IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 1, January 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

Twitter Spam Detection Using Machine Learning Algorithms

Sujata Bahadure, D.K. Chitre*

M. E Students, Department of Computer Science Engineering, Terna Engineering College, Nerul, Navi Mumbai, India

* Professor, Department of Computer Science Engineering, Terna Engineering College, Nerul, Navi Mumbai, India

ABSTRACT: With the trend that the internet is becoming more accessible and our devices being more mobile, people are spending an increasing amount of time on social networking sites. Out of these, twitter is one of the popular micro blogging service site. This popularity of twitter has attracted more & more spammers. Spammers send unwanted tweets to twitter users to promote websites or services. This leads to external phishing sites or malware downloads, which has become a huge issue for online safety & undetermined user experience. In order to stop spammers, although researchers have proposed number of mechanisms, the current solution fails to detect twitter spams precisely and effectively. In order to prevent this attacks, training tweets are added and for real time spam detection, 12 light weight features for tweet representation such as account age, number of followers, number of tweets, number of retweets are extracted. Spam detection mainly builds the classification model which includes the binary classification and further it can be solved by using machine learning algorithms. System reports the impact of the data related factors, such as spam to non-spam ratio, training data size, and data sampling, to the detection performance.. The System shows the spam detection is big challenge and it bridges the gap between the performance evaluation and mainly focus on the data, feature and model to identify the genuine user and report the spams.

KEYWORDS: Machine Learning, Parallel Computing, Spam Detection, Scalability, Twitter

I. INTRODUCTION

Online social networking sites like Twitter, Facebook, Instagram and some online social networking companies have become extremely popular in recent years [1]. People spend a lot of time in OSN making friends with people they are familiar with or interested in. Twitter, founded in 2006, has become one of the most popular micro blogging service sites. Around 200 million users create around 400 million new tweets a day for spam growth [2]. Twitter spam, known as unsolicited tweets containing malicious links that directs the victims to external sites containing the spread of malware, spreading malicious links, etc.[3] hit not only more legitimate users, but also the whole platform. Consider the example, during the election of the Australian Prime Minister in 2013,a notice confirming that his Twitter account had been hacked. Many of his followers have received direct spam messages containing malicious links [4]. The ability to order useful information is essential for the academic and industrial world to discover hidden ideas and predict trends on Twitter. However, spam generates a lot of noise on Twitter. To detect spam automatically, researchers applied machine learning algorithms to make spam detection a classification problem. There are three major types of feature – related solutions for Twitter spam detection. The first type is based on features of user account and tweets contents, such as account age, the number of followers/followings, the no of URLs contained in the tweet, etc. The second approach is to derive the features from social graph [5], which explores the relationship of sender and receivers. The third type of solution focuses on Tweets with URLs. Beside this twitter it has proposed some spam detection schemes to make Twitter as spam-free platform [6]. For instance, twitter has applied some rules to suspend accounts if they behave abnormally. Those accounts, which are frequently sending friend requests, sending duplicate contents or posting URL-only contents will be suspended by Twitter.

We propose a System which characterizes and detects spams on twitter effectively using Machine learning algorithms. It creates a big ground-truth for the research on spam tweet detection. System reports the impact of the data related factors, such as spam to non-spam ratio, training data size, and data sampling, to the detection performance. It xtracts12 lightweight features for streaming tweet spam detection and investigates machine learning algorithms to build up the tweet spam detection model

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

II. RELATED WORK

Twitter spam detection is an important topic in social network security. Many researchers have put their efforts for spam detection on the OSN. However, due to the constant adaption to the spam detection techniques, spammers are still active on the OSNs. Hence, to deal with the spread of spams, a series of methods and solutions have been proposed based on different types of features. Some Researchers used user profile features and message content features to identify spams; some proposed use of graph base features, typically the distance and connectivity of a social graph, and some others relied on embedded URLs as the means of spam detection features.

In this section, we briefly review the related work on Spam Detection and their different techniques.

