```
library(readxl)

UB <- read_excel("UniversalBank.xlsx",sheet=2)

UB <- UB[-1]

UB <- UB[-4]

table(UB$`Personal Loan`)

   0    1 
4520  480 


newrow <- c(40,10,84,2,2,2,0,1,0,0,1,1)

UB <- rbind(UB,newrow)

normalized <- function(x){

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

}

UB_N <-as.data.frame(lapply(UB[-8],normalized))

UB_train <- UB_N[1:3000,]

UB_validation <- UB_N[3001:5000,]

UB_C <- UB_N[5001:5001,]

UBTL <- UB[1:3000,]

UBVL <- UB[3001:5000,]

UBCL <- UB[5001:5001,]

UB_train_labels <- UBTL$`Personal Loan`

UB_validation_labels <- UBVL$`Personal Loan`

UB_C_labels<- UBCL$`Personal Loan`

library(class)

UB_validation_pred <- knn(train = UB_train, test= UB_C, cl=UB_train_labels, k=1)

library(gmodels)

CrossTable(x=UB_C_labels, y=UB_validation_pred, prop.chisq = FALSE)

UB_Validation_pred2 <- knn(train = UB_train, test= UB_validation, cl=UB_train_labels,k=5)

CrossTable(x=UB_validation_labels, y=UB_Validation_pred2, prop.chisq = FALSE)
```

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  2000

                     | UB_Validation_pred2
UB_validation_labels |         0 |         1 | Row Total |
---------------------|-----------|-----------|-----------|
                   0 |      1821 |         8 |      1829 |
                     |     0.996 |     0.004 |     0.914 |
                     |     0.954 |     0.088 |           |
                     |     0.910 |     0.004 |           |
---------------------|-----------|-----------|-----------|
                   1 |        88 |        83 |       171 |
                     |     0.515 |     0.485 |     0.086 |
                     |     0.046 |     0.912 |           |
                     |     0.044 |     0.042 |           |
---------------------|-----------|-----------|-----------|
        Column Total |      1909 |        91 |      2000 |
                     |     0.955 |     0.045 |           |
---------------------|-----------|-----------|-----------|
```

K values	False positive	False negative	Total error in validation set
1	56	25	81
2	71	31	102
3	79	12	91
4	83	11	94
5	88	8	96
6	90	7	97
7	99	7	106
8	99	5	104
9	104	7	111
10	107	7	114

```
UB_validation_pred3 <- knn(train = UB_train, test= UB_C, cl=UB_train_labels, k=5)

UB_train2 <- UB_N[1:2500,]

UB_validation2 <- UB_N[2501:4000,]

UB_test <- UB_N[4001:5000,]

UBTL2 <- UB[1:2500,]
```

```
UBVL2 <- UB[2501:4000,]

UBTT <- UB[4001:5000,]

UB_train_labels2 <- UBTL2$`Personal Loan`

UB_validation_labels2 <- UBVL2$`Personal Loan`

UB_test_labels <- UBTT$`Personal Loan`

UB_validation_pred4 <- knn(train= UB_train2, test=UB_validation2, cl= UB_train_labels2,k=5)
```

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  1500

                         | UB_validation_pred4
UB_validation_labels2    |         0 |         1 | Row Total |
-------------------------|-----------|-----------|-----------|
                       0 |      1353 |         6 |      1359 |
                         |     0.996 |     0.004 |     0.906 |
                         |     0.947 |     0.085 |           |
                         |     0.902 |     0.004 |           |
-------------------------|-----------|-----------|-----------|
                       1 |        76 |        65 |       141 |
                         |     0.539 |     0.461 |     0.094 |
                         |     0.053 |     0.915 |           |
                         |     0.051 |     0.043 |           |
-------------------------|-----------|-----------|-----------|
            Column Total |      1429 |        71 |      1500 |
                         |     0.953 |     0.047 |           |
-------------------------|-----------|-----------|-----------|
```

```
CrossTable(x=UB_validation_labels2, y=UB_validation_pred4, prop.chisq = FALSE)

UB_test_pred5 <- knn(train= UB_train2, test=UB_test, cl= UB_train_labels2,k=5)

CrossTable(x=UB_test_labels, y=UB_test_pred5, prop.chisq = FALSE)
```

