INAPPROPRIATE REVIEW RECOGNITION USING MACHINE LEARNING

[1]K Vinod Kumar Reddy [2]P.Rasi [3]Shaik Asfiya [4]Shaik Simra Naaz [5]Velamuri Sumanvitha
[1] Assistant Professor [2,3,4,5] U.G scholars
[1,2,3,4,5]Ravindra College of Engineering for Women

ABSTRACT

Social networking websites have interaction hundreds of thousands of customers across the world. The customers' interactions with these social web sites, consisting of Twitter and Facebook have a splendid effect and occasionally unwanted repercussions for each day life. The distinguished social networking web sites have changed into a goal platform for the spammers to disperse a big quantity of beside the point and deleterious information. Twitter, for example, has emerge as one of the maximum extravagantly sed structures of all instances and consequently lets in an unreasonable quantity of spam. Fake customers ship undesired tweets to customers to sell offerings or web sites that now no longer most effective have an effect on valid customers however also disrupt aid consumption. We make a clone of the twitter and we will update the tweets ,we can update the messages to the server and we can send even the friend requests. In this web application we are using the Django based server and we remove the spam messages and we will use special algorithms to filter the bot users.

I. Introduction

Online informal organization for example, Twitter, Face book, and some venture interpersonal organization, have gotten incredibly well known over the most recent couple of years. invest tremendous measures of energy in OSNs warming up to individuals who they know about or inspired by. Twitter, which was established in 2006, has gotten one of the most mainstream smaller scale blogging administration locales. These days, 200 million Twitter clients create more than 400 million new tweets for every day. The ubiquity of Twitter draws in an ever increasing number of spammers. Spammers drive superfluous tweets to twitter clients to advance sites or administrations, which are unsafe to typical clients. So as to stop spammers, analysts have proposed various systems. The focal point of late works is on the use of

AI methods into Twitter spam identification. Be that as it may, tweets are recovered in a streaming way, and Twitter gives the Streaming API to engineers and scientists to get to open tweets progressively. There does not have a presentation assessment of existing AI based streaming spam location techniques. This framework presents highlights which abuse the social entropy, profile attributes, spam examination for spammer's discovery in tweets. We adopt a managed strategy to the issue, yet influence existing hash labels in the Twitter information for building preparing information

Twitter is one such well known system where the short message correspondence (called tweets) has lured an enormous number of clients. Spammer tweets act either like commercials, tricks and help execute phishing assaults or the spread of malware through the inserted URLs. Spam is an issue all through the Internet, and Twitter isn't resistant. Likewise, Twitter spam is considerably more fruitful contrasted with email spam. Different techniques have been proposed by specialists to manage Twitter spam, for example, distinguishing spammers dependent on tweeting history or social traits, recognizing irregular conduct, and ordering tweet-implanted URLs. Bringing of twitters tweets for a specific hashtag. Each hashtag may have 1000 of remarks and new remarks are included each moment, so as to deal with such huge numbers of tweets we are utilizing twiter4j API and perform preprocessing by expelling cites, hash images and spam examination through URL, Number of Unique Mentions (NuMn), Unsolicited Mentions (UIMn), Duplicate Domain Names (DuDn) strategies and googlesafebrowsing API.

II. LITERATURE SURVEY

Twitter fake account detection

Authors: B. Erçahin, Ö. Aktaş, D. Kilinç, and C. Akyol

In the study, author present a classification method for detecting the fake accounts on Twitter. We have preprocessed our dataset using a supervised discretization technique named Entropy Minimization Discretization (EMD) on numerical features and analyzed the results of the Naïve Bayes algorithm.