M. Sangeetha, S. Nityhanathan and M. Jayanthi (2018) in their paper, "Comparison of twitter spam detection using various machine learning algorithms", compares the performance of different machine learning algorithms for detecting twitter spams in terms of accuracy, TPR/FPR and F-measure. By using account and tweet content based features, the performance is compared for various machine learning algorithms. Out of the compared machine learning algorithms (Random Forest, C 5.0, Naive Bayes, Gradient Boosting algorithm, KNN), the Random Forest and C 5.0 gets the high detection accuracy. Performance of Random Forest is more stable than other algorithms. [7]

P. Maragathavalli, B. Lekha, M. Girija and R. Karthikeyan (2018), in their paper "Trends Manipulation and Spam detection in twitter" proposes Naive Bayes algorithm for trend manipulation and Random Forest algorithm for spam detection. Twitter streaming API is used to gather over a million of tweets on hourly trending topics. Processing the raw data from API, the features are extracted from tweets (relevant to spam detection, e.g. twitter account, URL, specific keyword). From the collected tweets, retweets are extracted based on RT symbol. The endogenous factors are TF is calculated by considering both tweets ad retweets. It has been found that spam messages had URLs with much higher frequency, more numeric characters and hashtags. [8]

Isa Inuwa-Dutse, Mark Liptrott and Ioannis Korkontzelos, (2018) in their paper "Detection of spam-posting accounts on Twitter" presents a novel approach for distinguishing spam vs. non-spam social media posts and offers more insight into the behavior of spam users on Twitter. The approach proposes an optimized set of features independent of historical tweets, which are only available for a short time on Twitter. This take into account features related to the users of Twitter, their accounts and their pairwise engagement with each other. In this paper it experimentally demonstrates the efficacy and robustness of this approach and compare it to a typical feature set for spam detection in the literature, achieving a significant improvement on performance. In contrast to prior research findings, it is observed that an average automated spam account posted at least 12 tweets per day at well-defined periods. This method is claimed to be suitable for real-time deployment in a social media data collection pipeline as an initial preprocessing strategy to improve the validity of research data. [9]

III. PROPOSED SYSTEM:-

In proposed system, the process of Twitter spam detection by using machine learning algorithms analyses and classifies tweets as spam and non-spam. Before classification, a classifier that contains the knowledge structure should be trained with the prelabeled 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: learning and classifying. Features of tweets will be extracted and formatted as a vector. The class labels i.e. spam and non-spam could be getting via some other approaches. 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, and the training set is the vector. 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 will be labelled by the trained classification model.

Advantages of Proposed System:

1) The system implements a method that will use the ML mechanism to detect if the post is spam or not.
2) Implementation of system can also be hosted online for use and data will be archived and retrieved from the server.
3) The user with the maximum amount of spam can be blocked by the system.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

System Architecture:

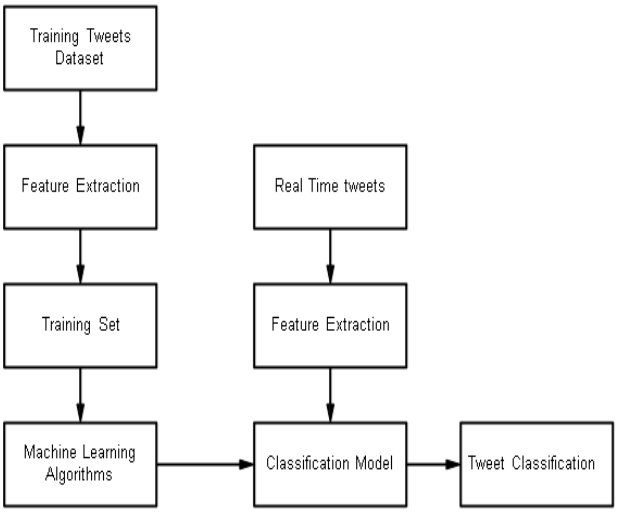

Figure 1:- proposed system

Methodology:

1) Feature Extraction: Extraction of 10-12 features and categories as Tag based features and URL based features. User-based features were extracted from the JSON object "user," such as account_age, which can be calculated by using the collection date minus the account created data. Other user-based features, like no_of followers, no_of followings, no_userfavourites, no_lists, and no_tweets, can be directly parsed from the JSON structure. Tweet-based features includes no_retweets, no_hashtags, no_usermentions, no_urls, no_chars, and no_digits. While no_chars and no_digits needs a little computing, i.e., counting them from the tweet text, others can also be straightforwardly extracted.

Feature Category	Feature Name	Description
Account-based features	account_age	The age of an account
	no_follower	The number of followers
	no_following	The number of followings
	no_userfavorites	The number of favourites this user received
	no_lists	The number of lists the user is a member of
	no_tweets	The number of a user posted tweets
Tweet content-based features	no_retweets	The number of times this tweet has been retweeted
	no_tweetfavorites	The number of favourites this tweet received
	no_hashtag	The number of hashtags in this tweet
	no_usermention	The number of times this tweet being mentioned
	no_urls	The number of URLs contained in this tweet
	no_char	The number of characters in this tweet
	no_digits	The number of digits in this tweet

Table 1: List of Extracted Features

2) Feature Statistics: System evaluate the spam detection performance on dataset by using machine learning algorithms.

