iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** I **Month of publication:** January 2021

DOI: https://doi.org/10.22214/ijraset.2021.32832

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV dec 2020- Available at www.ijraset.com

Spam Detection in Social Media using Machine Learning Algorithm

Yogita V Biyani[1], Dr. Rahat Afreen Khan[2]
[1]*Department Of Computer Science & Engineering, Deogiri Institute of Engg. & Management Studies, Aurangabad*
[2]*MTech 2nd Year Student, Asst. Professor/Dept of CSE, Deogiri Institute of Engg. & Management Studies, Aurangabad*

Abstract: ***Now a days, peoples are increasing most amount of time in social media and because of the popularity of online social media and increasing social networks, online malicious individuals which are called spammers has started spamming on these platforms for potential victims. Spams invite persons to external phishing sites or malware downloads huge security issue online and undermined the user experience. The increasing number of accounts in social networking sites is a serious and major threat to the internet users. To detect and avoid fake identities, it is need to understand the dynamic contamination. In this paper, tried to analyse various types of spamming attacks spam detection techniques, campaigns in Online Social site networks and information about spam detection.***

Keywords: Spam detection, Social Network Analysis, Machine Learning, fake accounts, data sets, social platforms, Online Social Networks.

I. INTRODUCTION

The platforms of social media have a great impact on many areas today. There is need to focus on to identify the fake and troll identities or persons in the social networks. There are many accounts that are dangerous and malicious to the people on internet. So, to identify the fake accounts in networking sites there are various machine learning techniques are given to overcome of these fake accounts. Using machine learning helps to find the fake accounts of many social media. This paper, discusses five different Spam Detection approach in Social Media such as Statistical Twitter Spam Detection Demystified [1], Data stream clustering [2], Spam detection of Data using Machine Learning Algorithms [3], Machine Learning Techniques [4] and Supervised Machine Learning for the Detection [5]. We have analysed these approaches to evaluate parameters like Performance, Stability and Scalability.

Social media are computer mediated technologies that facilitate the creation and sharing of information, ideas, career interests and other forms of expression via virtual communities and networks [1]. Social media use web-based technologies, desktop computers and mobile technologies to create highly interactive platforms through which individuals, communities and organizations can share, co-create, discuss and modify user generated content posted online. Social media differ from paper-based media or traditional electronic media in many ways including quality, reach, frequency, usability, immediacies and permanence. Various familiar social networking sites are Facebook, Twitter, LinkedIn, YouTube, What's App, Instagram, Skype, Pinterest, Snapchat *etc*. Users of these platforms freely generate and consume information leading to unprecedented amounts of data. Several domains have already recognized the crucial role of social media analysis in improving productivity and gaining competitive advantage. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in the entertainment industry to complement intuition and experience in business decisions and in politics to track election processes, promote wider engagement with supporters and predict poll outcomes. However, alongside the benefits, the rapid increase in social media spam contents questions the credibility of research based on analyzing this data. A report by Nexgate estimates that on average one spam post occurs in every 200 social media posts and a more recent study reports that approximately 15% of active Twitter users are automated bots. The growing volume of spam posts and the use of autonomous accounts (social bots) to generate posts raises many concerns about the credibility and representativeness of the data for research.

A. Spam and Spamming

Social spam is unwanted content appearing on social networking services [1]. The social spam can be in many ways including bulk messages, profanity, insults, hate speech, malicious links, fraudulent reviews, fake friend's etc. The major problem faced by users using these various networking sites is of Spam. However, some of these users, called as Spammers, started to misuse social media platforms by spreading misinformation, malicious links, unsolicited messages, fake news to legitimate users. Spam is any unwanted or prohibited behaviour that directly or indirectly violates the certain rules of any networking site.

