FML\_Assignment2

phani varshitha

2023-10-01

library(class)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(caret)

## Loading required package: ggplot2

## Loading required package: lattice

#loading data set  
dataset\_ub<-read.csv("C:/Users/varshitha/Downloads/UniversalBank.csv")  
head(dataset\_ub)

## ID Age Experience Income ZIP.Code Family CCAvg Education Mortgage  
## 1 1 25 1 49 91107 4 1.6 1 0  
## 2 2 45 19 34 90089 3 1.5 1 0  
## 3 3 39 15 11 94720 1 1.0 1 0  
## 4 4 35 9 100 94112 1 2.7 2 0  
## 5 5 35 8 45 91330 4 1.0 2 0  
## 6 6 37 13 29 92121 4 0.4 2 155  
## Personal.Loan Securities.Account CD.Account Online CreditCard  
## 1 0 1 0 0 0  
## 2 0 1 0 0 0  
## 3 0 0 0 0 0  
## 4 0 0 0 0 0  
## 5 0 0 0 0 1  
## 6 0 0 0 1 0

#removing unwanted columns i.e ID and Zip code  
dataset\_ub1<-dataset\_ub[,-1]  
head(dataset\_ub1)

## Age Experience Income ZIP.Code Family CCAvg Education Mortgage Personal.Loan  
## 1 25 1 49 91107 4 1.6 1 0 0  
## 2 45 19 34 90089 3 1.5 1 0 0  
## 3 39 15 11 94720 1 1.0 1 0 0  
## 4 35 9 100 94112 1 2.7 2 0 0  
## 5 35 8 45 91330 4 1.0 2 0 0  
## 6 37 13 29 92121 4 0.4 2 155 0  
## Securities.Account CD.Account Online CreditCard  
## 1 1 0 0 0  
## 2 1 0 0 0  
## 3 0 0 0 0  
## 4 0 0 0 0  
## 5 0 0 0 1  
## 6 0 0 1 0

dataset\_ub1<-dataset\_ub1[,-4]  
head(dataset\_ub1)

## Age Experience Income Family CCAvg Education Mortgage Personal.Loan  
## 1 25 1 49 4 1.6 1 0 0  
## 2 45 19 34 3 1.5 1 0 0  
## 3 39 15 11 1 1.0 1 0 0  
## 4 35 9 100 1 2.7 2 0 0  
## 5 35 8 45 4 1.0 2 0 0  
## 6 37 13 29 4 0.4 2 155 0  
## Securities.Account CD.Account Online CreditCard  
## 1 1 0 0 0  
## 2 1 0 0 0  
## 3 0 0 0 0  
## 4 0 0 0 0  
## 5 0 0 0 1  
## 6 0 0 1 0

#converting personal loan as factor  
dataset\_ub1$Personal.Loan=as.factor(dataset\_ub1$Personal.Loan)

#running is.na to check if there are any NA values  
head(is.na(dataset\_ub1))

## Age Experience Income Family CCAvg Education Mortgage Personal.Loan  
## [1,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [2,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [3,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [4,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [5,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [6,] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## Securities.Account CD.Account Online CreditCard  
## [1,] FALSE FALSE FALSE FALSE  
## [2,] FALSE FALSE FALSE FALSE  
## [3,] FALSE FALSE FALSE FALSE  
## [4,] FALSE FALSE FALSE FALSE  
## [5,] FALSE FALSE FALSE FALSE  
## [6,] FALSE FALSE FALSE FALSE

# Converting categorical variable into i.e education into dummy variables  
  
#converting education into character  
education<-as.character(dataset\_ub1$Education)  
  
dataset\_ub2<-cbind(dataset\_ub1[,-6],education)  
head(dataset\_ub2)

## Age Experience Income Family CCAvg Mortgage Personal.Loan Securities.Account  
## 1 25 1 49 4 1.6 0 0 1  
## 2 45 19 34 3 1.5 0 0 1  
## 3 39 15 11 1 1.0 0 0 0  
## 4 35 9 100 1 2.7 0 0 0  
## 5 35 8 45 4 1.0 0 0 0  
## 6 37 13 29 4 0.4 155 0 0  
## CD.Account Online CreditCard education  
## 1 0 0 0 1  
## 2 0 0 0 1  
## 3 0 0 0 1  
## 4 0 0 0 2  
## 5 0 0 1 2  
## 6 0 1 0 2

dummymodel<-dummyVars("~education",data = dataset\_ub2)  
educationdummy<-data.frame(predict(dummymodel,dataset\_ub2))  
head(educationdummy)