Detecting spammers on Twitter

Authors: F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida

In this paper authors evaluated the proposed set of features by exploiting very popular machine learning classification algorithms, namely kNearest Neighbor (k-NN), Decision Tree (DT), Naive Bayesian (NB), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost). The performance of these classifiers are evaluated and compared based on different evaluation metrics. We compared the performance of our proposed approach with four latest state of art approaches. The experimental results show that the proposed set of features gives better performance than existing state of art approaches

An integrated approach for malicious tweets detection using NLP

Authors: S. Gharge, and M. Chavan

The authors first collected the tweets related to many trending topics, labelling them on the basis of their content which is either malicious or safe. After a labelling process we extracted a many features based on the language models using language as a tool. We also evaluate the performance and classify tweets as spam or not spam. Thus our system can be applied for detecting spam on Twitter, focusing mainly on analyzing of tweets instead of the user accounts.

Twitter spam detection: Survey of new approaches and comparative study

Authors: T. Wu, S. Wen, Y. Xiang, and W. Zhou

Author worked out a new survey about Twitter spam detection techniques. This

survey includes three parts: 1) A literature review on the state-of-art: this part provides detailed analysis (e.g. taxonomies and biases on feature selection) and conversation (for example advantages and disadvantages on each run of the mill technique); 2) Comparative examinations: we need to analyze the presentation of different regular strategies on a widespread tried (for example same datasets and ground certainties) to give a quantitative comprehension of current strategies; 3) Open issues: the last part is to synopses the unsolved difficulties in current Twitter spam recognition procedures. Answers for these open issues are of incredible hugeness to both scholarly community and ventures. Perusers of this study may incorporate the individuals who do or don't have mastery around there and the individuals who are searching for profound comprehension of this field in order to develop new methods.

III. EXISTINGSYSTEM

- The proposed method combines characteristics with drawal from text content and information of social networks. The authors used matrix factorization to determine the underline feature matrix or the tweets and then came up with a social regularization with interaction coefficient to teach the factorization of the underline matrix. Subsequently, the authors combined knowledge with social regularization and factorization matrix processes, and performed experiments on the real-world Twitter dataset, i.e., UDI Twitter dataset.
- They described the Hidden Markov Model for filtering the spam related to recent time. The method supports the accessible and obtainable information in the tweet object to recognize spam tweets and the tweets that are handled previously related to the same topic.
- The system was analyzed the follow spam on Twitter as an alternative of dispersion of provoking public messages, spammers follow authorized users, and followed by authorized users. Categorization techniques were proposed that are used for the detection of follow spammers. The focus of the social relation is cascaded and formulated into two mechanism, i.e., social status filtering and trade significance
- profile filtering, where each of which uses two-hop sub networks that are centered at each other. Assemble techniques and cascading filtering are also proposed for combining the properties of both trade significance profile and social status. To check whether a user is fake or not, a two-hop social network for each user is focused to gather social information from social networks.

Disadvantages

- There is no filtering system based on a preprocessing schedule and on Naïve Bayes algorithm to discard the tweets containing inaccurate information,.
- Less security due No URL Based Spam Detection.

IV. PROPOSED SYSTEM

- In the proposed system, the system elaborates a classification of spammer detection techniques. We used machine learning algorithms to detect the spam and fake online messages. The proposed taxonomy is categorized into four main classes, namely, (i) fake content; (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Each category of identification methods relies on a specific model, technique, and detection algorithm.
- The principal class (counterfeit substance) incorporates different procedures, for example, relapse expectation model, malware cautioning framework, and Lfun conspire approach. In the subsequent classification (URL based spam discovery), the spammer is recognized in URL through various AI calculations. The third classification (spam in slanting themes) is recognized through and language model dissimilarity. The last classification (counterfeit client distinguishing proof) depends on identifying counterfeit clients through crossover strategies.

Advantages

- The average numbers of verified accounts that were either spam or non-spam and (ii) the number of followers of the user accounts.
- The fake content propagation was identified through the metrics that include: (i) social reputation,(ii) global engagement, (iii) topic engagement, (iv) liking ability, and (v) credibility. After that we utilized regression prediction model to ensure the overall impact of people who spread the fake content at that time and also to predict the fake content growth in future.

