

Survey Paper on Implementation of Abusive Language Detection in Chatting Application

Mr. Siddharth Shukracharya Rokade,
siddharthrokade0504@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of
Engineering And Research, Nashik

Mr. Shreyas Kailas More,
kartikmore1@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of
Engineering And Research, Nashik

Mr. Gaurav Sanjay Rajole,
gauravrajole15@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of
Engineering And Research, Nashik

Mr. Atharva Vivek Sandhan,
sandhanatharva4@gmail.com,
Department of Computer Engineering
KVN Naik Loknete Gopinathji Munde College of
Engineering And Research, Nashik

Abstract – Communication has become stronger due to exponential increase in the usage of social media in the last few years. People use them for communicating with friends, finding new friends, updating any important activities of their life, etc. Due to their growing popularity and deep reach, these mediums are infiltrated with huge volume of spam messages. Spam message randomly sent to multiple addressees by all sorts of groups, but mostly lazy advertisers and criminals who wish to lead you to phishing sites. In this we are using various machine learning techniques for detecting spam in the short text messages and also Google vision API for detecting spam images.

Key Words: Spam Filtration, Google Vision API, OWASP, Naive baice classifier, Dictionary Based Algorithm

1. INTRODUCTION

Spam may or may not be harmful to the intended person. It might range from just a funny text message to a deadly virus that may corrupt the entire machine or a code written to steal all the information on your machine. Initially, the spam started spreading with email, but with the increase in the use of the Internet and the advent of social media, they started to spread like an epidemic. These chatting apps include personal information and extra information about everyone. We should think about increasing cybercrimes through social media, spam messages that is fraud messages.

These apps are just for entertainment there are no extra features preventing cyber-crimes, vulgarity, obscene word, etc. Spam message randomly sent to multiple addressees by all sorts of groups, but mostly lazy advertisers and criminals who wish to lead you to phishing sites. It has a higher response rate as compared to email spam. Apart from emails, and SMS, social networking like Twitter, Facebook, instant messenger like WhatsApp are also contributing to a major chunk of spam over the network.

The Project App refers to irrelevant or unsolicited messages sent over the messengers for abusing or may harm someone's personal life. The spam may or may not be harmful to the

intended people. Message Protection is a tedious task and in the absence of automatic measure for filtering of message, the task of spam filtering is taken up with the person at the receiving end. Communication has become stronger due to exponential increase in the usage of social media in the last few years. People use them for communicating with friends, finding new friends, updating any important activities of their life, etc. Among different types of social media, most important are social networking sites and mobile networks. Due to their growing popularity and deep reach, these mediums are infiltrated with huge volume of spam messages.

1.1 MOTIVATION

The motivation for developing a project on "Implementation of Abusive Language Detection in Chatting Application" is driven by several important factors and benefits. The project has the potential to significantly improve spam which may or may not be harmful to the intended person. It might range from just a funny text message to a deadly virus that may corrupt the entire machine or a code written to steal all the information on your machine. Initially, the spam started spreading with email, but with the increase in the use of the Internet and the advent of social media, they started to spread like an epidemic. The increasing awareness about the email spam has decreased the spamming drastically, therefore traditional spammers are now using mobile and Internet technologies as a spam medium. So in our system, we are blocking spam messages, URL's and images too. When any user tries to send bad words or inappropriate text, images or phishing links then our system will block that and user will get logout of the system. That spam message will not reach to other user.

1.2 PROPOSED SYSTEM

This chatting application allows you to chat numerous users at once. The users of the system would be the people installing the application. He/She will use this system for chatting purpose and irrelevant content would be blocked if getting sent. The user's messages will get stored in an encrypted form on the firebase cloud. The users will have the following functions as they are using the chatting

application. The user need to create a new account. The second step that is verification should be done by using the phone number. The user will then get a list of registered users on the chatting application. Lastly the new user can start chatting with the selected user. This screen allows the user to chat with his/her friends.

II. PROBLEM STATEMENT

To develop a chatting application which blocks spam messages, URL's and images. Spam comments refer to the unwanted comments with rude words, advertisement, political or religious views. Massive spam comments seriously decrease users reading experience and hinder the healthy development of social media. Thus, it is essential to detect and Filter spam comments. It provides sensitivity to the client and adapts well to the future spam techniques. It increases Security and Control. This technique can also added to other apps as hike messenger, Instagram, etc.

III. LITERATURE SURVEY

According to a technical report by Ferris Research Group [3], it is stated that these types of mails occupy a chunk of bandwidth and storage space with the user wasting their precious time and energy in avoiding these types of mails. This has resulted in the financial strain on organizations, increased requirement of storage, spreading of offensive material like pornographic content and above all it violates the privacy of the people at the receiving end. Other mediums of spam are social networking sites, spam blogs, etc. which are used to send/ receive messages and the SMS which carry spam over mobile networks. The increasing awareness about the email spam has decreased the return rate drastically, therefore traditional spammers are now using mobile and Internet technologies as a spam medium. With the widespread availability of smart phone, there is an increase in the Volume of data exchanged over the network. SMS is a very cost effective method used for exchanging messages and therefore these can be used to send to the users. It has a higher response rate as compared to email spam. Apart from emails, and SMS [4], social networking like Twitter [5], Facebook, instant messenger like WhatsApp etc. are also contributing to a major chunk of spam over the network. Similar to Ad hoc classifiers, there is a rule based filtering, with the difference that rules are more formally written and can be deployed to a wide area of clients. A set of pre-defined rules are applied to an incoming messages and the message is marked as spam if the score of the test exceeds the threshold specified. Survey of anti-spam tools are provided in [6, 7]. This paper presents an overview of the road and traffic sign detection and recognition. It describes the characteristics of the road signs, the requirements and difficulties behind road signs

detection and recognition, how to deal with outdoor images, and the different techniques used in the image segmentation based on the color analysis, shape analysis. Many companies have additional checks in the form of white-listing, black-listing [8, 9] and grey-listing [10]. However, the success of these methods is limited and they need to be combined with automatic machine learning methods in order to get fairly good results. Machine learning algorithms comes under the category of content based classification technique since the properties and features are extracted from the text of the message.

IV. OUR PROPOSED MODEL

The proposed system will consist a GUI through which the user will get registerto application, interact with the system, providing functionalities such as welcome toapp, login to app, user would get verified through opt. New users would get added afterlogin to application. The proposed model will take the links, vulgar messages, vulgar images send by userand then perform blocking of those contents. After obtaining the results from the previous function, a small message saying 'inappropriate message' would get display. Hence the irrelevant content would get blocked. If the users send normal messages then it would get encrypted and stored on firebase cloud. In our system, we are blocking spam messages, URL's and images too. When any user tries to send bad words or inappropriate text, images or phishing links then our system will block that. That spam message will not reach to other user.

V. SYSTEM ARCHITECHURE

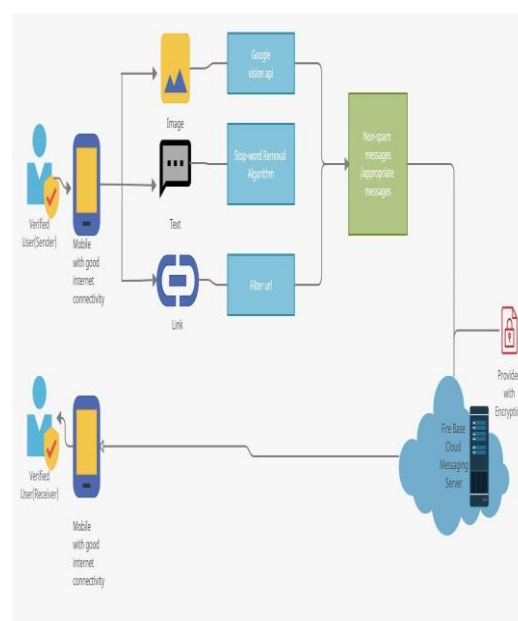


Figure 1: System Architecture

The above figure shows the system architecture of Spam Detection system which describes the overall working of the system. It is an overview of a system in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks. The start of the system is from the sender that is a verified user who is already registered to the chatting application in a mobile with good internet connectivity. The next step is divided in to three basic parts where the data sent by the user gets classified into image, text and links which gets filtered through the algorithm.