## education1 education2 education3  
## 1 1 0 0  
## 2 1 0 0  
## 3 1 0 0  
## 4 0 1 0  
## 5 0 1 0  
## 6 0 1 0

dataset\_ub\_dummy<-cbind(dataset\_ub2[,-12],educationdummy)  
head(dataset\_ub\_dummy)

## Age Experience Income Family CCAvg Mortgage Personal.Loan Securities.Account  
## 1 25 1 49 4 1.6 0 0 1  
## 2 45 19 34 3 1.5 0 0 1  
## 3 39 15 11 1 1.0 0 0 0  
## 4 35 9 100 1 2.7 0 0 0  
## 5 35 8 45 4 1.0 0 0 0  
## 6 37 13 29 4 0.4 155 0 0  
## CD.Account Online CreditCard education1 education2 education3  
## 1 0 0 0 1 0 0  
## 2 0 0 0 1 0 0  
## 3 0 0 0 1 0 0  
## 4 0 0 0 0 1 0  
## 5 0 0 1 0 1 0  
## 6 0 1 0 0 1 0

#dividing data into training and testing set  
set.seed(555)  
train<-createDataPartition(dataset\_ub\_dummy$Personal.Loan,p=0.60,list = FALSE)  
trainset<-dataset\_ub\_dummy[train,]  
nrow(trainset)

## [1] 3000

validationset<-dataset\_ub\_dummy[-train,]  
nrow(validationset)

## [1] 2000

testset<-data.frame(Age = 40, Experience = 10, Income = 84, Family = 2, CCAvg = 2, Mortgage = 0, Securities.Account = 0, CD.Account = 0, Online = 1,   
 CreditCard = 1,education1 = 0, education2 = 1, education3 = 0)  
  
  
summary(trainset)

## Age Experience Income Family   
## Min. :23.00 Min. :-3.00 Min. : 8.00 Min. :1.000   
## 1st Qu.:35.00 1st Qu.:10.00 1st Qu.: 40.00 1st Qu.:1.000   
## Median :45.00 Median :20.00 Median : 65.00 Median :2.000   
## Mean :45.31 Mean :20.08 Mean : 74.81 Mean :2.382   
## 3rd Qu.:55.00 3rd Qu.:30.00 3rd Qu.:100.00 3rd Qu.:3.000   
## Max. :67.00 Max. :43.00 Max. :224.00 Max. :4.000   
## CCAvg Mortgage Personal.Loan Securities.Account  
## Min. : 0.000 Min. : 0.00 0:2712 Min. :0.0000   
## 1st Qu.: 0.700 1st Qu.: 0.00 1: 288 1st Qu.:0.0000   
## Median : 1.500 Median : 0.00 Median :0.0000   
## Mean : 1.946 Mean : 56.32 Mean :0.1067   
## 3rd Qu.: 2.600 3rd Qu.:101.00 3rd Qu.:0.0000   
## Max. :10.000 Max. :635.00 Max. :1.0000   
## CD.Account Online CreditCard education1   
## Min. :0.00000 Min. :0.0000 Min. :0.000 Min. :0.0000   
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.0000   
## Median :0.00000 Median :1.0000 Median :0.000 Median :0.0000   
## Mean :0.06167 Mean :0.5963 Mean :0.297 Mean :0.4267   
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.000 3rd Qu.:1.0000   
## Max. :1.00000 Max. :1.0000 Max. :1.000 Max. :1.0000   
## education2 education3   
## Min. :0.00 Min. :0.0000   
## 1st Qu.:0.00 1st Qu.:0.0000   
## Median :0.00 Median :0.0000   
## Mean :0.28 Mean :0.2933   
## 3rd Qu.:1.00 3rd Qu.:1.0000   
## Max. :1.00 Max. :1.0000

summary(validationset)

## Age Experience Income Family   
## Min. :23.00 Min. :-3.00 Min. : 8.00 Min. :1.000   
## 1st Qu.:35.00 1st Qu.:10.00 1st Qu.: 38.00 1st Qu.:1.000   
## Median :45.50 Median :20.00 Median : 62.00 Median :2.000   
## Mean :45.38 Mean :20.14 Mean : 72.22 Mean :2.418   
## 3rd Qu.:55.00 3rd Qu.:30.00 3rd Qu.: 94.00 3rd Qu.:4.000   
## Max. :67.00 Max. :43.00 Max. :205.00 Max. :4.000   
## CCAvg Mortgage Personal.Loan Securities.Account  
## Min. :0.000 Min. : 0.00 0:1808 Min. :0.000   
## 1st Qu.:0.700 1st Qu.: 0.00 1: 192 1st Qu.:0.000   
## Median :1.500 Median : 0.00 Median :0.000   
## Mean :1.925 Mean : 56.77 Mean :0.101   
## 3rd Qu.:2.500 3rd Qu.:101.00 3rd Qu.:0.000   
## Max. :9.300 Max. :617.00 Max. :1.000   
## CD.Account Online CreditCard education1   
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.000   
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.000   
## Median :0.0000 Median :1.0000 Median :0.0000 Median :0.000   
## Mean :0.0585 Mean :0.5975 Mean :0.2895 Mean :0.408   
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.000   
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.000   
## education2 education3   
## Min. :0.0000 Min. :0.0000   
## 1st Qu.:0.0000 1st Qu.:0.0000   
## Median :0.0000 Median :0.0000   
## Mean :0.2815 Mean :0.3105   
## 3rd Qu.:1.0000 3rd Qu.:1.0000   
## Max. :1.0000 Max. :1.0000

summary(testset)

## Age Experience Income Family CCAvg Mortgage  
## Min. :40 Min. :10 Min. :84 Min. :2 Min. :2 Min. :0   
## 1st Qu.:40 1st Qu.:10 1st Qu.:84 1st Qu.:2 1st Qu.:2 1st Qu.:0   
## Median :40 Median :10 Median :84 Median :2 Median :2 Median :0   
## Mean :40 Mean :10 Mean :84 Mean :2 Mean :2 Mean :0   
## 3rd Qu.:40 3rd Qu.:10 3rd Qu.:84 3rd Qu.:2 3rd Qu.:2 3rd Qu.:0   
## Max. :40 Max. :10 Max. :84 Max. :2 Max. :2 Max. :0   
## Securities.Account CD.Account Online CreditCard education1  
## Min. :0 Min. :0 Min. :1 Min. :1 Min. :0   
## 1st Qu.:0 1st Qu.:0 1st Qu.:1 1st Qu.:1 1st Qu.:0   
## Median :0 Median :0 Median :1 Median :1 Median :0   
## Mean :0 Mean :0 Mean :1 Mean :1 Mean :0   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:1 3rd Qu.:1 3rd Qu.:0   
## Max. :0 Max. :0 Max. :1 Max. :1 Max. :0   
## education2 education3  
## Min. :1 Min. :0   
## 1st Qu.:1 1st Qu.:0   
## Median :1 Median :0   
## Mean :1 Mean :0   
## 3rd Qu.:1 3rd Qu.:0   
## Max. :1 Max. :0

#normalizing  
  
normvar<-c('Age',"Experience","Income","Family","CCAvg","Mortgage","Securities.Account","CD.Account","Online","CreditCard","education1","education2","education3")  
normalization\_values<-preProcess(trainset[,normvar],method = c('center','scale'))  
  
trainset.norm<-predict(normalization\_values,trainset)  
summary(trainset.norm)

## Age Experience Income Family   
## Min. :-1.95104 Min. :-2.0186 Min. :-1.4431 Min. :-1.2107   
## 1st Qu.:-0.90159 1st Qu.:-0.8817 1st Qu.:-0.7519 1st Qu.:-1.2107   
## Median :-0.02705 Median :-0.0072 Median :-0.2119 Median :-0.3344   
## Mean : 0.00000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.84749 3rd Qu.: 0.8673 3rd Qu.: 0.5441 3rd Qu.: 0.5418   
## Max. : 1.89694 Max. : 2.0042 Max. : 3.2226 Max. : 1.4180   
## CCAvg Mortgage Personal.Loan Securities.Account  
## Min. :-1.0976 Min. :-0.5527 0:2712 Min. :-0.3455   
## 1st Qu.:-0.7028 1st Qu.:-0.5527 1: 288 1st Qu.:-0.3455   
## Median :-0.2517 Median :-0.5527 Median :-0.3455   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.3687 3rd Qu.: 0.4385 3rd Qu.:-0.3455   
## Max. : 4.5418 Max. : 5.6790 Max. : 2.8935   
## CD.Account Online CreditCard education1   
## Min. :-0.2563 Min. :-1.2152 Min. :-0.6499 Min. :-0.8625   
## 1st Qu.:-0.2563 1st Qu.:-1.2152 1st Qu.:-0.6499 1st Qu.:-0.8625   
## Median :-0.2563 Median : 0.8226 Median :-0.6499 Median :-0.8625   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.:-0.2563 3rd Qu.: 0.8226 3rd Qu.: 1.5383 3rd Qu.: 1.1590   
## Max. : 3.9001 Max. : 0.8226 Max. : 1.5383 Max. : 1.1590   
## education2 education3   
## Min. :-0.6235 Min. :-0.6442   
## 1st Qu.:-0.6235 1st Qu.:-0.6442   
## Median :-0.6235 Median :-0.6442   
## Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 1.6033 3rd Qu.: 1.5519   
## Max. : 1.6033 Max. : 1.5519

validationset.norm<-predict(normalization\_values,validationset)  
summary(validationset.norm)

## Age Experience Income Family   
## Min. :-1.951044 Min. :-2.018590 Min. :-1.44310 Min. :-1.21067   
## 1st Qu.:-0.901594 1st Qu.:-0.881718 1st Qu.:-0.79509 1st Qu.:-1.21067   
## Median : 0.016675 Median :-0.007200 Median :-0.27668 Median :-0.33443   
## Mean : 0.006355 Mean : 0.004868 Mean :-0.05588 Mean : 0.03227   
## 3rd Qu.: 0.847489 3rd Qu.: 0.867317 3rd Qu.: 0.41453 3rd Qu.: 1.41805   
## Max. : 1.896939 Max. : 2.004190 Max. : 2.81218 Max. : 1.41805   
## CCAvg Mortgage Personal.Loan Securities.Account  
## Min. :-1.09759 Min. :-0.552664 0:1808 Min. :-0.34549   
## 1st Qu.:-0.70283 1st Qu.:-0.552664 1: 192 1st Qu.:-0.34549   
## Median :-0.25168 Median :-0.552664 Median :-0.34549   
## Mean :-0.01177 Mean : 0.004477 Mean :-0.01835   
## 3rd Qu.: 0.31226 3rd Qu.: 0.438506 3rd Qu.:-0.34549   
## Max. : 4.14705 Max. : 5.502307 Max. : 2.89348   
## CD.Account Online CreditCard education1   
## Min. :-0.25632 Min. :-1.215236 Min. :-0.64987 Min. :-0.86252   
## 1st Qu.:-0.25632 1st Qu.:-1.215236 1st Qu.:-0.64987 1st Qu.:-0.86252   
## Median :-0.25632 Median : 0.822611 Median :-0.64987 Median :-0.86252   
## Mean :-0.01316 Mean : 0.002377 Mean :-0.01641 Mean :-0.03774   
## 3rd Qu.:-0.25632 3rd Qu.: 0.822611 3rd Qu.: 1.53825 3rd Qu.: 1.15901   
## Max. : 3.90015 Max. : 0.822611 Max. : 1.53825 Max. : 1.15901   
## education2 education3   
## Min. :-0.62351 Min. :-0.6442   
## 1st Qu.:-0.62351 1st Qu.:-0.6442   
## Median :-0.62351 Median :-0.6442   
## Mean : 0.00334 Mean : 0.0377   
## 3rd Qu.: 1.60330 3rd Qu.: 1.5519   
## Max. : 1.60330 Max. : 1.5519

testset.norm<-predict(normalization\_values,testset)  
summary(testset.norm)

## Age Experience Income Family   
## Min. :-0.4643 Min. :-0.8817 Min. :0.1985 Min. :-0.3344   
## 1st Qu.:-0.4643 1st Qu.:-0.8817 1st Qu.:0.1985 1st Qu.:-0.3344   
## Median :-0.4643 Median :-0.8817 Median :0.1985 Median :-0.3344   
## Mean :-0.4643 Mean :-0.8817 Mean :0.1985 Mean :-0.3344   
## 3rd Qu.:-0.4643 3rd Qu.:-0.8817 3rd Qu.:0.1985 3rd Qu.:-0.3344   
## Max. :-0.4643 Max. :-0.8817 Max. :0.1985 Max. :-0.3344   
## CCAvg Mortgage Securities.Account CD.Account   
## Min. :0.03029 Min. :-0.5527 Min. :-0.3455 Min. :-0.2563   
## 1st Qu.:0.03029 1st Qu.:-0.5527 1st Qu.:-0.3455 1st Qu.:-0.2563   
## Median :0.03029 Median :-0.5527 Median :-0.3455 Median :-0.2563   
## Mean :0.03029 Mean :-0.5527 Mean :-0.3455 Mean :-0.2563   
## 3rd Qu.:0.03029 3rd Qu.:-0.5527 3rd Qu.:-0.3455 3rd Qu.:-0.2563   
## Max. :0.03029 Max. :-0.5527 Max. :-0.3455 Max. :-0.2563   
## Online CreditCard education1 education2   
## Min. :0.8226 Min. :1.538 Min. :-0.8625 Min. :1.603   
## 1st Qu.:0.8226 1st Qu.:1.538 1st Qu.:-0.8625 1st Qu.:1.603   
## Median :0.8226 Median :1.538 Median :-0.8625 Median :1.603   
## Mean :0.8226 Mean :1.538 Mean :-0.8625 Mean :1.603   
## 3rd Qu.:0.8226 3rd Qu.:1.538 3rd Qu.:-0.8625 3rd Qu.:1.603   
## Max. :0.8226 Max. :1.538 Max. :-0.8625 Max. :1.603   
## education3   
## Min. :-0.6442   
## 1st Qu.:-0.6442   
## Median :-0.6442   
## Mean :-0.6442   
## 3rd Qu.:-0.6442   
## Max. :-0.6442