V. PROPOSED METHODOLOGY:

This section describes the process of Twitter spam detection by using machine learning algorithms. Fig. 1 illustrates the steps involved in building a supervised classifier and detecting Twitter spam. Before classification, a classifier that contains the knowledge structure should be trained with the pre labeled tweets. After the classification model gains the knowledge structure of the training data, it can be used to predict a new incoming tweet. The whole process consists of two steps: 1) learning and 2) classifying. First, features of tweets will be extracted and formatted as a vector F = {f1,f2,...,fn}. The class labels (spam or non-spam) could be get via some other approaches (like manual inspection). Features and class label will be combined as one instance for training. One training tweet can then be represented by a pair containing one feature vector, which represents a tweet, and the expected result (F ,label), and the training sets the vector TS= {(F1,label1),(F2,label2),(Fn,labeln)}. The training set is the input of machine learning algorithm; the classification model will be built after training process. In the classifying

process, timely captured tweets T = {f1,f2,...,fn} will be labeled by the trained classification model.

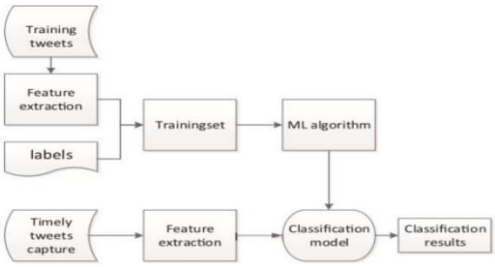


Figure 1 system design

B.Performance Metrics

In order to evaluate the performance of spam detection approaches, some metrics are
imported from information retrieval are widely used by the researchers. 1) Positives and Negatives: Suppose there is a tweet t and the spam class S. The
output of the classifier is whether t belongs to S or not. A common way to evaluate the classifier's performance is to use true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) . These metrics are defined as follows.

a) TP tweets of class S correctly classified as belonging to class S.

b) FP tweets not belonging to class S incorrectly classified as belonging to class S.

c) TN tweets not belonging to class S correctly classified as not belonging to class S.

In order to measure the ability to detect spam, we also import true positive rate (TPR)
and false positive rate (FPR).

a) TPR is defined as the ratio of those spam tweets correctly classified as belonging to class spam to the total number of tweets in class spam, it can be calculated by

$$\text{PRECISION} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

b) FPR is defined as the ratio of those non-spam tweets incorrectly classified as belonging to spam class S to the total number of non-spam tweets

2) Precision, Recall, and F-measure: Literature also uses precision, recall, and F-measure to evaluate per-class performance.

a) Precision is defined as the ratio of those tweets that truly belong class S to those identified as class S, it can be calculated by

b) Recall (which is also known as detection rate in the detection scenario) is defined as the ratio of those tweets correctly classified as belonging to class S to the total number of users in class S, it can be calculated by

$$\text{RECALL} = \frac{\text{TP}}{\text{TP+FN}}$$

c) F-measure is a combination of precision and recall, it is a widely adopt metric to evaluate per-class performance, it can be calculated by

$$\text{F-MEASURE} = \frac{2*\text{PRECISION}*\text{RECALL}}{\text{PRECISON+RECALL}}$$

C.Impact Of Spam To Non-Spam Ratio
In this section, the impact of spam to non-spam ratio of the above-mentioned machine learning algorithms on Datasets I and II is evaluated.

