Email Monitoring System

Atiya Kazi[1], Asavari Mhadgut[2], Sayali Nalawade[3], Pratiksha Ranshur[4]

[1]*Assistant Professor*
[2]*Student*
[3]*Student*
[4]*Student*

Department of Information Technology
Finolex Academy of Management and Technology, Ratnagiri, India

Abstract: *Nowadays, a huge amount of people rely on available email or messages sent by strangers. The possibility that anybody can leave an email or a message provides a golden opportunity for spammers to write spam messages about our different interests. Spam fills the inbox with a number of ridiculous emails. Degrades our internet speed to a great extent. Steals useful information like our details on our contact list. Identifying these spammers and also the spam content can be a hot topic of research and laborious tasks. Email spam is an operation to send messages in bulk by mail. Since the expense of the spam is borne mostly by the recipient, it is effectively postage due advertising. Spam email is a kind of commercial advertising which is economically viable because email could be a very cost effective medium for the sender. Using Email Monitoring System, spam and non-spam emails are classified using ML algorithm - Logistic Regression.*

Keywords: email, spam, non-spam, logistic regression, corpus, suspicious, fraud, unauthorized.

1. Introduction

Email has now become one of the best ways for advertisements due to which spam emails are generated. Spam emails are the emails that the receiver does not wish to receive a large number of identical messages are sent to several recipients of email. Spam usually arises as a result of giving out our email address on an unauthorized or unscrupulous website. There are many of the effects of Spam. Fills our Inbox with a number of ridiculous emails. Email Spam has become a major problem nowadays, with rapid growth of internet users. Spam emails are the messages sent to multiple addresses. People are using them for illegal and unethical conducts such as phishing and frauds. So, it is necessary to identify spam mails which are fraud. Using Email Monitoring System, spam and non-spam emails are classified using ML algorithm - Logistic Regression.

2. Literature Survey

We have reviewed some papers regarding our project “Email Monitoring System”.

In paper [1], authors have highlighted spam classification which is created using Bayes theorem and Naïve Bayes classifier and also IP addresses of the sender are often detected and its accuracy is 97%.

In paper [2], authors described cyber-attacks. Phishers and malicious attackers are frequently using email services to send false kinds of messages by which target users can lose their money and social reputation. It creates spam Email classification using Naïve Bayes Algorithm and the accuracy is 99%.

In paper [3], the system Uses Machine Learning techniques like Naïve Bayes, KNN, SVM and summarizes the overall scenario regarding Accuracy Rate.

In paper [4], It compares the implemented algorithm of XGboosting Classifier with the existing classifier of SVM, Naïve Bayes and its accuracy is 95% .

In paper [5], It implements spam detection by using Machine learning and deep Learning Techniques like Naïve Bayes, Decision Tree, Neural Networks and random forest.

3. Existing System

The Existing Email system has a weak spam detection Mechanism. This results in segregation of important emails into spam emails. This can lead to miscommunication or delay of messages. Sometimes, some unimportant or phishing emails are not segregated into spam folders which leads to unnecessary attention to unimportant emails.

Drawbacks of existing system are as follows:

It is the source of viruses. It is capable of harming one's computer and reading out a user's email address book and sending them to a number of people around the world. It can be the source of various spams. These spam mails can fill up the inbox and deletion of these mail consumes a lot of time. It is an informal method of communication. The documents that require signatures are not managed by email.

4. Proposed System

In our system to solve the problem of spam, the email monitoring system is implemented using a machine learning algorithm - Logistic Regression. Using our system, the problem of the existing system will be solved.

There are some advantages of our system:

Ensemble methods on the other hand proved to be useful as they used multiple classifiers for class prediction. Nowadays, lots of emails are sent and received and it is difficult as our project is only able to test emails using a limited amount of corpus. Our project, thus Email monitoring system, is proficient in filtering mails giving to the content of the email and not according to the domain names or any other criteria. Also, our system has good efficiency as well as greater accuracy.

5. Flow Diagram

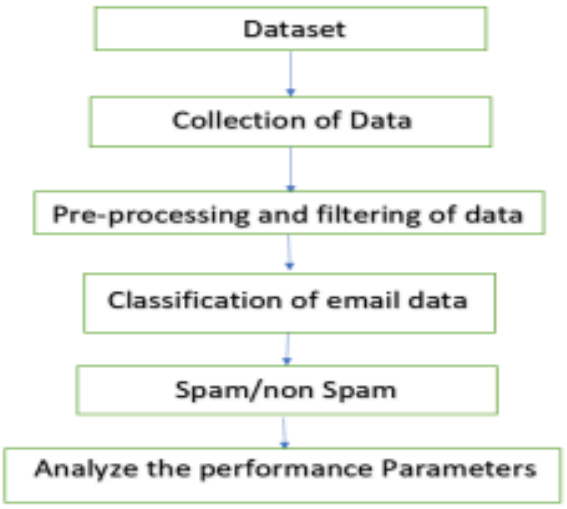

Figure 1. Flow Diagram

Figure 1 represents the flow chart of our system. In which emails are collected from the dataset and it is pre-processed and filtered and the email will be detected as spam or non-spam.

