**Introduction**

Email or electronic mail spam refers to the “using of email to send unsolicited emails or advertising emails to a group of recipients. Unsolicited emails mean the recipient has not granted permission for receiving those emails. “The popularity of using spam emails is increasing since last decade. Spam has become a big misfortune on the internet. Spam is a waste of storage, time and message speed. Automatic email filtering may be the most effective method of detecting spam but nowadays spammers can easily bypass all these spam filtering applications easily. Several years ago, mos t of the spam can be blocked manually coming from certain email addresses. Machine learning approach will be used for spam detection. Major approaches adopted closer to junk mail filtering encompass “text analysis, white and blacklists of domain names, and community-primarily based techniques”. Text assessment of contents of mails is an extensively used method to the spams. Many answers deployable on server and purchaser aspects are available. Naive Bayes is one of the utmost well-known algorithms applied in these procedures. However, rejecting sends essentially dependent on content examination can be a difficult issue in the event of bogus positives. Regularly clients and organizations would not need any legitimate messages to be lost. The boycott approach has been probably the soonest technique pursued for the separating of spams. The technique is to acknowledge all the sends other than those from the area/electronic mail ids. Expressly

boycotted. With more up to date areas coming into the classification of spamming space names this technique keeps an eye on no longer work so well. The white list approach is the approach of accepting the mails from the domain

names/addresses openly whitelisted and place others in a much less importance queue, that is delivered most effectively after the sender responds to an affirmation request sent through the “junk mail filtering system”.

*Spam and Ham:* According to Wikipedia “the use of electronic messaging systems to send unsolicited bulk

messages, especially mass advertis ement, malicious links etc.” are called as spam. “Unsolicited means that those things which you didn’t asked for messages from the sources. So, if you do not know about the sender the mail can be spam. People generally don’t realize they just signed in for those mailers when they download any free services, software or while updating the software. “Ham” this term was given by Spam Bayes around 2001 and it is defined as “Emails that are not generally desired and is not considered spam”.



Machine learning approaches are more efficient, a set of training data is used, these samples are the set of email which are pre classified. Machine learning approaches have a lot of algorithms that can be used for email filtering. These algorithms include “Naïve Bayes, support vector machines, Neural Networks, K-nearest neighbor, Random Forests etc.”