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/SMS CLASSIFIER (S.E.C)

Siddharth Patidar*1, Sanyam Sethiya*2, Satya Jain*3

*1,2,3Acropolis Institute Of Technology And Research, Indore, M.P., India.

ABSTRACT

The project is based on the spam classifier website, which is used to classify whether an email is spam or not. In the project, Flask is used, which is a framework for Python and a powerful tool for making dynamic and interactive webpages.

Keywords: Python, Streamlit.

I. INTRODUCTION

The internet has progressively assimilated into daily life. The number of people using email is growing daily as a result of increased internet usage. Spam, or unsolicited mass email, is an issue that has arisen as a result of the growing usage of email.

Due to email's current status as one of the greatest mediums for advertising, spam emails are produced.Emails that the recipient does not want to receive are referred to as spam. Multiple emailreceivers receive a lot of copies of the same message. When we disclose our email addresson an unofficial or dishonest website, spam frequently results .Spam has several negativeimpacts. fills our Inbox with a large amount of absurd emails. significantly reduces our Internet speed. stealing important data from your contacts list, such our contact information. any computer programme that modifies the search results you receive.

Spamis a major time waster for everyone and, if you get a lot of it, it can get downright annoying. It takes time to locate these spammers and their offensive information.

These emails could include links to phishing or malware-hosting websites known to steal sensitive data. Utilising various spam filtering techniques, this issue has been resolved. The spam filtering methods are used to keep our mailboxes free of unwanted emails.

Problem Formulation

Spam emails and phishing attacks have become a significant problem for email users, posing a threat to their privacy and security. Despite the availability of spam filters, many of these unwanted emails still manage to bypass such filters and reach the inbox, leading to an increasing risk of phishing attacks and identity theft. Therefore, the problem statement of this project is to develop a spam email detection system using machine learning algorithms that can automatically identify and filter out unwanted emails, thereby reducing the risk of falling victim to phishing attacks and other malicious activities. The proposed solution aims to provide an effective, reliable, and scalable approach to spam email detection that can improve the overall email experience and security of users.

II. LITERATURE REVIEW

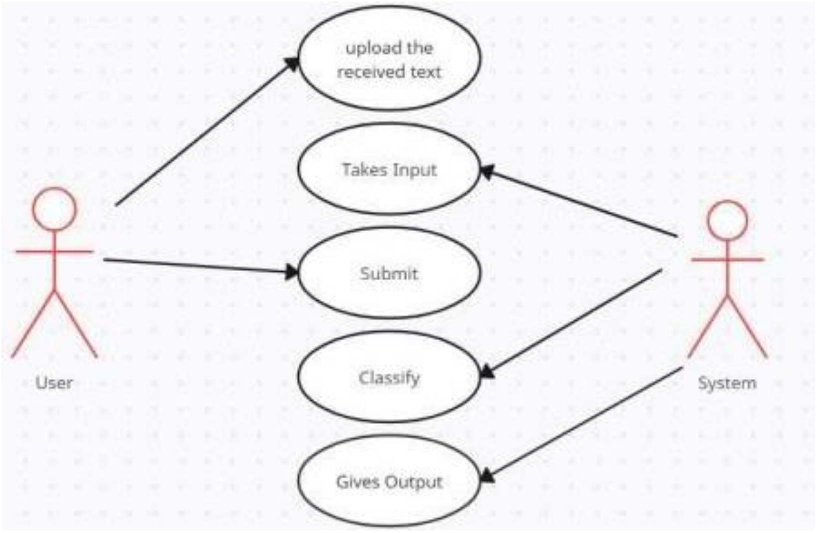
The first attempts to tackle the problem of spam emails involved creating rules-based filters that could identify spam emails based on specific keywords and patterns. These filters were relatively effective in the early days of spam, but spammers quickly adapted by using more sophisticated techniques, such as randomizing the text and using images to bypass the filters. In the early 2000s, machine learning techniques started to be appliedto the problem of email spam classification. The first machine learning-based spam filters used Bayesian algorithms, which were able to learn from the patterns of spam emails and make predictions based on probabilities. As spammers continued to evolve their techniques, more advanced machine learning algorithms were developed, such as support vector machines (SVMs) and decision trees. These algorithms were able to identify more complex patterns in spam emails and improve the accuracy of email spam classification. Today, most email providers use a combination of rules-based filters and machine learning algorithms to classify spam emails. These classifiers are continually updated and improved to keep up with the evolving tactics of spammers. Overall, the history of email spam classification shows how the problem has evolved over time and how technology has been used to develop increasingly sophisticated methods for identifying and filtering out unwanted spam emails

III. METHODOLOGY

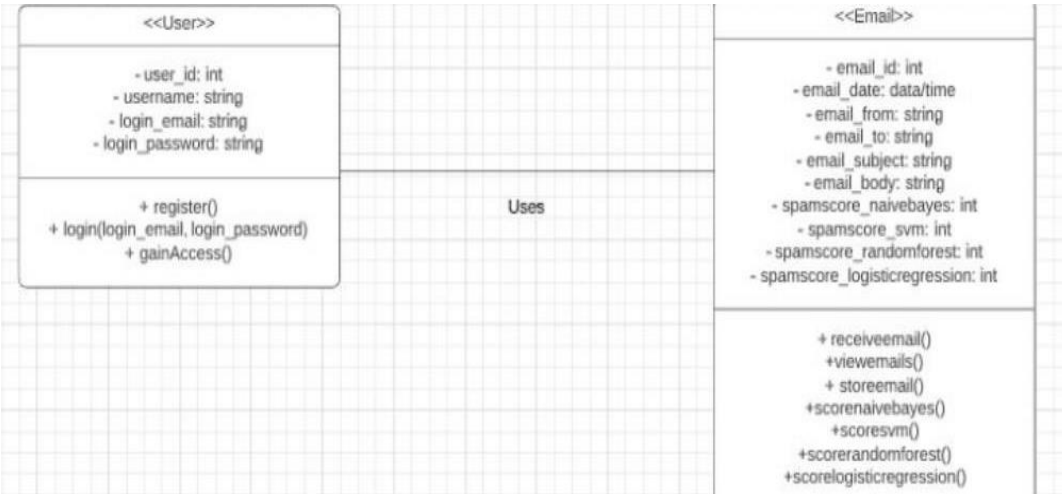
- For the purpose of detecting spam emails, many approaches may be applied. However, machine learning-based categorization is a widely utilised strategy. Here is a fundamental process for creating a machine learning model for spam email detection:
- Data Gathering: Gather both valid and spam emails as part of a sizable and varied dataset.
- Preprocessing Data: Remove any extraneous information from the data, such as HTML elements, punctuation, and stop words. To get a collection of pertinent characteristics, also do text normalization and feature extraction.
- 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 utilized characteristics.
- Model Training: On the preprocessed and feature-engineered dataset, train a machine learning model such as Naive Bayes, Support Vector Machines, or Random Forest.
- Model Evaluation : Evaluate the trained model's performance using a test dataset. To evaluate the model's performance, use measures like accuracy, precision, recall, and F1-score.
- Model tuning: To enhance the model's performance, fine-tune it by optimizing its hyperparameters.
- Deployment: To be used in spam email detection, deploy the finished model in a production environment.

UML Diagrams

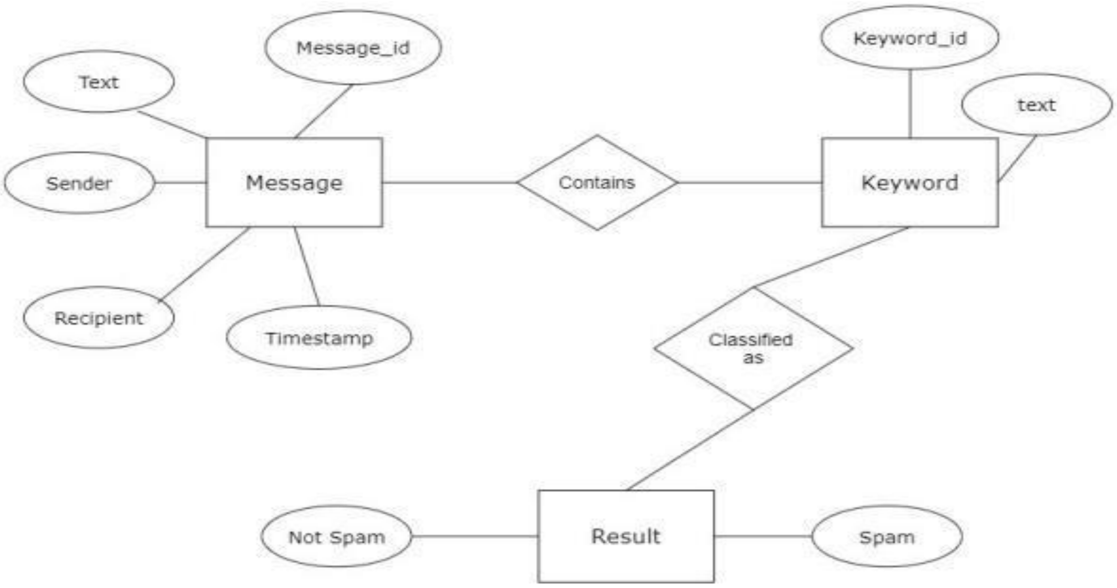
1. Use Case



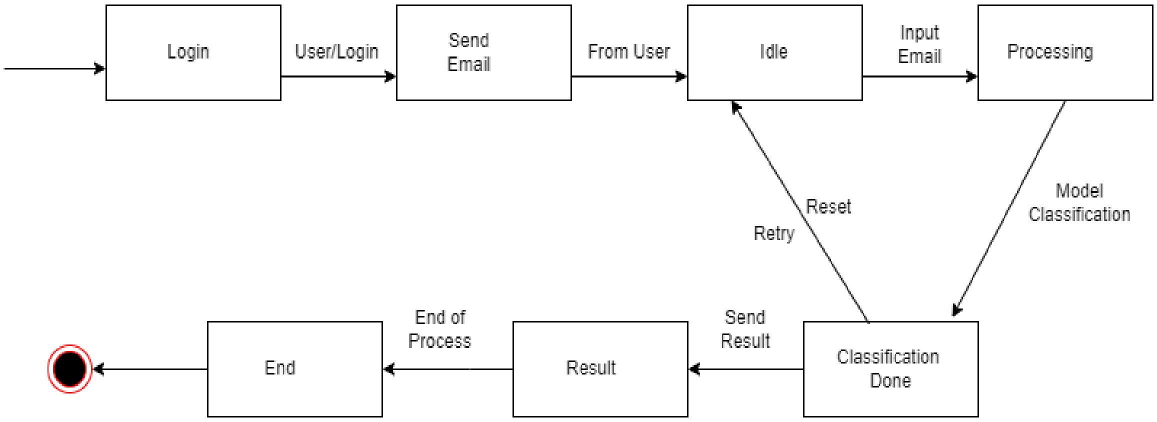
2. Class Diagram



3. ER-Diagram



4. State Diagram



IV. RESULT AND DISCUSSION

Spam detection classifiers are systems designed to identify and filter out unwanted or harmful messages, such as email spam or inappropriate content. These classifiers use various techniques to analyze messages and determine if they are spam or legitimate. Spam detection classifiers are like special filters for messages. They help sort out the bad messages (spam) from the good ones. They use some tricks to do this, like looking at words and patterns in the messages.

V. CONCLUSION

Today, email is the most significant form of communication since it allows for the delivery of any message anywhere in the globe thanks to internet connectivity. Every day, more than 270 billion emails are sent and received, of which 57% are spam. Spam emails, often referred to as "non-self," are unwanted commercial or harmful 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 might have connections to websites hosting phishing or malware intended to steal personal data. Spam is a severe problem that end consumers find bothersome but is also financially harmful and a security concern. Therefore, this system is created so that it can identify undesired and unsolicited emails and stop them, aiding in the decrease of spam messages, which would be extremely beneficial to both individuals and the business. In the future, this system may be developed using various algorithms, and it can also get new features added to it.

ACKNOWLEDGEMENT

We would like to express our gratitude towards Mr. Ambrish Srivastava for their support in accomplishment of our project on Spam Email/SMS Classifier (S.E.C) We would like to extend our deep appreciation to all my group members, without their support and coordination we would not have been able to complete this project.

VI. REFERENCES

[1] https://www.geeksforgeeks.org/

[2] https://www.researchgate.net/publication

[3] https://www.ncbi.nlm.nih.gov

[4] https://onlinelibrary.wiley.com/