iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: https://doi.org/10.22214/ijraset.2020.30337

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 VII July 2020- Available at www.ijraset.com

Spam Detection in Twitter

Mr. Girish Kumar D[1], Pavani K[2], Neha J[3], Aishwarya K[4], Amulya T. G. M[5]
[1]*Assistant professor, CSE Dept, BITM, Bellary, Karnataka, India.*
[2, 3, 4, 5]*B.E Student in computer science, BITM, Bellary.*

Abstract: ***In today's world humans have the ability to connect and communicate through social media independent of time or location. Since past two decades social media and internet has made a tremendous change in digital platform. Those were the old days where letters were written and posted. As technology is evolving digital media has changed the game over years, now there is Gmail, twitter, linked-in, what's- app, Instagram, telegram and many such applications which connects the world from one end to other where there is exchange of knowledge human to human connection and communication where association and exchange of ideas has become convenient and uncomplicated. Like everything created has its own pros and cons the digital media has its own. The improper use of social media has been misleading the bona fide users, because there is lot of personal and confidential information being shared on digital media such as twitter, Instagram and so on. The information is being hacked or leaked by the spammers. These spammers hack the genuine users account and post illegal information and abusive comments. To avoid this a methodology has been put forth which can be done by algorithms using machine learning. The performances of this model are measured using recall and F measure techniques. To detect spam we use Naïve bayes classification algorithm and support vector machine algorithm.***

Keywords: Machine learnig, Naïve-Bayes classifier, Random forest algorithm.

I. INTRODUCTION

The networking sites like Twitter, Instagram, facebook and many other digital networking sites permits to post a picture or share one's opinion and knowledge about the current affairs. Social media has been dominating human stability since past two decades, the success ratio can be measured by the increase in number of users every second. As per the survey for the past two years ie; in 2018 the number of facebook users in India went up to 243.3 M and in 2020 the graph has been rapidly increased to 346.2 M. In the first quarter of 2020 the monthly active users are of 2.3 B. Twitter is a social networking site where the conversational nature makes it look great to tweet and retweet stuff posted by users where business conversations can also be done in private. It is a microblogging system which allows us to share pictures, ideas, knowledge, articles on current affairs and so on. Because of the majority of users twitter has increased the characters for 140 to 280. Community groups can be formed between family, friends, business professionals, IT professionals where technical discussion can be interacted.

Top stars from film fraternity, business professionals, CEO's, users can follow their people from these respective professions and other successful people. The process of creating a twitter account is very convenient and uncomplicated because the users can switch on to any language required. Soon after filling their details a user name and password must be given, here the password should be of minimum 8 characters if less characters are given then it shows that the password strength is weak. This process is programmed in this way because of users security and privacy. The spammers create a fake account and mislead the genuine users by sending a link or by posting illegal information, these spammers hack the users account and are active on all networking sites.

Twitter has already come up many ways to report such fake accounts in order to avoid attacks from hackers and spammers. The user can post a tweet using a hash tag as a mark of reporting the spam accounts or can directly report the account. The accounts which are reported are suspended immediately. There are times where while filtrating or separating these spam accounts the real user accounts are also filtered. This issue should be avoided where the spam accounts and messages should be automatically filtered and detected.

The main objective of this model is to classify and detect the spam messages in twitter using machine learning algorithms. The primary and important goal of this project is to detect the spam messages using content based features then designing a model which analyse the approach of spam detection on features that are preferred. We also use SVM, Naïve bayes and these classifiers are compared by their performances. The final output exhibit the result of spam detection which shows the recall and F measure and accuracy of the model.

In abstract the discussion is about social media and how Twitter is playing its role on digital platform. In Introduction the discussion is about how twitter works and how the spammers misuse the users account. Further discussion will be on related work, proposed methodology, experimental results and the conclusion concluding the extension work of this model.

