Email Spam Detection System

Abstract:

E-mail is one of the most popular ways of communication due to its accessibility, low sending cost and fast message transfer. However, Spam emails appear as a severe problem affecting this application of today's Internet. Filtering is an important approach to isolate those spam emails. In this paper, an approach for filtering spam email is proposed, which is based on classification techniques.

The approach analyses the body of email messages and assigns weights to terms (features) that can help identify spam and clean (ham) emails. An adaptation is proposed that tries to reduce the dimensionality of the extracted features, in which only determined (meaningful) terms are regarded by consulting a dictionary. A thorough comparative study has been studied among different classification algorithms that prove the efficiency of the filtering approach proposed.

Existing System:

The upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust anti-spam filters. Machine learning methods of recent are being used to successfully detect and filter spam emails. We present a systematic review of some of the popular machine learning based email spam filtering approaches. Our review covers surveys of the important concepts, attempts, efficiency, and the research trend in spam filtering.

The preliminary discussion in the study background examines the applications of machine learning techniques to the email spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use of machine learning techniques was done. Our review compares the strengths and drawbacks of existing machine learning approaches and the open research problems in spam filtering. We recommended deep learning and deep adversarial learning as the future techniques that can effectively handle the menace of spam emails.

Proposed System:

To effectively handle the threat posed by email spams, leading email providers such as Gmail, Yahoo mail and Outlook have employed the combination of different machine learning (ML) techniques such as Neural Networks in its spam filters. These ML techniques have the capacity to learn and identify spam emails and phishing messages by analysing loads of such messages throughout a vast collection of computers. Since machine learning has the capacity to adapt to varying conditions, Gmail and Yahoo mail spam filters do more than just checking junk emails using pre-existing rules. They generate new rules themselves based on what they have learnt as they continue in their spam filtering operation.

The machine learning model used by Google has now advanced to the point that it can detect and filter out spam and phishing emails with about 99.9 percent accuracy. The implication of

this is that one out of a thousand messages succeed in evading their email spam filter. Statistics from Google revealed that between 50-70 percent of emails that Gmail receives are unsolicited mail. Google's detection models have also incorporated tools called Google Safe Browsing for identifying websites that have malicious URLs. The phishing-detection performance of Google has been enhanced by the introduction of a system that delays the delivery of some Gmail messages for a while to carry out additional comprehensive scrutiny of the phishing messages since they are easier to detect when they are analysed collectively. The purpose of delaying the delivery of some of these suspicious emails is to conduct a deeper examination while more messages arrive in due course of time and the algorithms are updated in real time. Only about 0.05 percent of emails are affected by this deliberate delay.

Though there are several email spam filtering methods in existence, the state-of-the-art approaches are discussed in this paper. We explained below the different categories of spam filtering techniques that have been widely applied to overcome the problem of email spam.

Software Tools:

1. VS Code
2. Jupyter Notebook
3. Anaconda
4. Python3
5. Scikit-Learn
6. Pandas
7. NumPy
8. Matplotlib

Hardware Tools:

1. Laptop
2. Operating System: Windows-11
3. RAM: 16GB
4. ROM: 4GB
5. Fast Internet Connectivity

Applications:

1. This application can be integrated with third party email services like Zoho, Hostgator, Hostinger, etc.
2. Multi Category of Emails can also be done through this architecture.