#question 1: Classifying the given customer  
set.seed(555)  
new\_grid<-expand.grid(k=c(1))  
new\_model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=new\_grid)  
  
new\_model

## k-Nearest Neighbors   
##   
## 3000 samples  
## 13 predictor  
## 2 classes: '0', '1'   
##   
## No pre-processing  
## Resampling: Bootstrapped (25 reps)   
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.9518741 0.6936177  
##   
## Tuning parameter 'k' was held constant at a value of 1

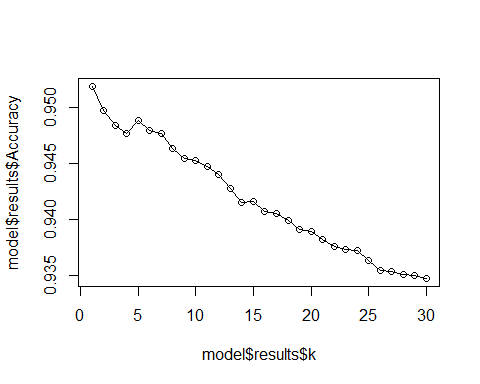
predict\_test<-predict(new\_model,testset.norm)  
predict\_test

## [1] 0  
## Levels: 0 1

#question 2: identifying the best k  
set.seed(555)  
searchGrid <- expand.grid(k=seq(1:30))  
model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=searchGrid)  
model

## k-Nearest Neighbors   
##   
## 3000 samples  
## 13 predictor  
## 2 classes: '0', '1'   
##   
## No pre-processing  
## Resampling: Bootstrapped (25 reps)   
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...   
## Resampling results across tuning parameters:  
##   
## k Accuracy Kappa   
## 1 0.9518741 0.6936177  
## 2 0.9497284 0.6783892  
## 3 0.9483786 0.6611715  
## 4 0.9476472 0.6493192  
## 5 0.9488503 0.6502041  
## 6 0.9479069 0.6389555  
## 7 0.9477101 0.6312418  
## 8 0.9463695 0.6188154  
## 9 0.9454200 0.6064940  
## 10 0.9452489 0.6023107  
## 11 0.9447388 0.5956424  
## 12 0.9439812 0.5885615  
## 13 0.9427742 0.5771545  
## 14 0.9415347 0.5630486  
## 15 0.9416088 0.5628185  
## 16 0.9407328 0.5548557  
## 17 0.9404893 0.5516391  
## 18 0.9399027 0.5455684  
## 19 0.9391046 0.5359012  
## 20 0.9389587 0.5339743  
## 21 0.9381946 0.5253688  
## 22 0.9375805 0.5184377  
## 23 0.9373295 0.5160644  
## 24 0.9372150 0.5151960  
## 25 0.9363069 0.5052569  
## 26 0.9354303 0.4956116  
## 27 0.9353960 0.4944564  
## 28 0.9350620 0.4895966  
## 29 0.9350298 0.4882462  
## 30 0.9347369 0.4839273  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was k = 1.

plot(model$results$k,model$results$Accuracy, type = 'o')



#finding the best k  
best\_k <- model$bestTune[[1]]  
best\_k

## [1] 1

#question3:confusion matrix  
library(gmodels)  
  
train\_label<-trainset.norm[,7]  
validation\_label<-validationset.norm[,7]  
test\_label<-testset.norm[,7]  
  
predicted\_validationlabel<-knn(trainset.norm,validationset.norm,cl=train\_label,k=5)  
  
CrossTable(x=validation\_label,y=predicted\_validationlabel,prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 2000   
##   
##   
## | predicted\_validationlabel   
## validation\_label | 0 | 1 | Row Total |   
## -----------------|-----------|-----------|-----------|  
## 0 | 1803 | 5 | 1808 |   
## | 0.997 | 0.003 | 0.904 |   
## | 0.971 | 0.035 | |   
## | 0.901 | 0.002 | |   
## -----------------|-----------|-----------|-----------|  
## 1 | 54 | 138 | 192 |   
## | 0.281 | 0.719 | 0.096 |   
## | 0.029 | 0.965 | |   
## | 0.027 | 0.069 | |   
## -----------------|-----------|-----------|-----------|  
## Column Total | 1857 | 143 | 2000 |   
## | 0.928 | 0.071 | |   
## -----------------|-----------|-----------|-----------|  
##   
##

