# K-Nearest Neighbor Algorithm

Problem Statement:

Prepare a model for glass classification using KNN

Data Description:

RI : refractive index

Na: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)

Mg: Magnesium

AI: Aluminum

Si: Silicon

K:Potassium

Ca: Calcium

Ba: Barium

Fe: Iron

Type: Type of glass: (class attribute)

1 -- building\_windows\_float\_processed

2 --building\_windows\_non\_float\_processed

3 --vehicle\_windows\_float\_processed

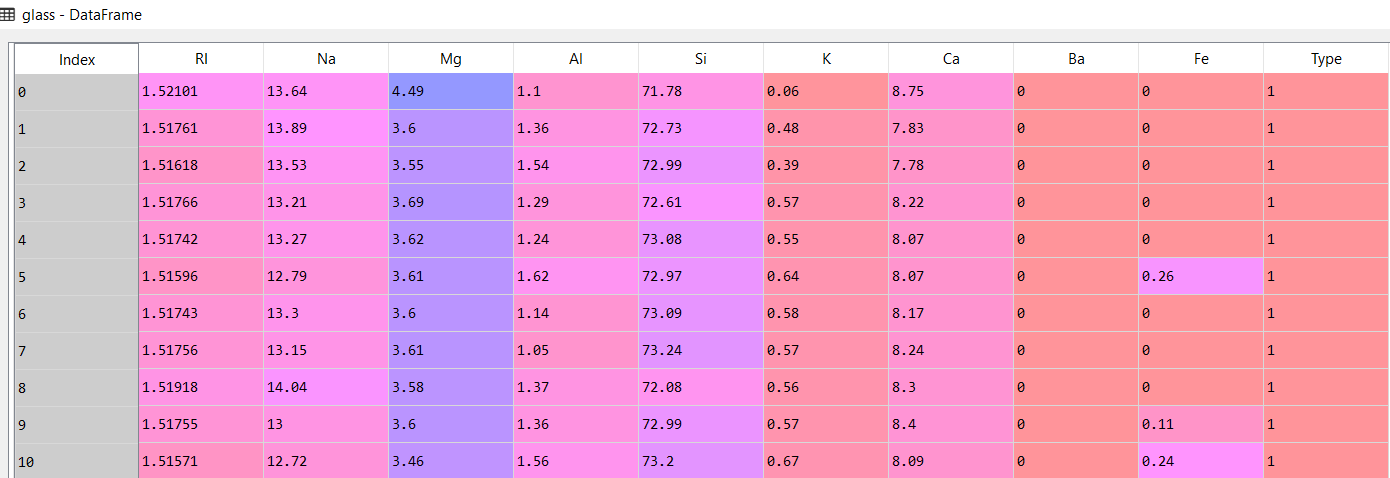
4 --vehicle\_windows\_non\_float\_processed (none in this database)

5 --containers

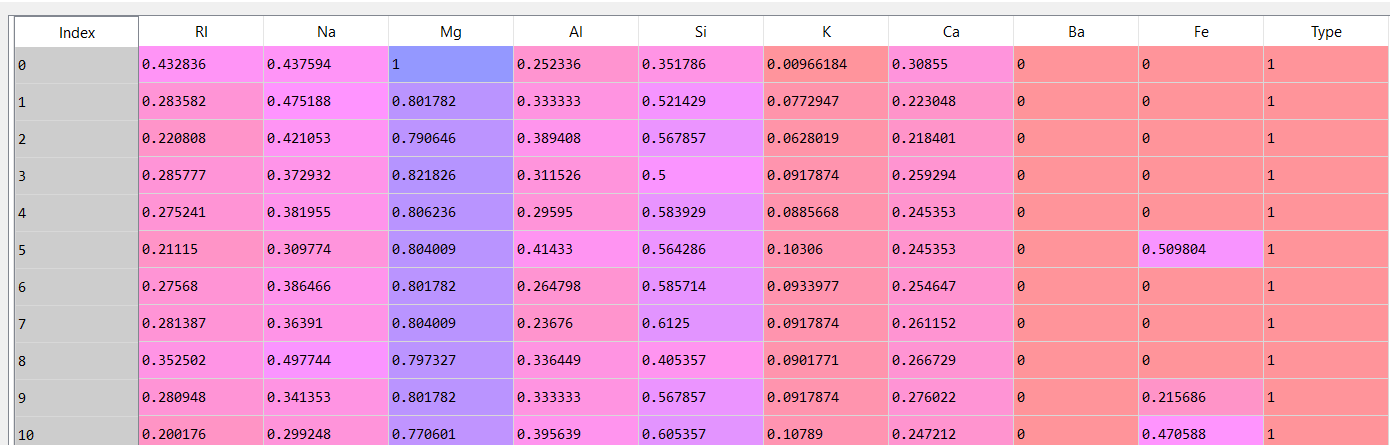
6 --tableware

7 –headlamps

The following is the dataset of glass dataframe:



In the above dataframe the features are of different scales. To resolve this problem we normalize the data as follows:



We then split the data into test and train datasets

Train data:



Test data:



We execute the algorithm for 3 nearest neighbors

The following are the results:

Train accuracy: 0.8245614035087719

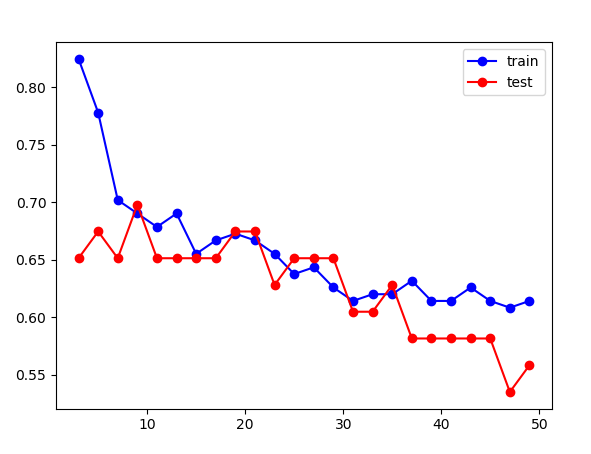
Test accuracy: 0.6511627906976745

We execute the algorithm for 5 nearest neighbors

Train accuracy: 0.7777777777777778

Test accuracy: 0.6744186046511628

By running KNN algorithm for 3 to 50 nearest neighbours(odd numbers) and storing the accuracy values. We get a plot as follows:



In the above plot the accuracy of the test and train data when K = 9 is about 69%. The test data accuracy seem to be hovering around 69% for the initial set of K values but he accuracy of the train data drastically falls beyond K = 9