Journal of Advances in Management, Engineering and Science (JAMES)
Volume 01, Issue 01, March 2023

SPAM DETECTION FRAMEWORK USING MACHINE LEARNING ALGORITHMS

Sai Chandhu Kanaparthi
Masters in computer science, university of Dayton, United States
Email: saichandukanaparthi@gmail.com

Abstract: As social media is a cheap and actively employed part of information sharing, its present use has generated unimaginable volumes of social data. Today, a significant portion of people base their decisions on the information available on social media like comments and suggestions about a subject or product. The ability to share knowledge with many users has swiftly caused social spammers to take advantage of the network of trust to spread spam messages and support personal forums, advertising, phishing, frauds, and other illegal activities. Although numerous tests have lately been carried out with the goal of identifying these spammers and spam material, to date the approaches are only just able to identify spam feedback, and none have been successful.

Keywords: *Social Media, Spam Detection, Machine Learning Algorithms, Feedback*

1. INTRODUCTION

With the advancement of technology and the digitization of everything, we make some of our judgments based on the content of information that we see available on the internet in order to make the sensible or ideal selection to maximize the benefits attainable when making a choice. We frequently read product reviews before making purchases, from electronics to food and even healthcare supplies, to determine which is the most dependable and trustworthy option [1]. This generally works out for the better, but occasionally a bogus review or spam message may deceive, take attention away from reliable products, expose recipients to danger, or even defraud naïve individuals. When a review is flagged as spam by platform users, designated staff members manually detect the spam. This is beneficial in that it prevents situations when the algorithm mistakenly interprets a user's genuine review in a different language as spam and immediately deletes it, but it also means that many more potentially spammy reviews could end up on the site unless they are reported [2]. Some spam reviews are written perfectly to sound like a legitimate review and are then copied and pasted over

the internet [3]. the amount of spam reviews that are deceptive or fraudulent can be significantly decreased by automating spam identification utilising a clear machine learning framework. Our system integrates NLP techniques with machine learning algorithms including Random forest, Bayes Network, Naive Bayes, K-nearest neighbour, and support vector machine to locate and eradicate spam as well as identify the spammer [4-6].

2. EXISTING SYSTEM

The current systems of spam detection are only depended on three methods: -

Linguistic Based Methods

Robots are taught some language in order to assist them comprehend linguistic constructions since unlike humans, machines cannot understand linguistic constructs and their interpretations. These methods are employed by search engines to identify the next word in an incomplete sentence. They are divided into two Bigrams and two Unigrams (Words one at a time) (Words two at a time). This strategy is not as effective and takes more time because each term must be recalled[7-8].

Behavior Based Methods

It is powered by metadata. Users that utilise this approach must develop a set of laws and be knowledgeable about them. Because spam's features evolve over time, rules must be revised to reflect these changes. Because of this, it is mostly user-dependent, and people still need to look at additional specifics[9].

Graph Based Methods

By using graph-based anomaly detection methods for graphical representation and combining many, disparate elements into a single graphical representation, this method finds anomalous patterns in the data that indicate spammer behaviors[10]. It is difficult to identify incorrect beliefs using this method since it is unreliable. Spam characteristics cannot be engineered, are not pre-installed, are not statistically dependant, depend mostly on the commercial allure of words, and are fully content-oriented. The considerable degradation of the system is caused by both of these elements [11-12].

3. PROPOSED SYSTEM

The approach suggested in this research uses NLP principles in conjunction with the supervised classification algorithm random forest to categorize and identify spam reviews among all other reviews on the TWITTER dataset. The algorithm uses four key components, including eight NLP concepts in figure 1.

