IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

TWITTER SPAM DETECTION USING MACHINE LEARNING MECHANISM

Aswathi k[1], Dr. D Sathya Srinivas[2]

[1]M.sc, Department of Computer science and Engineering, Dr. MGR Educational and Research Institute, Chennai, India

[2]Faculty of Center of Excellence in Digital Forensics, Dr. MGR Educational and Research Institute, Chennai, India

ABSTRACT:

Social networking sites has become very popular in nowadays. Users use them to find new friends and updates their day to day basis activities and latest thoughts. Social medias such as Facebook, Instagram, Whatsapp, Twitter became most popular among everyone. Among these Twitter has become the most popular one. We suggest, by replacing the former methods of Euclidean distance by set of classifiers in order to assign the incoming samples to most relative microcluster with arbitrary distribution. Here, a set of Incremental Naive Bayes (Classifier used) is trained for microclusters whose outcome become greater than the threshold value. INB can give mean and boundary of the microclusters , While the Euclidean distance mainly focusses on the mean value and acts inappropriate for the big assymetric microclusters. In this paper, DenStream was improved by proposed sytem, here as INB Denstream. Performance was calculated in terms of purity , general precision , general recall, F1 measure, parameter sensitivity, and computational complexity by the application of methods such as DenStream, CluStream, StreamKM++ to Twitter datasets to show the effectiveness of INB DenStream. Then, our results shows that among the different classifiers used, multinomial NB produces the best result and can achieve 98.8 % accuracy.

Key words: Spam, Detection, Ham, INB, Accuracy, ML.

INTRODUCTION:

Social media platforms such as Twitter, Facebook, Instagram and also WhatsApp, have become the preferred online platform for interaction and communication. Especially social medias namely Twitter which is the most common platform. When focused on the latest statistics report, Twitter has approximately 200 million accounts. Moreover 400 millions tweets per day. Most of the tweets may be of fraudulent links or messages, fishing attack, advertising messages, malware links etc, Generally spam tweets have common indications such as shortened URLs, “Hashtags” “Mentions”. But the every tweets with those kind of common inidcators may not be always spam tweets. For instances , number of tweets received by an twitter account holder is greater than the connected friend in twitter, then there is more probability of all messages received are of spams. When considering the existence of spam indicators in most of the spam tweets received , it became very easy to detect spam or ham tweets using the machine learning algorithms. So far both supervised and unsupervised methods have been deployed, but the supervised ones provides the better outcome. In this paper , we have developed a framework where the former methods of Euclidean distance had been replaced by incremental NB classifiers in order to enhance the accuracy of spam detection methods. Here the performance was determined on the basis of F1 measure , general recall, accuracy , precision etc.

REVIEW OF LITERATURE:

Anisha Rodrigues, Roshan fernandes and et al., [6] had proposed the system for real time twitter spam detection and sentimental analysis using ML and deep learning techniques. In that proposed model they had utilized the ML learning algorithms namely LSTM, naïve bayes , and support vector machine classifier. This paper aims to extract information from the tweets received, and by analysing the tweet obtained sentimentally and detecting whether is spam or ham. Features in need are collected through vectorizers such as TF -IDF and Bag of Words model and LSTM achieved the highest validation accuracy of 98.74%.

Saleh beyt sheikh ahmad, Mahnaz rafie and et al.,[1] had proposed the system for Spam detection on twitter using support vector machine and users features by identifying their interactions. This method includes two important steps mainly preprocessing and feature extraction. They had proposed the enhanced method by using the five classes of features for eg: account information features, user profile features, Again for learning purposes subsets of these features were used. They had utilized MLP (Multi layer perceptron) , NB (Naive bayes), RF (Random forest), KNN (K nearest neighbours). They had achieved better performance overall with highest accuracy.

Tingmin wu, Shigang liu and et al.,[2] had proposed the Twitter spam detection based on deep learning. In that they had proposed the deep learning models for the detection. They had also conduct the performance evaluation and comparison for various deep learning algorithms namely (LSTM, GRU, LSTM-ATT, CNN, BLSTM, CNN-LSTM, LSTM-TCN) on different dataset. According to their research, existing system have achieved the accuracy of 80%. Moreover machine learning cant effectively detect spam or nonspam when large dataset was given and accuracy rate was also too less. They have experimented using different deep learning classifiers and further compare it with existing system , finally they came to conclusion that proposed system was more accurate than the existing one.

