SPAM HAM DETECTION

[1]Samruddhi Pansare, [2]Joy Kaplol, [3]Shubhda Jadhav, [4]Shreyash Bhogade, [5]Ahetesha Inamdar
Department of Computer Engineering, Bhivrabai sawant Polytechnic, Pune, Maharashtra, India[1,2,3,4,5]

INTRODUCTION

Spam is any kind of unwanted, unsolicited digital communication that gets sent out in bulk. Often spam is sent via email, but it can also be distributed via text messages, phone calls, or social media.

While it may not be possible to avoid spam altogether, there are steps you can take to help protect yourself against falling for a scam or getting phished from a spam message

What does the word "Ham" mean, in the context of anti-spam?

Nowadays, it's likely that everyone knows what Spam means, in the context of e-mail. The use of the word Ham: It is an email that is not Spam. In other words, "non-spam", or "good mail". It should be considered a shorter, snappier synonym for "non-spam". Its usage is particularly common among anti-spam software developers, and not widely known elsewhere; in general it is probably better to use the term "non-spam", instead. By detecting unsolicited and unwanted emails, we can prevent spam messages from creeping into the user's inbox, thereby improving user experience.

LITERATURE REVIEW/RELATED WORK

The goal of this survey is to undertake a thorough literature evaluation on approaches for detecting and classifying spam content in social media. There are several sources of textual data on social media platforms such as Facebook, Twitter, E-mail, and YouTube. A variety of ways have been used to detect and regulate spam text. Our efforts are primarily motivated by a desire to learn more about different spam text detection and categorization algorithms. This section discusses the survey methodology that we used to conduct our detailed spam detection review.Based on our research objective, the initial search keywords were carefully chosen. Following an initial search, new words discovered in several related articles were used to generate several keywords. These keywords were later trimmed to fit the research's objectives. We chose certain search keywords based on the goal of our survey work, and after performing an initial search on those words, several keywords were derived from selected articles. The number of keywords is then reduced in order to meet our research goal.

EXISTING SYSTEM:

WhatsApp

WhatsApp is an internationally available freeware, cross-platform, centralised instant messaging and voice-over-IP service owned by American company Meta Platforms. It allows users to send text and voice messages, make voice and video calls, and share images, documents, user locations, and other content.

In the existing system there is no such notification/ Indication given by the developer in the application. Which can help the user to classify if the sent link is HAM or SPAM

PROPOSED SYSTEM:

To classify the link is is HAM or SPAM we will be adding a small indicator just next to the link which can help the user to identify whether the link is HAM or SPAM

ADVANTAGES

- Adding an indicator will be more beneficial for the users to identify the link
- It will help the user to stay away of the spam or online scams
- Protects against malware
- Website/Application safety check
- Keeps the user safe

DISADVANTAGES

- Reading the encrypted message would be more challenging

LIMITATIONS

- end to end encryption will not allow us to read the chats / message we can only detect the spam if the link is forwarded

APPLICATION

This application will be used for detecting the Spam messages or emails. Once the model is trained and done it is packed into a docker image. This image is then uploaded to AWS and runs in its own container as a Lambda Function. This type of serverless deployment has many advantages due to being very easy to scale and update as well as allowing the model to be invoked anytime from anywhere without worrying about servers and infrastructures.

For the case of our application, the model is invoked whenever a message request arrives to our messaging service. If the content of the message exceeds the minimum length then it is sent to the model, which will determine the language of the message and run the spam predictor on its content. The model will then return a json file with its outputs including language, timestamp, debugging features (if requested) and finally the spam probability.

SOFTWARE AND HARDWARE REQUIREMENTS AND SPECIFICATION:

Hardware requirements:

1.i3 Processor Based Computer

2.4 GB-RAM

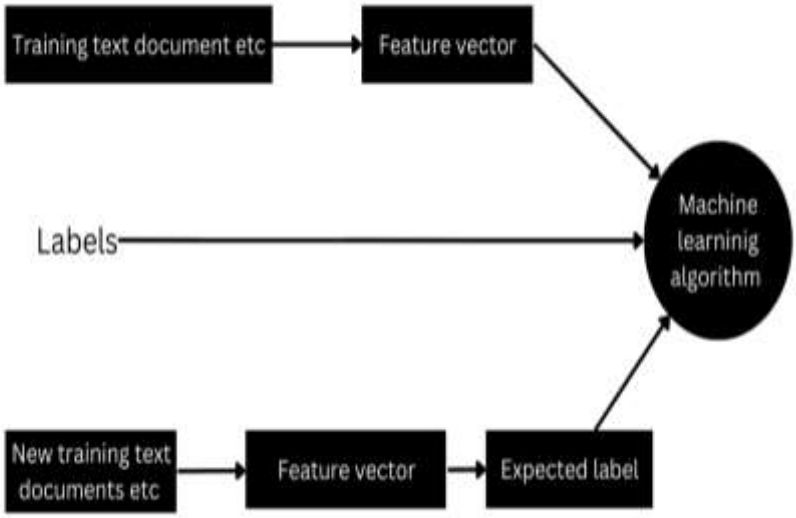
3.64 GB Hard Disk

Software requirements:

