Prepare a model for glass classification using KNN

**Ans :**

**R Code :**

## KNN

########## Glass Data Set #########

glass\_r <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\KNN\\glass\_r.csv')

# table of diagnosis

table(glass\_r$Type)

# table or proportions with more informative labels

round(prop.table(table(glass\_r$Type)) \* 100, digits = 1)

# create normalization function

normalize <- function(x) {

return ((x - min(x)) / (max(x) - min(x)))

}

# normalize the wbcd data

glass\_n <- as.data.frame(lapply(glass\_r[1:9], normalize))

glass\_nl<-cbind(glass\_n,glass\_r$Type)

# create training and test data

glass\_train <- glass\_n[1:150, ]

glass\_test <- glass\_n[151:214, ]

# create labels for training and test data

glass\_train\_labels <- glass\_r[1:150, 10]

glass\_test\_labels <- glass\_r[151:214, 10]

#---- Training a model on the data ----

# load the "class" library

library(class)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test,

cl = glass\_train\_labels, k=20)

##--------Evaluating model performance ----

# load the "gmodels" library

library(gmodels)

# Create the cross tabulation of predicted vs. actual

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred,

prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

#############################################################

## Improving model performance ----

# use the scale() function to z-score standardize a data frame

glass\_z <- as.data.frame(scale(glass\_r[-10]))

# create training and test datasets

glass\_train <- glass\_z[1:150, ]

glass\_test <- glass\_z[151:214, ]

# re-classify test cases

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test,

cl = glass\_train\_labels, k=21)

# Create the cross tabulation of predicted vs. actual

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred,

prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

# try several different values of k

glass\_train <- glass\_n[1:150, ]

glass\_test <- glass\_n[151:214, ]

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=1)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=5)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=11)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=15)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=21)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=27)

CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

**Results :**

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test,

+ cl = glass\_train\_labels, k=20)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred,

+ prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 5 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|

1 | 19 | 2 | 0 | 0 | 21 |

| 0.297 | 0.031 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

2 | 8 | 14 | 1 | 0 | 23 |

| 0.125 | 0.219 | 0.016 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

3 | 0 | 5 | 0 | 0 | 5 |

| 0.000 | 0.078 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 1 | 2 | 1 | 4 |

| 0.000 | 0.016 | 0.031 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 0 | 0 | 3 | 3 |

| 0.000 | 0.000 | 0.000 | 0.047 | |

------------------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 27 | 22 | 3 | 12 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test,

+ cl = glass\_train\_labels, k=21)

> # Create the cross tabulation of predicted vs. actual

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred,

+ prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 5 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|

1 | 21 | 0 | 0 | 0 | 21 |

| 0.328 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

2 | 13 | 10 | 0 | 0 | 23 |

| 0.203 | 0.156 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

3 | 1 | 4 | 0 | 0 | 5 |

| 0.016 | 0.062 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 3 | 1 | 0 | 4 |

| 0.000 | 0.047 | 0.016 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

6 | 1 | 0 | 0 | 2 | 3 |

| 0.016 | 0.000 | 0.000 | 0.031 | |

------------------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 36 | 17 | 1 | 10 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=1)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 3 | 5 | 6 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

1 | 12 | 6 | 3 | 0 | 0 | 0 | 21 |

| 0.188 | 0.094 | 0.047 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

2 | 7 | 11 | 1 | 2 | 1 | 1 | 23 |

| 0.109 | 0.172 | 0.016 | 0.031 | 0.016 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

3 | 0 | 5 | 0 | 0 | 0 | 0 | 5 |

| 0.000 | 0.078 | 0.000 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 2 | 0 | 2 | 0 | 0 | 4 |

| 0.000 | 0.031 | 0.000 | 0.031 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 1 | 0 | 1 | 1 | 0 | 3 |

| 0.000 | 0.016 | 0.000 | 0.016 | 0.016 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Column Total | 19 | 25 | 4 | 5 | 2 | 9 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=5)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 3 | 5 | 6 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

1 | 14 | 6 | 1 | 0 | 0 | 0 | 21 |

| 0.219 | 0.094 | 0.016 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

2 | 5 | 14 | 0 | 1 | 3 | 0 | 23 |

| 0.078 | 0.219 | 0.000 | 0.016 | 0.047 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