In this dissertation, we have done a survey of research papers. This work aims to define the various types of techniques used to detect spam in various social networks. This work also aims to give a review of how these techniques have been implemented by various researchers. In this study, we focus on Twitter and we proposed effective approach to detect and filter unwanted tweets, complementing earlier approaches in this direction. Previous studies rely on historical features of tweets that are often unavailable on Twitter after a short period of time, hence not suitable for real- time use.

B. Need of Spam Detection

After various researches, the researchers have concluded that there are many works have been conducted about social spam detection; however, most previous work on social spam has concentrated on the methods and techniques for spam detection and prevention on a single social network. It has been identified that these works are done either for Facebook or MySpace or Twitter. To understand the approach, it is noted that various methods have been discussed in these researches such as collaborative filtering, friend graph analysis, classification, behavioural analysis *etc*. Authors have taken into consideration the key findings from the previous researches while proposing the new framework. Different classifiers proposed by various researchers have been tested in spam detection earlier and it is found that it is a big challenge to choose the right one for the same purpose.

C. Objectives

As stated earlier, the purpose of this work is to detect the spam on social networking sites and to prevent users from spamming.
The scope of this work mainly deals with data to improve the following key issues;
To use efficient and practically reasonable method of combination.
To detect the spam on social networking sites like Twitter, Facebook, Skype, *etc*.
To Improve the Accuracy of the work which was carried out.

II. LITERATURE SURVEY

In research literature, many spam detection methods have been studied to provide various malware detection schemes and improve the performance, stability, scalability.
The [1] have worked on evaluation of the spam detection performance on dataiu0hu 8i90set by using machine learning algorithms. The process of Twitter spam detection is done by using machine learning algorithms efficiently. Before classification, a classifier that contains the knowledge structure should be trained with the pre-labelled tweets.
The [2] has proposed A Review on Spam Detection this study we conclude that there are different approaches in spam detection. Some approaches used features to detect spam with the Machine learning algorithms. Futures involved were content features, user-based features, URL-based features, features based on social graph.
The [3] has worked on Identification of the Human or Bots Twitter Data that aiming at to identify the malicious activities in social contact using machine learning engineered techniques. The has worked on Twitter Spammers Detection. The proposed method compares the performance of three different Machine Learning algorithms in tackling this spam detection task. The experimental session involves a publicly available dataset.
The [4] have proposed Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network that presents a methodology to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to a real profile within the same network by analysing the content of comments generated by both profiles.
[3] in their paper ‘Using Social Network Analysis for Spam Detection’ describes about the use of centrality in the social graph of a social networking site to predict spam detection such as the probability of a user is likely to post spam in a social network. In another research about Twitter, Wang mentioned about another technique which is the use of graph-based metrics to improve spam classification on a microblogging platform.
[2] presented a scheme utilized for identifying spam URLs in social sites which have been used to protect users from links that are related with malware and other low-quality suspicious text. The behavior has been analyzed using two different schemes (i) initially, study the links posted by public on Twitter; (ii) secondly is how these links are accessed by the user.
[10] proposed a classifier for detecting the spam. To analyze the outcome of the designed algorithm the experiment has been carried out on two different datasets such as “YouTube Comments dataset” and YouTube spam collection dataset.
[9] proposed a system to differentiate between genuine and suspicious behavior of the message. The performance for precision, F-measure has been obtained and concluded that Bayesian classifier works well among other existing algorithms.

[3] presented a hybrid machine learning algorithm that comprises of SVM as a classifier and whale as an optimization algorithm. These algorithms have been utilized for recognizing spammers in OSNs. presented an automatic detection system for providing security to twitter users against spammers. In this paper, the author used three classification techniques named as SVM, random forests and decision tree. From the experiments, it has been observed that when random forest used with decision tree algorithm the overall accuracy up to 95% have been obtained. In the case when the random forest is used with behavioral feature the accuracy of the detection system increases and become 97.877%.