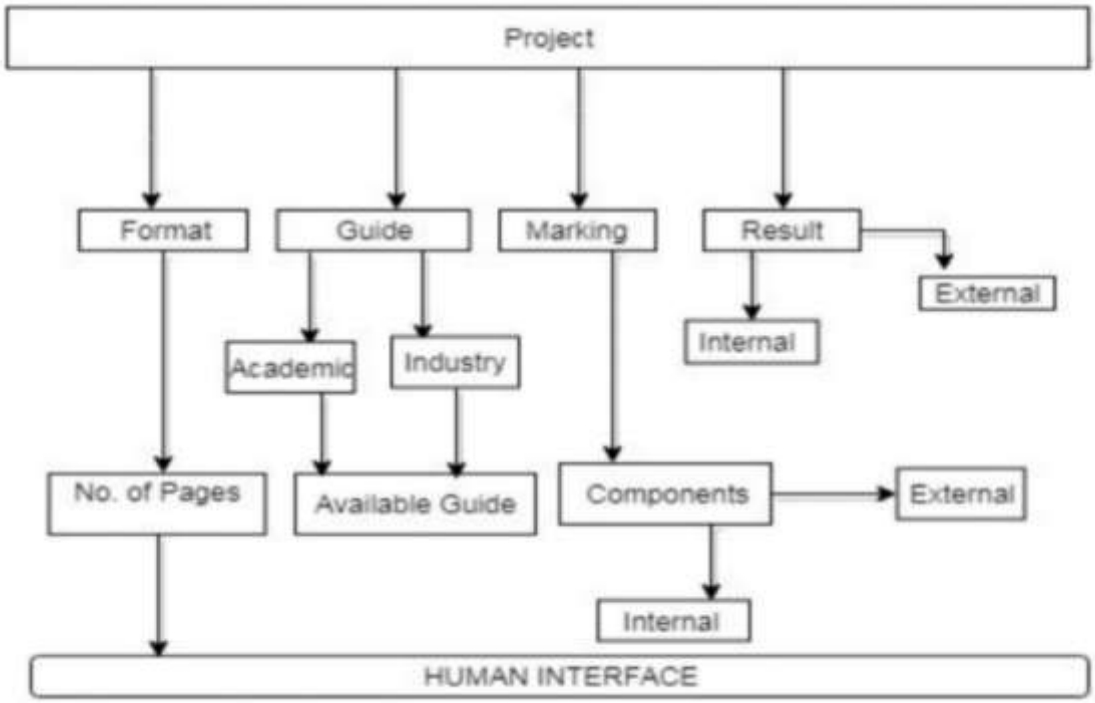
1.Windows XP,Windows7(ultimate & enterprise)