M. McCord, M. Chuah and et al., had proposed the spam detection on twitter using traditional classifiers. In that proposed model they had used four different types of algorithms. Among the most common social media platforms such as twitter has become more popular , its popularity attracts many illegal activities, by using the content based features and user based features. Then , they have used these features to facilitate spam detection . Also they had evaluated the features with algorithms , and finally they have came to conclusion that Random forest classifier produces the best outcomes and they can achieve 95.7% of precision and 95.7 % F1 measure using RF. And evaluated the existing model with proposed system ,they could see proposed system was more accurate.

Malik Mateen, Muhammad aleem and et al., [2] had proposed the hybrid approach for spam detection for twitter. Nowadays online social media platforms such as twitter has becom e more popular , and users use them for finding new friends and also for sharing the activities and latest thoughts in twitter, so that it can easily attracts many of illegitimate users. Illegitimate activities include the malicious links or advertising attachments and all , also some users may have receive more tweets than the number of friends they have connected to. Then , we can say it is spam or ham .when compared to the existing methods , here they have used hybrid techniques for spam detection and arrived at a conclusion that proposed system have more accuracy and detection rate than the existing system.

Sreekanth madisetty, Maunendra sankar desarkar had proposed the neural network based ensemble approach for spam detection in twitter. In that proposed model they had used the deep learning and traditional feature based model using multi layer neural network which acts as a meta classifier. In twitter platform spam messages are common , most current techniques for spam filtering focuses on the spam detection methods and how to block the further spam. Here, they have developed various deep learning models based on convolutional neural networks. With the use of content based and user based features , finally they have came to result as proposed method accuracy is greater than the existing one.

Yi xie, Chao chen and et al., [2] had proposed the performance evaluation of machine learning based streaming spam tweets detection. In this proposed model they had evaluated the impact of different factors on spam detection performance, which included spam to non spam ratio, feature discretization, training data size and data sampling. As per their research on existing methods , they have mostly applied machine learning or deep learning algorithms to detect spam. Finally came to conclusion that stream spam detection methods is still a big challenge and robust detection technique should taken into account and evaluated their sampling with robust techniques and found that proposed system is more accurate.

RESEARCH METHODOLOGY:

To figure out or detect the spam detection using the machine learning algorithm namely as NB navie bayes. In this model detecting the spam messages which maybe in different forms such as malicious links, advertising messages or tweets for financial fraud etc. This model had mainly focused on the Twitter social media platform because it is the most popular one. First in this model we have extracted the dataset, and use the extracted data sample for the detection purposes. Before the data extraction the data pre-processing had done. The data pre-processing steps are one of the most inevitable stage and it should be done before data collection. The data pre-processing is simply the cleaning of data with the use of Natural language processing, by this method data has been cleaned like removal of stop words and vectorization etc. From that pre-processed dataset we had took the particular word for our deployment model And the next level is the training the dataset for the detection of spam tweets or messages. In this training step the ML techniques had used. To train the dataset the different types of ML algorithms used. We have experimented with different algorithms and obtained best accuracy for Multinomial NB which is simple and efficient. After training the model , evaluation process had done, and proposed model performance checked through certain parameters.