#question4:Classifying the given customer with best k  
set.seed(789)  
bestk\_grid<-expand.grid(k=c(best\_k))  
bestk\_model<-train(Personal.Loan~.,data=trainset.norm,method="knn",tuneGrid=bestk\_grid)  
bestk\_model

## k-Nearest Neighbors   
##   
## 3000 samples  
## 13 predictor  
## 2 classes: '0', '1'   
##   
## No pre-processing  
## Resampling: Bootstrapped (25 reps)   
## Summary of sample sizes: 3000, 3000, 3000, 3000, 3000, 3000, ...   
## Resampling results:  
##   
## Accuracy Kappa   
## 0.951444 0.6896332  
##   
## Tuning parameter 'k' was held constant at a value of 1

bestk\_test<-predict(bestk\_model,testset.norm)  
bestk\_test

## [1] 0  
## Levels: 0 1

#question5:confusion matrix for validation and training sets  
#dividing dataset into traning, validation and testing set  
set.seed(555)  
train1<-createDataPartition(dataset\_ub\_dummy$Personal.Loan,p=0.50,list = FALSE)  
trainset\_2<-dataset\_ub\_dummy[train1,]  
middleset<-dataset\_ub\_dummy[-train1,]  
nrow(middleset)

## [1] 2500

train2<-createDataPartition(middleset$Personal.Loan,p=0.6,list = FALSE)  
validationset\_2<-middleset[train2,]  
testset\_2<-middleset[-train2,]  
  
nrow(trainset\_2)

## [1] 2500

nrow(validationset\_2)

## [1] 1500

nrow(testset\_2)

## [1] 1000

#normalizing trainset\_2,validationset\_2,testset\_2  
  
normvar<-c('Age',"Experience","Income","Family","CCAvg","Mortgage","Securities.Account","CD.Account","Online","CreditCard","education1","education2","education3")  
normalization\_values\_2<-preProcess(trainset\_2[,normvar],method = c('center','scale'))  
  
trainset.norm\_2<-predict(normalization\_values\_2,trainset\_2)  
summary(trainset.norm\_2)

## Age Experience Income Family   
## Min. :-1.93768 Min. :-2.009123 Min. :-1.4553 Min. :-1.2004   
## 1st Qu.:-0.89130 1st Qu.:-0.873828 1st Qu.:-0.7568 1st Qu.:-1.2004   
## Median :-0.01932 Median :-0.000524 Median :-0.2111 Median :-0.3216   
## Mean : 0.00000 Mean : 0.000000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.85266 3rd Qu.: 0.872780 3rd Qu.: 0.5747 3rd Qu.: 0.5571   
## Max. : 1.89903 Max. : 2.008075 Max. : 3.1285 Max. : 1.4359   
## CCAvg Mortgage Personal.Loan Securities.Account  
## Min. :-1.1142 Min. :-0.5617 0:2260 Min. :-0.3435   
## 1st Qu.:-0.7136 1st Qu.:-0.5617 1: 240 1st Qu.:-0.3435   
## Median :-0.1987 Median :-0.5617 Median :-0.3435   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.0000   
## 3rd Qu.: 0.3735 3rd Qu.: 0.4160 3rd Qu.:-0.3435   
## Max. : 4.0353 Max. : 5.4080 Max. : 2.9097   
## CD.Account Online CreditCard education1   
## Min. :-0.2454 Min. :-1.2093 Min. :-0.652 Min. :-0.8648   
## 1st Qu.:-0.2454 1st Qu.:-1.2093 1st Qu.:-0.652 1st Qu.:-0.8648   
## Median :-0.2454 Median : 0.8266 Median :-0.652 Median :-0.8648   
## Mean : 0.0000 Mean : 0.0000 Mean : 0.000 Mean : 0.0000   
## 3rd Qu.:-0.2454 3rd Qu.: 0.8266 3rd Qu.: 1.533 3rd Qu.: 1.1558   
## Max. : 4.0742 Max. : 0.8266 Max. : 1.533 Max. : 1.1558   
## education2 education3   
## Min. :-0.6315 Min. :-0.634   
## 1st Qu.:-0.6315 1st Qu.:-0.634   
## Median :-0.6315 Median :-0.634   
## Mean : 0.0000 Mean : 0.000   
## 3rd Qu.: 1.5828 3rd Qu.: 1.577   
## Max. : 1.5828 Max. : 1.577

validationset.norm\_2<-predict(normalization\_values\_2,validationset\_2)  
summary(validationset.norm\_2)