[8] presents that simple unsupervised algorithm can be used in spam detection. This algorithm uses statistical properties of effective spam profiles. It says that these properties help to deliver extremely accurate and speedy algorithm for detecting spam. Studies shows that due to the advancement of technology, social networks such as Facebook, MySpace, LinkedIn, Friendster and Tickle have large number of members, almost millions, who use them as both social networking as well as business networking. Latest studies are carried out to influence social network into email spam discovery according to the Bayesian likelihood algorithm. The concept this algorithm is to use social relationship between sender and recipient to decide proximity and trust value. The next step is augment or decrease Bayesian probability according to these obtained values.

III. PROPOSED SYSTEM

Spam detection is a machine learning task where we want to determine which the general detection of a given document is using machine learning techniques and natural language processing, we can extract the subjective information of a document and try to classify it according to its polarity such as positive, neutral or negative. It is a really useful analysis since we could possibly determine the overall opinion about a selling object, or predict stock markets for a given company like, if most people think positive about it, possibly its stock markets will increase, and so on. Sentiment analysis is actually far from to be solved since the language is very complex (objectivity/subjectivity, negation, vocabulary, grammar) but it is also why it is very interesting to working on. In this project I choose to try to classify tweets from Twitter into “positive” or “negative” detection by building a model based on probabilities. Twitter is a microblogging website where people can share their feelings quickly and spontaneously by sending a tweet limited by 140 characters. You can directly address a tweet to someone by adding the target sign “@” or participate to a topic by adding a hashtag “#” to your tweet. Because of the usage of Twitter, it is a perfect source of data to determine the current overall opinion about anything.

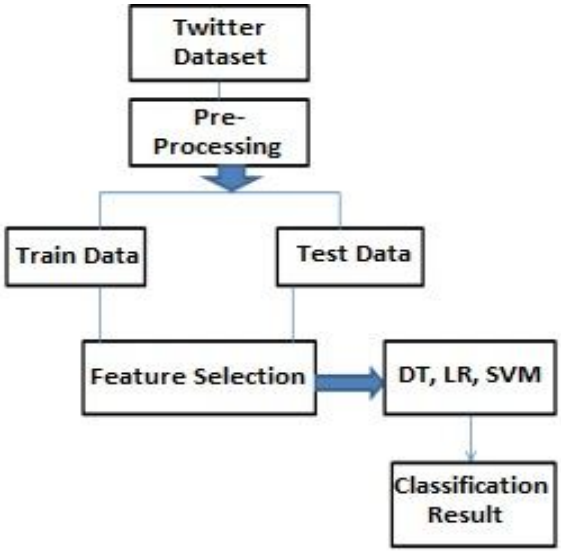


Figure 1: Block Diagram of Proposed System

A. Dataset and Outline

We use one public dataset Social Honeypot dataset to validate the effectiveness of our proposed features. The dataset used for the current project work is called Social Honeypot dataset. The social honeypot dataset was collected from Twitter on Kaggle. The dataset was used for experiments in the paper. The dataset contains 22,223 content polluters, their number of followings over time, 2,353,473 tweets, and 19,276 legitimate users, their number of followings over time and 3,259,693 tweets.

B. Data Preprocessing

Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data Pre-processing is a technique that is used to convert the raw data into a clean data set. In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.

There are 4 main important steps for the preprocessing of data.

1) Splitting of the data set in Training and Testing sets
2) Taking care of Missing values
3) Taking care of Categorical Features
4) Normalization of data set

C. Feature Selection

Feature selection is a process where you automatically select those features in your data that contribute most to the prediction variable or output in which you are interested.

Having irrelevant features in your data can decrease the accuracy of many models, especially linear algorithms like linear and logistic regression.

Three benefits of performing feature selection before modeling your data are:

1) *Reduces Overfitting:* Less redundant data means less opportunity to make decisions based on noise.
2) *Improves Accuracy:* Less misleading data means modelling accuracy improves.
3) *Reduces Training Time:* Less data means that algorithms train faster.

