

## Read Me

Question 1 : Naive Bayes to predict whether the mail is spam or not

<https://www.kaggle.com/muditr97/assignment-5-question-1?scriptVersionId=12497429>

Question 2 : Naive Bayes Classifier to predict whether river or non river using satellite image

<https://www.kaggle.com/muditr97/assignment-5-question-2?scriptVersionId=12497436>

Data Set:

1. Spam Image
2. Hooghly River

The Assignment is on Jupyter Notebook

FileName for Naive bayes to predict whether the mail is spam or not:

**ICM2015502\_Assignment 5\_question\_1**

FileName for Naive bayes to predict whether river or not :

**ICM2015502\_Assignment\_Question\_2**

Open the file which ( is public ) add this to Notebook, Data set Name :

“assignment6 “ and run the code, for question 1.

This will predict the accuracy of our model to classify the given mail whether it is ham or spam.

Open the file which ( is public ) add this to Notebook, Data set Name :

“Assignment\_Hoogly “ and run the code, for question 2.

The code will produce two images black and white that is 0 and 255

Libraries used :

1. NumPy
2. SciPy
3. Pandas
4. Os

5. Sklearn
6. Matplotlib
7. PIL
8. Random
9. csv

## Analysis for Assignment for Naive Bayesian classifier

Train\_Test Split = 70 % and 30 % randomly using sklearn train test split tts()

```
# probability for some value of y, for some jth word of dictionary
def x_probability(phi, x_val) :
    val1 = phi ** x_val
    val2 = (1 - phi) ** (1 - x_val)

    return val1 * val2
```

This function will return the probability for x data values when the word is present in the dictionary.

The average accuracy of the model is about 86%.

### **How I Implemented**

- Four satellite Images of Kolkata (Rband, Gband, Bband and Iband) are given to you with equal image size (512 \* 512).
- The feature vector dimension is 4
- Each pixel location we have four values.
- Two Classes are given (River and NonRiver)
- Take 50 sample points (Pixel location's corresponding pixel values) from river class for training for each band
- Take 100 sample points (Pixel location's corresponding pixel values) from non river class for training for each band.
- Take (512 \* 512) sample points (Pixel location's corresponding pixel values) for testing for each band.
- Apply baye's decision rule to classify all the test sample either in river or nonriver class denoting 0 and 255 at corresponding pixel locations.
- Show the result in image form with black and white image (either 0 and 255)