## Age Experience Income Family   
## Min. :-1.93768 Min. :-2.009123 Min. :-1.4553 Min. :-1.20039   
## 1st Qu.:-0.80410 1st Qu.:-0.786498 1st Qu.:-0.7841 1st Qu.:-1.20039   
## Median :-0.01932 Median :-0.000524 Median :-0.2766 Median :-0.32163   
## Mean : 0.02532 Mean : 0.021076 Mean :-0.0367 Mean : 0.02578   
## 3rd Qu.: 0.85266 3rd Qu.: 0.872780 3rd Qu.: 0.4601 3rd Qu.: 0.55714   
## Max. : 1.89903 Max. : 2.008075 Max. : 3.2595 Max. : 1.43590   
## CCAvg Mortgage Personal.Loan Securities.Account   
## Min. :-1.11415 Min. :-0.56174 0:1356 Min. :-0.343541   
## 1st Qu.:-0.71364 1st Qu.:-0.56174 1: 144 1st Qu.:-0.343541   
## Median :-0.25592 Median :-0.56174 Median :-0.343541   
## Mean :-0.01726 Mean :-0.05339 Mean :-0.007374   
## 3rd Qu.: 0.31624 3rd Qu.: 0.36193 3rd Qu.:-0.343541   
## Max. : 4.60742 Max. : 4.97559 Max. : 2.909692   
## CD.Account Online CreditCard education1   
## Min. :-0.24535 Min. :-1.20933 Min. :-0.6520 Min. :-0.86484   
## 1st Qu.:-0.24535 1st Qu.:-1.20933 1st Qu.:-0.6520 1st Qu.:-0.86484   
## Median :-0.24535 Median : 0.82658 Median :-0.6520 Median :-0.86484   
## Mean : 0.03398 Mean :-0.01086 Mean :-0.0169 Mean :-0.01347   
## 3rd Qu.:-0.24535 3rd Qu.: 0.82658 3rd Qu.: 1.5331 3rd Qu.: 1.15582   
## Max. : 4.07419 Max. : 0.82658 Max. : 1.5331 Max. : 1.15582   
## education2 education3   
## Min. :-0.63153 Min. :-0.63401   
## 1st Qu.:-0.63153 1st Qu.:-0.63401   
## Median :-0.63153 Median :-0.63401   
## Mean :-0.03513 Mean : 0.04981   
## 3rd Qu.: 1.58282 3rd Qu.: 1.57663   
## Max. : 1.58282 Max. : 1.57663

testset.norm\_2<-predict(normalization\_values\_2,testset\_2)  
summary(testset.norm\_2)

## Age Experience Income Family   
## Min. :-1.93768 Min. :-2.00912 Min. :-1.45534 Min. :-1.20039   
## 1st Qu.:-0.89130 1st Qu.:-0.96116 1st Qu.:-0.80050 1st Qu.:-1.20039   
## Median : 0.06787 Median : 0.08681 Median :-0.25480 Median :-0.32163   
## Mean : 0.01294 Mean : 0.01144 Mean :-0.04307 Mean : 0.09491   
## 3rd Qu.: 0.93985 3rd Qu.: 0.87278 3rd Qu.: 0.40549 3rd Qu.: 1.43590   
## Max. : 1.89903 Max. : 1.83341 Max. : 2.82295 Max. : 1.43590   
## CCAvg Mortgage Personal.Loan Securities.Account   
## Min. :-1.114153 Min. :-0.56174 0:904 Min. :-0.343541   
## 1st Qu.:-0.713643 1st Qu.:-0.56174 1: 96 1st Qu.:-0.343541   
## Median :-0.255917 Median :-0.56174 Median :-0.343541   
## Mean :-0.000843 Mean :-0.07284 Mean :-0.008458   
## 3rd Qu.: 0.316241 3rd Qu.: 0.35958 3rd Qu.:-0.343541   
## Max. : 4.607421 Max. : 4.95679 Max. : 2.909692   
## CD.Account Online CreditCard education1   
## Min. :-0.24535 Min. :-1.20933 Min. :-0.65203 Min. :-0.8648   
## 1st Qu.:-0.24535 1st Qu.:-1.20933 1st Qu.:-0.65203 1st Qu.:-0.8648   
## Median :-0.24535 Median : 0.82658 Median :-0.65203 Median :-0.8648   
## Mean : 0.02678 Mean : 0.04479 Mean :-0.02272 Mean :-0.0687   
## 3rd Qu.:-0.24535 3rd Qu.: 0.82658 3rd Qu.: 1.53306 3rd Qu.: 1.1558   
## Max. : 4.07419 Max. : 0.82658 Max. : 1.53306 Max. : 1.1558   
## education2 education3   
## Min. :-0.631532 Min. :-0.63401   
## 1st Qu.:-0.631532 1st Qu.:-0.63401   
## Median :-0.631532 Median :-0.63401   
## Mean : 0.001772 Mean : 0.07339   
## 3rd Qu.: 1.582817 3rd Qu.: 1.57663   
## Max. : 1.582817 Max. : 1.57663

#confusion matrix  
library(gmodels)  
  
train\_label\_2<-trainset.norm\_2[,7]  
validation\_label\_2<-validationset.norm\_2[,7]  
test\_label\_2<-testset.norm\_2[,7]  
  
predicted\_validationlabel\_2<-knn(trainset.norm\_2,validationset.norm\_2,cl=train\_label\_2,k=best\_k)  
  
predicted\_testlabel\_2<-knn(trainset.norm\_2,testset.norm\_2,cl=train\_label\_2,k=best\_k)  
  
confusionmatrix\_1<-CrossTable(x=validation\_label\_2,y=predicted\_validationlabel\_2,prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1500   
##   
##   
## | predicted\_validationlabel\_2   
## validation\_label\_2 | 0 | 1 | Row Total |   
## -------------------|-----------|-----------|-----------|  
## 0 | 1354 | 2 | 1356 |   
## | 0.999 | 0.001 | 0.904 |   
## | 0.974 | 0.018 | |   
## | 0.903 | 0.001 | |   
## -------------------|-----------|-----------|-----------|  
## 1 | 36 | 108 | 144 |   
## | 0.250 | 0.750 | 0.096 |   
## | 0.026 | 0.982 | |   
## | 0.024 | 0.072 | |   
## -------------------|-----------|-----------|-----------|  
## Column Total | 1390 | 110 | 1500 |   
## | 0.927 | 0.073 | |   
## -------------------|-----------|-----------|-----------|  
##   
##

confusionmatrix\_2<-CrossTable(x=test\_label\_2,y=predicted\_testlabel\_2,prop.chisq = FALSE)

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Row Total |  
## | N / Col Total |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 1000   
##   
##   
## | predicted\_testlabel\_2   
## test\_label\_2 | 0 | 1 | Row Total |   
## -------------|-----------|-----------|-----------|  
## 0 | 901 | 3 | 904 |   
## | 0.997 | 0.003 | 0.904 |   
## | 0.979 | 0.037 | |   
## | 0.901 | 0.003 | |   
## -------------|-----------|-----------|-----------|  
## 1 | 19 | 77 | 96 |   
## | 0.198 | 0.802 | 0.096 |   
## | 0.021 | 0.963 | |   
## | 0.019 | 0.077 | |   
## -------------|-----------|-----------|-----------|  
## Column Total | 920 | 80 | 1000 |   
## | 0.920 | 0.080 | |   
## -------------|-----------|-----------|-----------|  
##   
##

validation\_table<-table(validation\_label\_2,predicted\_validationlabel\_2)  
confusionMatrix(validation\_table)

## Confusion Matrix and Statistics  
##   
## predicted\_validationlabel\_2  
## validation\_label\_2 0 1  
## 0 1354 2  
## 1 36 108  
##   
## Accuracy : 0.9747   
## 95% CI : (0.9654, 0.982)  
## No Information Rate : 0.9267   
## P-Value [Acc > NIR] : 2.894e-16   
##   
## Kappa : 0.8368   
##   
## Mcnemar's Test P-Value : 8.636e-08   
##   
## Sensitivity : 0.9741   
## Specificity : 0.9818   
## Pos Pred Value : 0.9985   
## Neg Pred Value : 0.7500   
## Prevalence : 0.9267   
## Detection Rate : 0.9027   
## Detection Prevalence : 0.9040   
## Balanced Accuracy : 0.9780   
##   
## 'Positive' Class : 0   
##

test\_table<-table(test\_label\_2,predicted\_testlabel\_2)  
confusionMatrix(test\_table)

## Confusion Matrix and Statistics  
##   
## predicted\_testlabel\_2  
## test\_label\_2 0 1  
## 0 901 3  
## 1 19 77  
##   
## Accuracy : 0.978   
## 95% CI : (0.9669, 0.9862)  
## No Information Rate : 0.92   
## P-Value [Acc > NIR] : 2.68e-15   
##   
## Kappa : 0.863   
##   
## Mcnemar's Test P-Value : 0.001384   
##   
## Sensitivity : 0.9793   
## Specificity : 0.9625   
## Pos Pred Value : 0.9967   
## Neg Pred Value : 0.8021   
## Prevalence : 0.9200   
## Detection Rate : 0.9010   
## Detection Prevalence : 0.9040   
## Balanced Accuracy : 0.9709   
##   
## 'Positive' Class : 0   
##

# on comparing the confusion matrix of validation set and testing set it can be seen that accuracy and sensitivity of validation is slightly greater than test set.