**Email spam detection model(semester III)**

Nowadays, emails are used in almost every field, from business to education. Emails have two subcategories - ham and spam. The increasing volume of unsolicited spam email (bulk) has generated a need for reliable anti-spam filters. Machine learning techniques now days used to automatically filter the spam e-mail in a very successful rate. Email spam, also called junk emails or unwanted emails, is a type of email that can be used to harm any user by wasting his/her time, computing resources, and stealing valuable information or a subset of electronic spam involving nearly identical messages sent to numerous recipients by email. The ratio of spam emails is increasing rapidly day by day. Automatic filtering of spam emails becomes essential feature for a good email service provider. This paper surveys the machine learning techniques used for spam filtering techniques used in email and IoT platforms by classifying them into suitable categories. The most of the existing datasets were collected and prepared a long back and the spammers have been changing the content to evade the filters trained based on these datasets.

In this project I reviewed some of the most popular machine learning methods (Naive Bayes classification, k-NN, SVMs logistic regression, decision tree, random forest) and of their applicability to the problem of spam Email classification. Descriptions of the algorithms are presented, and the comparison of their performance on the Spam corpus is presented. Electronic mail has eased communication methods for many organizations as well as individuals. This method is exploited for fraudulent gain by spammers through sending unsolicited emails. Extensive research was done to implement machine learning models using Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree on an email dataset, along with feature extraction and pre-processing. The different algorithms like Naïve Bayes, Support Vector Machine, Random Forest, Decision Tree were implemented to optimize the performance of classifiers. Multinomial Naïve Bayes Algorithm performed the best overall. A comprehensive comparison of these techniques is also made based on accuracy, precision etc. The comparison of our results with other machine learning models to show the best suitable model is also discussed.

**Tools used**

Machine learning algorithms in Jupyter

**Role**

Programmer