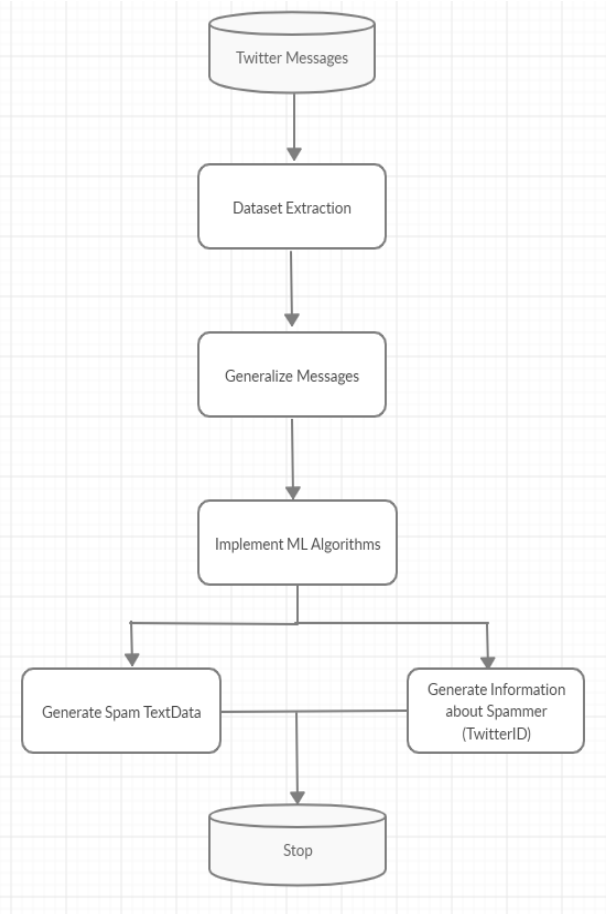


Figure 1 : Data flow diagram

I. MODULE DESCRIPTION

A. Dataset Extraction

The first step is to collect data from the dataset, which in our case is Twitter messages. After collecting the data, it is cleansed by removing extra spaces, duplicates, and other errors.

B. Collecting Metadata

The cleaned dataset is used to implement the RB features. First, the message's time frame is determined. After determining the time frame, it is compared to the threshold rating deviation to determine the spammer's diversity and variance. As a result, metadata about the spam

message and spammer is gathered.

C. Generalize Messages

Regardless of whether they are spam or not, all Twitter messages are collected and aggregated. Much time can be saved by generalizing the messages.

D. Implementing ML Algorithms

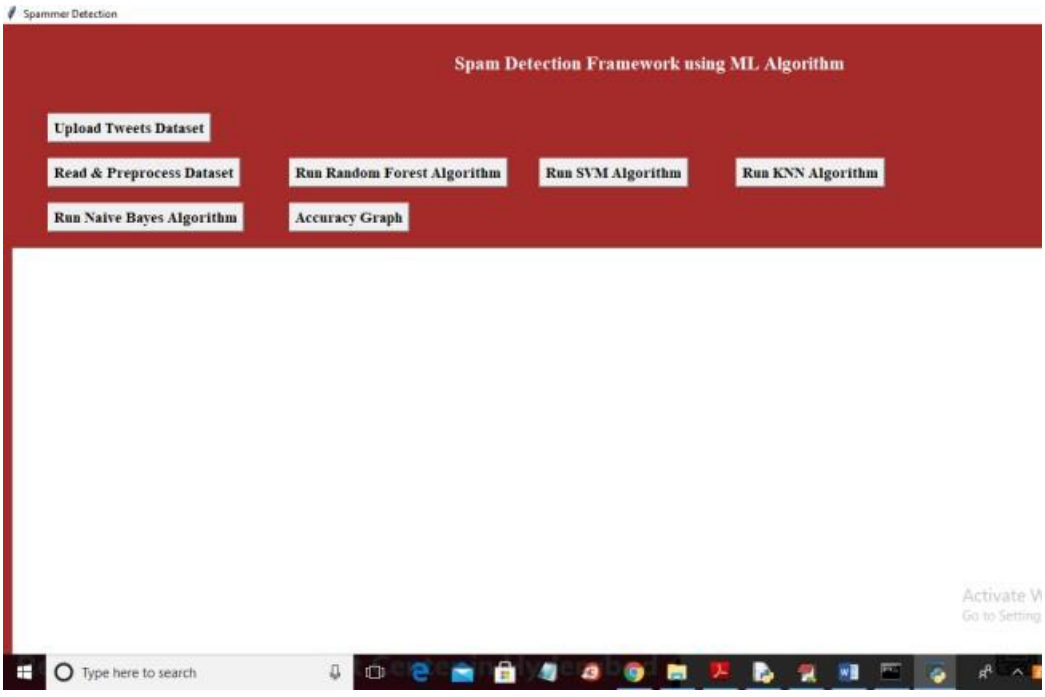
In this stage, the ML algorithms are implemented by categorizing the messages as spam or original content. Random forest, Bayes Network, Nave Bayes, K-nearest neighbour, and support vector machine are among the ML algorithms used.

