Description of the Project ("Naïve Bayes Email Classification")

For this project, I worked on an email classification problem using one of the simplest techniques called naive Bayes classification, which we learned about in class. Here, I have classified the email whether it is spam or ham as discussed in our class. Naive Bayes classifiers are the classifiers that are based on Bayes' theorem, a theorem that gives the probability of an event based on prior knowledge of conditions related to the event. It can be used to build a naive but good enough spam classifier.

I tried to provide a complete step-by-step pythonic implementation of naive Bayes without the use of any frameworks or Scikit learn, and by keeping in mind the mathematical and probabilistic difficulties I usually encounter when trying to delve deep into the algorithmic insights of machine learning algorithms. I developed the method from scratch while studying Naive Bayes classifiers to help solidify my understanding of math.

I think the project that I have done closely demonstrates the familiarity with a specific AI topic that was covered in class. The Project is running in a pretty good way delivering promising results and better accuracy on the test dataset.

Testing Accuracy:90.22%

Testing Error:9.77%

The Dataset used for this project is obtained from:

http://archive.ics.uci.edu/ml/machine-learning-databases/spambase/spambase.data

****The dataset, code, and demo video are included in the zip file.***

"Contributions"

Contribution # 1: Data Preprocessing and Normalizing

Contribution # 2: Done everything from scratch without using Scikit Learn for model building and to find the accuracy of the model.

Contribution # 3: Created a simpler and easy-to-understand version of the sklearn.naive Bayes estimator.