II. RELATED WORK

Twitter is a social networking site which is popular like other sites including instagram, facebook where people voice their opinions using tweets. In recent days twitter has got immense popularity due to its easy usage and the fact you can interact with anyone in any part of the world has added to its positive side. In Twitter, user's uniqueness is identified by their user ID. User can follow other users whom he/she finds interesting to follow and can allow certain users to follow them, that way privacy is maintained. One can tweet and use # in his tweets and others can re-tweet to the original tweet. All the users can protect his accounts using the privacy guidelines given by twitter, certain users create chaos by posting a lot of unsolicited tweets to seek attention, it is important to filter spam tweets to create spam-free environment. to avoid this many have come up with their methodologies.

Title	Methodology used	Results Accuracy
Twitter" [1]		
"Who is Tweeting on Twitter: Human, Bot, or Cyborg?"	Bayesian networks	TPR with 90.7%
"CATS: Characterizing Automation of Twitter Spammers"	Random ,decorat forest naïve e, bayes.	92.6%
"SPAM: A Framework for Social Profile Abuse Monitoring"	Random Forest, SVM	88%
"Spam Detection on Twitter Using Traditional Classifiers"	Naïve- KNN, Bayesian, Random Forest	89.7%
"A study of effecti features for ve long- detecti surviving ng spam Twitter accounts"	Decision tree	Precision with 84%
"Twitter Spammer Profile Detection"	SVM, Naïve Bayes	89.6%
"Detecting Spammers on Social Networks"	Random Forest Algorithm	90.93%
"Making the Most of Tweet-Inherent Features for Social Spam Detection on Twitter"	Naïve Bayes, SVM, KNN.	Precision with 90.6%
"Detecting malicious tweets in trending topics using a statistical analysis of language"	Decision tree, KNN, Random forest.	94.5%

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

III. PROPOSED METHODOLOGY

We collected tweets from websites like kaggle in accordance to user account ID and in order to classify the tweet as spam or non-spam we use certain words as keywords and collect the data necessary.

US = cluster of exceptional words in spam tweets' data.

UNS = Cluster of exceptional words in the non- spam tweets' data.

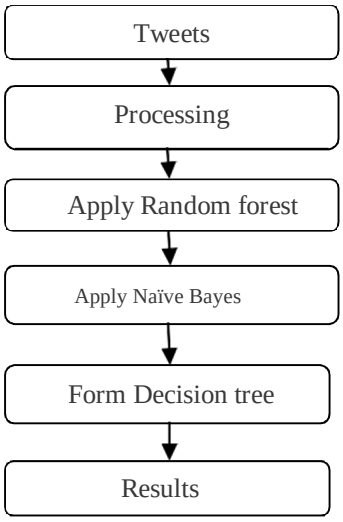
For every word T in US and UNS we figure the accompanying probability esteems:

$$P(\mathcal{T}|\mathcal{U}_{\mathcal{S}}) = \frac{\text{\# of Spam tweets that contain } \mathcal{T}}{\text{total \# of Spam tweets}}$$

$$P(\mathcal{T}|\mathcal{U}_{\mathcal{NS}}) = \frac{\text{\# of Non-Spam tweets that contain } \mathcal{T}}{\text{total \# of Non-Spam tweets}}$$

$$\gamma_{\mathcal{T}} = \left| \frac{P(\mathcal{T}|\mathcal{U}_{\mathcal{S}})}{P(\mathcal{T}|\mathcal{U}_{\mathcal{NS}})} \times \log_{10}\left[\frac{P(\mathcal{T}|\mathcal{U}_{\mathcal{S}})}{P(\mathcal{T}|\mathcal{U}_{\mathcal{NS}})}\right] \right|$$

We sort words in diminishing request dependent on their γT. We take the main 15 words from every one of the US and UNS utilizing above computation. We combine these words to get our list of top 30 words that will be used. The advantage of considering these words depending on their entropy score is that we had an option of decreasing the vulnerability in the forecast results as they have an alternate effect of recurrence including in spam and non-spam tweets. Subsequently we consider these 30 words assist us to classify the tweets precisely for every class.



