Reflection Report

In the homework week 09, I have learned about Naïve Bayes Classifier which assume that features are independent and applies Bayes' Theorem to classify data using a probabilistic model. I explored more about three Naive Bayes classifiers, BernoulliNB, GaussianNB, and MultinomialNB. I studied about the differences of these three classifiers for the homework and the appropriate usage and differences of each classifier. As I need to compare the result, I have to recall my memory and search again for the precision, recall, F1sore, False Negative and False Positive which I always forget.

While trying to run for the code given, I found out that the MultinomialNB cannot run the data from fraud detection. Upon further research, I have learned that MultinomialNB is designed for count data and it cannot handle negative values. So, I have transformed the data using minmax scaler. For the Bernoulli model it required binary data but since I could run the code without error, I look for reason and discovered that automatically binarize continuous or count features by default but this sometimes affects precision and recall when using BernoulliNB on non-binary data.

Based on the results from the code, I found that GaussianNB performed best for fraud detection because it achieved a high recall, allowing most fraud cases to be detected, while maintaining a reasonable number of false positives, which is important in fraud detection cases.