TEST Engineering & Management

July-August 2020
ISSN: 0193-4120 Page No. 4417 - 4421

SPAMMER DETECTION USING MACHINE LEARNING ALGORITHM

P.Hareesh, S.Sridhar
UG Scholar, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
Assistant Professor, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India.
pendelahareesh01121999@gmail.com, 007sridol@gmail.com

Article Info
Volume 83
Page Number: 574 - 585
Publication Issue:
July-August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *10 August 2020*

Abstract

Due to the growth of communication technology common people also uses internet. E-mail is one of the important applications of internet. E-mail is used send and receives any type of information within a short period by using their Mail-Id. This is the easiest way to create a communication to the people in all over the world. Sometimes the unwanted mails are available in our inbox. These unwanted E-mails are called spam. The main purposes of spam are send advertisements regarding their services and products to the multiple persons at the same time. But rarely the unwanted email contains malwares. The malwares are spoiling the user's stored data on the local system. Now the world is moving from traditional technologies to the current communication technology such as IoT (Internet of Things). The applications of IoT are very large. It is used in various domains. Most of the industries are also move from their existing system to IoT concept. The main aim of the IoT is connecting various devices and objects with the help of internet. In production industries also used the concept IoT to connect various devices in an organization and mobile cloud computing concept. It provides the actual task of machineries and optimizes their production. But sometimes this option can be used by spammers to spoil the power of manufacturing industries. Due to these spammers the production of the company will be spoiled. To avoid this problem various type of machine learning algorithms are used. Finally the various machine learning algorithms are compared with their accuracy, true positive rate and false positive rate.

Keywords: *Industrial mobile network, Internet of Things, spammers, intelligent identification, machine learning;*

I. INTRODUCTION

The technology has been developed day by day. Common peoples are also used internet in their daily activities. E-mail is one of the important features of internet. Using this E-mail feature everyone create the communication between others quickly. Most of the industries are transferred from their existing process to the current technology such as IoT. This technology is very useful in current production organizations. It is one of major module of current communication technology. It is used create a communication between various devices

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

July-August 2020
ISSN: 0193-4120 Page No. 4417 - 4421

and objects with the help of internet. This IoT technology creates a connection between human beings and objects, entity to entity or machine to machine. It is extensively by various domains like manufacturing companies and military enquiry. Here smart perception, detection technique and persistent techniques are used.

Sometimes the spammers spread the malware to the various group of peoples at the time. The common people and company information's are stolen by using malware. This malwares are spoiling the entire data of the particular organization. The following diagram shows the spammers in mobile cloud computing technique. The neuron like diagram with images in the left portion shows that mobile cloud computing concept. The right portion of the picture shows the data performance of every user.

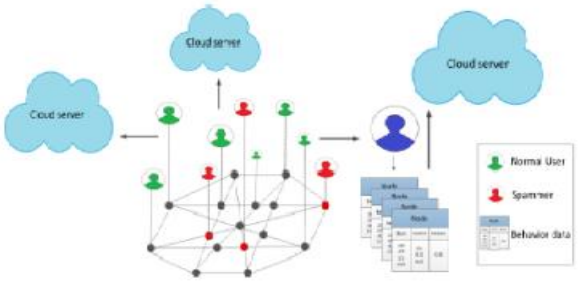

Figure 1 Spammers in mobile cloud computing technique

Due to this manufacturing industries are facing heavy financial loss. Here spam e-mail identification is very difficult. It is looks like a normal e-mail. To avoid such kind of problem machine learning algorithms are used. These algorithms are used to easily identify the normal mail and the spam mail. The final stage of this work various machine learning algorithms characteristics are compared.

This research paper is divided into V parts. Section II describes about various machine learning algorithms are used to detect the spam message. Section III describes with proposed frame work used to detect spam message. Section IV shows the result and discussion part. Section V discuss about the conclusion part.

II LITERATURE SURVEY

Tie Qiu et al., proposed a new system to solve spam problem in industries mobile communication and decrease computing difficulty by using huge cloud data. In this paper proposed a new model using Gaussian model. In this model take out the characters connected to labels from the data set. Based upon the data the user comes under to any one of the distribution. The characteristics are divided into three parts, and divide the data into two divisions. At last the simulation model was executed and assesses the value of the proposed model. This proposed model was executed based binary classification concept [1].

Prof.Satish Manje, et al., says that IoT is the important component of the current world. It provided the assurance of machine capacity in manufacturing industries. Spammer's means the unwanted mails are spoil the entire content of the local system. To avoid spammers in our system the authors proposed a new system using machine learning algorithms [2].

Zhang Bin, et al., said that in current scenario spammer identification is the difficult task in SMS. The major issue in the spam detection is providing user security. Using existing system detect the spam messages by using reserved words and issuing occurrence. In this system the user behaviors are used to detect the spam message. Here to explore the user's behaviors by using machine learning algorithms such as decision tree concept, random forest technique and SVM. The output of the above mentioned algorithms was compared [3].

