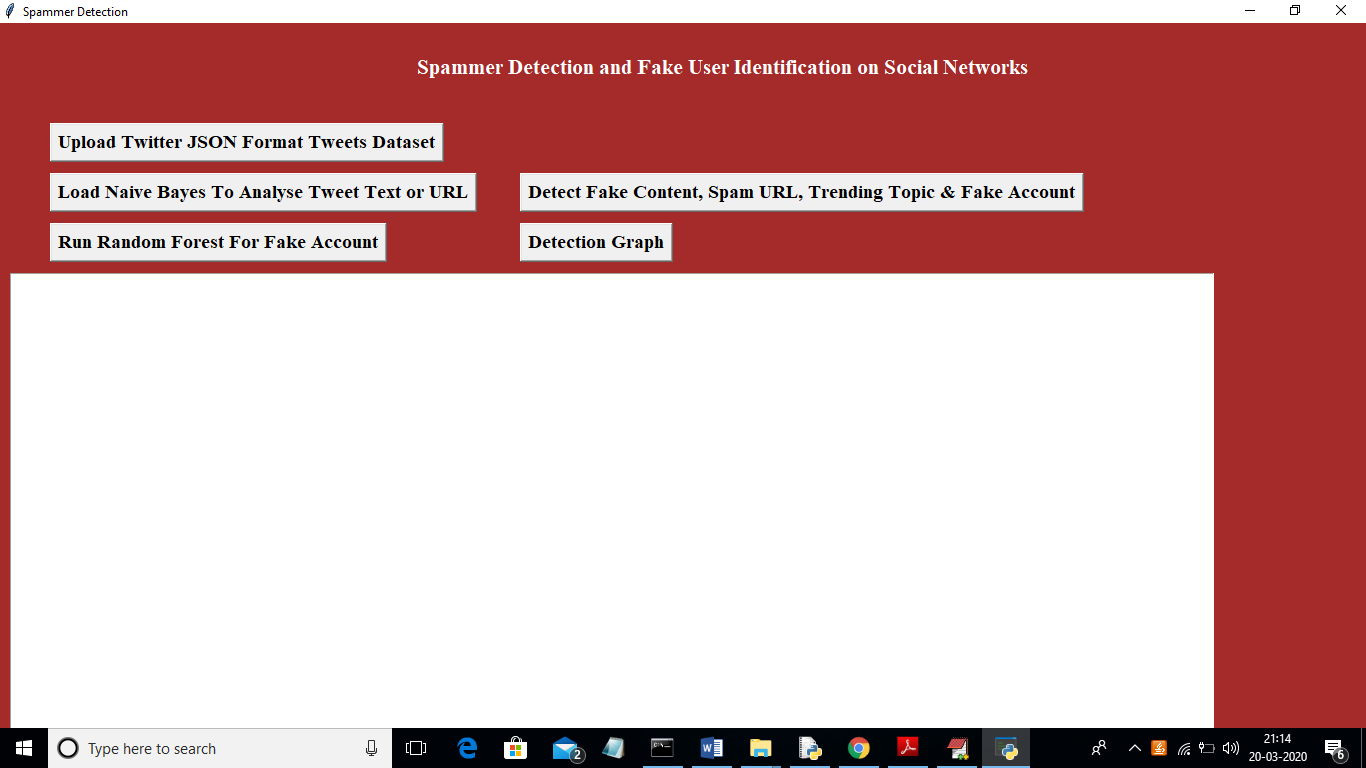


Fig.3Upload Twitter JSON Format Tweets Dataset

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

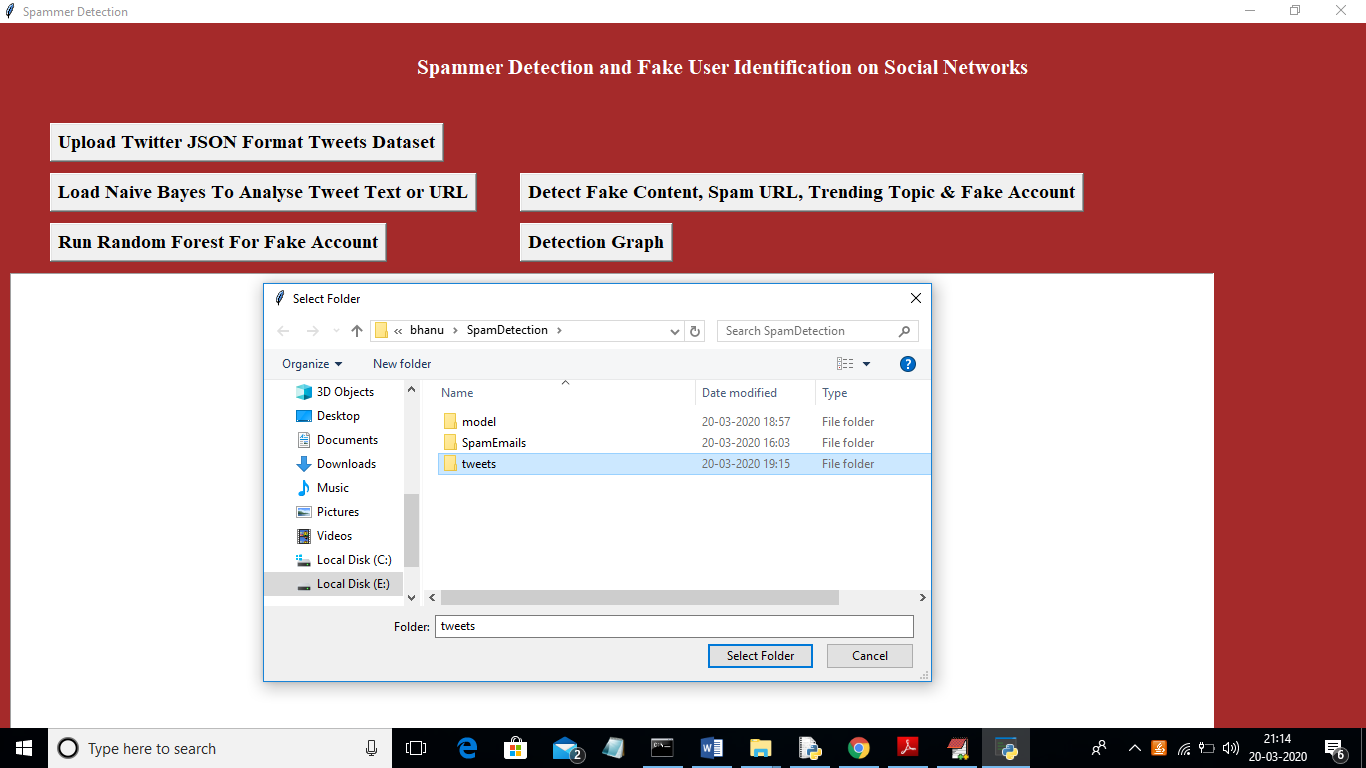


Fig.4. Uploading ‘tweets’ folder

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

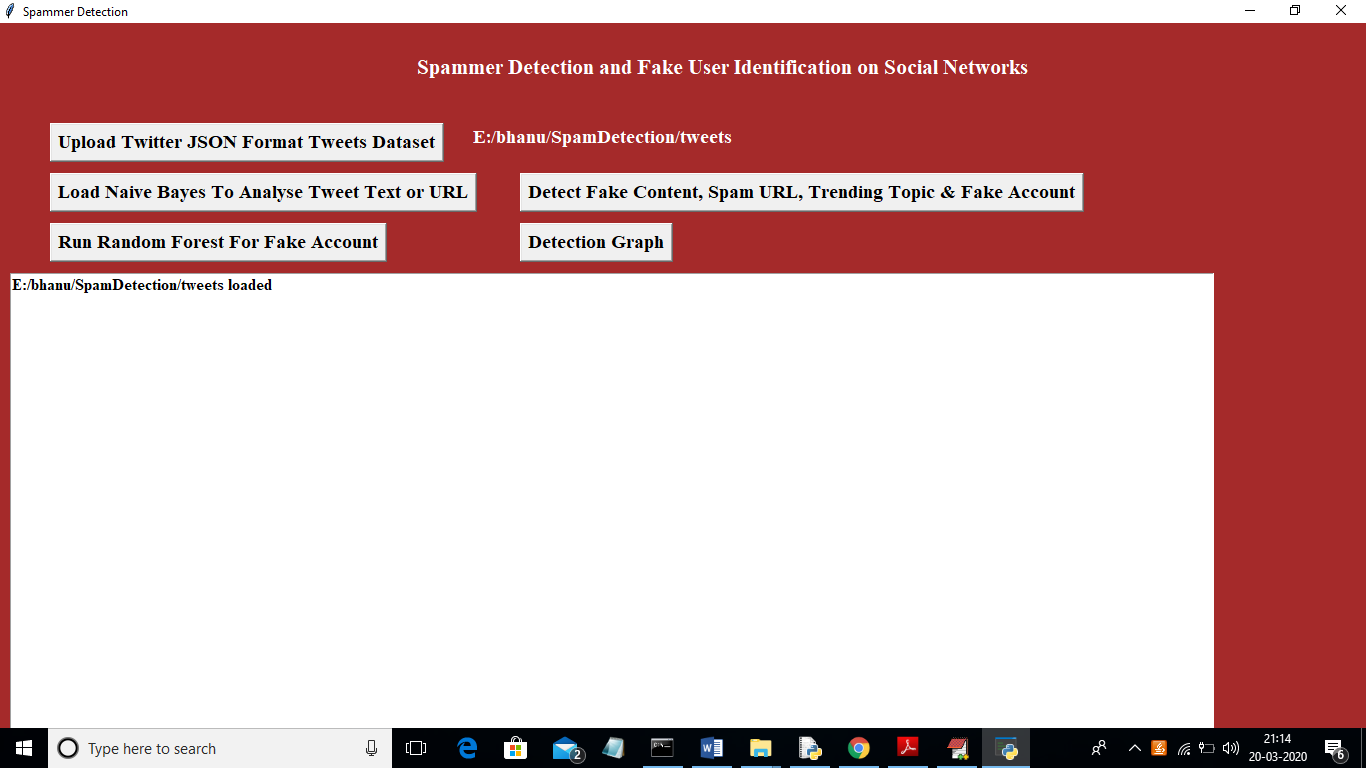


Fig.5. Load Naive Bayes

In above figure 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

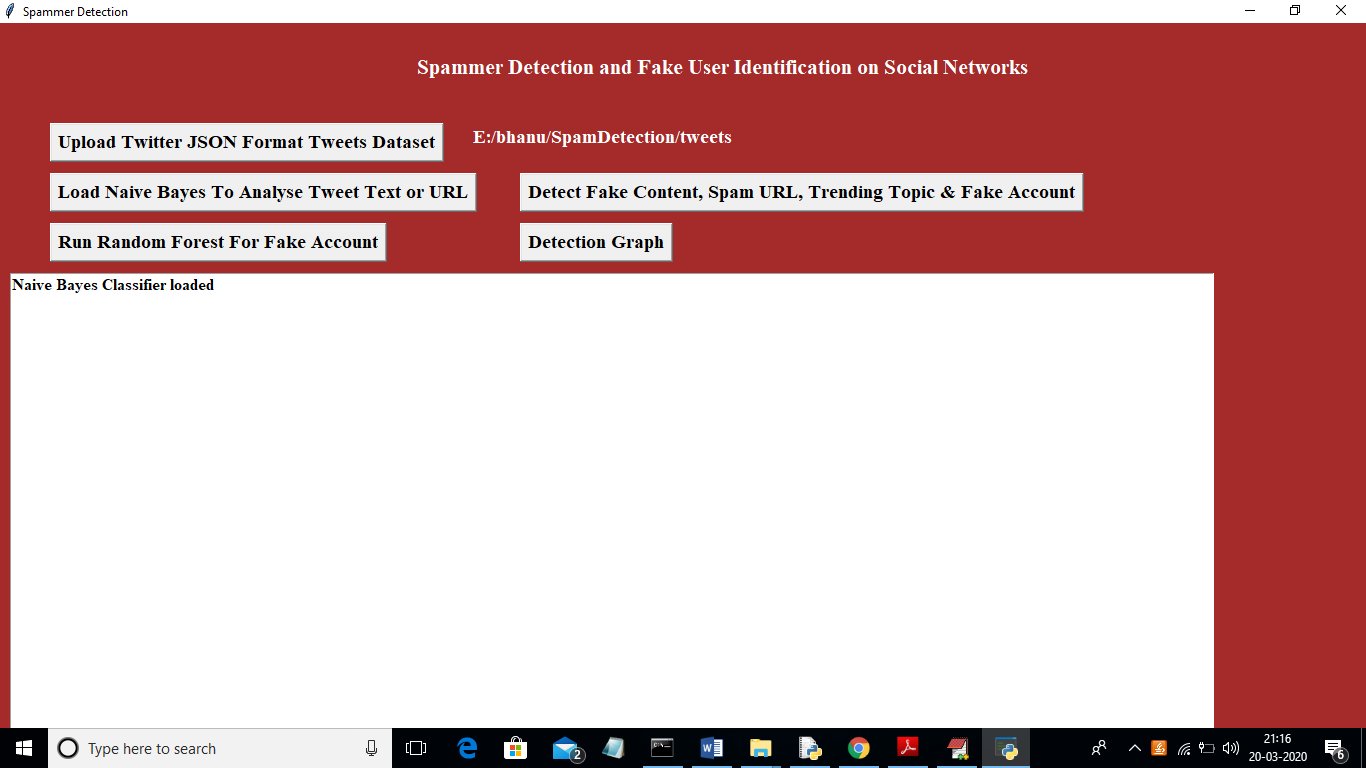


Fig.6. Detect Fake Content

In above figure 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

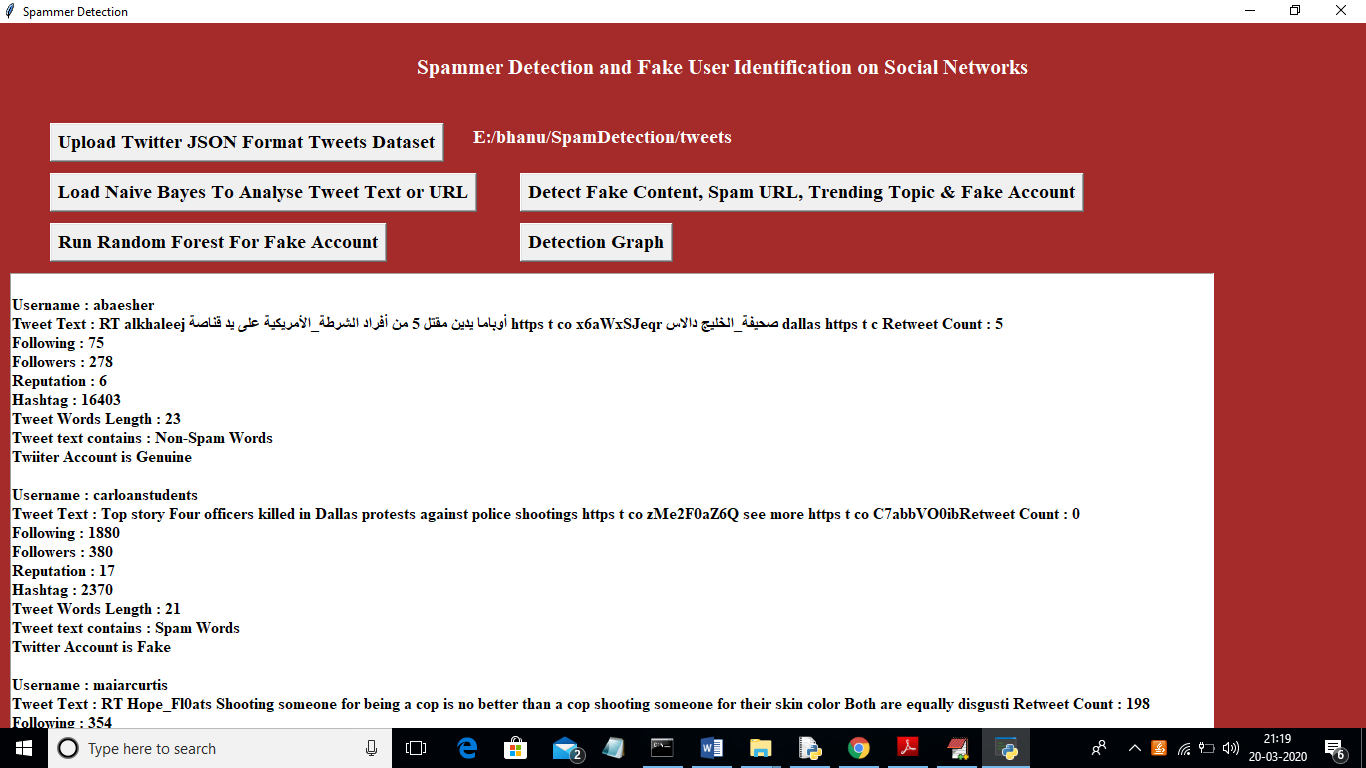