In our approach, we used Random forest and Naïve Bayes algorithm. The datasets that are collected are taken into two steps for learning i.e., data pre-processing and processing. In data pre- processing we train the machine and also clean the machine. And once we train the machine using our algorithms, we process the data and acquire the objectives. pre-processing remembers the planning of data for expected organization for execution. The data ought to be spotless, no noise and consistency. The pre-processing of data is used for decreasing the measure of data into useful group of data group. To examine the nature of data, pre-processing step plays a prominent role. In this procedure # subtelites are analysed and only text comments are taken into count.

Naïve Bayes Classifier: This is one of the proved effective classifier and is used to classify a tweet as spam or non-spam. It uses probabilistic learning technique which depends on Bayesian hypothesis. This classifier is likewise utilized in notion investigation, spam identification and content information arrangement. Based on posterior probability the tweets are classified as spam or non-spam.

Random forests are a mixture of random tree predictors as all trees depend on the values of random vector sampled autonomously. The random forest algorithm works as described below:

A. Draw n tree bootstrap using original sample data.
B. Produce an unpruned classification for every bootstrap sample by modifying every node. Rather than selecting the most preferable split amongst the predictions, ideally sample m tries all the predictions and selects most effective split among them.
C. Anticipate new data by cumulating the predictions of n trees using the majority votes for classification.

IV. RESULT AND ANALYSIS

Result is the final information obtained by the actions or events which is a quantity or quality obtained by calculation. Performance analysis is the analysis of an operation which is a set of basic quantitative relation between performance quantities. Software testing is a concept or an activity to verify the results obtained with the results that are expected. Software testing also involves execution of a software system/component to evaluate the required interest. Software testing is also used to identify any kind of error, gaps or any missing statements or requirements corresponding to the actual requirements. This process can be done using automated tools or even manually. Software testing methods include Black Box testing and White Box testing.

In our system, we have done manual testing and also, stress testing to check the breakpoint of our network. The manual testing is done by manually using the pre-known outputs. The first testing was done at the first module i.e., Data Pre-processing which is used to make sure that the dataset doesn't contain any unknown r missing value. The CSV file is taken as input to our system and data cleaning is performed.

The second and third testing was done in the second module i.e., Feature Extraction in reduction the complexity of the dataset. The pre-processed CSV file is taken as input and the algorithms are applied successfully and separately to obtain the reduced feature dataset.

The formula used to obtain the accuracy is:

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ number\ of\ samples}$$

And, the output obtained is:

Out of 100 tests	True results	False results
Spam	89	11
Non-spam	93	7

After performing the Analysis on the results obtained, the approach we have obtained has 91.2% accuracy when compared to other approaches.

V. CONCLUSION

The trend for social media has grown intensely and twitter is one of the most used social media. By this we see spamming tweets every day. Hence, through our system, "spam detection on twitter" which is built using machine learning, we collected a huge number of public tweets and based on the text in the tweets' we extracted the words which gave us the highest information gain so as to classify the tweets. As the tweets' feature keeps changing frequently in the real world, we keep on updating the bag-of-words using the self-learning algorithm. hence, this is the most effective ways to detect the spam tweets.

REFERENCES

[1] S. Yardi, "detecting spam on twitter network", 2019
[2] G. Sthringhini, G.Vigna, "Detecting spammers on social media",ACM, 2019
[3] A.H Wang, "spam detection in twitter, security and cryptography", 2018
[4] M. Antonakais, W. feamster, "building system for spam detection", 2019
[5] B. Zhu, Y. Luo, "cost effective pam filtering", 2018

