**Abstract:**

In this project we are going to detect spam and ham comment of you tube by using naïve bayes algorithm.We have a dataset.We will use this dataset to find spam and ham comments.We are using cloud base run time which is actually called google colab.It allows jupyter notebooks on the cloud.To perform this project we used python.By observing the output we found that we get a better result.By using this algorithm we train our data to find a good accuracy.

**Introduction:**

Naïve bayes algorithm:Naïve bayes algorithm is a machine learning algorithm which is used for classification problems.It is used for text classification.It involves high dimentional train data.

**Methodology:**

First of all we import nessesary modules like pandas as pd ,zipfile which allow to unzip dataset,pickle.For processing data we import sklearn.model\_selection library.It’s gonna be train\_test\_split to split in our test set.We will import sklearn.feature\_extraction.test.It is a factorisation technique.We will import naïve base algoritm.We import multinpmial nb algorithm.We will import some fuction from matrix model which will allow us to analyes performance of the model once we train it.We will import confusion matrix and classification report.Then we will read the dataset and drop comment\_id,date,and author form the data set because we will be working with the content of the comment of the datatset as well as the class label.After that we will call train\_test\_split fuction which will take the actual content which gonna be the comments and second arguments will be the class which is the labels.This fuction will return four values which will be the x\_train values,X\_test sets,Y\_train labels,Y test labels.It will perform 75% train sets and 25% test sets which is split by defaluts.We will vectorize the texts and fit the train data set which gives us numerical values so that machine can understand it.Then we will train the model.After prediction it will give an array where 0 menas ham comments and 1 means spam.For getting the performance of the model well will call confusion matrix.For accurary we will call classification\_report.

**Result analysis:**