Fig. 7. Features extracted from tweets

In above figure 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

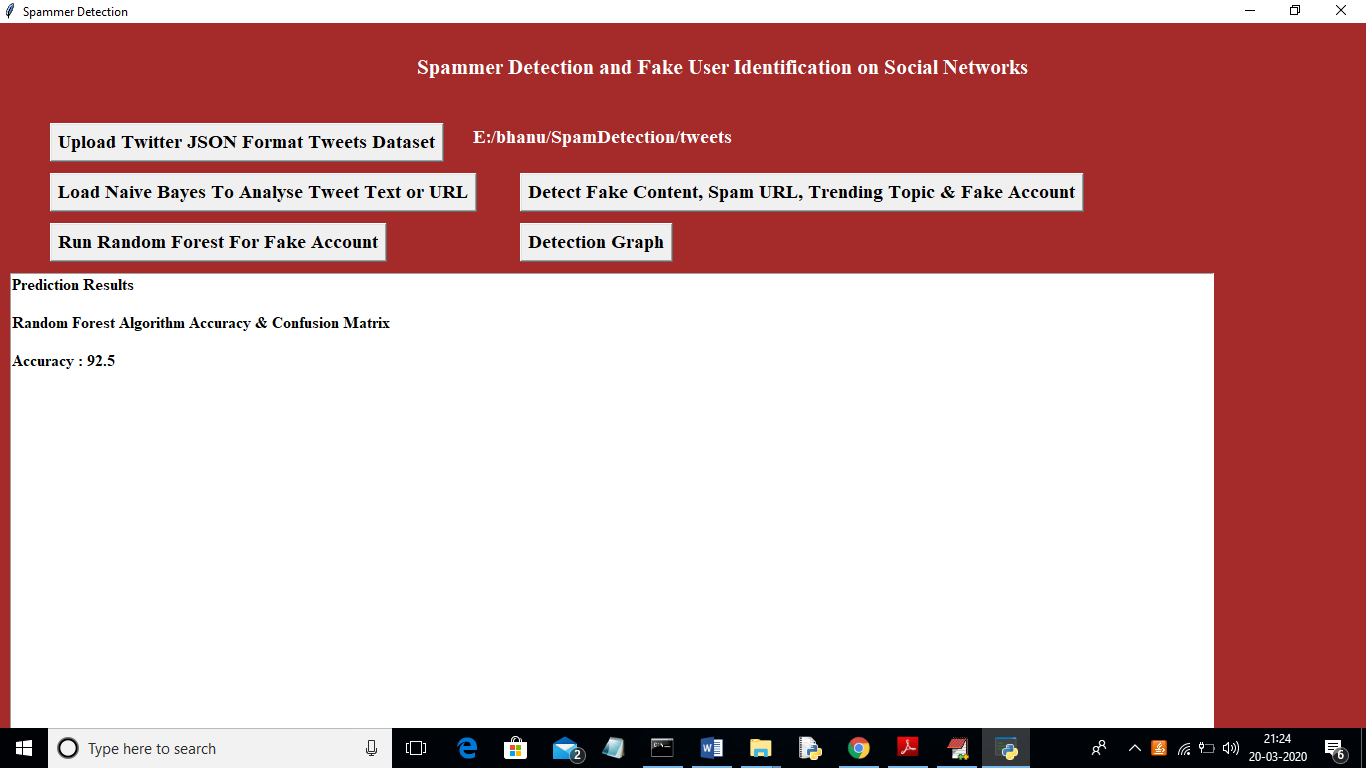


Fig.8. Random Forest Prediction Accuracy as 92%

In above figure we got random forest prediction accuracy as 92%, now click on ‘Detection Graph’ button to know total tweets and spam and fake account graph

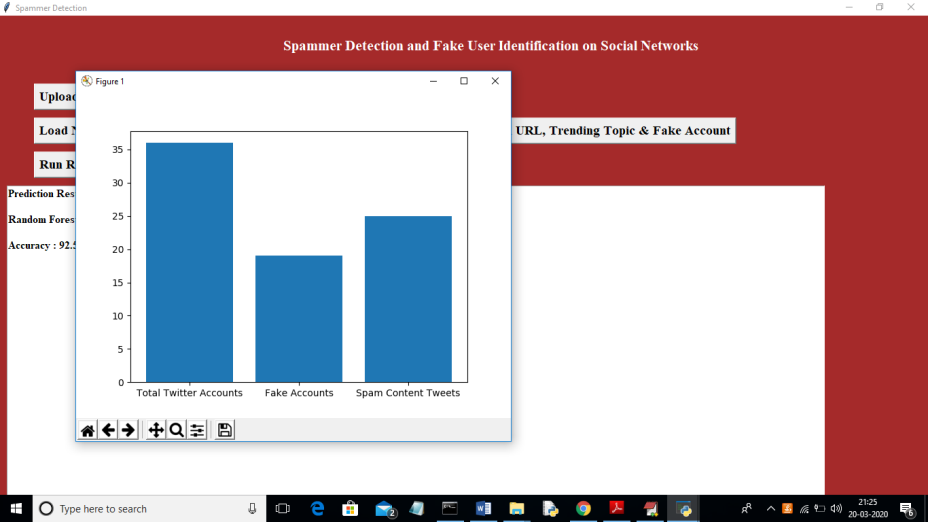


Fig.9. Total Tweets

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

**Extension Outcomes:**

Using the EML algorithm to do extension work

As extra work, I have added the EML algorithm to this project. Here are the details of the algorithm. For details about the base paper, you can look at the previous figure shots. This one only has information about the extension.

Extreme machine learning (EML) are feed-forward neural networks for classification, regression, clustering, sparse approximation, compression, and feature learning with a single layer or multiple layers of hidden nodes, where the parameters of hidden nodes (not just the weights connecting inputs to hidden nodes) do not need to be tuned. These hidden nodes can be given at random and never changed (this is called a random projection with nonlinear transforms), or they can get them from their ancestors without changing. Most of the time, learning the output weights of hidden nodes is done in a single step, which is basically the same as learning a linear model. The main person who made these models, Guang-Bin Huang, called them "extreme learning machines" (ELM).

The people who made these models say that they can generalise well and learn thousands of times faster than networks that were trained with backpropagation. It is also shown in the literature that these models can do a better job than support vector machines (SVM) and other classifiers.