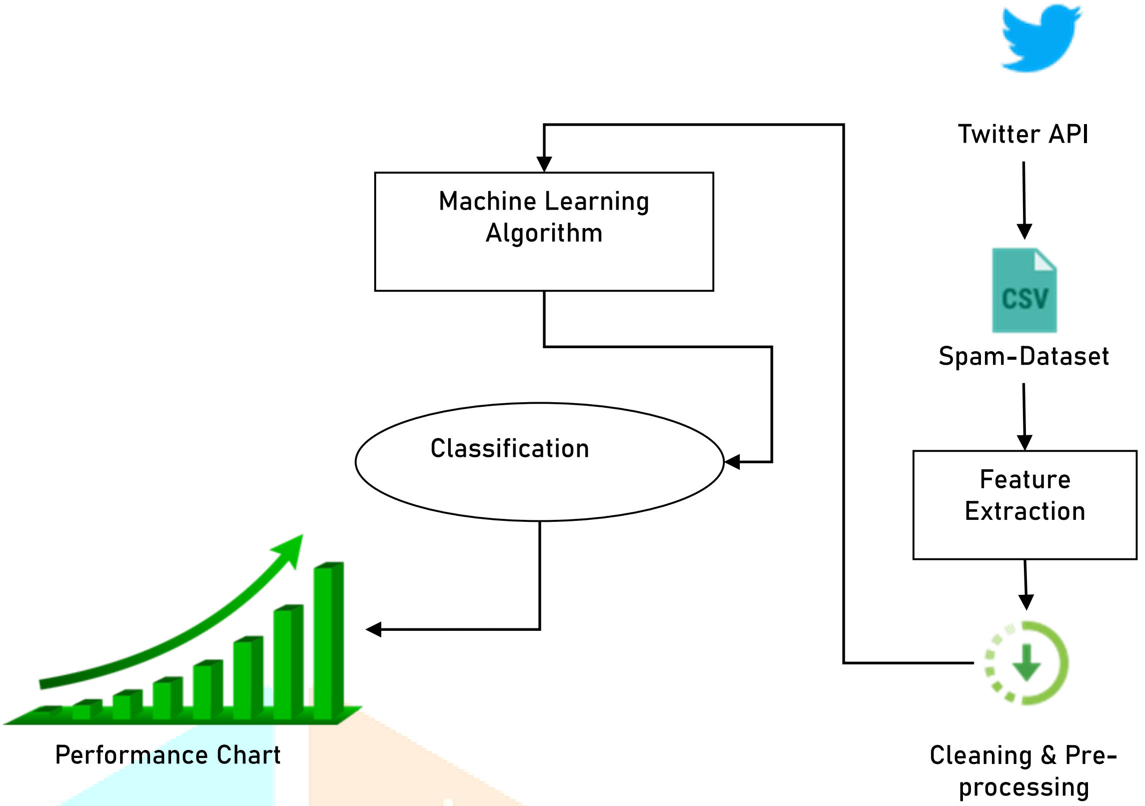


Figure .1: Visualization depicting the interconnected components and interactions within the proposed system architecture

The figure 1 shows the complete step by step process about the proposed model. In this model if the illegitimate user has sent any spam tweets or messages into legitimate account by tricking the people as look like normal tweet. Here for this model we have extracted a large dataset of both spam and ham. And the step by step process starts from the cleaning and preprocessing stage where after the extraction the features are maintained properly. For this model we had used the jupyter notebook. Jupyter note book is the web based environment that allows users for live coding. In that jupyter notebook we had run our ML code. Once the code is running the jupyter notebook redirect to the tweets under control, and the word cloud in every spam or ham tweets are entered into the trained anaconda prompt website, and it was detected by the NB (Naive Bayes) and its predicted that as spam or ham. And after the classification we are able to create a performance chart of accuracy, precision and F1 measure. Accuracy rate was better for NB when compare to the other algorithms with which we have experimented, also when compared to existing one performance chart shows that this proposed model outperforms the existing others.

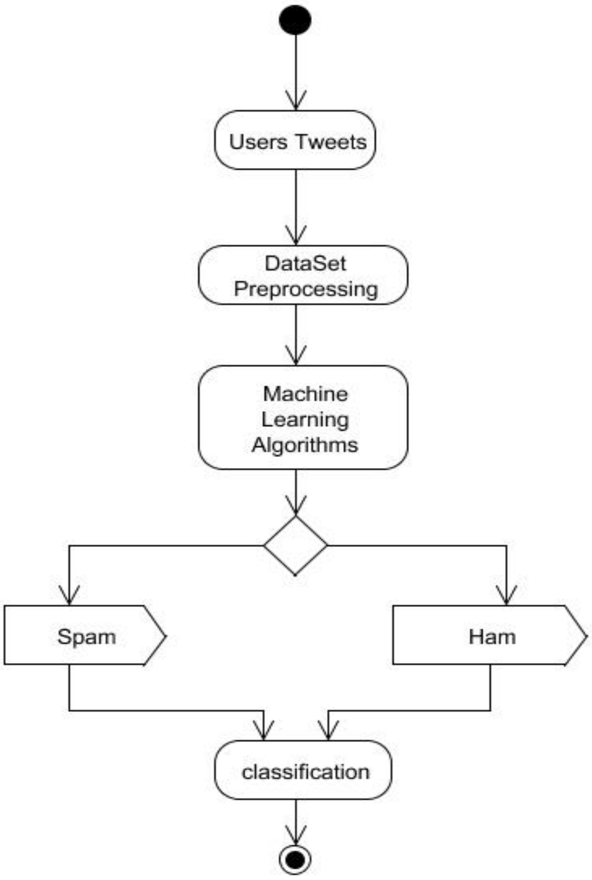


Figure. 2: Spam or ham detection step

The figure 2 shows that preprocessed dataset when we have trained with ML algorthim naive bayes , it really helps us to find out whether the tweets or messages that we have received is spam or ham with better performance level.

