Assignement-7

**Introduction**

K-Nearest Neighbors is one of the most basic classification algorithms in Machine Learning. KNN can be used for both classification and regression predictive However, it is more widely used in classification problems in the industry. To evaluate any technique, we generally look at 3 important aspects:

1. Ease to interpret output

2. Calculation time

3. Predictive Power

**KNN Implementation steps**

1. Load the data and split data into Train and Test datasets
2. Chosen k number of neighbors

3. Calculate the distance between two data instances

4. Sort the ordered collection of distances and indices from smallest to largest by the distances

5. Pick the first K entries from the sorted collection

6. Get the labels of the selected K entries

7. If classification, return the mode of the K labels

8. Summarize the Accuracy of prediction values

**Problems Encountered**:

1. One problem choosing the best distance calculation methods because there are four types

* Manhattan distance
* Euclidean distance
* Hamming Distance
* Minkowski Distance

1. Other problem is Choosing best K value is quite difficult.
2. To select the best K-Value, we run the KNN algorithm several times with different values of K and choose the best K-Value that reduces the number of errors we encounter while maintaining the algorithm’s ability to accurately make predictions on prediction data.

We need to choose best K-value based on the accuracy of the data prediction. In this I uses K-folds Cross-validation method and also used some predefined functions like cross\_val\_score, accuracy\_score in the Sklearn library.From this predefined functions I calculated the best K-value and its efficiency. Based on my train data and test data best K-vaue is 7 and its efficiency is 0.98.