AUTHORS PROFILE

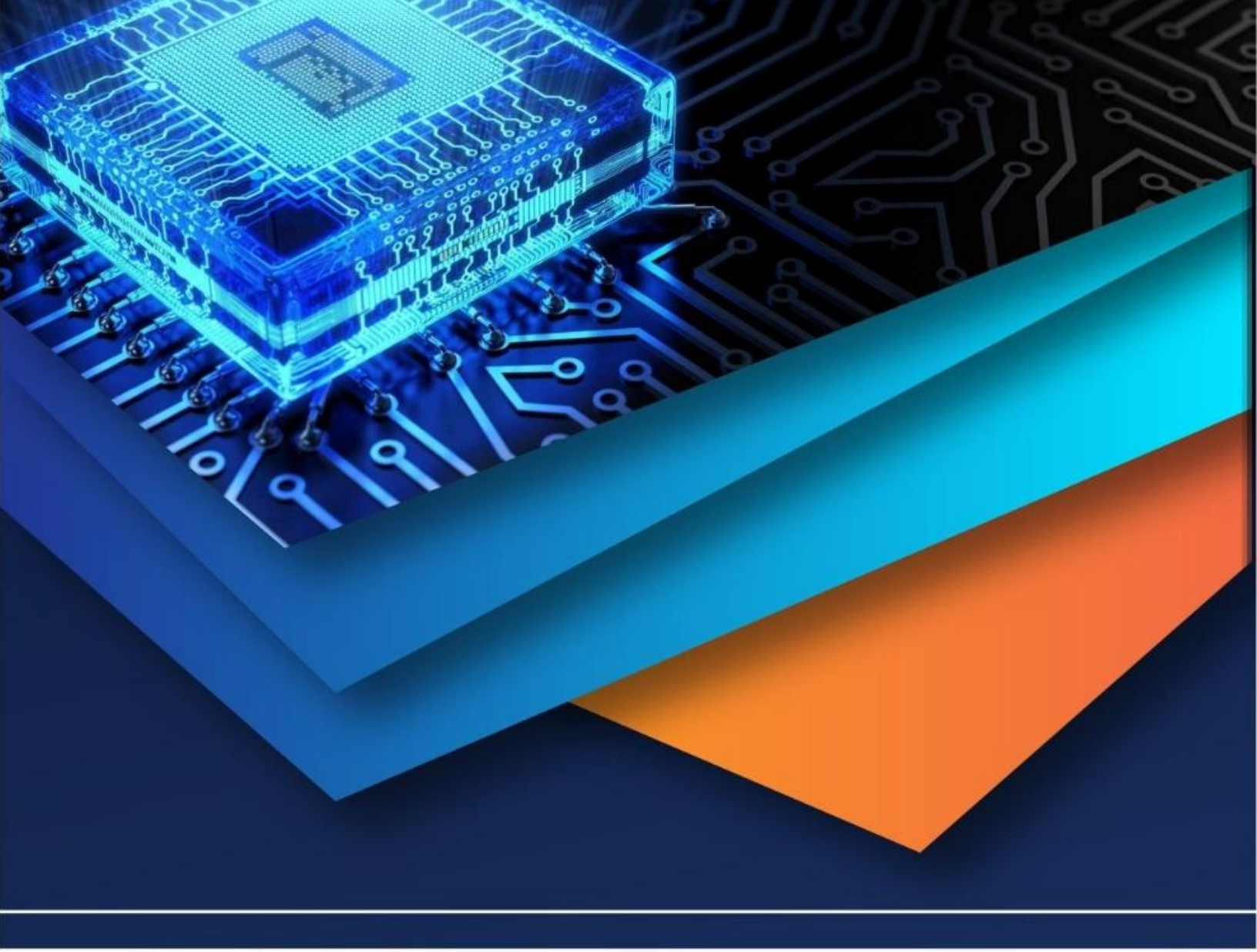
First Author Mr. Girish Kumar D,
Assistant professor, CSE dept,
BITM, Bellary

Second Author Pavani.K, B.E
in computer science, BITM,
Bellary

Third Author Neha.J, B.E in computer science, BITM, Bellary

Fourth Author Aishwarya K, B.E in computer science, BITM, Bellary.

Fifth Author Amulya T.G.M, B.E in computer science, BITM, Bellary.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY