WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Spam Email/SMS Classifier (S.E.C)

Atharva Puranik[1], Atharva Pagare[2], Bhumika Patidar[3]

[1,2,3]Department of Computer science, Acropolis Institute of Technology and research, Indore

ABSTRACT

Today, a sizable portion of individuals rely on freely accessible email or communications provided by strangers. Because anybody may send an email or leave a note, spammers have an excellent chance to compose 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 our contact information, from us. Finding these spammers and the spam content may be difficult work and a popular research area. Spam email is the practise of sending many communications using postal mail.Spam is basically postage due advertising because the recipient bears the majority of the cost. Spam email is a form of commercial advertising that is commercially feasible due to email's potential as a very cost-effective medium for senders. Using Bayes' theorem and Naive Bayes' Classifier, the suggested model allows the supplied message to be classified as spam or not.

Keywords— Term Frequency, Naive Bayes' Classifier.

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 email receivers receive a lot of copies of the same message. When we disclose our email address on an unofficial or dishonest website, spam frequently results .Spam has several negative impacts. 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.Spam is 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.

PROPOSED 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:

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

2) 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 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 a test dataset. To evaluate the model's performance, use measures like accuracy, precision, recall, and F1-score.

6) Model tuning: To enhance the 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.

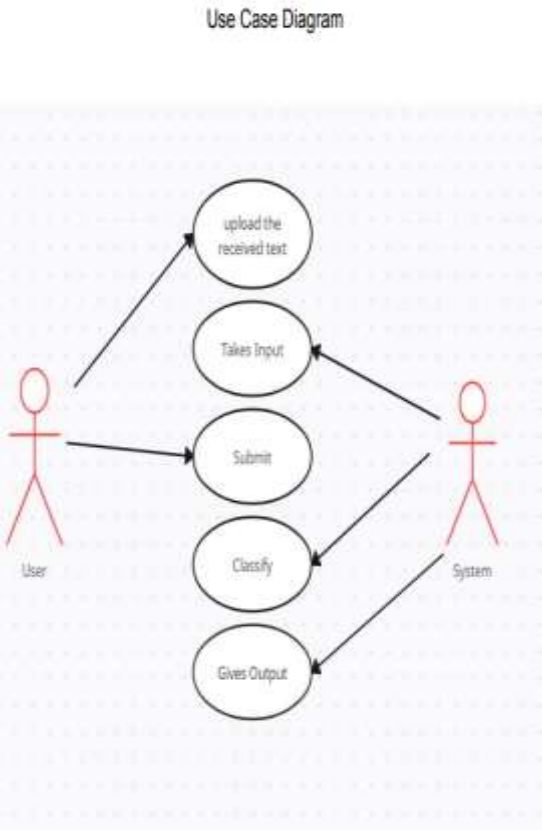
It's crucial to remember that the aforementioned technique is only a general outline, and the specifics may change based on the demands of the unique application. access to the attendance.

Design :

After the needs have been gathered and assessed, they must be given a proper structure. from this phase, the project's architecture will be created utilising the requirements acquired from the phase before as a reference.

In this stage, several architectural diagrams, such as ER diagrams, use case diagrams (Fig. 1), etc., are designed. The ER diagram demonstrates how the various entities are related and dependent on one another.

Fig 1: Usecase Diagram (SEC)



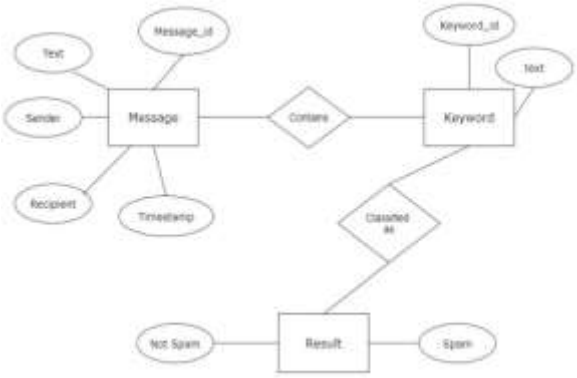


Fig2 : ER Diagram (SEC

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

Fig 3.1: Screen Shot of Project

Fig 3.2: Screen Shot of Project (Spam)

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.

Future Scope:

As the amount of spam emails increases and spammers' techniques advance, the future potential of spam email detection models seems pretty bright. Here are some probable directions for advancement in spam email detection algorithms going forward:Deep Learning-based models: Deep Learning techniques like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated promising results in natural language processing tasks, and they may also be helpful for spam email identification.Hybrid Models: Combining rule-based and machine learning-based techniques may produce superior outcomes to each technique used alone. For instance, a model that combines keyword-based criteria with a machine learning algorithm could be better at spotting spam emails. Real-time Detection: As email continues to be a primary mode of communication, there is a growing need for real-time spam email detection. Machine learning models that can quickly and accurately classify emails as spam or legitimate in real-time will be in high demand.Contextual Understanding: Spammers often use contextual cues to evade detection, such as using misspellings, substituting characters with symbols, or embedding images in emails. Models that can understand the context of an email and identify such cues will be more effective in detecting spam.

Acknowledgement

An individual may have some limitations, but with the association and cooperation of thought-provoking people, he can achieve his otherwise difficult dreams. The exchange of ideas generates a new object to work in a better way. Whenever a person is helped or operated by others, his heart is bound to pay gratitude to them.I would like to express my heartfelt thanks and high level of respect to my project guide, **Prof. Preeti Shukla**, whose constant support, vast knowledge, and experience have been a tremendous source of strength in my endeavour.

References

[1]"Spam Filtering: An Overview" by Andrzej M. J. Skulimowski: This paper provides an overview of various spam filtering techniques and their performance.

https://ieeexplore.ieee.org/abstract/document/4457923

[2]"A Survey of Techniques for Email Spam Filtering" by A. Sahami, S. Dumais, D. Heckerman, and E. Horvitz: This paper provides a comprehensive survey of the various techniques used for spam filtering.

https://www.microsoft.com/en-us/research/publication/a-survey-of-techniques-for-email-spam-filtering/

[3]"A Review of Machine Learning Techniques for Spam Email Detection" by N. K. Singh and A. K. Garg: This paper reviews various machine learning techniques that have been used for spam email detection. https://link.springer.com/article/10.1007/s10796-018-9906-1

[4]"Spam Detection Using Machine Learning Techniques: A Review" by S. Rawat, S. Singh, and V. Kumar: This paper provides a review of various machine learning techniques used for spam detection. https://www.sciencedirect.com/science/article/pii/S2405452618313148

[5]"Real-Time Email Spam Detection Using Machine Learning Techniques" by A. R. Siddique and A. O. E. Abu-Ein: This paper describes a real-time email spam detection system using machine learning techniques. https://ieeexplore.ieee.org/abstract/document/7985203