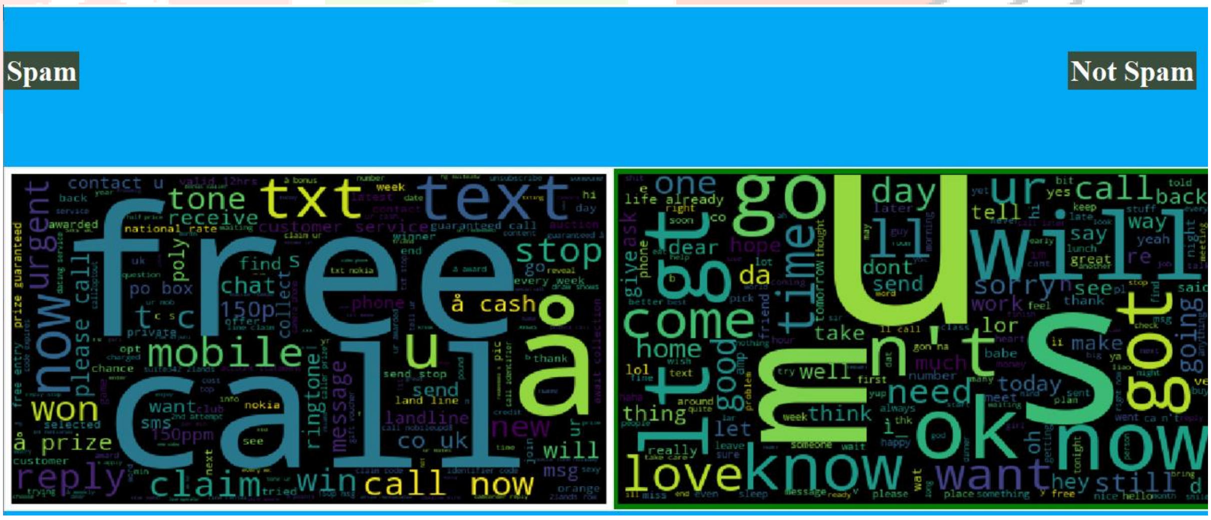
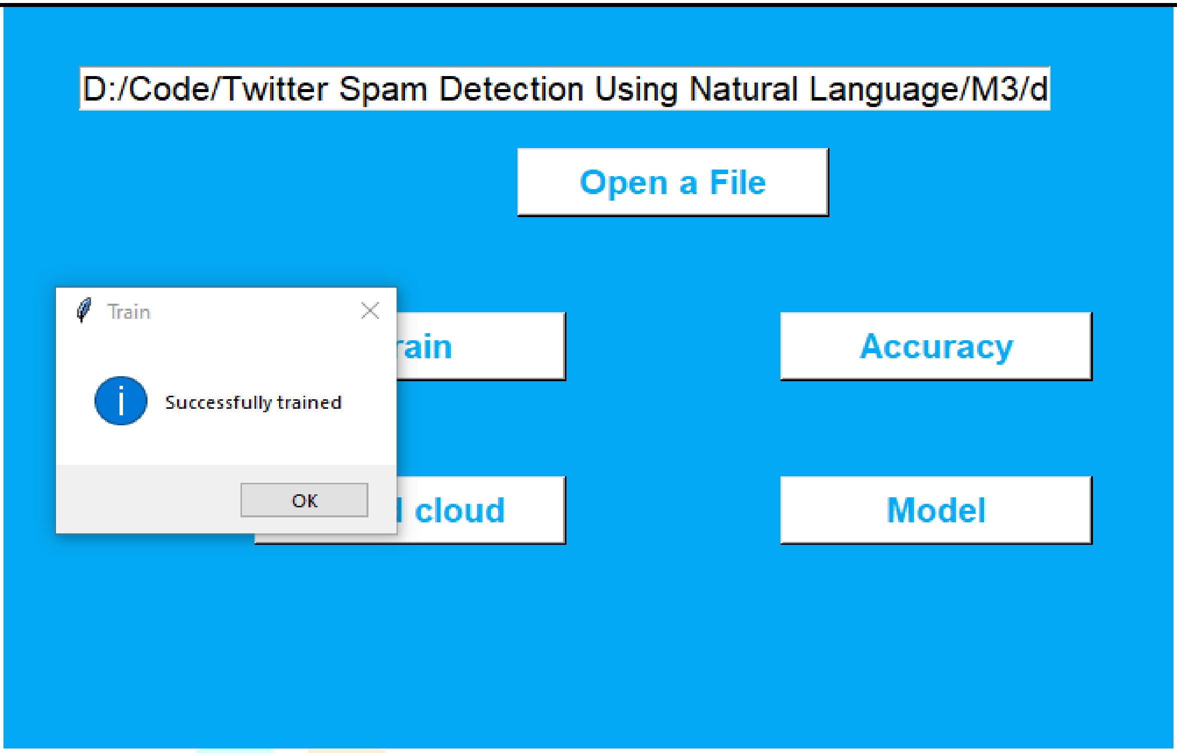


Figure. 3: Word cloud for spam and ham

The figure 3 shows that the word cloud format which specifies that spam messages or tweets are mostly generated of certain types of words, as well as of nonspam mesaages. Here we can easily predict based on the word set whether it is spam or ham, because spam usually contains certain words such as Call, won, Free etc, and ham contains Good, Know etc.



