e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

SPAM EMAIL CLASSIFIER

Prof. Shraddha Sharma[*1], Meenal D. Amritphale[*2], Krish Vijayvargiya[*3], Harshwardhan Akhand[*4], Jinisha Kataria[*5]

[*1,2,3,4,5]Computer Science And Engineering, Acropolis Institute Of Technology And Research, Indore, M.P., India.

ABSTRACT

Today, a sizable portion of people rely on freely accessible email or communications provided by strangers. Because anyone may send an email or leave a note, spammers have an excellent opportunity to write spam messages regarding our various interests. Spam overflows email inboxes with absurd emails. severely reduces the speed of our internet. stealing vital information, such as contact information, from us. Finding these spammers and the spam content can be difficult work and a popular research area. Spam email is the act of sending many messages via postal mail. Spam is effectively postage due advertising because the recipient bears the majority of the cost.

I. INTRODUCTION

The internet is becoming a necessary component of daily life. Users of email are growing daily due to increased internet usage. Unsolicited mass email messages, or "Spam," have become an issue due to the growing usage of email. Spam emails are produced because email has become one of the best platforms for advertising. The recipient does not want to receive emails that are labelled as spam. Emails are delivered to numerous recipients in a high number of similar messages. Giving up our email address on an unofficial or dishonest website almost always results in spam. The consequences of spam are numerous. numerous crazy emails into our Inbox. These might also include links to websites hosting malware or phishing attacks, which have been known to steal sensitive data. Different spam filtering methods are employed to address this issue. Our mailbox is guarded against spam using spam filtering algorithms.

II. METHODOLOGY

Email Spam Detection is a project aimed at designing and implementing a system to identify and filter out unwanted spam emails from a user's inbox. The following are key components of this project:

- **Data collection:** This involves collecting a large dataset of both spam and legitimate emails for training and testing purposes.
- **Feature extraction:** This involves extracting relevant features from the collected emails, such as the sender's address, subject line, and email content.
- **Model training:** A machine learning model, such as a Naive Bayes classifier, SVM, or neural network, will be trained on the extracted features to learn the patterns of spam emails.
- **Model evaluation:** The trained model will be evaluated on a test set of emails to determine its accuracy and effectiveness in detecting spam.
- **Integration:** The spam detection system will be integrated with email platforms like Gmail, Yahoo, or custom email servers to filter out spam emails in real-time.
- **User interface:** The project will also include the development of a user interface to allow users to customize the spam detection system and manage their spam folder.

The end goal of the Email Spam Detection project is to build a robust, accurate, and efficient system to protect users from unwanted spam emails, while minimizing the risk of legitimate emails being marked as spam.

III. MODELING AND ANALYSIS

For the purpose of detecting spam emails, many approaches may be applied. However, machine learning-based categorization is a widely utilized strategy. Here is fundamental process for creating a machine learning model for spam email detection:

1) Data Gathering: Gather both valid and spam emails as part of sizable and varied dataset.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

2) Preprocessing Data: Remove any extraneous information from data, such as HTML elements, punctuation, and stop words. To get a collection of pertinent characteristics, also do text normalisation and feature extraction.

3) Feature engineering: Feature engineering is the process of choosing the most effective set of features that may be used to categorize emails as authentic or spam. The length of the email, the amount of hyperlinks, and the presence of specified characters are a few often utilised characteristics.

4) Model Training: On the preprocessed and feature-engineered dataset, train a machine learning model such as Naive Bayes, Support Vector Machines, or Random Forest.

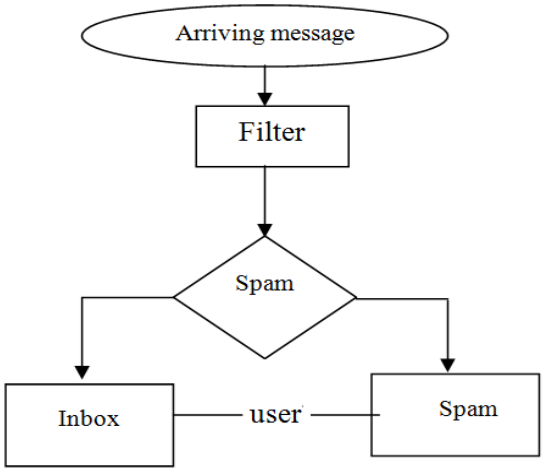
5) Model Evaluation: Evaluate the trained model's performance using test dataset. To evaluate the model's performance, use measures like accuracy, precision, recall, and F1-score.

