CSEN 1083 – Data Mining
Assignment #2
(Due on: April 4 at mid-night)
(This assignment can be done in groups of 2 students each)

You are required to design a Naïve Bayes classifier that can recognize scanned images of the 26 lower-case characters provided in the file "Assignment 2 Dataset.zip". The zip file contains two folders: "Train" and "Test". The "Train" folder contains 7 images for each lower-case character while the "Test" folder contains 2 images for each lower-case character. The images in the "Train" folder should be used to train a classifier for each character. After the classifiers are trained, test each classifier using the images given in the "Test" folder. Assume that each pixel is distributed according to a Gaussian distribution whose parameters are identified in the training phase of the Naïve Bayes classifier.

Deliverables:

- Your code.
- A plot of the number of images classified correctly for each character. The x-axis should show the character (a, b, …, z) while the y-axis should show the count. Name the plot "Accuracy.jpg".

Notes:

- It might be useful to divide the value of each pixel by 255 before training the classifier or testing it if the maximum brightness in images is 255.
- In case the probabilities computed during the testing phase are very small or equal to $0/0$, set the probability to a very small value (say 0.1).