3) Machine learning algorithm for Twitter Spam Detection:-

i) Naive Bays: Naive Bayes classifiers are a collection of classification algorithms based on Bayes Theorem**.** It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other. The fundamental Naive Bayes assumption is that each feature makes an:

- Independent
- Equal contribution to the outcome.

Naive Bayes finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem is stated mathematically as the following equation:

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

$$P(A/B) = \frac{P(B/A) * P(A)}{P(B)}$$

Where,

P(A/B) :- Probability (conditional probability) of occurrences of event “A” given that event “B” is true. It is called Posterior.

P(A) & P(B) :- Probabilities of occurrence of events A & B. P(A) is prior probability of proposition and P(B) is prior probability of evidence.

P(B/A) :- Probability of occurrence of event “B” given the event “A” is true. It is called as likelihood.

Random Forest Algorithm:

It is a supervised learning algorithm which builds multiple decision trees & merges them together to get more accurate & stable prediction. As the name suggest, this algorithm creates the forest with a number of trees.

In general, the more trees in the forest the more robust the forest looks like. In the same way in the Random Forest classifier, the higher the number of trees in the forest gives the high accuracy results.

For splitting a node, it searches for the best feature among random subset of features. It is used for both classification & regression. It randomly selects observations & features to build several decision trees and then averages the results.

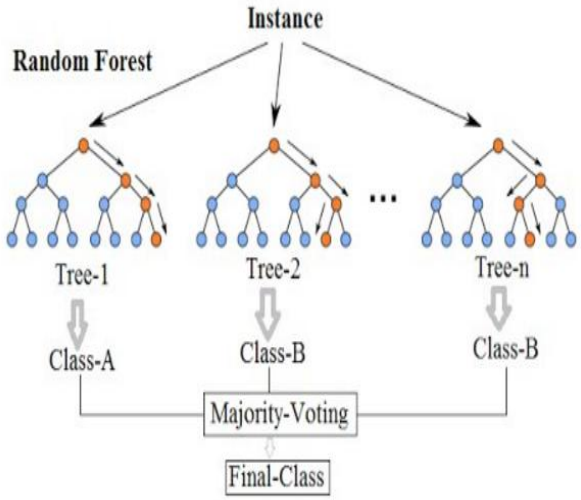

Figure 2: Random Forest Algorithm

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

Working of Random Forest Algorithm:-

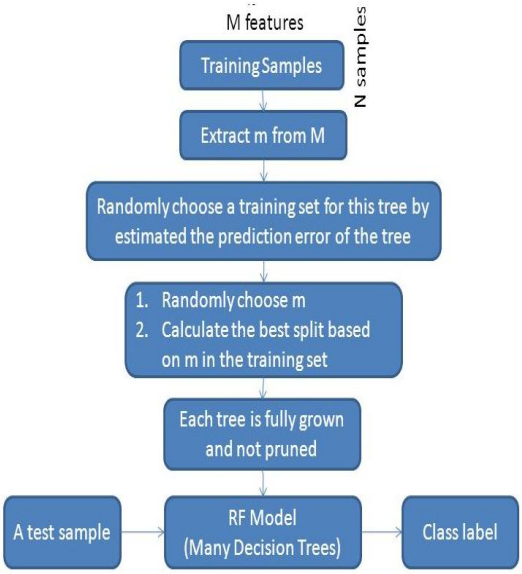

Figure 3: Flowchart for Random Forest Algorithm

As per above flow chart working of Random Forest Algorithm is as given below :

i.Read N Training Samples with M features.

ii.Out of these M features extract m best features randomly .

iii calculate the best split based on m features and construct a decision tree for every sample. Algorithm get the prediction result for every decision tree.

iv.Voting will be performed for every predicted result and at last the most voted prediction will be selected as final predicted result.