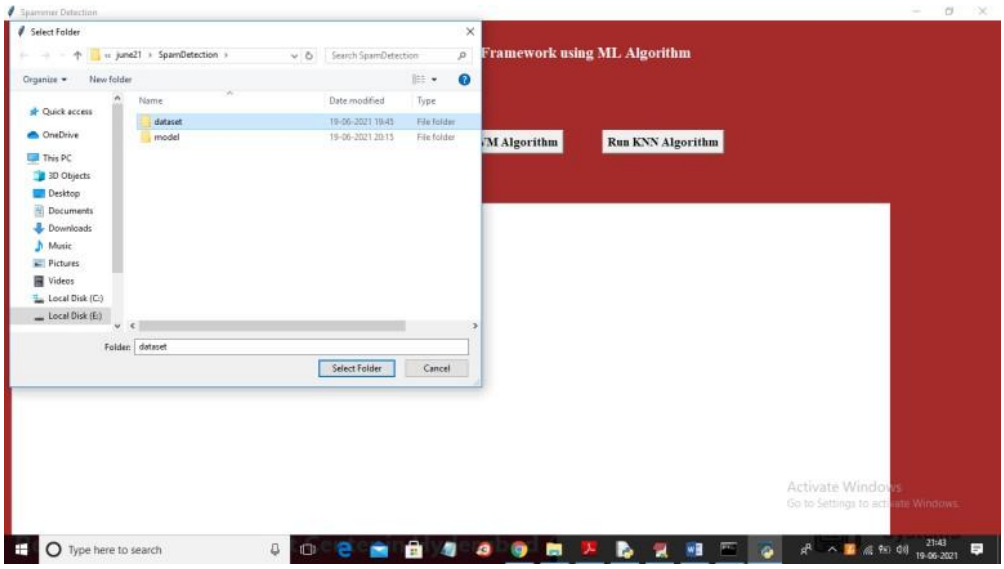
E. Generating Spam Text Data andinformation about the Spammer

Following the implementation of the ML algorithms, spam messages are identified and obtained, as well as information about the spammer who wrote the spam message. This information allows access to the spammer's entire history and analysis of all his messages.