**FIGURE SHOTS**

To run project double click on ‘run.bat’ file to get below figure

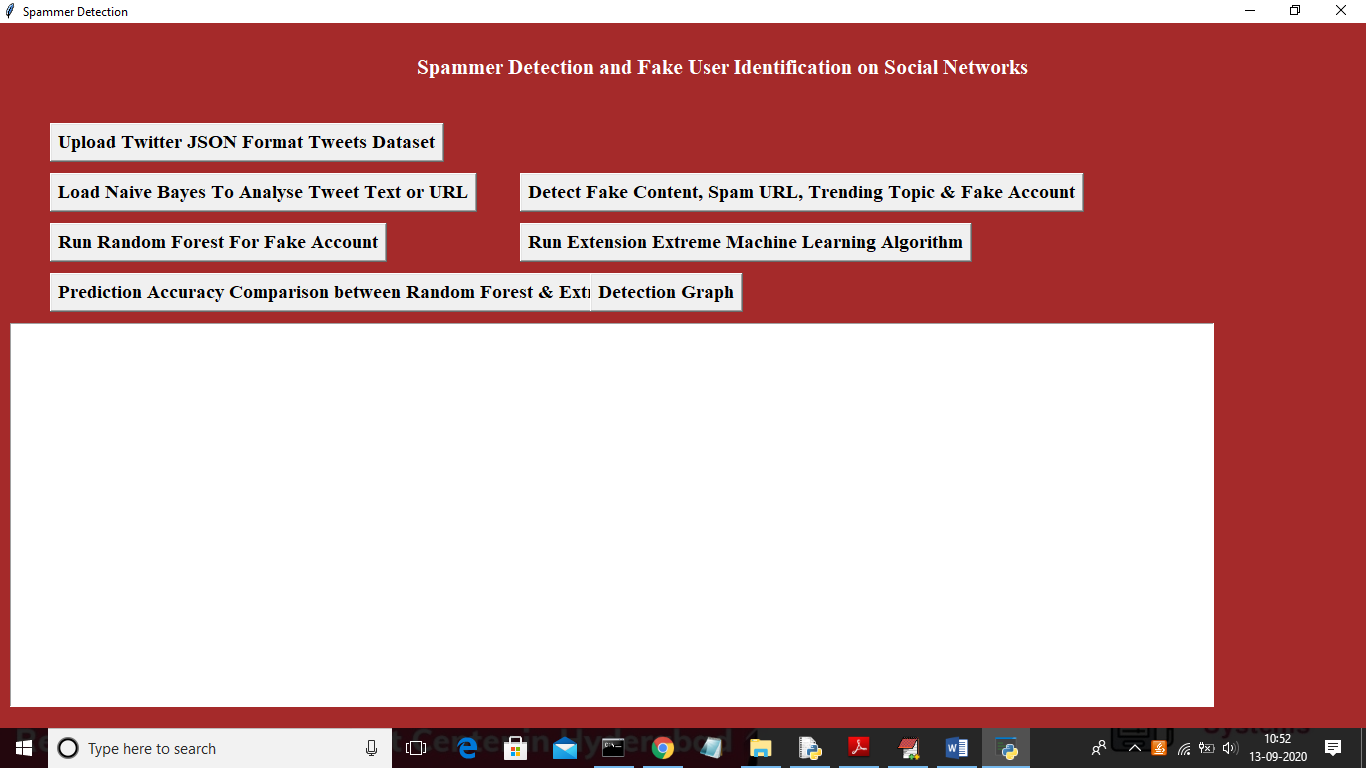


Fig.10

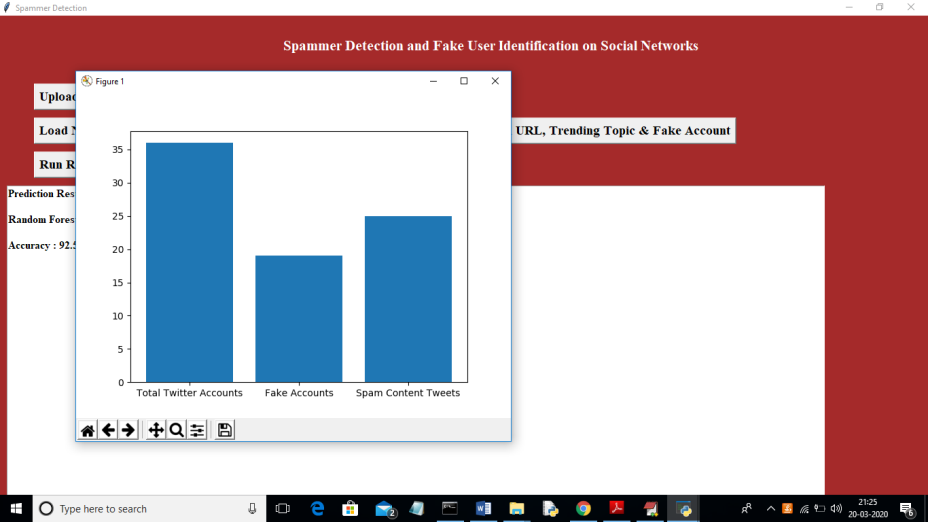


Fig.11. Upload dataset

In above figure upload dataset by clicking on ‘Upload Twitter JSON Format Tweets Dataset’ button

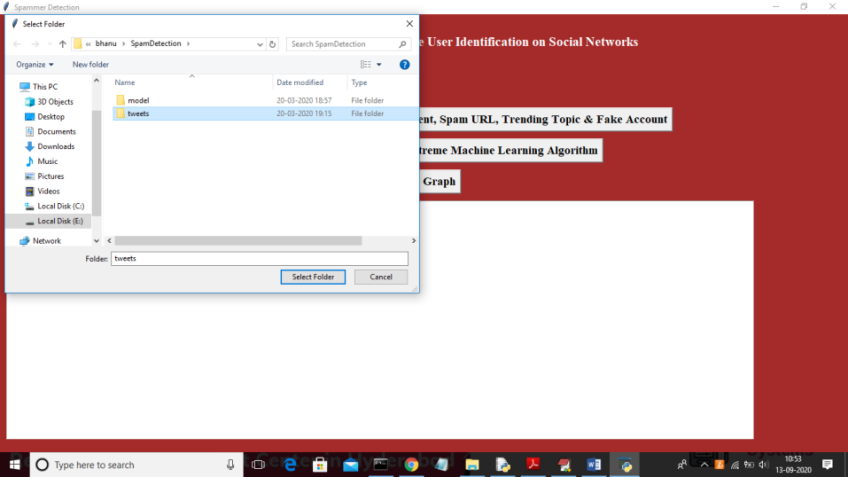


Fig.12. Select Folder

In above figure after selecting tweets folder click on ‘Select Folder’ to load tweets dataset and to get below figure



Fig.13. Load Naïve Bayes

Now click on ‘Load Naïve Bayes ToAnalyse Tweet Text or URL’ button to analyse tweet text and generate machine learning features

