SPAM EMAIL CLASSIFIER

Harshvardhan Sisodiya
Computer Science and Engineering
Acropolis Institute Of Technology And Research
Indore, India
harshvardhan20295@acropolis.in

Dhruv Jain
Computer Science and Engineering
Acropolis Institute Of Technology And Research
Indore, India
dhruvjain20414@acropolis.in

Hemant Sisodiya
Computer Science and Engineering
Acropolis Institute Of Technology And Research
Indore, India
hemantsisodiya20450@acropolis.in

Prof. Krupi Saraf
Computer Science and Engineering
Acropolis Institute Of Technology And Research
Indore, India
krupisaraf@acropolis.in

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 our 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.**

Keywords—Machine Learning(ML), Dataset, Parkinson's Disease, Artificial Intelligence(AI), Support Vector Machine(SVM**).**

I. INTRODUCTION

In 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.

Spam has several negative impacts. It fills our Inbox with a large number 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. Utilizing various spam filtering techniques, this issue has been resolved. The spam filtering methods are used to keep our mailboxes free of unwanted emails.

II. PROBLEM STATEMENT

Unwanted spam emails can be a significant issue for individuals and organizations, leading to wasted time and reduced productivity, as well as potential security risks. The objective of a spam email classifier is to automatic ally identify and filter out unwanted spam emails, thereby improving the efficiency of email communication and enhancing email security.

The exponential growth of electronic communication has led to a significant increase in the volume of emails being exchanged daily. Unfortunately, this surge has also resulted in a proportional rise in spam emails, which are not only a nuisance but also pose security risks such as phishing attacks and malware distribution. Identifying and filtering out spam emails has become a critical challenge to ensure the integrity and security of email communication.

III. LITERATURE REVIEW

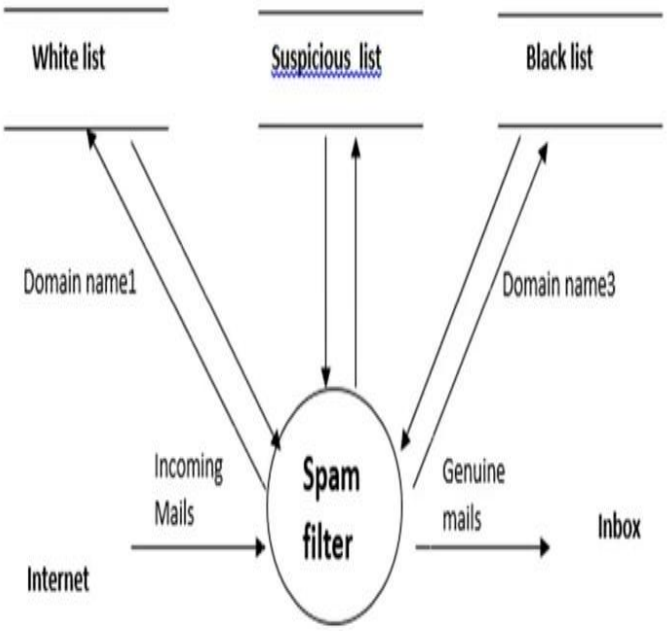
The first attempts to tackle the problem of spam emails involved creating rules-based filters that could identify spam emails based on specific keywords and patterns. These filters were relatively effective in the early days of spam, but spammers quickly adapted by using more sophisticated techniques, such as randomizing the text and using images to bypass the filters. In the early 2000s, machine learning techniques started to be applied to 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 rulesbased 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.

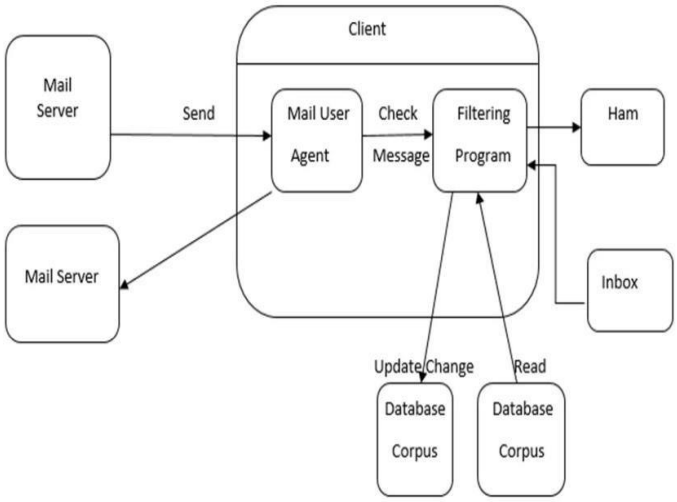
IV. METHODOLOGY

The methodology for email spam detection typically involves the following steps:

Preprocessing: This step involves cleaning and preprocessing the data by removing unwanted characters, converting text to lowercase, and removing stop words.
Feature extraction: This step involves extracting relevant features from the emails, such as the sender's address, subject line, email content, and header information. These features are used as inputs to the machine-learning model.
Model training: This step involves training a machine learning model, such as a Naive Bayes classifier, a Support Vector Machine (SVM), or a Neural Network, on the extracted features to learn the patterns of spam emails.
Model evaluation: This step involves evaluating the trained model on a test set of emails to determine its accuracy in detecting spam. This step is crucial to avoid overfitting and to ensure the model generalizes well to new data.
Model deployment: This step involves deploying the trained model in a real-world setting, such as by integrating it with email platforms or custom email servers to filter out spam emails in real-time.
Performance monitoring: This step involves monitoring the performance of the deployed model and updating it periodically based on feedback and changing spam patterns. In summary, the methodology for email spam detection involves preprocessing the data, extracting relevant features, training a machine learning model, evaluating its performance, deploying it in a real-world setting, and monitoring its performance over time.

V. SYSTEM DESIGN





VI. 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.

VII. REFERENCES

[1] “Machine Learning for Beginners”: by Harsh Bhasin
[2] Online Research: National Library Medicine
[3] https://www.ncbi.nlm.nih.gov/
[4] https://iopscience.iop.org/book/mono/978-1-64327-720-2
[5] https://en.wikipedia.org/wiki/Predictive_medicine
[6] “Machine Learning: Algorithms, Real-World Applications and Research Directions”: by Iqbal H. Sarker
[7] Parkinson's disease book by John Mitrofanis
[8] “THands-On Machine Learning with Scikit-Learn, Keras.