Published by: The Mattingley Publishing Co., Inc.

July-August 2020
ISSN: 0193-4120 Page No. 4417 - 4421

Shivangi Gheewala et al., described that social network media was used to create a communication between peoples and maintain social dealings. Millions of peoples are participated in public connectivity. The main aim of the social media was create social communication, commercial activities, and amusement activities and transfer the data. At the same time many people's with naughty activities are also available on the social media. These activities lead to the huge economic loss. To avoid these spam activities the authors introduced a new system to detect malwares. Spam data is the important part used in this research work [4].

Harjot Kaur et al., expressed as E-mail has take a important place in common people life. E-mails are the major source of data transferring. It will send and receive the data within the short period of time. The spam mails are sometimes threaten the account holders. To avoid this situation various spam filters are used. The spam filers categorize the email based on content or header. In this proposed work filter are used to identify the malwares [5].

K Subba Reddy et al., described that peoples communications are developed by using social media. The users are divided into two types. They are — Now a day's human relations are maintained by social media networks. They are genuine users and spammers. The genuine users are misguided by the spam users. The genuine users are getting more unwanted mails in their inbox. To overcome this situation the authors proposed a new methodology by using various machine learning concepts [6].

Hassan Najadat1 et al., says that from past few years the usage of internet becomes very high. Peoples gave more importance of internet usage in their daily routine life. Spam message gave threat to various types of clients. Bayesian Filters are applied to detect the spam messages. In this proposed system the authors were improved the existing Naïve Bayes Classifier concept. The result part shows that the improved Naïve Bayes classifier performance is good. It was compared with traditional methods [7].

M. Nivaashin et al., says that in current scenario the usage of digital devices was increased. The business peoples are sending a SMS service to number of peoples for advertise their product or service. Most of the peoples are not like this type of spam message. Here the authors proposed a new system to filter the spam SMS messages by using Restricted Boltzmann Machine (RBM) with deep learning concept[8].

Alexy Bhowmick et al., was presented a complete evaluation of the most successful and efficient e-mail spam filtering techniques using content based concept. The authors concluded in their survey machine learning concepts are used to fight against the spam messages [9].

W.A. Awad et al., says that the volume of spam mails are increased day by day. To avoid the spam mails anti-spam filters are used. The new computing concept machine learning techniques are used to easily filter the spam mails successfully. In this paper the authors reviewed the famous machine learning algorithms are used to filter spam e-mails and classify it. [10].

III PROPOSED METHOD OF AIR POLLUTION MONITORING

The following figure 2 is used to present the proposed architecture for Spammer message detection in Industrial Mobile Cloud Computing concept.

Published by: The Mattingley Publishing Co., Inc.

July-August 2020
ISSN: 0193-4120 Page No. 4417 - 4421

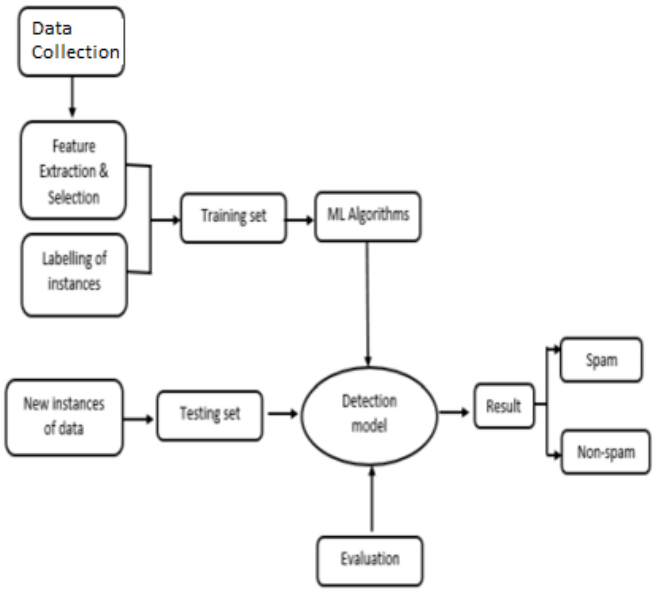

Figure 2: Architecture Diagram

According this proposed architecture the data are collected from various sources. Then the requirement features are selected and put the label of the instances from the dataset. All the features are not used for trained the data. Features are used to display more efficiency for providing proper result. It is called as a training data set. In this training dataset the various machine learning algorithms are applied. From that algorithm the detection model is developed. At last detection models are evaluate using their parameters accuracy rate identification rate etc.