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
```

```
|          N / Table Total |
|--------------------------|

Total Observations in Table:  1000

                 | UB_test_pred5
UB_test_labels   |         0 |         1 | Row Total |
-----------------|-----------|-----------|-----------|
               0 |       915 |         2 |       917 |
                 |     0.998 |     0.002 |     0.917 |
                 |     0.954 |     0.049 |           |
                 |     0.915 |     0.002 |           |
-----------------|-----------|-----------|-----------|
               1 |        44 |        39 |        83 |
                 |     0.530 |     0.470 |     0.083 |
                 |     0.046 |     0.951 |           |
                 |     0.044 |     0.039 |           |
-----------------|-----------|-----------|-----------|
    Column Total |       959 |        41 |      1000 |
                 |     0.959 |     0.041 |           |
-----------------|-----------|-----------|-----------|
```

UB_train_pred6 <- knn(train= UB_train2, test=UB_train2, cl=UB_train_labels2,k=5)

CrossTable(UB_train_labels2, y=UB_train_pred6, prop.chisq = FALSE)

```
   Cell Contents
|--------------------------|
|                        N |
|            N / Row Total |
|            N / Col Total |
|          N / Table Total |
|--------------------------|

Total Observations in Table:  2500

                  | UB_train_pred6
UB_train_labels2  |         0 |         1 | Row Total |
------------------|-----------|-----------|-----------|
                0 |      2240 |         4 |      2244 |
                  |     0.998 |     0.002 |     0.898 |
                  |     0.967 |     0.022 |           |
                  |     0.896 |     0.002 |           |
------------------|-----------|-----------|-----------|
                1 |        77 |       179 |       256 |
                  |     0.301 |     0.699 |     0.102 |
                  |     0.033 |     0.978 |           |
                  |     0.031 |     0.072 |           |
------------------|-----------|-----------|-----------|
    Column Total  |      2317 |       183 |      2500 |
                  |     0.927 |     0.073 |           |
------------------|-----------|-----------|-----------|
```

Interpretation:- a) This customer would be classified as 0 (loan rejected)

b) We compared the error rate for k=1 to k=10 there is a sharp dip in error rate around k=6 so we take k=5 as the lower the value of k will have better sensitivity to local characteristics and we don't go lower to prevent overfitting.

d). The customer is classified as 0 again (loan rejected)

e) the training set has 3.24% error rate, test set has 4.6% and validation set has 5.4% error rate.

The reason for less error in training set is because the model got trained on the training set, so it has better prediction than the validation set, and the validation set has more error than the training set, the error in the test set is reduced because from multiple models we select the best model in validation set and finally the improved model gets used. Here as we compare both as the test set the results might be random for the validation set and the test set.

```
library(readxl)

BH <- read_xlsx("BostonHousing.xlsx",sheet=2)

normalized <- function(x){

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

}


BH_N <-as.data.frame(lapply(BH,normalized))

BH_train <- BH_N[1:304,]

BH_validation <- BH_N[305:506,]

BHS <- read_xlsx("BostonHousing.xlsx",sheet=2)

BHTL <- BHS[1:304,]

BHVL <- BHS[305:506,]

BH_train_labels <- BHTL$`CAT. MEDV`

BH_validation_labels <- BHVL$`CAT. MEDV`


library(class)

BH_validation_pred1 <- knn(train = BH_train[,1:13], test = BH_validation[,1:13], cl=BH_train_labels,
k=1)

library(gmodels)

CrossTable(x=BH_validation_labels, y=BH_validation_pred1, prop.chisq = FALSE)


   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  202

                     | BH_validation_pred1
BH_validation_labels |         0 |         1 | Row Total |
---------------------|-----------|-----------|-----------|
```

```
                     0 |         192 |           0 |         192 |
                       |       1.000 |       0.000 |       0.950 |
                       |       0.985 |       0.000 |             |
                       |       0.950 |       0.000 |             |
---------------------|-----------|-----------|-----------|
                     1 |           3 |           7 |          10 |
                       |       0.300 |       0.700 |       0.050 |
                       |       0.015 |       1.000 |             |
                       |       0.015 |       0.035 |             |
---------------------|-----------|-----------|-----------|
          Column Total |         195 |           7 |         202 |
                       |       0.965 |       0.035 |             |
---------------------|-----------|-----------|-----------|
```

#Error = 3

BH_Validation_pred2 <- knn(train = BH_train[,1:13], test= BH_validation[,1:13], cl=BH_train_labels,k=2)

CrossTable(x=BH_validation_labels, y=BH_Validation_pred2, prop.chisq = FALSE)

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  202

                      | BH_Validation_pred2
BH_validation_labels |         0 |         1 | Row Total |
---------------------|-----------|-----------|-----------|
                     0 |         192 |           0 |         192 |
                       |       1.000 |       0.000 |       0.950 |
                       |       0.985 |       0.000 |             |
                       |       0.950 |       0.000 |             |
---------------------|-----------|-----------|-----------|
                     1 |           3 |           7 |          10 |
                       |       0.300 |       0.700 |       0.050 |
                       |       0.015 |       1.000 |             |
                       |       0.015 |       0.035 |             |
---------------------|-----------|-----------|-----------|
          Column Total |         195 |           7 |         202 |
                       |       0.965 |       0.035 |             |
---------------------|-----------|-----------|-----------|
```