Table 1: Features of dataset

Feature No.	Feature Description
1	Id
2	Created_At
3	Viewed_At
4	Friends_Count
5	FollowersCount
6	Status_count
7	LenScreenName
8	LenProfile

D. Classification Algorithms

1) *Support Vector Machine (SVM):* The support vector machine was proposed to interpret the pattern recognition issues. The data is mapped in an exorbitant dimensional input space using this approach and then designs an ideal disjointed hyperplane within this expanse. A quadratic software issue is mostly used, whereas for neural network architectures, an inclined tutoring approach, on the one hand, deteriorates from the genuineness of most of the native minima. An SVM model is basically a representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated in an iterative manner by SVM so that the error can be minimized. The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH).

The followings are important concepts in SVM −

1) *Support Vectors:* Datapoints that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points.
2) *Hyperplane:* As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
3) *Margin:* It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

The main goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane (MMH) and it can be done in the following two steps −
First, SVM will generate hyperplanes iteratively that segregates the classes in best way.
Then, it will choose the hyperplane that separates the classes correctly.

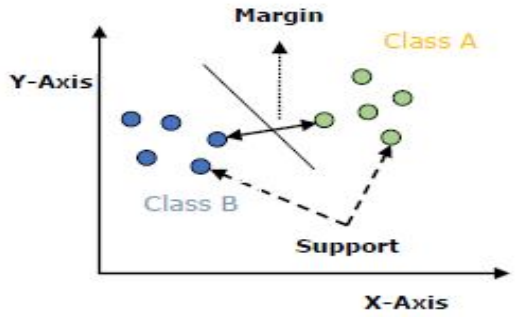


Figure 2: Support Vector Machine

A hyperplane or an array of hyperplanes betwixt classes is construed by a support vector machine if the classes are properly disjointed from one another; a decent disjoining created by SVM, since the overall greater the separation betwixt data points and the hyperplane, the less the fallacy can be gotten by the classifier.
A Support Vector Machine (SVM) is a supervised machine learning algorithm that can be employed for both classification and regression purposes. SVMs are more commonly used in classification problems and as such, this is what we will focus on in this post. SVMs are based on the idea of finding a hyperplane that best divides a dataset into two classes, as shown in the image below.

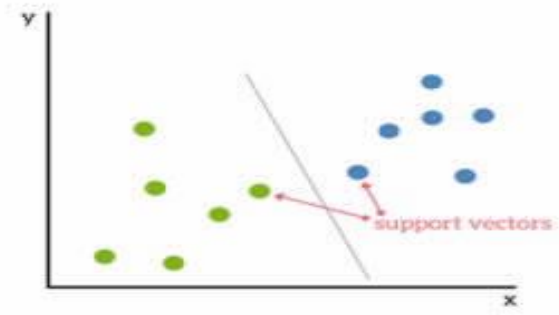


Figure 3: Finding Hyper plane in SVM

2) *Identify the right hyper-plane (Scenario-1):* Here, we have three hyper-planes (A, B and C). Now, identify the right hyper-plane to classify star and circle.

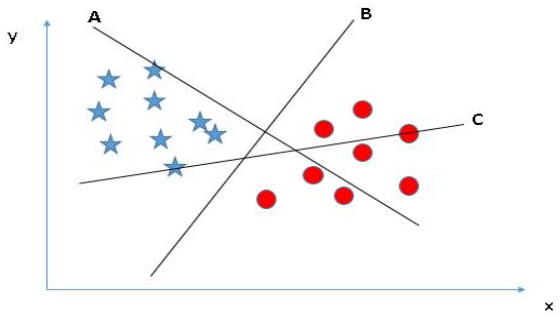


Figure 4: Finding Hyperplane in SVM (Scenario-1)

You need to remember a thumb rule to identify the right hyper-plane: “Select the hyper-plane which segregates the two classes better”. In this scenario, hyper-plane “B” has excellently performed this job.

