**K- Nearest Neighbour**

**Business Problem:** Prepare a model for glass classification using KNN

**Data:**

Note that the value of k effects the results, its ideal to test the model for different values of k for better results and there by a better model.

Glass Database contains 10 attributes including id. The response is glass type which has 7 discrete values.

### Attributes

* **Id**: 1 to 214 (removed from CSV file)
* **RI**: refractive index
* **Na**: Sodium (unit measurement: weight percent in corresponding oxide, as are attributes 4-10)
* **Mg**: Magnesium
* **Al**: Aluminum
* **Si**: Silicon
* **K**: Potassium
* **Ca**: Calcium
* **Ba**: Barium
* **Fe**: Iron
* **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

**Analysis:**

* It’s doesn’t have NA’s. Data are in cleaned format. So I’m proceeding further.
* Data is not in same scale. So standardize the data using scale function.
* To Test properly, split the data into Training and Test.

**Build the K-NN Model:**

predicted.type <- knn(train[1:9],test[1:9],train$Type,k=1)

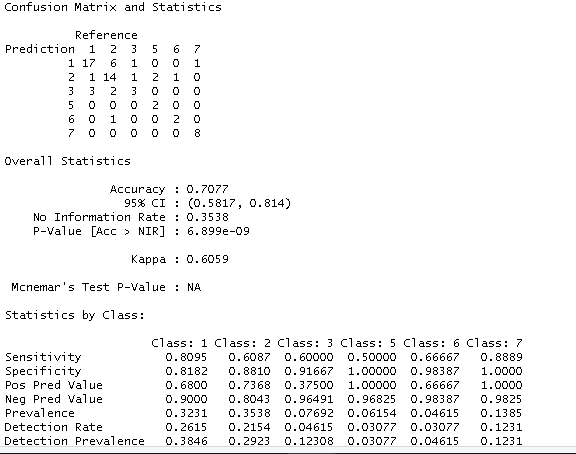
#Error in prediction

error <- mean(predicted.type!=test$Type)

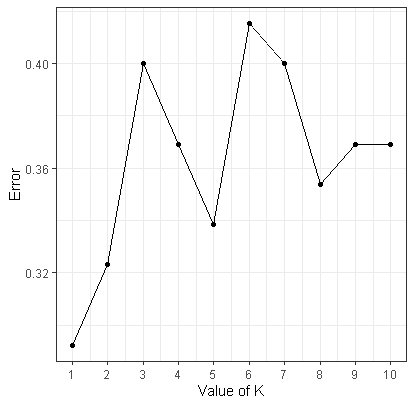
#Confusion Matrix

confusionMatrix(predicted.type,test$Type)

**Output:**



**Choosing K value using Visualization:**

****

By looking the Plot we can go with K value as 3.

**Building the module by using K value as 3 :**

predicted.type <- knn(train[1:9],test[1:9],train$Type,k=3)

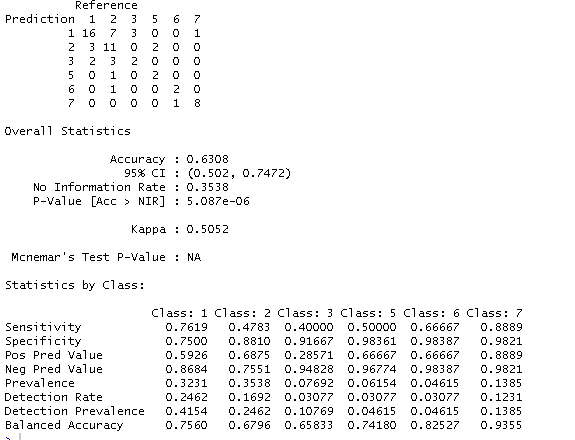
#Error in prediction

error <- mean(predicted.type!=test$Type)

#Confusion Matrix

confusionMatrix(predicted.type,test$Type)

**Output:**



**Conclusion:** **K = 3** is the better value for Glass database.

**Business Problem:** Implement a KNN model to classify the animals into categories.

**Data:**

This dataset consists of 101 animals from a zoo. There are 18 variables with various traits to describe the animals. The 7 Class Types are: Mammal, Bird, Reptile, Fish, Amphibian, Bug and Invertebrate.

The purpose for this dataset is to be able to predict the classification of the animals, based upon the variables.

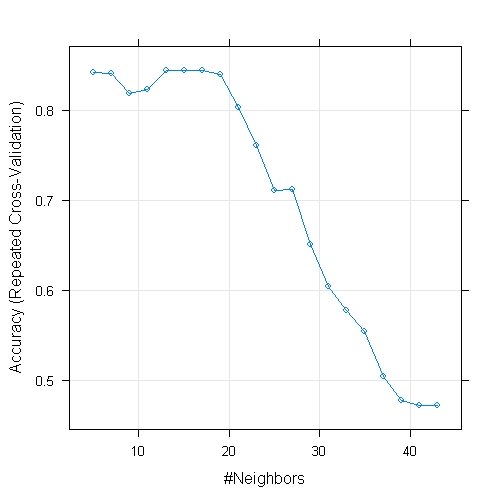
**Animal feature and Type of data:**

* hair Boolean
* feathers Boolean
* eggs Boolean
* milk Boolean
* airborne Boolean
* aquatic Boolean
* predator Boolean
* toothed Boolean
* backbone Boolean
* breathes Boolean
* venomous Boolean
* fins Boolean
* legs Numeric (set of values: {0,2,4,5,6,8})
* tail Boolean
* domestic Boolean
* catsize Boolean
* class\_type Numeric

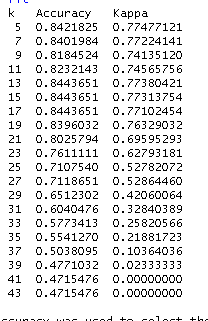
**Analysis:**

* It’s doesn’t have NA’s. Data are in cleaned format. So I’m proceeding further.
* Data is not in same scale. So standardize the data using scale function.
* To Test properly, split the data into Training and Test.

**Data visualization:**

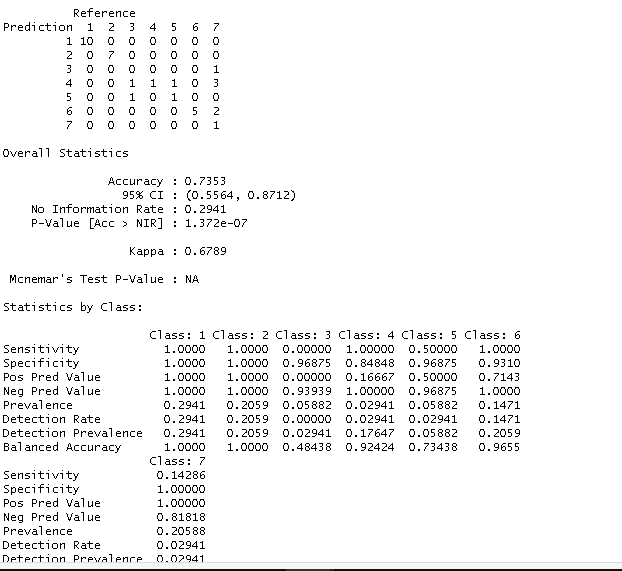


**Accuracy for different K values:**

****

By looking the above data **k = 5** will be good value.

**Output:**

****

**Conclusion:** **K = 5** is the better value for Zoo database.