**ANALYIZATION**

After Analyzation for different values of K from 1 to 10, There is Partial increase and Decrease of Accuracy Values.

training data points: 1968
validation data points: 189
testing data points: 540
k=1, accuracy=97.35%
k=3, accuracy=96.83%
k=5

The KNN smooths out the data as K grows. This is because a larger K value reduces the edginess of the model by including more data, lessening the model's overall complexity and adaptability.

As can be seen in the output file, increasing the value of K raises the score until a certain point, after which it begins to fall again.

When splitting a data set there are two competing concerns.

1. Your parameter estimate will have a lot of variance if you just have a little training data, and your performance statistic will have a lot of volatility if you only have a little testing data.
2. The Data should be divided in so that neither of them is excessively high, which is more dependent on the amount of data you have.
3. If your data is too small then no split will give you satisfactory variance so you will have to do cross validation but if your data is too high this doesn’t matter whether you choose 80:20 or 90:10, split.