SYSTEM ARCHITECTURE:

DATA FLOW DIAGRAM:

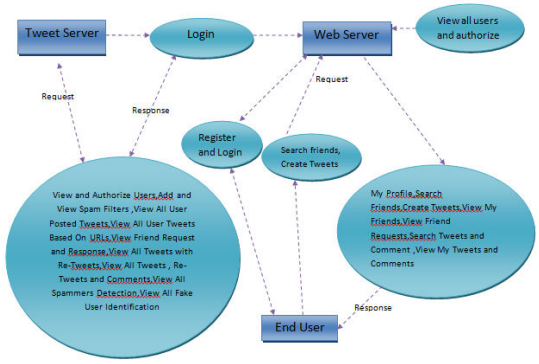


Figure 2 data flow diagram

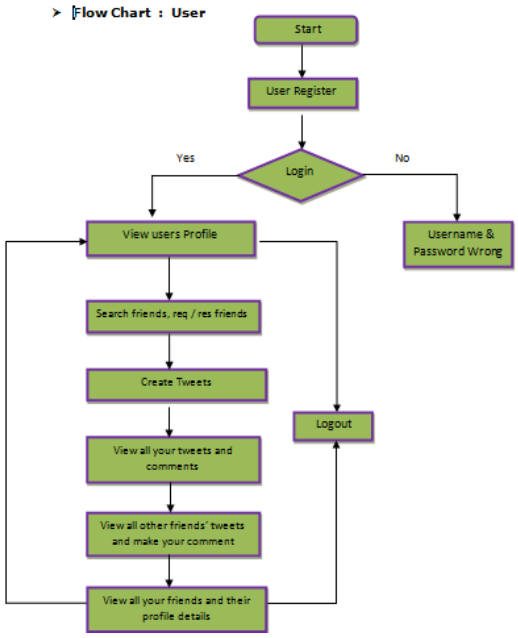


Figure 3 user flow diagram

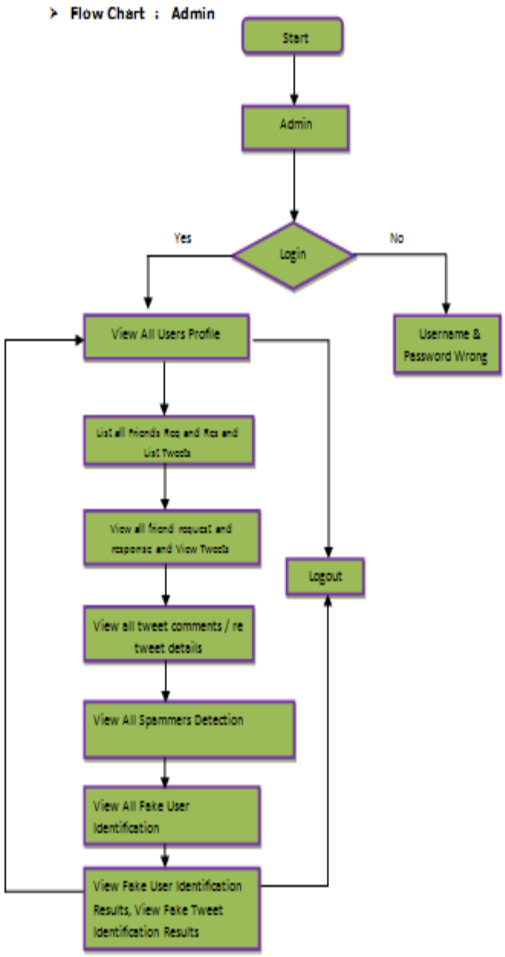
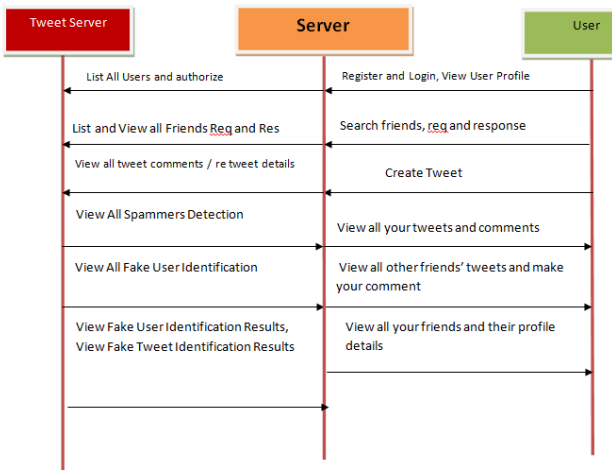