#Error = 3

BH_Validation_pred3 <- knn(train = BH_train[,1:13], test= BH_validation[,1:13], cl=BH_train_labels,k=3)

CrossTable(x=BH_validation_labels, y=BH_Validation_pred3, prop.chisq = FALSE)

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  202

```

BH_validation_labels	BH_Validation_pred3 = 0	BH_Validation_pred3 = 1	Row Total
0	192	0	192
	1.000	0.000	0.950
	0.990	0.000	
	0.950	0.000	
1	2	8	10
	0.200	0.800	0.050
	0.010	1.000	
	0.010	0.040	
Column Total	194	8	202
	0.960	0.040	

#Error = 2

BH_Validation_pred4 <- knn(train = BH_train[,1:13], test= BH_validation[,1:13], cl=BH_train_labels,k=4)

CrossTable(x=BH_validation_labels, y=BH_Validation_pred4, prop.chisq = FALSE)

```
   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  202
```

```
                     | BH_Validation_pred4
BH_validation_labels |         0 |         1 | Row Total |
---------------------|-----------|-----------|-----------|
                   0 |       192 |         0 |       192 |
                     |     1.000 |     0.000 |     0.950 |
                     |     0.980 |     0.000 |           |
                     |     0.950 |     0.000 |           |
---------------------|-----------|-----------|-----------|
                   1 |         4 |         6 |        10 |
                     |     0.400 |     0.600 |     0.050 |
                     |     0.020 |     1.000 |           |
                     |     0.020 |     0.030 |           |
---------------------|-----------|-----------|-----------|
        Column Total |       196 |         6 |       202 |
                     |     0.970 |     0.030 |           |
---------------------|-----------|-----------|-----------|
```

#Error = 4

BH_Validation_pred5 <- knn(train = BH_train[,1:13], test= BH_validation[,1:13], cl=BH_train_labels,k=5)

CrossTable(x=BH_validation_labels, y=BH_Validation_pred5, prop.chisq = FALSE)

```
Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|

Total Observations in Table:  202

                     | BH_Validation_pred5
BH_validation_labels |         0 |         1 | Row Total |
---------------------|-----------|-----------|-----------|
                   0 |       192 |         0 |       192 |
                     |     1.000 |     0.000 |     0.950 |
                     |     0.970 |     0.000 |           |
                     |     0.950 |     0.000 |           |
---------------------|-----------|-----------|-----------|
                   1 |         6 |         4 |        10 |
                     |     0.600 |     0.400 |     0.050 |
                     |     0.030 |     1.000 |           |
                     |     0.030 |     0.020 |           |
---------------------|-----------|-----------|-----------|
        Column Total |       198 |         4 |       202 |
                     |     0.980 |     0.020 |           |
---------------------|-----------|-----------|-----------|
```

#Error = 6

#Best K = 3

```
BHnewrow <- c(0.2,0,7,0,0.538,6,62,4.7,4,307,21,10)

BHaddon <- rbind(BH,BHnewrow)

BH_N_ADDON <-as.data.frame(lapply(BHaddon[,1:13],normalized))

BH_ADDON <- BH_N_ADDON[507:507,]

BH_ADDONL <- BHaddon[507:507,]

BH_ADDON_train_labels <- BHTL$`MEDV`

BH_ADDON_lable <- BH_ADDONL$`MEDV`

BH_ADDON_validation_pred <- knn(train = BH_train[,1:12], test = BH_ADDON[,-13],
cl=BH_ADDON_train_labels, k=3)

BH_ADDON_validation_pred
```

```
BH_ADDON_validation_pred
[1] 20.4
177 Levels: 11.8 12.7 13.1 13.2 13.3 13.4 13.5 13.6 13.8 13.9 14 14.3 14.4 ... 50

>
```

Interpretation:-a) The best k is 3

b) The MEDV is 20.4

c) The model was trained on the training data and the data is well organised as it makes correct classification according to parameters. That's why the error of the training data is zero.

d) The data quality of the validation set is good and different models are used to see which model fits the validation set better . So the model selected may not be that accurate with new data as we don't know the quality of new data and if the selected model fits the new data as well.

e) K-NN is a lazy learner the time consuming computation is deferred to the time of prediction. For every record to be predicted we calculate the distances from the entire set of training records only at the time of prediction. So for a large number of records this becomes problematic.

Find the nearest neighbour or neighbors to the record to be predicted

The record is predicted as the average response value of k neighbours