## KNN Classification

## Problem Statement

Need to engineer dataset so as to be able to predict quality with a high accuracy, similar to when predicting type.

## Duration of Training

### Model 1:

Predicting for **Type** & **Quality**. Using binning only for Wine Type, with Red = 0, White = 1.

### Model 2:

Predicting for **Type** & **Quality**. Using binning for Wine Type (Red = 0, White = 1) & Quality Score (Low score of 0 to 4 =0, Medium score of 4 to 7 = 1, High score of 7 to 10 = 2)

### Model 3:

Predicting for **Quality**. Using dimension reduction through PCA to select features for x input, using the same binning techniques from Model 2

## Shared Process for all Models

### Data Cleaning

<image of orginal wine.head()>

Drop rows with null values,

Dataset

## Results

Model 1:

**Predicting Type (Red = 0, White = 1)**  
**\* quality column dropped**

print(accuracy\_score(y\_test, y\_pred))

0.9344059405940595

**Predicting Quality (with type Red = 0, White = 1)**

print(accuracy\_score(Y2\_test, Y2\_pred))

0.4542079207920792

**KNN – Test Data Accuracy**

from sklearn.neighbors import KNeighborsClassifier

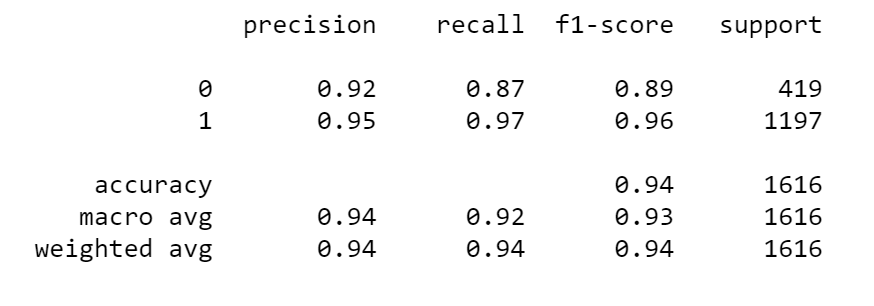
from sklearn.metrics import classification\_report,accuracy\_score

knn = KNeighborsClassifier()

knn.fit(X\_train,y\_train)

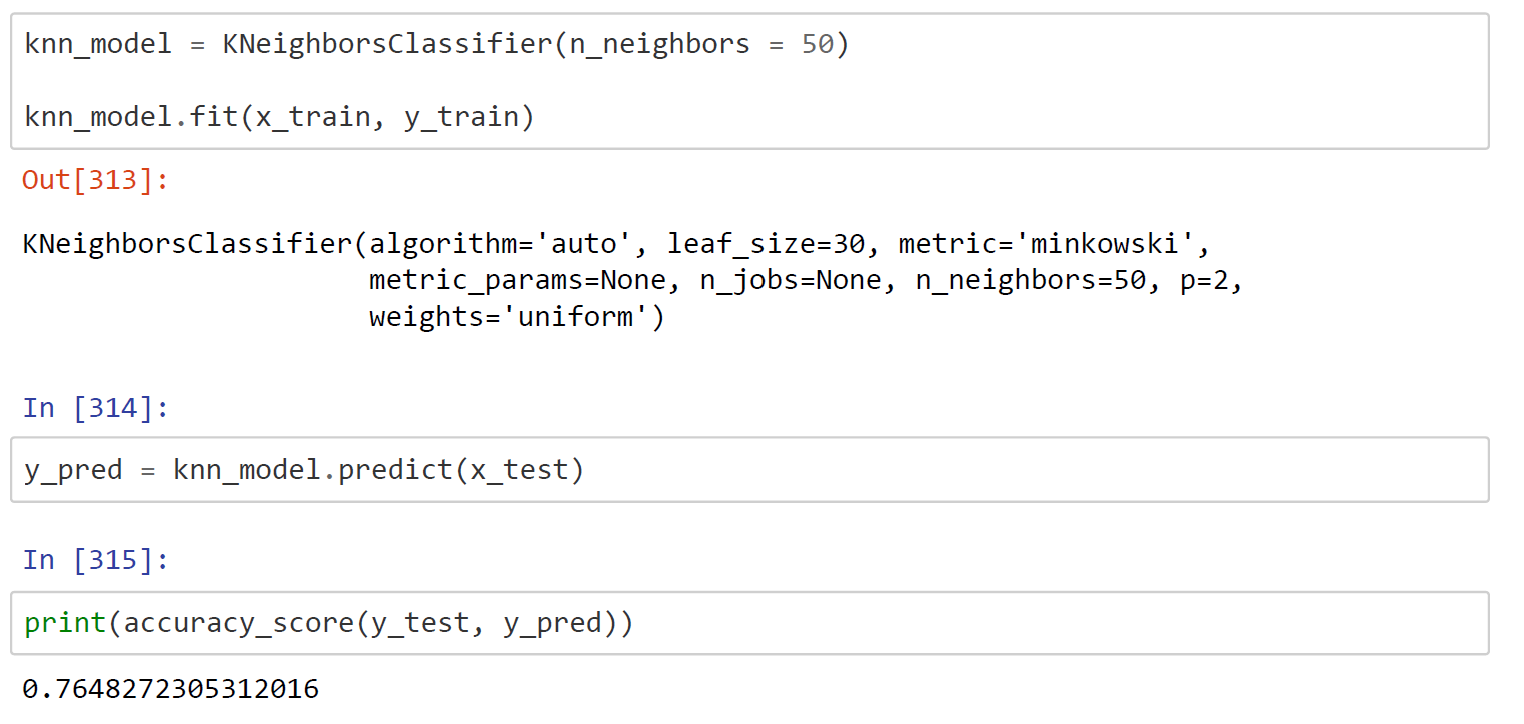
pred\_knn=knn.predict(X\_test)

print(classification\_report(y\_test, pred\_knn))



Model 2:

Model 3:



Observations

Despite reducing dimensions by containing x input to only 5 features,