Figure 4 administration flow diagrasm

SEQUENCE DIAGRAM:



VI. MODULES:

Admin

In this module, the Admin has to login by using valid user name and password. After login successful he can do some operations such as View and Authorize Users, Add and View Spam Filters ,View All User Posted Tweets, View All User Tweets Based On URLs, View Friend Request and Response, View All Tweets with Re-Tweets, View All Tweets , Re-Tweets and Comments, View All Spammers Detection, View All Fake User Identification, View Fake User Identification Results, View Fake Tweet Identification Results

User

In this module, there are n numbers of users are present. User should register before doing some operations. After registration successful he has to wait for admin to authorize him and after admin authorized him. He can login by using authorized user name and password. Login successful he will do some operations like My Profile, Search Friends ,Create Tweets, View My Friends, View Friend Requests ,Search Tweets and Comment ,View My Tweets and Comments, View Friend's Retweets and Give Comments.

VII. ALGORITHM:

SVM ALGORITHM:

Support Vector Machine or SVM algorithm is a basic yet ground-breaking Supervised Machine Learning calculation that can be utilized for building both relapse and order models. SVM calculation can perform well with both straightly distinct and non-directly divisible datasets. Indeed, even with a constrained measure of information, the help vector machine calculation doesn't neglect to show its enchantment

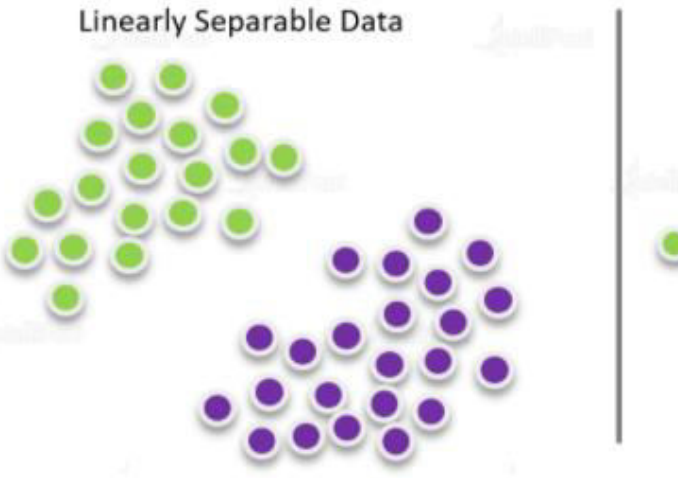


Figure 5 Linearly non separable data

SVM algorithm was designed under the concept of 'decision planes', where hyper planes are used to classify a set of given objects. Let us have examples of support vector machine algorithm. As we can see in Figure 5, we have two sets of data. These datasets can be separated easily with the help of a line, called a **decision boundary**.

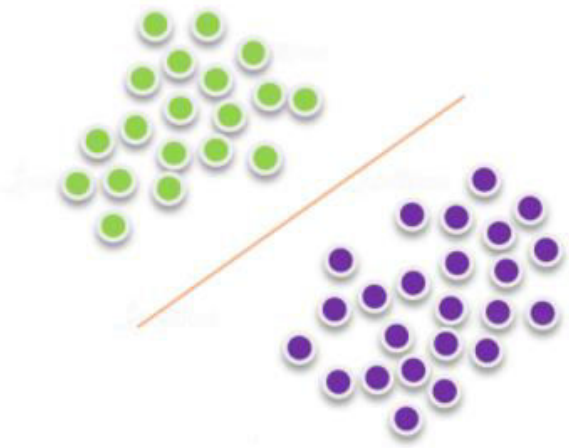


Figure 6 svm decision boundary

However, there can be a few choice limits that can isolate the information focuses with no mistakes. For instance, in Figure 6, all choice limits characterize the datasets effectively.