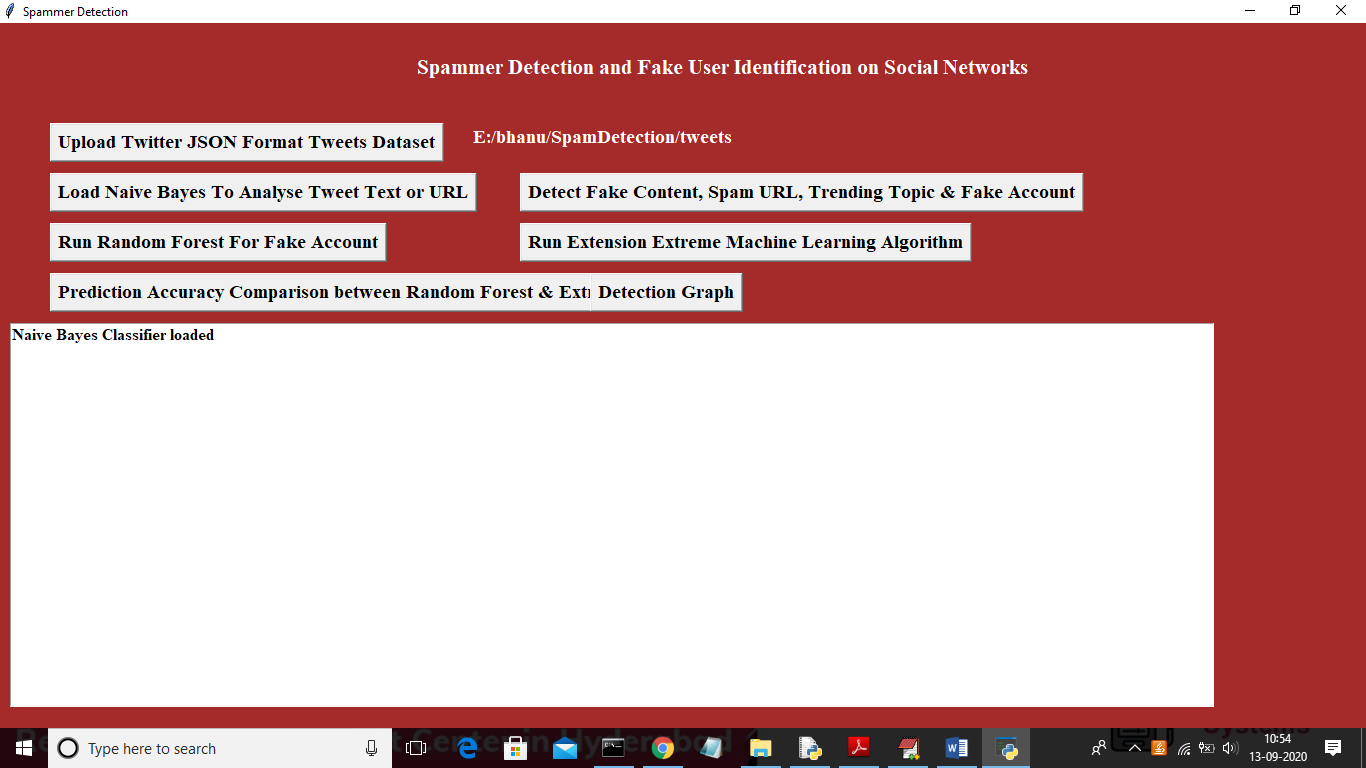


Fig.14. Naïve Bayes classifier loaded

In above figure naïve bayes classifier loaded to analyse tweet text and now click on ‘Detect Fake Content, Spam URL, Trending Topic & Fake Account’ to detect spam accounts and create machine learning features

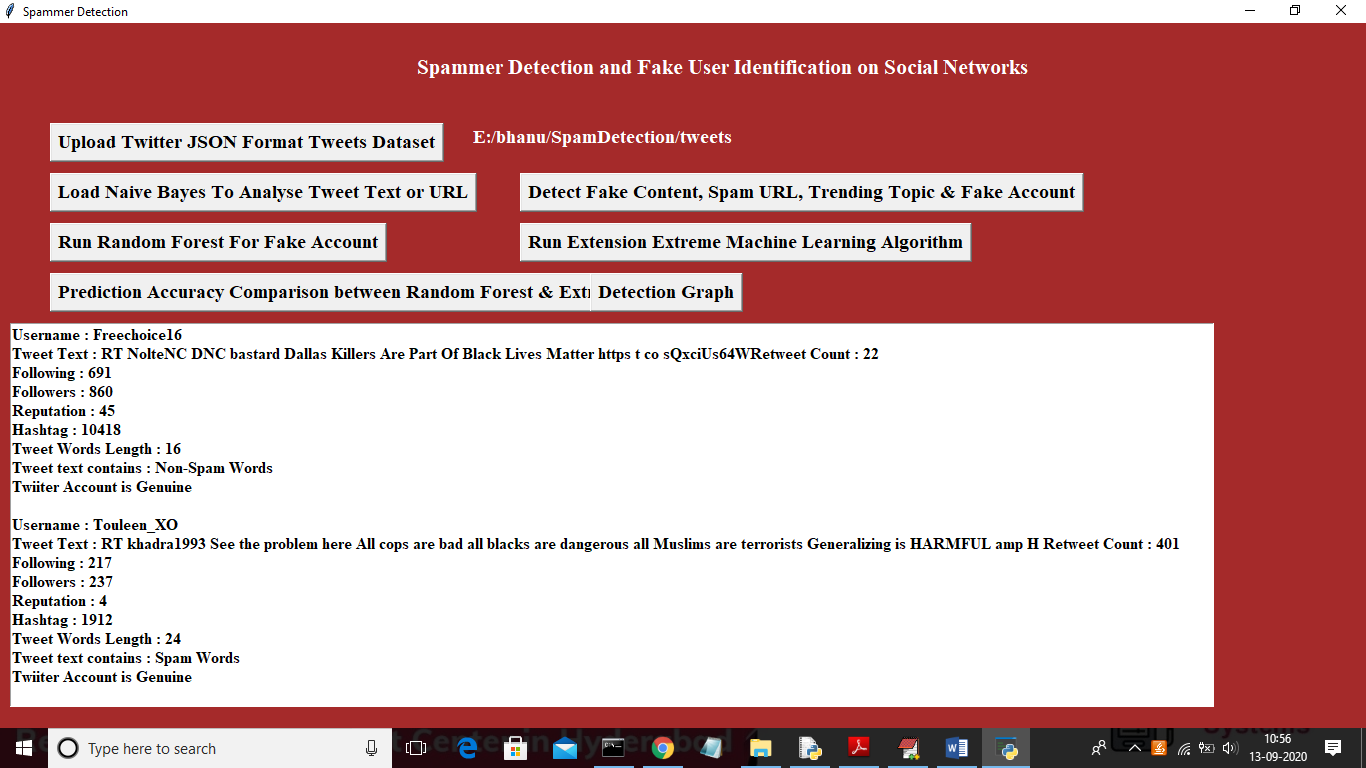


Fig.15. Analyze all users

In above figure we analyse all users accounts and then identify whether account is normal or contains spam content using Naïve Bayes algorithm and now click on ‘Run Random Forest For Fake Account’ button to build machine learning model on above data so in future by applying new account details we can predict whether account is normal or fake.

