Email Spam Detection

Problem Statement

We were asked to build a system to identify spam emails.

Method

Initially,as a 1st step,we had imported required libraries and given data(ie,csv file).Now,we can see and analyse data.The issue is that its so huge.So,we had tried to reduce it to most required attributes.

Then,we had replaced empty spaces,white spaces and punctuations also.Stopwords is used to perform NLTK operations and to clean the data.

Tools Used

For solving the problem, we had used below tools:

NLTK, Wordcloud, stopword, seaborn, matplotlib(for visualization)

TfidfVectoriser,MultinomialNB etc is used to perform spam detection action on the data.train_test_split method is used to split data into training and testing set.

Accuracy_score,Confusion_matrix and Classification_report are used to check whether we are performing right actions in the project

Result

We built a model in order to improve the spam detection for company with accuracy of 86.1%