1. **LITERATURE SURVEY**

In this section, related work has been depicted with existing works.

Billions of peoples are using emails and the major problem they are facing is inevitable spam emails which contains malwares and wrong information. This results loss of time and money and it is estimated that by 2025 there will be loss of $10.5 trillions because of spam emails[17]. Such spams can be detected using spam filter which differenciate the spam emails with ham emails based on the content of emails, email header details, the route that it takes form its original sender. Some major types of spam filter for email spam detection are given below[6]:

**Bayesian Filters**

These filter is based on Bayes theorem which determines the probability that a message is spam or ham based on the words and phrases it contains.

**Content-Based Filters**

Content-based filters examine the actual content of messages for certain keywords, phrases, or patterns commonly associated with spam.

**Header Analysis Filters**

Header Analysis filters analysis the email header information like sender’s email address, IP address, subject line etc to distinguish between spam and ham emails.

**Machine learning Filters**

These filters are based on machine learning algorithms like decision tree, random forest, logistic regression, neural networks, support vectors etc. Machine learning filters is trained using data set and along with the time the accuracy of this filters can be improved.

**Sender Policy Framework(SPF)**

SPF is an email authentication protocol that helps verify the legitimacy of the sender's domain. Spam filters can check SPF records to determine if an email is coming from an authorized source.

Although there are many other spam filters, the spam traps will still be there. So, many organizations and researchers are aiming to build robust and efficient filters to stop this problem. Many models based on machine learning algorithms show excellent performance to detect whether an email is spam or ham. Many researchers have got above 90% accuracy in their related algorithms along with that they have consider features like email text, header information, DNS record.They have used popular machine learning algorithms like logistic regression, SVM, Naive Bayes, Decision tree for training their model.

Spam email can also be detected based on the link present on it. For example, if mail contains the phishing websites then it is obvious that the email is a spam.Detecting a phishing websites can be determined based on the address bar features, abnormal based features, HTML and javascript based features, Domain based features[5].

Saleh et al.[18] presents spam filters that is based on Negative Selection Algorithms that has a high performance and low false detection rate. In this study, the system operates by elimination through Negative Selection similar to the functionality of T-cells in biological systems.Ahmed, Naeem Ahmed [3] surveys the machine learning techniques used for spam filtering techniques used in emails and IoT platforms by classifying them into suitable categories.

Thashina Sultana [1] uses email spam detection using machine learning approach (Naive Bayes) using IP address and email text as parameter. Similarly, focusing on data preprocessing procedure, Nikhil Kumar [2] used different steps like data cleaning, data integration,data transformation and data reduction. Yaseen, Qussai [8], concludes that spam email wastes millions of dollars annually. In this research paper, they uses Deep neural network model for such spam detection. Kartik [7] uses machine learning approach and For result comparison kappa coefficient, matthew correlation coefficients are used in this research work. Debnath, Kingshuk, and Nirmalya Kar [4] have used enron email dataset in deep learning models like LSTM and BERT. NLP was used to analyze data preprocessing of text of email.