**Part 1 : Digit Recognition**

**a) Status and stopping point-**

In the first dataset for optical recognition of digits, the maximum accuracy percentage is around ~88.9% which can be considered as an acceptable value. Multinomial Naive Bayes Classifier is used in this case. Based on the varying values for alpha parameter, the accuracy varies. Different values of alpha values has been tested which varies from 0.005 to 1 but the prediction accuracy is not very much affected by it. That could be considered as the stopping point.

**b) Additional functions and analysis:**

For digit recognition multinomial naive bayes seems to be an effective model which produces almost 88.9% of the accuracy. It is followed by Bernoulli naive bayes and then Gaussian Naive Bayes.

**c)** The most challenging part is the data formation which followed by the accuracy calculation. When the parameters is changed with varied values and the prediction percentage doesn’t change much.

**Part 2: Amazon Data Set**

1. Status and stopping point -

With Amazon Data Set, the maximum accuracy percentage that I get is around ~65.5%. Which is less compared to other classifiers used so far. Even after changing values of alpha. This classifier is slower and accuracy wasn’t that great either.After changing the value of alpha upto a lot of extent the accuracy is still the same. The accuracy is just changing by a tiny factor.

1. Additional functions and analysis -

After analysing data I realized that it's necessary to clean the data as data set had rows with no reviews. To get rid of those rows I used dropna() method on data frame.

Then review itself had some html markup words. To counter that I used Beautiful soup. And then in order to get rid of stop words such as ‘is’,’and’ I used nltk package.

If I don’t perform all these operations the accuracy of the code is lower than 54%.

1. Challenges -

Amazon data set is really huge and playing with such huge data was challenging for me. Also figuring out suitable values for alpha is a challenge, as various combinations give almost similar accuracy.