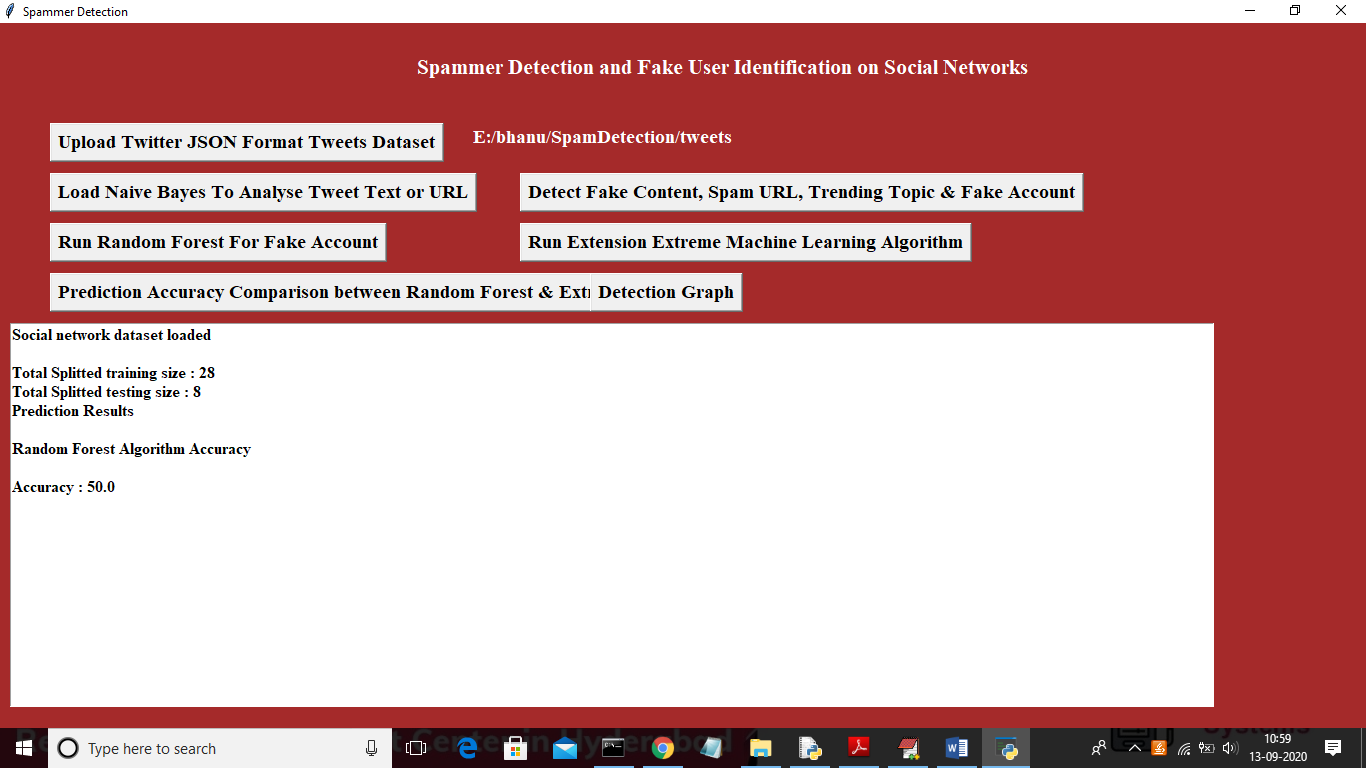


Fig.16. Details of dataset

In above figure we can see total dataset contains 36 accounts and application using 80% records (28) for training and 20% records (8) for testing and with random forest we got test data prediction accuracy as 50%. Now click on ‘Run Extension Extreme Machine Learning Algorithm’ button to generate training model on above data and to get prediction accuracy on test data

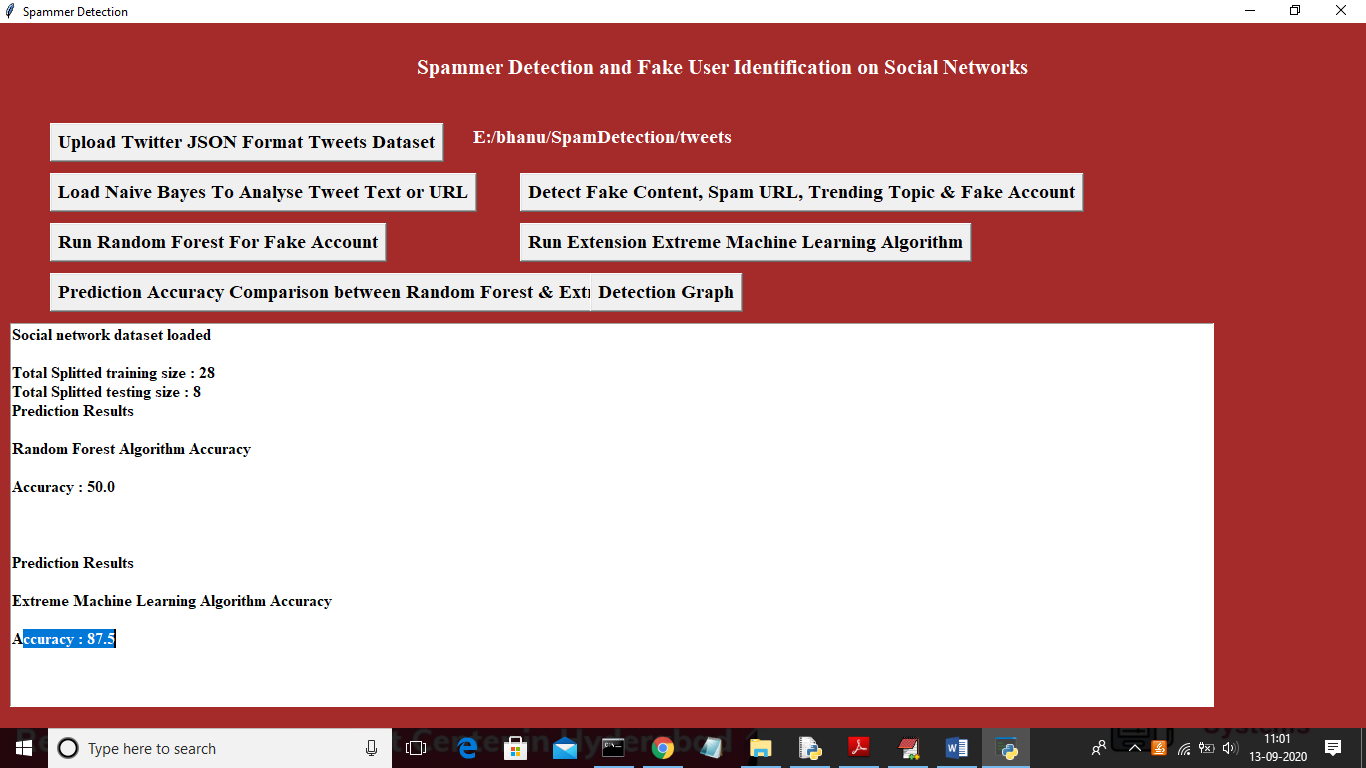


Fig.17.Extreme Machine Learning algorithm we got 87.5% accuracy

In above figure with extension extreme machine learning algorithm we got 87.5% accuracy and its better in prediction fake or normal account compare to random forest. Here dataset will be splitted to training and testing part randomly so accuracy may vary for each run. Now click on ‘Prediction Accuracy Comparison between Random Forest & Extreme ML’ button to get accuracy comparison between random forest and extreme machine learning