IV. PERFORMANCE MEASUREMENT

In order to evaluate the performance of spam detection approaches, following metrics which are imported from information retrieval are used

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.

d) FN tweets of class S incorrectly classified as not belonging to class S

The relationship of TP, FP, TN and FN in social spam detection is shown in below table:

	T(Spam)	F(Non-spam)
T (Spam)	TP	FN
F (Non-Spam)	FP	TN

a) TPR:- 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

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

$$TPR = \frac{TP}{TP + FN}$$

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

$$FPR = \frac{FP}{FP + TN}$$

Precision:- It is ratio of those tweets that truly belong to class S to those which are predicted as class S

$$\mathbf{P}recision = \frac{TP}{TP + FP}$$

Recall:- It is ratio of those tweets correctly classified as belonging to class S to the total number of tweets in class S. It determines how many objects in a class are wrongly classified.

$$\mathrm{Re}\,call = \frac{TP}{TP + FN}$$

F Measure:- It is combination of precision and recall

$$\text{F Measure} = \frac{2\,X \quad \mathrm{Pr}\,ecision\;X \quad \mathrm{Re}\,call}{\mathrm{Pr}\,ecision + \mathrm{Re}\,call}$$

Accuracy:- It is percentage of correctly identified cases (Spam & Non Spam) in total number of examined cases.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

V. RESULTS AND DISCUSSIONS

A. Offline Dataset Results

Naïve Bayes & Random Forest Performance:

For the offline data set results, we collected 200 tweets from the kaggle and UCI library and applied Naive Bayes and Random Forest algorithm for spam detection and compared their performance using performance measuring metrics explained earlier.

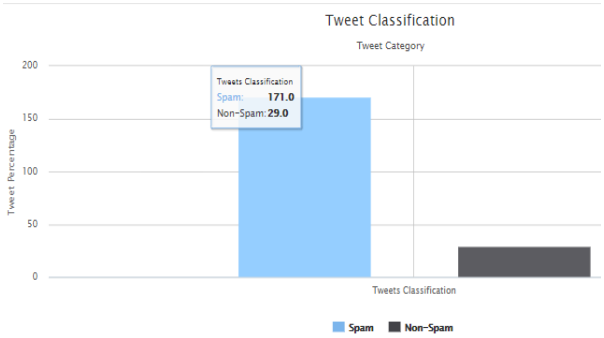

Figure 4: Offline Tweet Classification using Naive Bayes Algorithm

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

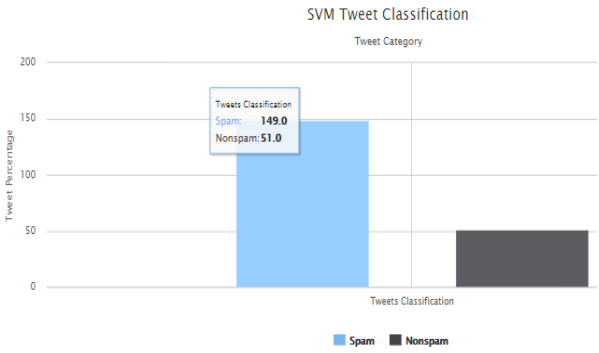

Figure 5: Offline Tweet Classification using Random Forest Algorithm

As shown in Figure 4, Naive Bayes classifies 171 tweets as Spam and 29 tweets as non-spam whereas the Random Forest Algorithm classifies 149 tweets as Spam and 51 tweets as non-spam as shown in Figure 5.

We also measured the TPR, FPR, Precision, Recall, F-measure and Accuracy. As shown in below Figure 6, the Accuracy of Naive Bayes is 90.7% whereas the Accuracy of Random Forest is 79.4 as shown in Figure 7. The values of TPR, FPR, Precision, Recall, F-measure and Accuracy are given in below Table 2 for Offline Tweets Data.

It is very clear from the data obtained that the Accuracy of Naive Bayes Algorithm is much higher than Random Forest Algorithm in offline tweet data set. This higher accuracy of Naive Bayes compared to Random Forest could be attributed to the small data size of tweets as Naive Bayes performs well for small data set.