II. RESULTS:

To run project double on „run.bat“ file to get below screen



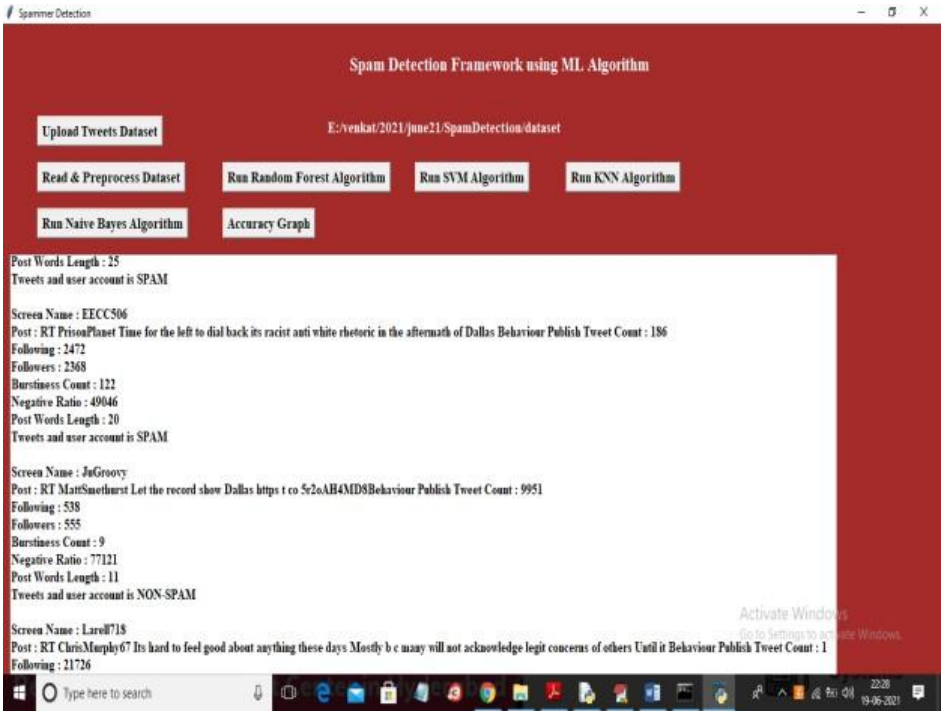


In above screen click on „Upload TweetsDataset" button to upload tweets

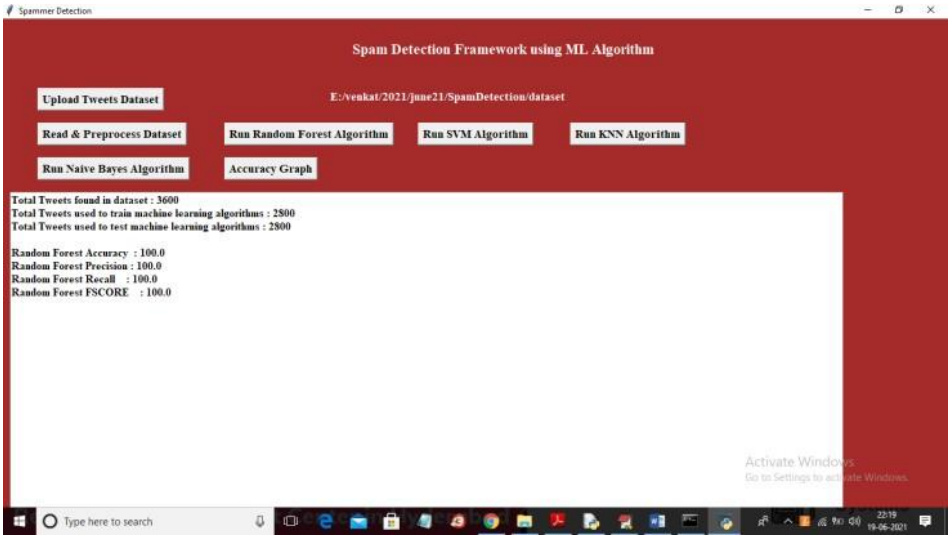
In above screen selecting and uploading entire „dataset" folder to upload tweets and to get below screen

In above screen dataset loaded and now click on „Read & Preprocess Dataset" button to read tweets and then extract 4 features from each tweet

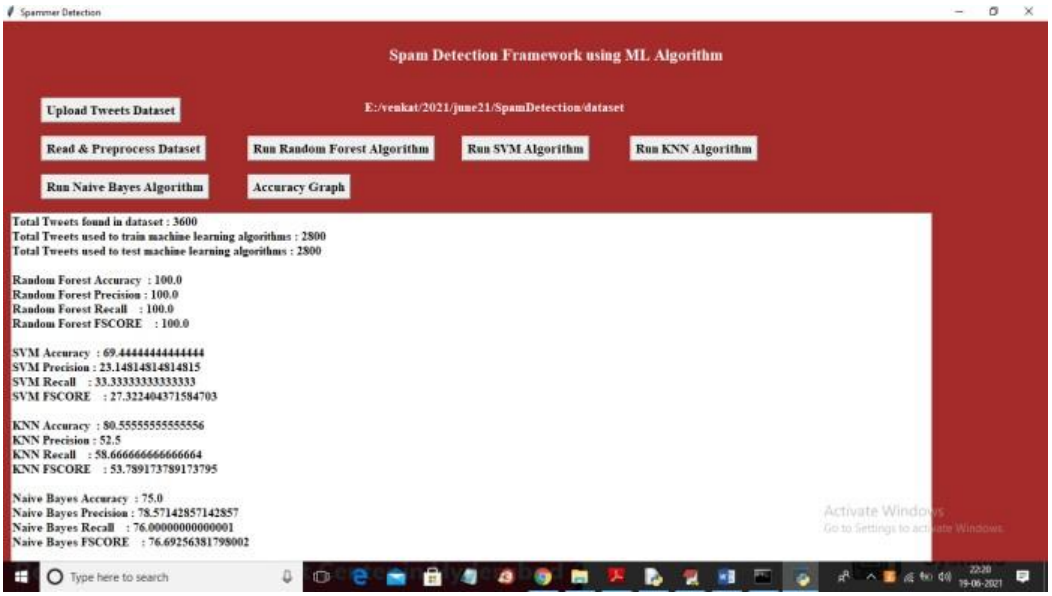
In above screen from each user we extract all 4 features and then based on features tweets and user account is consider as SPAM and NON- SPAM and you can scroll down above text area to view all tweets.