D:/Code/Twitter Spam Detection Using Natural Language/M3/d

Successfully trained

Figure. 4: Displaying of proposed model after data has trained with NB

The figure 4 shows that proposed model where the preprocessed dataset had trained with NB algorthms and in anaconda prompt created model for detecting spam or ham from the collected dataset.

CONCLUSION

In this proposed twitter spam detection model using NB naive bayes in machine learning ensures that the spam messages or tweets that had clearly detected and predicted. And Multinomial NB has performed well when compared to already existing model with various other algorithms. NB was simple and most efficient when it was used along with Anaconda prompt. The combined effect of both these algorithms and platform helped to reach out to a better outcome. In the future this proposed model have been added to the other social media platforms. For example we can use this proposed model with the use of API keys. In the future it will helpful to detect whether any fraudulent activities may hide in tweets or messages that we have received through social medias especially twitter and it would be much helpful to the investigators, private detectives and even normal peoples to understand whether the received tweets are spam or ham and to solve the twitter spam related cases. And also in the future this proposed model will help to reduce the ratio of the people who are being affected by the spam.

REFERENCES

[1] W. W. Chi, T. Y. Tang, N. M. Salleh, M. Mukred, H. AlSalman and M. Zohaib, "Data Augmentation With Semantic Enrichment for Deep Learning Invoice Text Classification," in IEEE Access, vol. 12, pp. 57326-57344, 2024

[2] A. Qazi, N. Hasan, R. Mao, M. E. M. Abo, S. K. Dey and G. Hardaker, "Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review," in IEEE Access

[3] R. Agarwal et al., "A novel approach for spam detection using natural language processing with AMALS models," in IEEE Access

[4] P. Jain, S. Singh and C. K. Saxena, "Detecting Email Spam with NLP: A Machine Learning Approach," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 393-398

[5] M. Salman, M. Ikram and M. A. Kaafar, "Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models," in IEEE Access, vol. 12, pp. 24306-24324, 2024

[6] M. Khalid et al., "Novel Sentiment Majority Voting Classifier and Transfer Learning-Based Feature Engineering for Sentiment Analysis of Deepfake Tweets," in IEEE Access, vol. 12, pp. 67117-67129, 2024

[7] R. Rani, K. K. Yogi and S. P. Yadav, "Tech Innovations & Dataset Analysis to Combat Fake Accounts in Digital Communities," 2024 2nd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2024, pp. 1679-1684

[8] A. Seetha, S. S. Chouhan, E. S. Pilli, V. Raychoudhury and S. Saha, "DiEvD-SF: Disruptive Event Detection Using Continual Machine Learning With Selective Forgetting," in IEEE Transactions on Computational Social System

[9] Babu, R., Kannappan, J., Krishna, B.V. *et al.* An efficient spam detector model for accurate categorization of spam tweets using quantum chaotic optimization-based stacked recurrent network. *Nonlinear Dyn* **111**, 18523–18540 (2023)

[10] A. M. Al-Zoubi, A. M. Mora and H. Faris, "A Multilingual Spam Reviews Detection Based on Pre-Trained Word Embedding and Weighted Swarm Support Vector Machines," in IEEE Access, vol. 11, pp. 72250-72271, 2023

[11] P. Jain, S. Singh and C. K. Saxena, "Detecting Email Spam with NLP: A Machine Learning Approach," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 393-398

[12] M. Salman, M. Ikram and M. A. Kaafar, "Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models," in IEEE Access, vol. 12, pp. 24306-24324, 2024,

[13] Z. Zhang, R. Hou and J. Yang, "Detection of Social Network Spam Based on Improved Extreme Learning Machine," in IEEE Access, vol. 8, pp. 112003-112014, 2020

[14] G. Lingam, R. R. Rout, D. V. L. N. Somayajulu and S. K. Ghosh, "Particle Swarm Optimization on Deep Reinforcement Learning for Detecting Social Spam Bots and Spam-Influential Users in Twitter Network," in IEEE Systems Journal, vol. 15, no. 2, pp. 2281-2292, June 2021

[15] X. Liu, H. Lu and A. Nayak, "A Spam Transformer Model for SMS Spam Detection," in IEEE Access, vol. 9, pp. 80253-80263, 2021