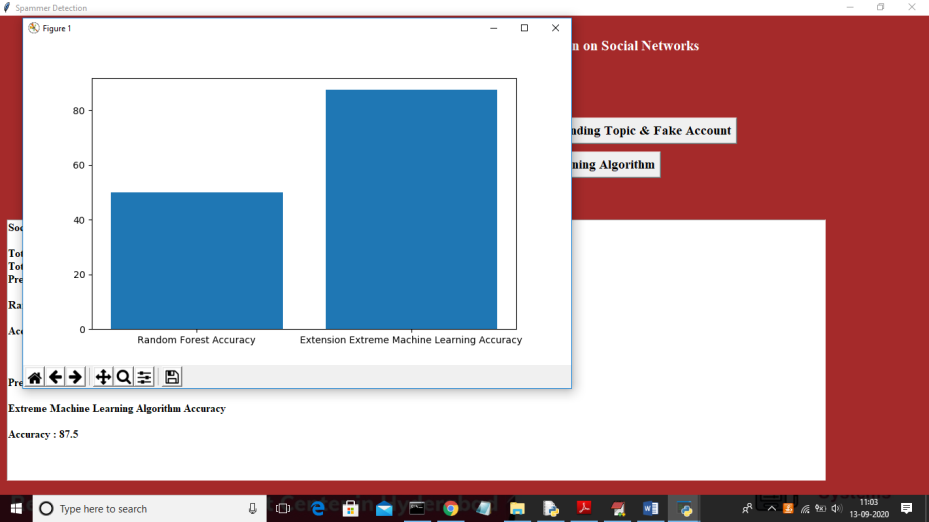


Fig.18Name vs Accuracy

In the above graph, the x-axis shows the names of the algorithms and the y-axis shows how accurate those algorithms are. From the above graph, we can see that extension work is getting more accurate, and we can now click on the "Detection Graph" button to see the total number of accounts, both fake and real. graph

In the above graph, the x-axis shows the types of accounts, and the y-axis shows how many of each type of account there are.

**CONCLUSION**

In this paper, we looked at the different ways that spammers can be found on Twitter. We also showed a taxonomy of ways to find Twitter spam and put them into four groups: fake content detection, spam detection based on URLs, spam detection in trending topics, and fake user detection techniques. We also compared the techniques shown based on user features, content features, graph features, structure features, and time features, among others. Also, the techniques were compared based on their goals and the datasets they used. It is hoped that the presented review will make it easier for researchers to find information on the most up-to-date ways to find spam on Twitter.

**REFERENCES**

[1] B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol, ‘‘Twitter fake account detection,’’ in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388–392.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ‘‘Detecting spammers on Twitter,’’ in Proc. Collaboration, Electron. Messaging, AntiAbuse Spam Conf. (CEAS), vol. 6, Jul. 2010, p. 12.

[3] S. Gharge, and M. Chavan, ‘‘An integrated approach for malicious tweets detection using NLP,’’ in Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT), Mar. 2017, pp. 435–438.

[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ‘‘Twitter spam detection: Survey of new approaches and comparative study,’’ Comput. Secur., vol. 76, pp. 265–284, Jul. 2018.

[5] S. J. Soman, ‘‘A survey on behaviors exhibited by spammers in popular social media networks’’, in Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCPCT), Mar. 2016, pp. 1–6.

[6] A. Gupta, H. Lamba, and P. Kumaraguru, ‘‘1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter,’’ in Proc. eCrime Researchers Summit (eCRS), 2013, pp. 1–12.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ‘‘Twitter analysis for real-time malware discovery,’’ in Proc. AEIT Int. Annu. Conf., Sep. 2017, pp. 1–6.

[8] N. Eshraqi, M. Jalali, and M. H. Moattar, ‘‘Detecting spam tweets in Twitter using a data stream clustering algorithm,’’ in Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK), Nov. 2015, pp. 347–351.

[9] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ‘‘Statistical features-based real-time detection of drifted Twitter spam,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 4, pp. 914–925, Apr. 2017.

[10] C. Buntain and J. Golbeck, ‘‘Automatically identifying fake news in popular Twitter threads,’’ in Proc. IEEE Int. Conf. Smart Cloud (SmartCloud), Nov. 2017, pp. 208–215.

[11] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, ‘‘A performance evaluation of machine learning-based streaming spam tweets detection,’’ IEEE Trans. Comput. Social Syst., vol. 2, no. 3, pp. 65–76, Sep. 2015.

[12] G. Stafford and L. L. Yu, ‘‘An evaluation of the effect of spam on Twitter trending topics,’’ in Proc. Int. Conf. Social Comput., Sep. 2013, pp. 373–378.

[13] M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, ‘‘A hybrid approach for spam detection for Twitter,’’ in Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST), Jan. 2017, pp. 466–471.

[14] A. Gupta and R. Kaushal, ‘‘Improving spam detection in online social networks,’’ in Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP), Mar. 2015, pp. 1–6.

[15] F. Fathaliani and M. Bouguessa, ‘‘A model-based approach for identifying spammers in social networks,’’ in Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2015, pp. 1–9.

[16] V. Chauhan, A. Pilaniya, V. Middha, A. Gupta, U. Bana, B. R. Prasad, and S. Agarwal, ‘‘Anomalous behavior detection in social networking,’’ in Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2017, pp. 1–5.

[17] S. Jeong, G. Noh, H. Oh, and C.-K. Kim, ‘‘Follow spam detection based on cascaded social information,’’ Inf. Sci., vol. 369, pp. 481–499, Nov. 2016.

[18] M. Washha, A. Qaroush, and F. Sedes, ‘‘Leveraging time for spammers detection on Twitter,’’ in Proc. 8th Int. Conf. Manage. Digit. EcoSyst., Nov. 2016, pp. 109–116.

[19] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, ‘‘Making the most of tweet-inherent features for social spam detection on Twitter,’’ 2015, arXiv:1503.07405. [Online]. Available: <https://arxiv.org/abs/1503.07405>

[20] M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan, S. Din, A. Ahmad, G. Jeon, and A. G. Reddy, ‘‘Towards ontology-based multilingual URL filtering: A big data problem,’’ J. Supercomput., vol. 74, no. 10, pp. 5003–5021, Oct. 2018.