6. Methodology

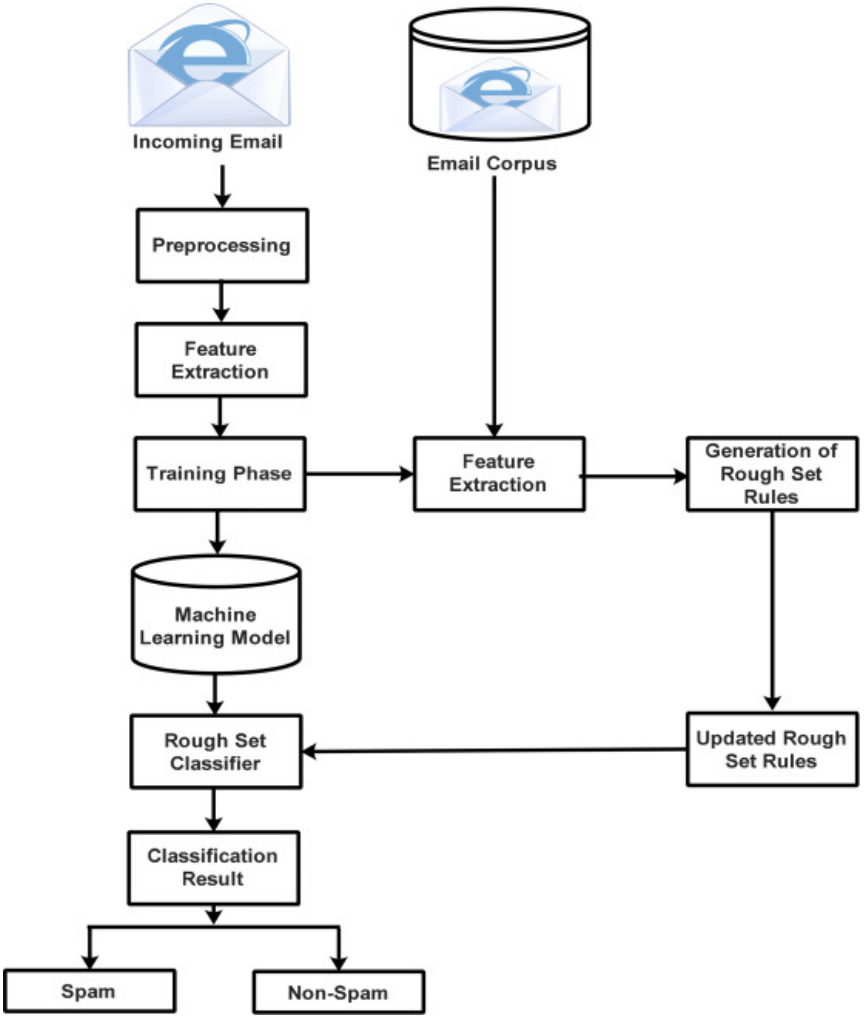

Figure 2. System Architecture

In our system, according to Figure 2, incoming Emails will be pre-processed and features are extracted and the features extracted are given to the training phase. And Email corpus is the text file of the email from which features are extracted and rough set rules are generated. Then using the machine learning model - Logistic Regression, emails are classified as spam and non-spam.

7. Result

Label	Meaning
1	ham
1	ham
0	spam
1	ham
1	ham
1	ham
1	ham
1	ham
0	spam
0	spam
1	ham
0	spam
0	spam
1	ham
1	ham
0	spam
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham
1	ham

8. Conclusion

Email has been the most important medium of communication nowadays; through internet connectivity any message can be delivered to all over the world. More than 270 billion emails are exchanged daily, about 57% of these are just spam emails. Spam emails, also known as non-self, are undesired commercial or malicious emails, which affects or hacks personal information like bank, related to money or anything that causes destruction to a single individual or a corporation or a group of people.

Hence this system is designed in such a way that it detects unsolicited and unwanted emails and prevents them hence helping in reducing the spam message which would be of great benefit to individuals as well as to the company. In the future this system can be implemented by using different algorithms and also more features can be added to the existing system. We will implement an email monitoring system using Machine Learning Algorithm - Logistic Regression.

9. References

[1] “Email based Spam Detection”
Thashina Sultana, K A Sapnaz, Fathima Sana, Mrs. Jamedar Najath, Dept. of Computer Science and Engineering, Yenepoya Institute of Technology Moodbidri, India.

[2] “Email Spam Detection using Naive Bayes Classifier”
Megha Tope, ME Student, Computer Science and Engineering, CSMSS College of Engineering, Aurangabad, India.

[3] “A Survey of Existing E-Mail Spam Filtering Methods”
Considering Machine Learning Techniques By Hanif Bhuiyan, Akm Ashiquzzaman, Tamanna Islam Juthi, Suzit Biswas & Jinat Ara, Southeast University.

[4] “Email Spam Filtering Using Machine Learning Based Xgboost Classifier Method”,
1p.U. Anitha, 2dr.C.V. Guru Rao, 3dr. D. Suresh Babu, 1PhD Scholar, Dept of Computer Science and Engineering, JNTU, Hyderabad 2DIRECTOR, S.R. Engineering College, Warangal, Telangana, India, 3H.O.D, Department of Computer Science and Engineering, Kakatiya Government College.

[5] “Machine Learning Techniques for Spam Detection in Email and IoT Platforms: Analysis and Research Challenges”,
Naeem Ahmed, Rashid Amin, Hamza Aldabbas, Deepika Koundal, Bader Alouffi, and Tariq Shah.

[6] “Identification of Spam Email Based on Information from Email Header”,
Shukor Bin Abd Razak, Ahmad Fahrulrazie Bin Mohamad 13th International Conference on Intelligent Systems Design and Applications (ISDA), 2013.

[7] “E-Mail Spam Detection Based on Part of Speech Tagging”,
Mohammed Reza Parsei, Mohammed Salehi 2nd International Conference on Knowledge Based Engineering and Innovation (KBEI), 2015.

[8] "Email classification research trends: Review and open issues",
Mujtaba, Ghulam, et al. IEEE Access 5 (2017).

[9] "Spam/ham e-mail classification using machine learning methods based on bag of words technique",
Sahın, Esra, Murat Aydos, and Fatih Orhan, 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE, 2018.