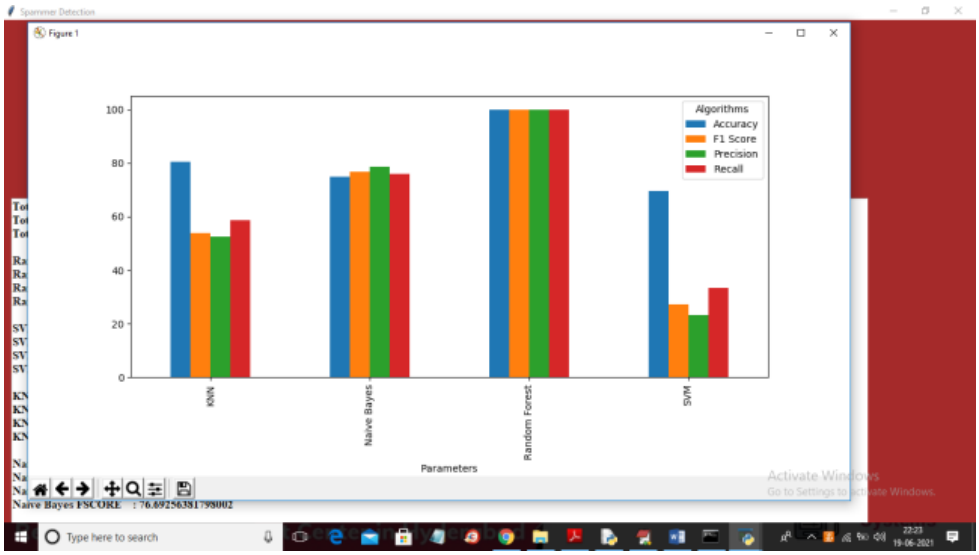
Now click on „Run Random Forest Algorithm“ button to train above dataset with Random Forest Algorithm



In above screen we trained above dataset with random forest and we got accuracy as 100% and similarly click on all buttons to train other 3machine learning algorithms



In above screen all 4 algorithms are trained with tweeter dataset and in all algorithms random forest has given better accuracy and prediction result. Now click on „Accuracy Graph‟ button toget below comparison graph



In above graph x-axis represents algorithms names and y-axis represents accuracy, precision,recall and FSCORE. In all algorithms Random forest has given better performance

4. CONCLUSION:

In this study, we identified spams and spammers in a Twitter dataset using machine learning techniques and NLP principles. In order to identify additional spams, spammers, and their message writing style, it is helpful to study the spam in order to obtain and show all of the spammer's data. We considered two kinds of attributes: content and user behavior. Exclamation sentence ratio, first personal pronoun ratio, maximum and average content similitude, and exclamation sentence ratio are used to analyze the content. Properties like written reviews and an average of unfavorable ratios are used to predict user behavior. It is thus a very reliable and accurate spam detection system.

REFERENCES

1. Nurul Fitriah Rusland, Norfaradilla Wahid, Shahreen Kasim, Hanayanti Hafit, “Analysis of Naive Bayes Algorithm for Email Spam Filtering across Multiple Datasets”.
2. J. Rout, S. Singh, S. Jena, and S. Bakshi, “Deceptive Review Detection Using Labeled and Unlabeled Data”.

3. Feng Qian, Abhinav Pathak, Y. Charlie Hu, Z. Morley Mao, and Yinglian Xie, “A Case for Unsupervised-Learning-based Spam Filtering”.
4. Shrawan Kumar Trivedi, “A Study of Machine Learning Classifiers for Spam Detection”. W.A. Awad, S.M. ELseuofi, “Machine Learning Methods for Spam E-mail Classification”
5. S. Gharge, and M. Chavan``An integrated approach for malicious tweets detection using NLP," in Proc. Int. Conf. Inventive Commun. Comput.Technol. (ICICCT), Mar. 2017, pp. 435_438.
6. 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.
7. M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, ``A hybrid approach for spam detection for Twitter," in Proc. 14th Int. Bhurban Conf. Appl. Sci.Technol. (IBCAST), Jan. 2017, pp. 466_471.
8. F. Fathaliani and M. Bouguessa, ``A model- based approach for identifying spammers in social networks," in Proc. IEEE Int. Conf. Data Sci. Adv.Anal. (DSAA), Oct. 2015, pp. 1_9.
9. Saeedreza Shehnepoor, Mostafa Salehi*, Reza Farahbakhsh, Noel Crespi, “NetSpam: a Network-based Spam Detection Framework for Reviews in Online Social Media “
10. G. Jain, M. Sharma, and B. Agarwal, ``Spam detection in social media using convolutional and long short term memory neural network," Ann. Math. Artif. Intell., vol. 85, no. 1, pp. 21_44, Jan. 2019.
11. C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, ``A machine learning approach for Twitter spammers detection," in Proc. Int. Carnahan Conf.Secur. Technol. (ICCST), Oct. 2014, pp. 1_6.