2.Python / Java

3.Graphical Q

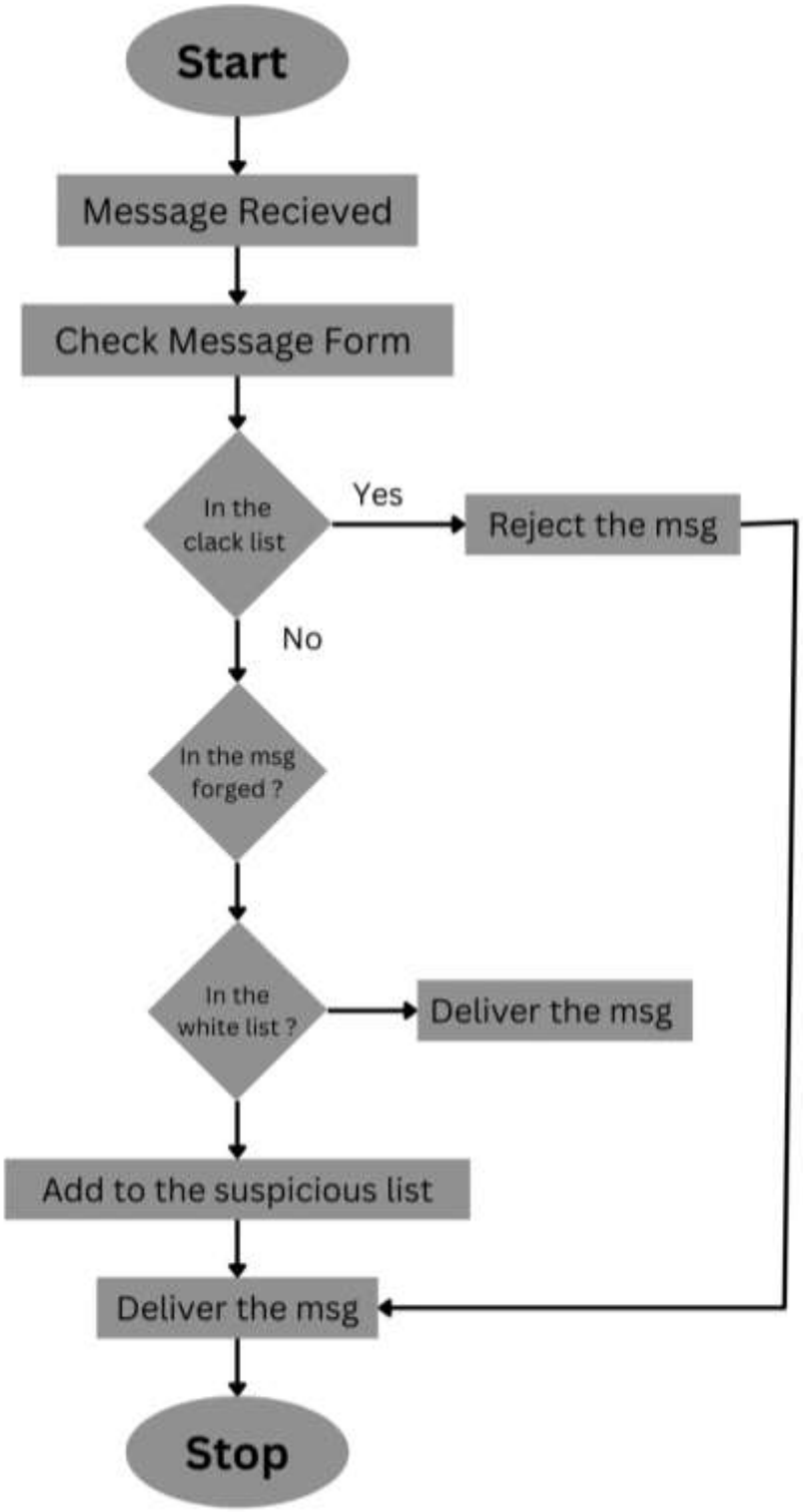


DFD



USE CASE DIAGRAM

FLOW CHART


Start
Message Recieved
Check Message Form
In the clack list
Yes
Reject the msg
No
In the msg forged ?
In the white list ?
Deliver the msg
Add to the suspicious list
Deliver the msg
Stop

CONCLUSION

Automatic labelling based on frequency allows for a reasonable creation of a labelled dataset.

- Spam detection can be re-framed as a regression problem and the added structure of message spam probability provides a more subtle classification.
- The model successfully detects new spam patterns not seen on the training set.
- Results improve considerably from better labels.
- Spam probabilities should not be interpreted in a vacuum but rather in the context of the problem and the analysed dataset.
- Running the model on the cloud is very cheap due to its speed and low memory usage.

The results of multiple classification models applied to the SMS Spam dataset are shown in table IV. From simulation results, multinomial naive Bayes with laplace smoothing and SVM with linear kernel are among the best classifiers for SMS spam detection. The best classifier in the original paper citing this dataset is the one utilising SVM as the learning algorithm, which yields overall accuracy of 97.64% . Next best classifier in their work is boosted naive Bayes with overall accuracy of 97.50%. Compared to the result of previous work, our classifier reduces the overall error by more than half. Adding meaningful features such as the length of messages in number of characters, adding certain thresholds for the length, and analysing the learning curves and misclassified data have been the factors that contributed to this improvement in results

REFERENCES

[1] https://en.wikipedia.org/wiki/Messaging_spam
[2] https://en.wikipedia.org/wiki/Anti-spam_techniques
[3] https://en.wikipedia.org/wiki/Spamming
[4] https://www.geeksforgeeks.org/sms-spam-detection-using-tensorflow-in-python/
[5] https://medium.com/@sharadjoshi/sms-spam-detection-45a9d6564d90