E. Classification Result

Classification is performed using Support Vector Machine, Decision Tree, Logistic Regression Machine Learning algorithm. Input to classifier is input and trained data. Classifier compares input with trained data and predicts result.

F. Evaluation Metrics

Several performance measurements are suggested to estimate the classification model, and they include recall, error rates, accuracy, precision, and so on. Six metrics have been considered, and they rely on the outcomes of designing chaotic matrix. Accuracy, precision, recall, specificity, and f1 are therefore utilized as a performance matrix for this piece of research.

1) *Accuracy:* The most unlearned performance measure is the Accuracy. It is the rate of accurately classified tags on every forecast which can also be estimated by employing the formula below;

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

2) *Precision:* The accuracy that is part of the most popularly utilized performance measure is the Precision. The rate of the accurate affirmative tag over all the accurate forecasts, encompassing accurate inspections which are inaccurate, can be termed as Precision. It can be calculated using the formula below;

$$precision = \frac{TP}{TP + FP}$$

3) *Recall:* The recall is the entirety that is likewise defined as responsiveness or verifiable affirmative ratio. It is the rate of accurately forecasted affirmative tags. The denominator of recall formula calculates all affirmative activity, notwithstanding where they were forecasted accurately by the prototype.

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

4) *F-1 Score:* Fist we calculate precision and recall. Precision is the ratio of selected accounts that are spammers. Recall is the ratio of spammers that are detected so. F1-score is the harmonic mean of precision and recall.

$$f1 - score = \frac{2 \times precision \times recall}{precision + recall}$$

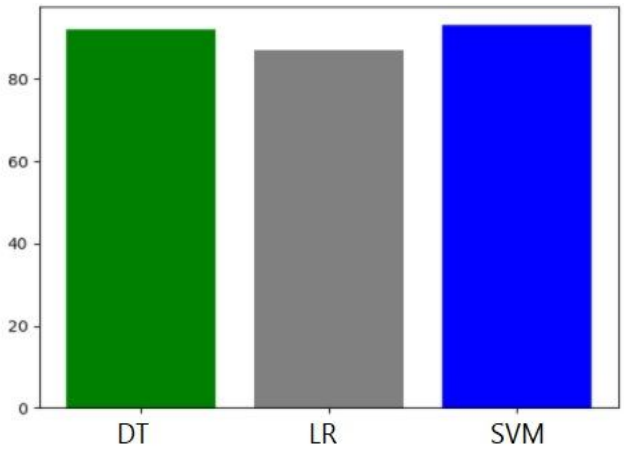
IV. RESULT ANALYSIS

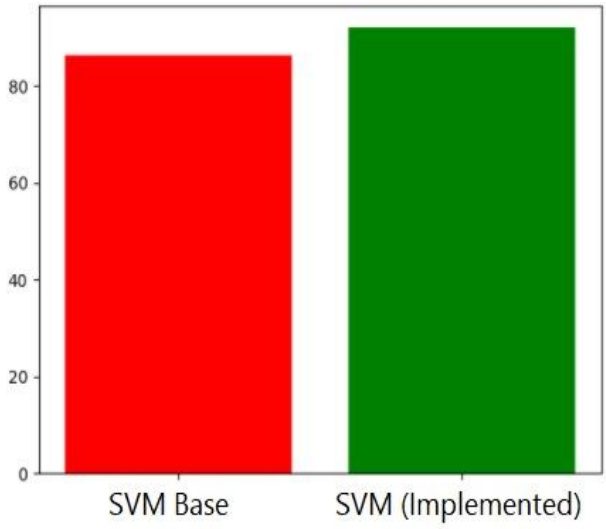
From the result, it is evident that Support Vector Machine classifier provides the best possible outcome on the parameter tweets in the dataset which is highest among the other used classification algorithms. The Support Vector Machine was selected as the main algorithm for developing learning model and it was then modified to yield more accurate results. Proposed system is compared to the previous techniques for analysing which system is best suit for prediction of Spam Detection.