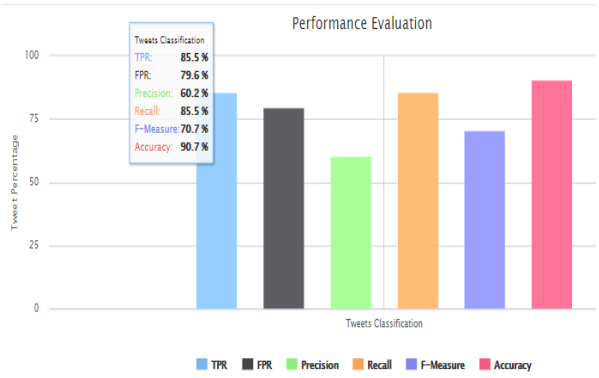

Figure 6: Offline Tweet Performance Evaluations using Naive Bayes Algorithm

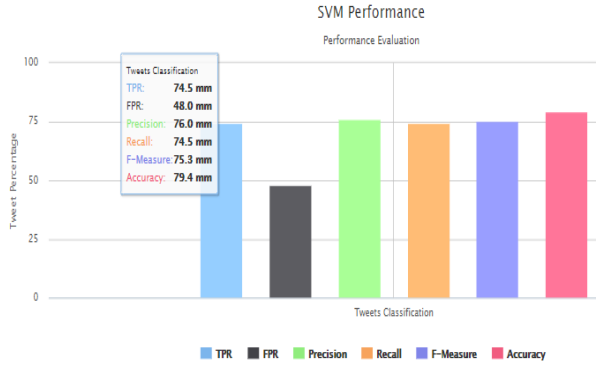

Figure 7: Performance Evaluation of Offline Tweets using Random Forest Algorithm

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

Parameters	Percentage	
	Naive Bayes	Random Forest
TPR	85.5	74.5
FPR	79.6	48.0
Precision	60.2	76.0
Recall	85.5	74.5
F-Measure	70.7	75.3
Accuracy	90.7	79.4

Table 2: Performance Evaluation of Offline Tweets Using Naive Bayes & Random Forest Algorithm

It is also clear from Table 2 that all other performance parameters like TPR, FPR, Precision, Recall and F measure are also better for Naive Bayes compared to Random Forest Algorithm due to the reason explained above.

B. **Online Dataset Results**

Naïve Bayes & Random Forest Performance:

For the online data set results, we collected the dynamic real time data from twitter and applied Naive Bayes and Random Forest algorithm for spam detection and compared their performance using performance measuring metrics explained earlier.

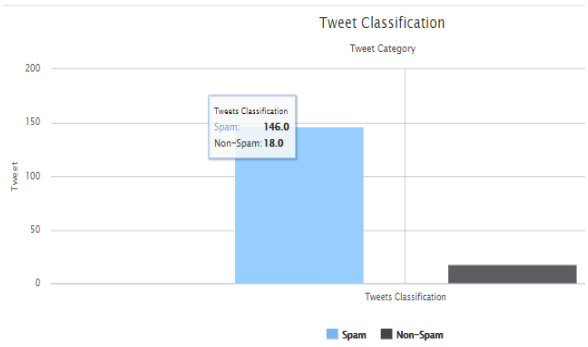

Figure 8: Online Tweet Classification using Naive Bayes Algorithm

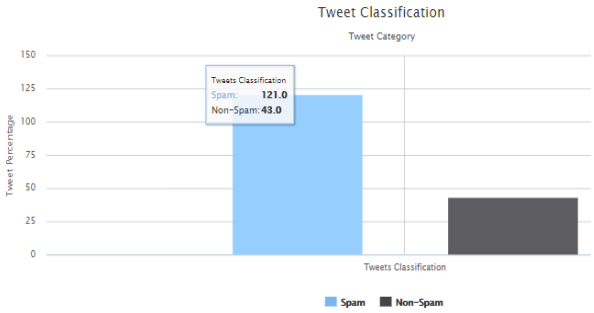

Figure 9: Online Tweet Classification using Random Forest Algorithm

As shown in Figure 8, Naive Bayes classifies 146 tweets as Spam and 18 tweets as non-spam whereas the Random Forest Algorithm classifies 121 tweets as Spam and 48 tweets as non-spam as shown in Figure 9.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

We also measured the TPR, FPR, Precision, Recall, F-measure and Accuracy. As shown in below Figure 10, the Accuracy of Naive Bayes is 93.4% whereas the Accuracy of Random Forest is 78.4 as shown in Figure 11. The values of TPR, FPR, Precision, Recall, F-measure and Accuracy are given in below Table 3 for Online Tweets Data.