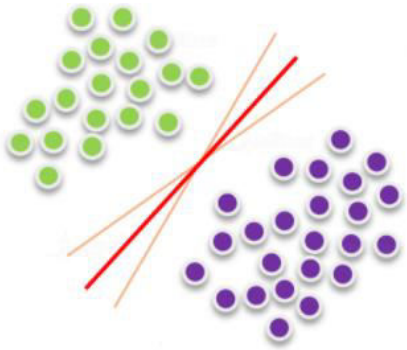


Figure 7 possible decision boundaries

the best decision boundaryis the one which has most extreme good ways from the closest purposes of these two classes, as appeared in Figure 7

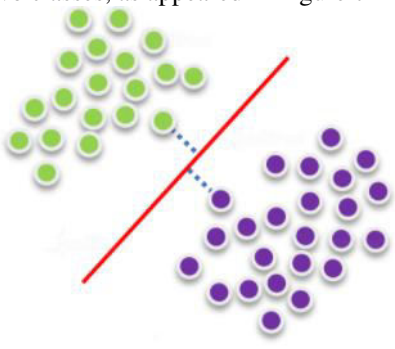


Figure 8 max distance from boundaries

that the nearest points from the optimal decision boundary that maximize the distance are called **support vectors**.

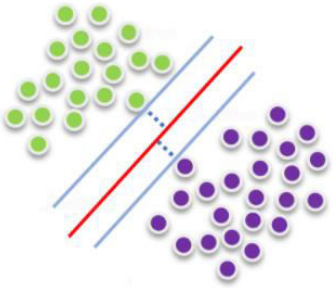


Figure 9 margin and margin classifier

The locale that the nearest focuses characterize around the choice limit is known as the edge.

That is the reason the choice limit of a help vector machine model is known as the most extreme edge classifier or the greatest edge hyper plane. At the end of the day, here's the means by which a help vector machine calculation model works:

First, it discovers lines or limits that accurately group the preparation dataset.

Then, from those lines or limits, it picks the one that has the most extreme good ways from the nearest information focuses.

Okay, in the above help vector machine model, the dataset was straightly distinguishable. Presently, the inquiry, how would we group non-directly distinguishable datasets as appeared in Figure 6

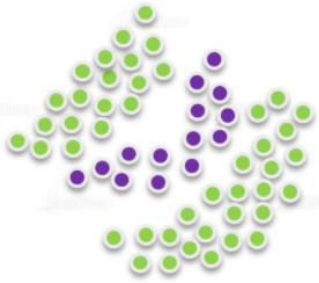


Figure 10 linearly separable data

Clearly, straight lines can't be used to classify the above dataset. That is where Kernel SVM comes into the picture.

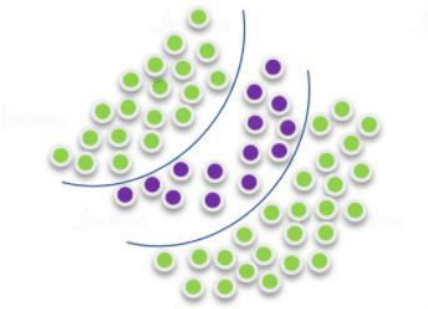


Figure 11 after using SVC classifier

Advantages of Support Vector Machine Algorithm

- Accuracy
- Works very well with limited datasets
- Kernel SVM contains a non-linear transformation function to convert the complicated non-linearly separable data into linearly separable data.