Table 2: Comparision with the existing system

Sr. No	Method	Parameters			
		Accuracy	Precision	Recall	F-1 Score
1	DT	90.33 %	85 %	95 %	89.79%
2	LR	85.70 %	75.90 %	87%	77.98%
3	SVM	92.04 %	90.3 %	92.8%	91.88%

The above table shows the comparisons of the various techniques using the respective methods and the graphical representation of the evaluation parameters are given below.





The above graph shows the accuracy of the proposed system is greater than the previous implemented algorithms.

V. CONCLUSION

Nowadays, Spam Detection is a hot topic in machine learning. We are still far to detect the spams of the corpus of texts very accurately.In this project I tried to show the basic way of classifying tweets into real or spam category using as baseline and how language models are related to the algorithm and can produce better results. The python code was used to run the machine learning algorithms which are support vector machine (SVM), decision tree, and the logistic regression to find out which algorithm is the suitable for Twitter to identify the spam and real accounts. From the results received, it is identified that the support vector machine algorithm provides more accurate results compared to other algorithms. Moreover, the python program identified eight best features that can be used to identify the account, whether it is spam or not they are Id, Created_At, Viewed_At, Friends_Count, Followers_Count, Status_count, Len_Screen_Name, Len_Profile.

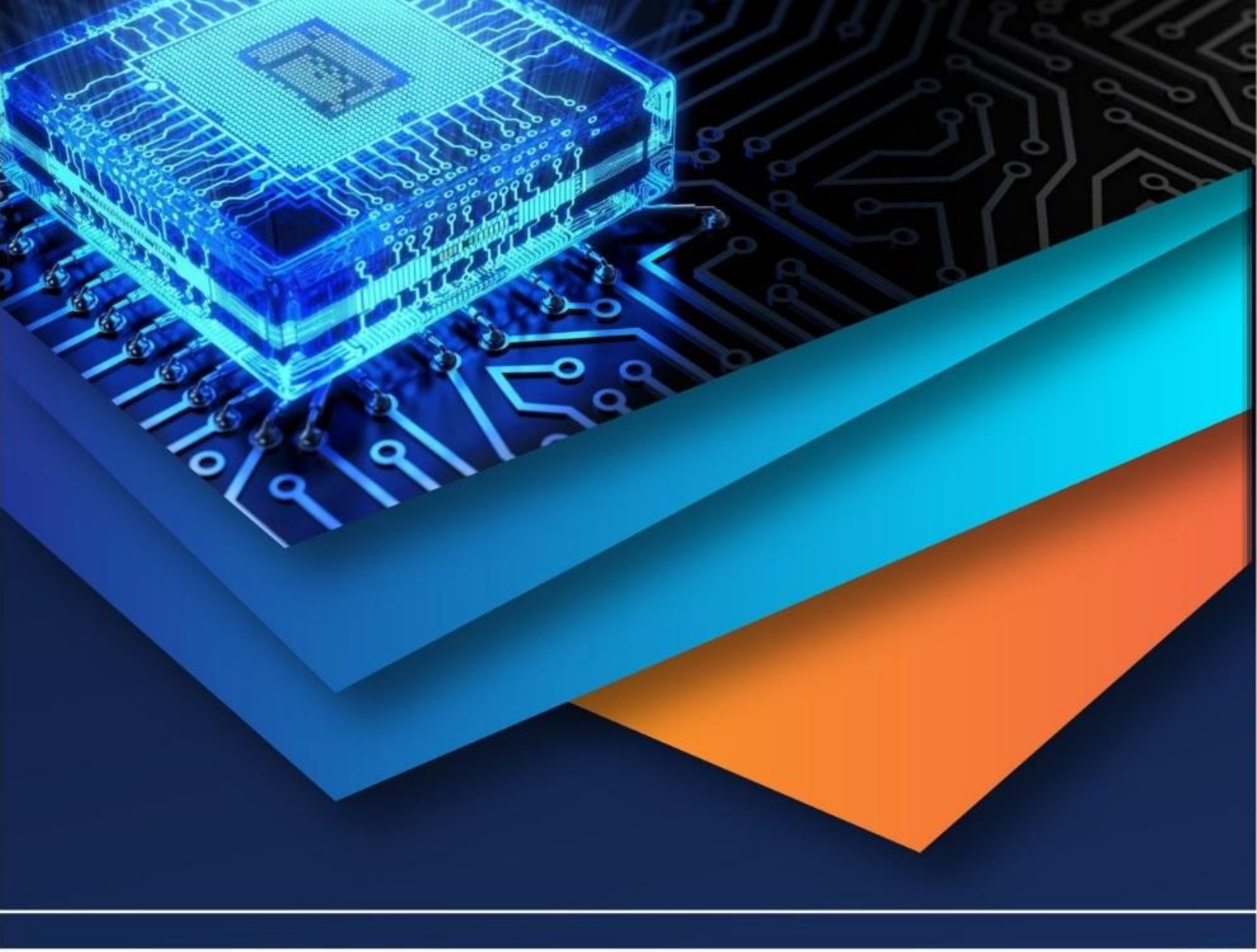
VI. FUTURE WORK

The future works that are suggested for this project are provided below:

A. Collect more twitter account details using Twitter API.
B. Run the Classification algorithm more than once to take an accurate value for the performance measurement**.**
C. Create an excel model that will automatically classify the collected account into real and spam accounts.

REFERENCES

