**CSE 537 – Artificial Intelligence**

**Project 5**

**Submitted by**

**Sreepradhana Thirumalaiswami - 110349668**

**Sadhana Kumaravel – 110338732**

**1.Naïve Bayesian Classifier:**

The classifier takes the presence of ‘+’ or ‘#’ pixel in each possible position as 1 feature each to determine the number given the handwritten digits. The classifier is trained on the given training dataset. The prior probabilities of each of the digits - 0 to 9 are computed followed by computation of conditional probabilities of presence of ‘+’,’#’ pixels at a position given the numberThe classifier is then tested using the given test dataset consisting of 1000 records. . The classifier labels a handwritten a digit as a number for which the maximum Posteriori Probability is obtained. The obtained classification is then compared with the the labels given for each record of the test dataset.

**Execution note:**

python Naive\_Bayes.py --traininginput trainingimages.txt --traininglabels traininglabels.txt --testinginput testimages.txt --testinglabels testlabels.txt --trainingnum 5000 --testingnum 1000

option argument

--traininginput path corresponding to file containing training images

--traininglabels path corresponding to file containing training labels

--testinginput path corresponding to file containing testing images

--testinglabels path corresponding to file containing testing labels

--trainingnum number of records in training set

--testingnum number of records in testing set

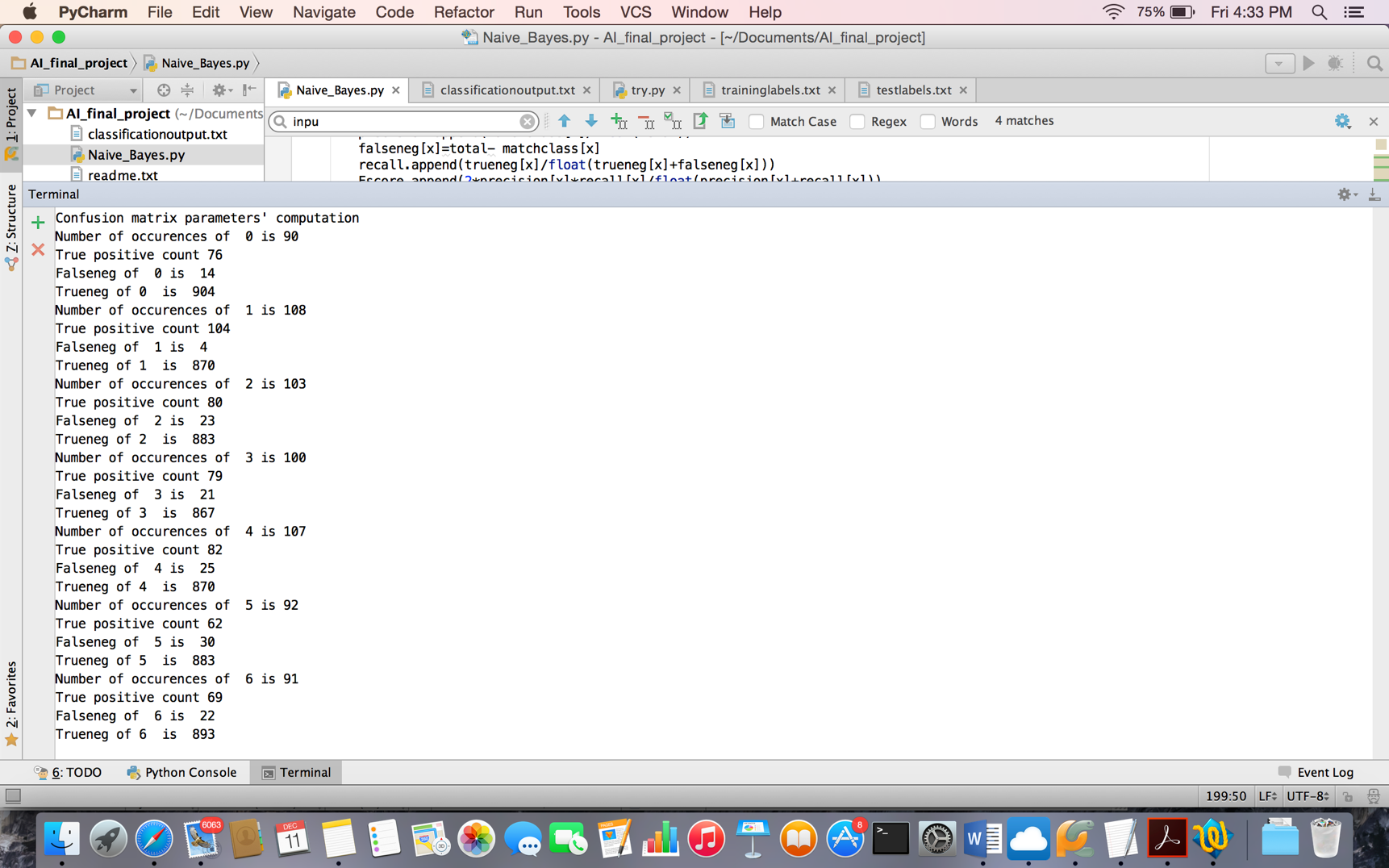
The output of the classification is stored in classificationoutput.txt in the same folder where the program resides.

**Evaluation metrics used:**

**Accuracy**: Proportion of total records classified correctly by the classifier.

The confusion matrix parameters True positive, False negative, True negative are computed based on results of test data classification.

**Value obtained: 77%**

****

**Precision or positive predictive value**: The number of correct positive results divided by the number of all positive results. This is the ratio of true positive to the sum of true positive and false negative counts.

**Recall or sensitivity**: The number of correct positive results divided by the number of positive results that should have been returned. This is computed as the ratio of the true negative count to the sum of true negative and false negative.

**F-score**: It is the weighted average of precision and recall.