VIII. RESULTS:

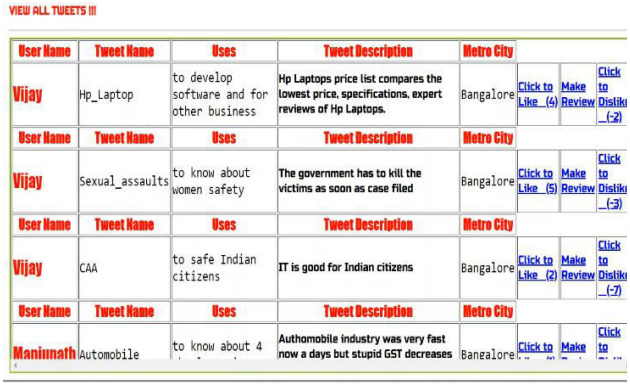


Figure 12 TWEETS

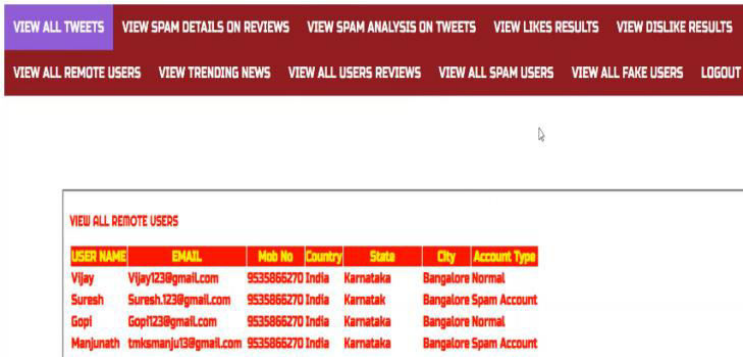


Figure 13 USERS IDENTIFICATION

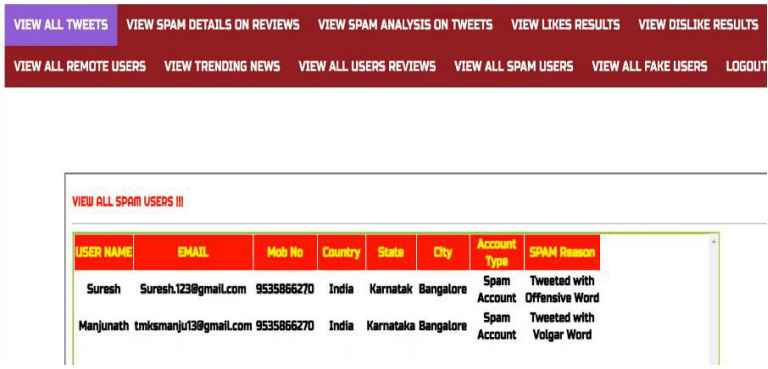


Figure 14 SPAM USERS

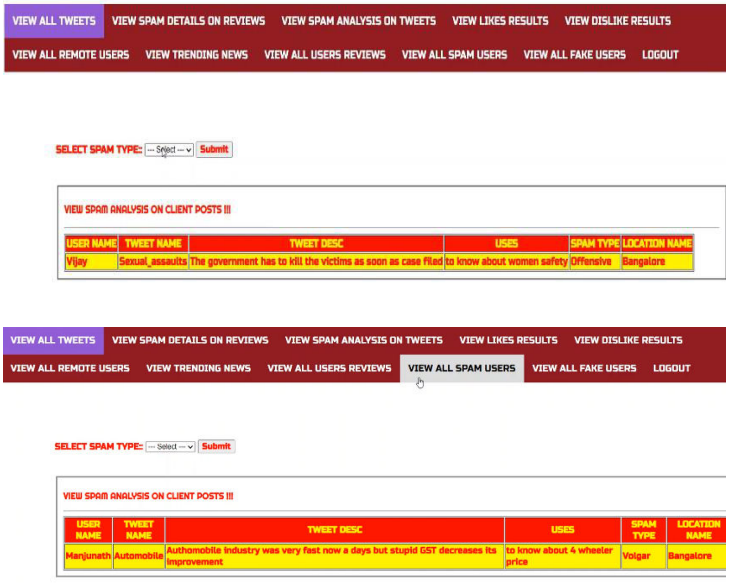


Figure 15 SELECTION OF SPAM TYPE

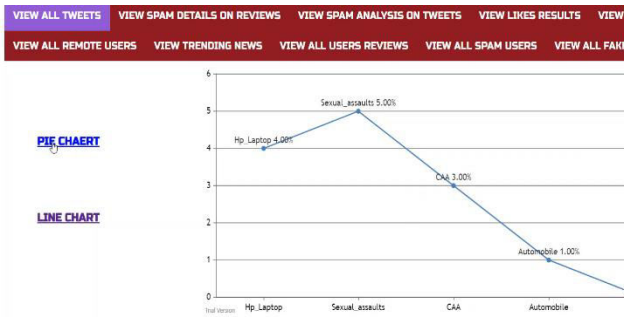


Figure16 LINE GRAPH OF ANALYSED DATA

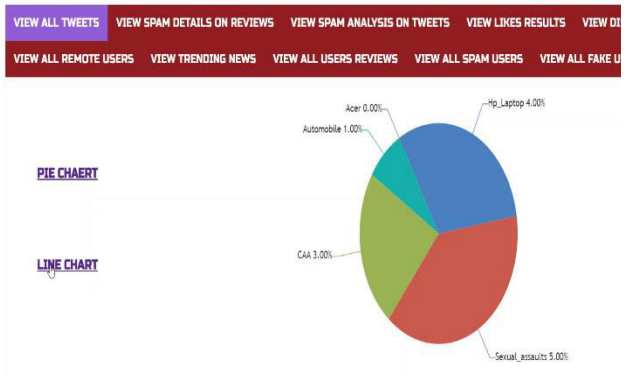


Figure 17 ANALYSIS OF TWEETS

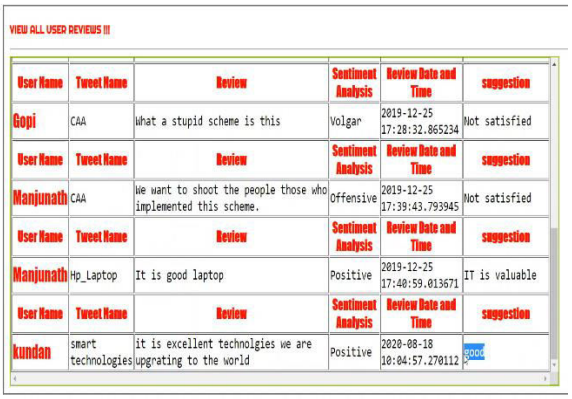


Figure 18 USER POSTS

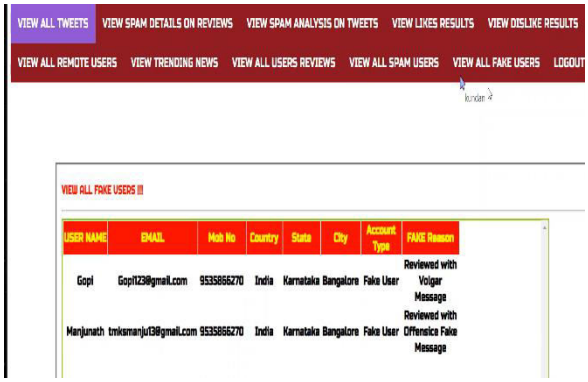


Figure 19 FAKE USERS

IX. CONCLUSION:

By this project we conclude that the spam users and fake online reviews can be efficiently identified using the svm algorithm. It provides the highest accuracy for the real time data.

X. REFERENCES :

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

[2] **2.** F. Benevenuto, G. Magno,T. Rodrigues and V. Almeida, "Detecting spammers on Twitter", Proc. Collaboration Electron. Messaging Anti-Abuse Spam Conf. (CEAS), vol. 6, pp. 12, Jul. 2010.

[3] **3.** S. Gharge and M. Chavan, "An integrated approach for malicious tweets detection using NLP", Proc. Int. Conf. Inventive Common. Computer. Technol. (ICICCT), pp. 435-438, Mar. 2017.

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

[6] A. Gupta, H. Lamba and P. Kumaraguru, "1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter", Proc. Crime Researchers Summit (CRS), pp. 1-12, 2013