6) Model tuning: To enhance model's performance, fine-tune it by optimising its hyperparameters.

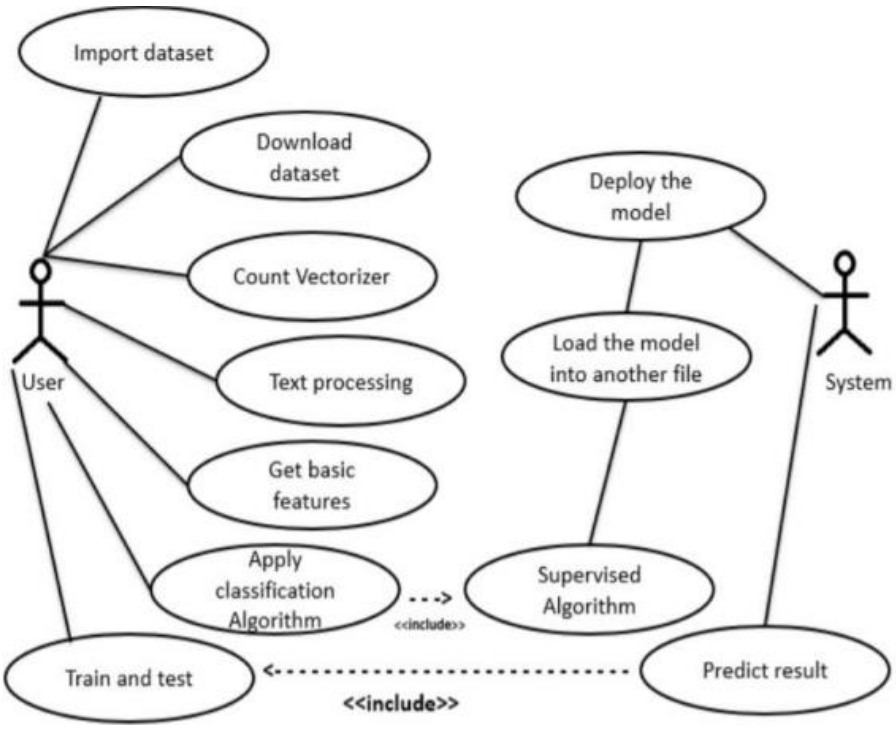
7) Deployment: To be used in spam email detection, deploy the finished model in a production environment.

It's crucial to remember that the aforementioned technique is only a general outline, and the specifics may change based on the demands of the unique application. access to the attendance.

Block Diagram for SEC:

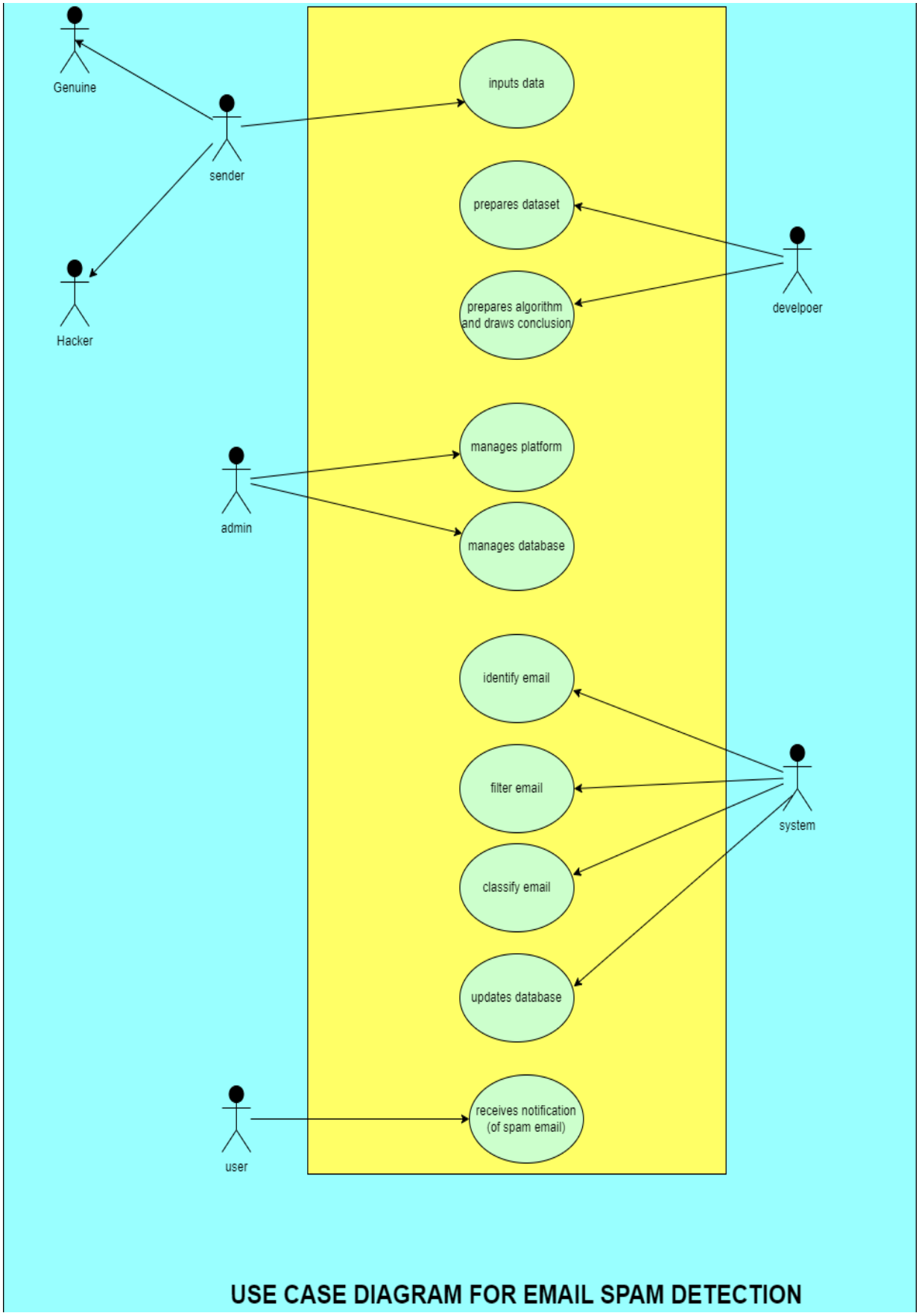


System Design for SEC:



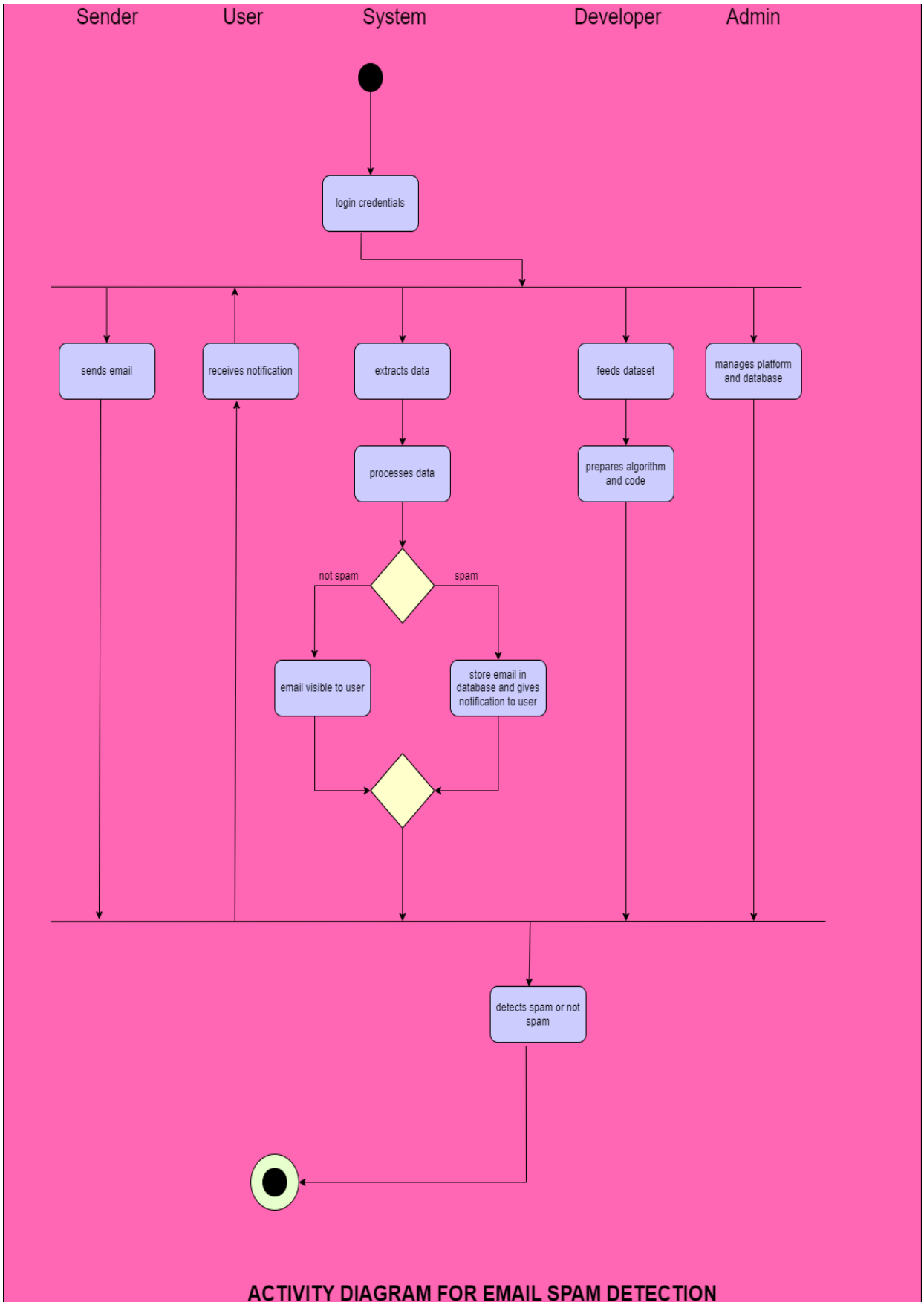
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Use Case Diagram for SEC:



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Activity Diagram for SEC:



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

Here are some of the results of the application. Figures below show how screen looks for the users:

IV. CONCLUSION

Today, email is the most significant form of communication because it allows for the delivery of any message anywhere on the globe thanks to internet connectivity. Every day, more than 270 billion emails are sent and received, of which 57% are spam. Spam emails, also referred to as "nonself," are unwanted commercial or malicious emails that damage or hack personal information like bank accounts, information relating to money, or anything else that causes harm to a single person, a business, or a group of people. In addition to advertisements, they could have connections to websites hosting phishing or malware created to steal personal data. Spam is a serious problem that not only annoys end users but is also financially damaging and a security risk. Therefore, this system was created so that it could identify unwanted and unsolicited emails and stop them, aiding in the decrease of spam messages, which would be extremely beneficial to both individuals and businesses. In the future, this system can be developed using various algorithms, and it can also get new features added to it.

V. REFERENCES

[1] Review of the literature: Bahman Rahimian et al., "A Survey of Email Spam Filtering Techniques" (2010)

[2] W. Nick Street and W. John K. Hsu's study, "Support Vector Machines for Spam Classification" (2003)

[3] Apache Spam Assassin is an open source implementation (https://spamassassin.apache.org/).

[4] “How to Build a Spam Filter Using Machine Learning" is a blog post by Jason Brownlee that can be found at https://machinelearningmastery.com/.