The image will be classified using Google Vision API. The text sent in the chatting will be classified using Stop Word Removal Algorithm. Lastly the links will be filtered and the remaining data will be forwarded to the next step. Provided with the encryption no one else who have the chatting application will the messages inside it for the security purpose. Each stage plays a crucial role in ensuring that the data is not only well-processed and analyzed but also effectively utilized in the application for the end to end user to easily use the potential of system. This process is a testament to how data-driven approaches are revolutionizing the abusive language detection which does not allows any inappropriate messages to correlate with the user using the chatting application.

4.2 Mathematical Model

Let 'S' be the system

S = I, O, F, Success Failure where,

I = Input

I = Message/links/images

F = Function

F = stop word removal, pattern matching, immaga API

O = Output

O = non spam messages, Decent messages

Hardware Requirement:

CPU = 13 6th generation and above

RAM – 4GB

Hard Disk – 500 GB

Software Requirement:

OS – Windows 10

Android Studio

Imagga – inbuilt machine learning platform

Database – Management System

Success- All functions executed successfully.

Failure- Some words may not get detected due to lack of data.

Some images and URL'S may not get filtered due to lack of datasets.

V. RESULTS AND DISCUSSIONS

Spam is a serious issue that is not just annoying to the end-users, but also financially damaging and a security risk. Message containing vulgar words or Images will not be sent from the messenger apps. Without spam filtering and security people can lack interest and loose trust on such applications, our application is secured. Also, this pandemic resulted to lots of cybercrimes through social media which would be avoided. The chatting application provides secure end to end encryption. Descent communication between users as only appropriate words would get sent. Descent group communication in future. Also, protecting the children from accessing adult content which is most important in today's world. If some illegal information is tried to send the information of that user can be given to government to stop any illegal action. Need internet connectivity every time for accessing the chatting application. Need to scan the database every time. Application will be useful only for android users. Hence it can't be operated with other operating systems.

Since the system is running on real time it should be predict the results of the spam detection through images, url's and words. The system should not be fail during the journey time. To avoid from failing the training model should be trained with enough number of database url's. The database is a software used to organize and manage data efficiently. Naive Bayes algorithms is designed to give most accurate and efficient result. Considering the security, we are adding AES algorithm which is standard encryption. In contrast, our proposed approach demonstrated its effectiveness in providing accurate blocking inappropriate messages or illegal data that is sent from a user for a wide range of conditions, retrieving the information and showcasing its adaptability and comprehensiveness. The integration of the trained model into an application enables chatting experience for the user in real-world scenarios, empowering and individuals to make informed decisions regarding inappropriate content or text assessment and management.

VI. CONCLUSION

Spam is a serious issue that is not just annoying to the end-users, but also financially damaging and a security risk. Message containing vulgar words or Images will not be sent from the messenger apps. Without spam filtering and security people can lack interest and loose trust on such applications, our application is secured. Also, this pandemic resulted to lots of cybercrimes through social media which would be avoided.

Apart from the promising results presented, there are different aspects for future research. As a future work, Group chatting feature will be added. Even in group chatting spam would be detected to restrict the spam messages getting viral. Or in group chat one should be targeted of trolls. It is a real time chatting application which can be used for professional level and also for common people for descent chatting and children can also use it as everything sent is appropriate in this application.

REFERENCES

- [1] L. F. Cranor, & B. A. La Macchia, "Spam!", Communications of the ACM, Vol 41, Issue 8, pp.74-83. 1998.
- [2] G. Jain, & M. Sharma, "Social Media: A Review", In Information Systems Design and Intelligent Applications, Springer India, pp. 387-395. 2016
- [3] M. Nelson, "Spam Control: Problems and Opportunities", Ferris Research, India, pp.23-82, 2003.
- [4] G.V. Cormack, J.M.G. Hidalgo, E.P. S  n  z, "Feature engineering for mobile (SMS) spam filtering", In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, US, pp. 871-872, 2007.