[1] Rohit V.Adagale, Aniket C.Sanap, Anil V.Gitte, Prof. R. H. Kulkarni, "A Survey on Statistical Twitter Spam Detection Demystified: Performance, Stability and Scalability", International Journal of Interdisciplinary Innovative Research & Development (IJIIRD), ISSN: 2456-236X Vol. 02 Issue 02 | 2018.
[2] Z. Miller, B. Dickinson, W. Deitrick, W. Hu, and A. H. Wang, "Twitter spammer detection using data stream clustering," Inf. Sci., vol. 260, pp. 64–73, Mar. 2014.
[3] Nambouri Sravya, Chavana Sai praneetha, S. Saraswathi, "Identify the Human or Bots Twitter Data using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), Volume: 06 Issue: 03 | Mar 2019.
[4] CLAUDIA MEDA, FEDERICA BISIO, PAOLO GASTALDO, RODOLFO ZUNINO DITEN, "Machine Learning Techniques applied to Twitter Spammers Detection", Recent Advances in Electrical and Electronic Engineering, ISBN: 978- 960-474-399-5, AUGUST, 2016.
[5] Patxi Gal´an-Garc´ıa, Jos´e Gaviria de la Puerta, Carlos Laorden G´omez, Igor Santos and Pablo Garc´ıa Bringas, "Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Cyberbullying", IET Software, Vol. 6, Iss. 6, MAY 2014.
[6] D. Kim, Y. Jo, I.-C. Moon, A. Oh, Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users, in: CHI 2010 Work. Microblogging What How Can We Learn From It, Atlanta, Georgia, USA, 2010. doi:10.1.1.163.7391.
[7] Using Twitter lists, Twitter. (2017). https://support.twitter.com/articles/76460 (accessed February 5, 2017).
[8] Verma, M., & Sofat, S. (2014). Techniques to detect spammers in twitter-a survey. International Journal of Computer Applications, 85(10).
[9] Wang, D., Irani, D., & Pu, C. (2011, September). A social-spam detection framework. In Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (pp. 46-54). ACM.
[10] Isa Inuwa-Dutse*, Mark Liptrott, Ioannis Korkontzelos "Detection of spam-posting accounts on Twitter" 6, AUGUST , 2018



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY