CSE 303 Computer Networks Project

Topic: Anti-Spam Techniques for E-mail Communication

Abstract:

In the present era of technology, the Internet is playing a major role in the field of communication. E-mail (Electronic mail) has become one of the most preferred tools for exchanging information and ideas. Communication and Marketing products or services by email is a fast, flexible, and cost-effective way and it allows to send targeted and personalized messages. Several companies and government departments use E-mail as a communication tool to advertise their products and for inter and intra-organizational communication by employing spam techniques, which makes an abusive use of e-mails that lead to the spreading of unwanted information (messages) and malicious content to the Internet users. According to a survey on spam reports in 2021, it was estimated that nearly 320 billion emails were sent daily, out of which 85% of emails were detected as spam. These spam emails can lead to cyber-attacks, spread viruses to PC'S, and waste bandwidth, time, and memory.
Spam achieves this by including interesting subject lines that make us believe that they are coming from authorized sources. So, it becomes difficult to differentiate spam mail.

Therefore, with increasing in spam it is equally important to achieve security and filter spam mail from large volumes of data. This paper proposes an approach for spam filtering using some machine learning techniques and also some identification approaches for finding spam mail like Header filters. Although several machine learning algorithms have been employed in anti-spam e-mail filtering, like Support Vectors, Neural networks, and k-nearest neighbor, in this paper, we will discuss the Naive Bayes method which is a very popular open-source anti-spam e-mail filtering model. Bayesian filtering works by evaluating the probability of different words appearing in spam mail and then classifying them based on those probabilities. This document describes in detail how other common spam filtering techniques like List-based filters such as black list, white list, and grey list and Content-based filters like heuristic filter work, how effective they are at stopping spam, and their strengths, and weaknesses.

Keywords: Machine learning; Bayesian filtering; List-based filters; Content-based filters;

Kavyasri Paladugula AP20110010074

Hima Sri Tethala AP20110010081

Prameela Nannapaneni AP20110010120

Jyothika Reddy Palli AP20110010126