It is very clear from the data obtained that the Accuracy of Naive Bayes Algorithm is much higher than the Random Forest Algorithm in online tweet data set. The Naive Bayes algorithms performs well for small data set and in this experiment we have applied Naive Bayes and Random algorithm for small data set and hence we have achieved better performance for Naive Bayes compared to Random Forest Algorithm,

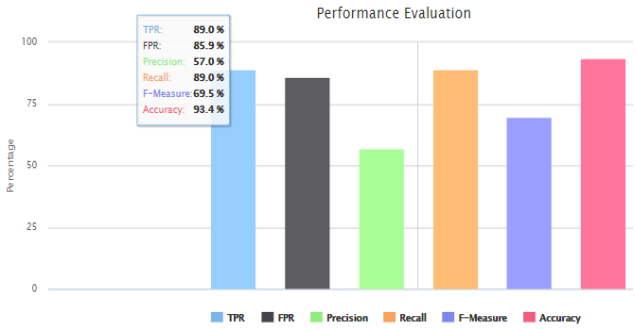

Figure 10: Online Tweet Performance Evaluations using Naive Bayes Algorithm

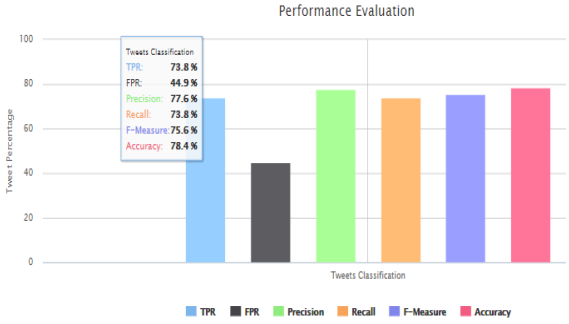

Figure 11: Online Tweet Performance Evaluations using Random Forest Algorithm

Parameters	Percentage	
	Naive Bayes	Random Forest
TPR	89.0	73.8
FPR	85.9	44.9
Precision	57.0	77.6
Recall	89.0	73.8
F-Measure	69.5	75.6
Accuracy	93.4	78.4

Table 3: Performance Evaluation of Online Tweets Using Naive Bayes & Random Forest Algorithm

Table 3 shows that all other performance parameters like TPR, FPR, Precision, Recall and F measure are also better for Naive Bayes compared to Random Forest Algorithm,

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001075

VI. CONCLUSION

In this paper tweet spam classification is done by extracting features from tweets and applying machine learning algorithms like Naive Bayes and Random Forest. For result analysis, we consider different performance metrics like TPR, FPR, Accuracy.

Result analysis shows that Naive Bayes has higher accuracy as compared to the Random Forest Algorithm. System also identifies that feature extraction plays vital role in the classification of tweets as spam and non-spam. Again it has been observed that increase in training only can not bring more benefits to detect tweeter spams after certain number of training samples. In future se will use different classifiers to evaluate performance and feature selection algorithm can be used to reduce redundancy of data set.

REFERENCES

[1] C. P.-Y. Chin, N. Evans, and K.-K. R. Choo, "Exploring factors influencing the use of enterprise social networks in multinational professional service firms," J. Organizational. Computing and Electron. Commerce, vol. 25, no. 3, pp. 289–315, 2015.

[2] H. Tsukayama, "Twitter turns 7: Users send over 400 million tweets per day," Washington Post, Mar. 2013 [Online]. Available: http://articles.washingtonpost.com/2013-03-21/business/37889387_1_tweets-jack-dor sey-twitter.

[3] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammer on Twitter," presented at the 7th Annual. Collaboration. Electron. Messaging Anti-Abuse Spam Conf., Redmond, WA, USA, Jul. 2010.

[4] L. Timson, "Electoral commission Twitter account hacked, voters asked not to click," Sydney Morning Herald, Aug. 2013 [Online]. Available: http://www.smh.com.au/it-pro/security-it/electoral-commission-twitteraccount-hacked-voters-asked-not-to-click-20130807-hv1b5.html.

[5] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," IEEE Transactions on Information Forensics and Security, vol. 8, no. 8, pp. 1280–1293, 2013.

[6] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time url spam filtering service," in 2011 IEEE Symposium on Security and Privacy. IEEE, 2011, pp. 447–462.

[7] M. Sangeetha, S. Nityhanathan, M. Jayanthi, "Comparison of twitter spam detection using various machine learning algorithms", International Journal of Engineering and Technology, 7(1.3) (2018), 61-65.

[8] Dr. P. Maragathavalli, B. Lekha, M. Girija, R.Karthikeyan, "Trends Manipulation and Spam detection in twitter"' International journal for research in applied science and engineering technology (IJRASET) Volume 6, Issue 4, April 2018.

[9] Isa Inuwa-Dutse, Mark Liptrott, Ioannis Korkontzelos, "Detection of spam-posting accounts on Twitter", Neurocomputing 315 (2018) 496–511.

[10] Abdullah Talha Kabakus, Resul Kara, "A Survey of Spam detection methods on twitter", International journal of advanced computer science and applications, volume 8, No. 3, 2017.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details