3 | 3 | 2 | 0 | 0 | 0 | 0 | 5 |

| 0.047 | 0.031 | 0.000 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 1 | 0 | 2 | 0 | 1 | 4 |

| 0.000 | 0.016 | 0.000 | 0.031 | 0.000 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 1 | 0 | 1 | 0 | 1 | 3 |

| 0.000 | 0.016 | 0.000 | 0.016 | 0.000 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

Column Total | 22 | 24 | 1 | 4 | 3 | 10 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=11)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 5 | 6 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

1 | 19 | 2 | 0 | 0 | 0 | 21 |

| 0.297 | 0.031 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

2 | 11 | 8 | 0 | 4 | 0 | 23 |

| 0.172 | 0.125 | 0.000 | 0.062 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

3 | 2 | 3 | 0 | 0 | 0 | 5 |

| 0.031 | 0.047 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 2 | 1 | 0 | 1 | 4 |

| 0.000 | 0.031 | 0.016 | 0.000 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 0 | 1 | 0 | 2 | 3 |

| 0.000 | 0.000 | 0.016 | 0.000 | 0.031 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

Column Total | 32 | 15 | 2 | 4 | 11 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=15)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 5 | 6 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

1 | 21 | 0 | 0 | 0 | 0 | 21 |

| 0.328 | 0.000 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

2 | 11 | 10 | 1 | 1 | 0 | 23 |

| 0.172 | 0.156 | 0.016 | 0.016 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

3 | 1 | 4 | 0 | 0 | 0 | 5 |

| 0.016 | 0.062 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 1 | 2 | 0 | 1 | 4 |

| 0.000 | 0.016 | 0.031 | 0.000 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 0 | 1 | 0 | 2 | 3 |

| 0.000 | 0.000 | 0.016 | 0.000 | 0.031 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

Column Total | 33 | 15 | 4 | 1 | 11 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=21)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 5 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|-----------|

1 | 21 | 0 | 0 | 0 | 21 |

| 0.328 | 0.000 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

2 | 6 | 16 | 1 | 0 | 23 |

| 0.094 | 0.250 | 0.016 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

3 | 1 | 4 | 0 | 0 | 5 |

| 0.016 | 0.062 | 0.000 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|-----------|

5 | 0 | 1 | 2 | 1 | 4 |

| 0.000 | 0.016 | 0.031 | 0.016 | |

------------------|-----------|-----------|-----------|-----------|-----------|

6 | 0 | 0 | 0 | 3 | 3 |

| 0.000 | 0.000 | 0.000 | 0.047 | |

------------------|-----------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|-----------|

Column Total | 28 | 21 | 3 | 12 | 64 |

------------------|-----------|-----------|-----------|-----------|-----------|

> glass\_test\_pred <- knn(train = glass\_train, test = glass\_test, cl = glass\_train\_labels, k=27)

> CrossTable(x = glass\_test\_labels, y = glass\_test\_pred, prop.chisq=FALSE,prop.c = FALSE, prop.r = FALSE)

Cell Contents

|-------------------------|

| N |

| N / Table Total |

|-------------------------|

Total Observations in Table: 64

| glass\_test\_pred

glass\_test\_labels | 1 | 2 | 7 | Row Total |

------------------|-----------|-----------|-----------|-----------|

1 | 19 | 2 | 0 | 21 |

| 0.297 | 0.031 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|

2 | 7 | 16 | 0 | 23 |

| 0.109 | 0.250 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|

3 | 1 | 4 | 0 | 5 |

| 0.016 | 0.062 | 0.000 | |

------------------|-----------|-----------|-----------|-----------|

5 | 0 | 1 | 3 | 4 |

| 0.000 | 0.016 | 0.047 | |

------------------|-----------|-----------|-----------|-----------|

6 | 0 | 0 | 3 | 3 |

| 0.000 | 0.000 | 0.047 | |

------------------|-----------|-----------|-----------|-----------|

7 | 0 | 0 | 8 | 8 |

| 0.000 | 0.000 | 0.125 | |

------------------|-----------|-----------|-----------|-----------|

Column Total | 27 | 23 | 14 | 64 |

------------------|-----------|-----------|-----------|-----------|

**Inference :**

We have got more accuracy when k=5.