

# Concept of Email Spam

**Pritee Patil<sup>1</sup> and Ashwani Patil<sup>2</sup>**

Assistant Professor, Department of IT<sup>1</sup>

Student, P.G. Department of IT<sup>2</sup>

Veer Wajekar ASC College, Phunde, Uran

**Abstract:** Spam emails are usually called junk mail and bulk unsolicited emails are delivered to the inbox. Emails used in advertising are regularly sent to the user via a subscription email that they may not have requested. Users of the spam email problem regularly experience it. Recently, Russia produced the largest share of 23.52 percent of the total spam emails in the world. One event claims that Google has registered 2,145,013 sensitive patent sites since Jan. 17, 2021. This has increased from 1,690,000 in Jan. 19, 2020 (increased by 27% in 12 months). As it is a well-known fact that 91% of all online attacks start with spam emails and about security issues, 97% of users fail to identify spam emails so finding a solution by filtering spam email can help reduce the risk of being a network. threats to business or employees. The "Spam Mail Recovery" project is a model built using a machine-readable learning method. It starts with data collection and continues with the previous processing method using NLP (Indigenous Language Processing Strategy). To select features on the website, the TF-IDF (Term Frequency and Inverse Document Frequency) method and vectorizer feature are used. The main stage is to build a model based on those features using machine learning algorithms such as Decision Tree, Naïve Bayes, Support Vector Class, Logistic Regression, KNearest Neighbor, Random Forest, Ada Boost, Bagging, Extra Tree Classifier, Gradient Boost Classifier and XG Boost all of these algorithms we have used on the website and based on comparative analysis of different machine learning models, the Extra Tree Classifier model provides 98% high accuracy of spam email detection.

**Keywords:** Machine learning, Natural Language Processing, Spam detection, Spam emails, Supervised learning

