

## SPAM DETECTION

### Problem Statement:-

We were asked to build system for SPAM Detection.

### Method:-

We Import libraries for our use and then we import given data. Now we can see data and analyze but because data is huge we will use some basics steps to replace email addresses, web-addresses 10 digit phone number, numbers with “email”, “webaddress”, “phonenumbr”, “numbr” and at last we also replace punctuation, whitespaces with “ ” empty spaces respectively. To perform NLTK operation and then clean data with the help of “stopwords”.

### Tools:-

For Solving our problem we use NLTK, wordcloud, stopwords, seaborn and matplotlib (for visualisation).

We need to perform spam detection action on data and that's the reason we will use here few models like TfidfVectorizer, MultinomialNB and then split data into train and test dataset using train\_test\_split,, we also use accuracy\_score, confusion\_matrix and classification\_report to see we are performing right action for project.

### Result:-

We make a model in order to improve the spam detection for company with accuracy of 88.39%