IV RESULTS AND DISCUSSIONS

Most of the industries transfer their existing task to computerized task. All the manufacturing process will be executed automatically by using the current communication topic IoT. It means every objects and machines are communicated with each other. The spammers are send the unwanted mails to the industries mail purposefully. The malwares are collapse the entire system data. It leads to heavy data and finance loss. To avoid this kind of situation machine learning algorithms are used to filter the spam mails. In this proposed system DNN, RF, NB,SVM and KNN algorithms are used to filter the spam mail. The following table 1 shows the accuracy, true positive rate and false positive rate of the machine learning algorithms.

Models	Accuracy	TP Rate	FP Rate
DNN	98.18	98.73	2.27
RF	97.7	97.4	3.3
NB	76.26	76.3	4.5
SVM	76.26	76.3	7.5
KNN	77.27	76.6	7.3

Table 1 PERFORMANCE COMPARISON

The following figure 3 shows the graphical representation of different machine learning algorithms are used to filter spam mails in industries.

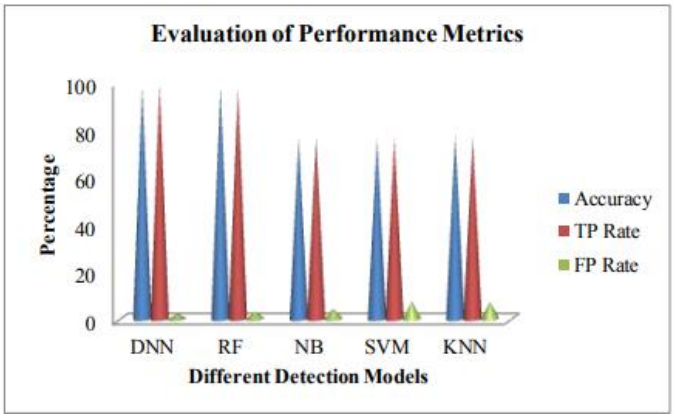

Figure 3: Comparison of various Spam detection models

V CONCLUSION

In this proposed research work various machine learning algorithms are used to filter spam mails. For this task the data is collected from various sources and extract their valid features. From that extracted features the detection model is developed. After that the proposed model is compared with the new data. Finally the performance of the various

Published by: The Mattingley Publishing Co., Inc.

July-August 2020
ISSN: 0193-4120 Page No. 4417 - 4421

machine learning algorithms are analyzed. From that analysis the DNN approach only produce more accurate result compared with other machine learning algorithms.

REFERENCES

1. Qiu, T., Wang, H., Li, K., Ning, H., Sangaiah, A. K., & Chen, B. (2018), SIGMM: A Novel Machine Learning Algorithm for Spammer Identification in Industrial Mobile Cloud Computing", IEEE Transactions on Industrial Informatics, pp 1-10.
2. Satish Manje, Nikita Palav, Sarika Aher & Payal Jage(2019), "Machine Learning Algorithm For Spammer Identification In Industrial Mobile Cloud Computing", International Journal for Research in Engineering Application & Management (IJREAM), ISSN : 2454-9150, pp.82-86
3. Bin, Z., Gang, Z., Yunbo, F., Xiaolu, Z., Weiqiang, J., Jing, D., & Jiafeng, G. (2016), " Behavior Analysis Based SMS Spammer Detection in Mobile Communication Networks", IEEE First International Conference on Data Science in Cyberspace (DSC), pp.538-543.
4. Gheewala, S., & Patel, R. (2018), " Machine Learning Based Twitter Spam Account Detection: A Review", 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), pp 79-84.
5. Harjot Kaur & Prince Verma(2017), "Survey On E-Mail Spam Detection Using Supervised Approach With Feature Selection ",International Journal of Engineering Sciences & Research Technology, Vol. 6, No. 4, ISSN: 2277-9655, pp. 120-128.
6. K Subba Reddy & E. Srinivasa Reddy(2019), "Detecting Spam Messages in Twitter Data by Machine learning Algorithms using Cross Validation", International Journal of Innovative Technology and Exploring Engineering (IJITEE) , Vol.8, No. 12, ISSN: 2278-3075, pp 2941-2946.
7. Hassan Najadat & , Ismail Hmeidi(2008), "Web Spam Detection Using Machine Learning in Specific Domain Features ", Journal of Information Assurance and Security, pp 221-228.
8. M. Nivaashini, R.S.Soundariya, A.Kodieswari & P.Thangaraj(2018), " SMS Spam Detection using Deep Neural Network", International Journal of Pure and Applied Mathematics, Vol. 119 No. 18, pp.2425-2436.
9. Alexy Bhowmick ·& Shyamanta M. Hazarika (2016), "Machine Learning for E-mail Spam Filtering: Review, Techniques and Trends".
10. W.A. Awad & S.M. ELseuofi(2011), "Machine Learning Methods For Spam E-Mail Classification " International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, pp.173-184.

Published by: The Mattingley Publishing Co., Inc.