setwd("E:\Acadgild\R\project2")

data<-read\_excel("Churn.xls",sheet = 1)

head(data)

class(data)

data<-as.data.frame(data)

head(data)

str(data)

#to find na's

sum(is.na(data))

#to order columns

library(tibble)

colnames(data)

my\_data<-data[,c(19,20,21,1,2,3,4,5,6,7,9,10,11,12,13,14,15,16,17,18,8)]

head(my\_data)

str(my\_data)

my\_data$Phone<-gsub("-", "", paste(my\_data$Phone))

table(my\_data$State)

my\_data$Phone<-as.numeric(my\_data$Phone)

my\_data$State<-as.factor(my\_data$State)

my\_data$Churn<-as.factor(my\_data$Churn)

ind <- sample(2, nrow(my\_data), replace = TRUE, prob = c(0.8, 0.2))

train <- my\_data[ind==1,]

test <- my\_data[ind==2,]

head(train)

#cross validation

library(caret)

control <- trainControl(method = 'repeatedcv',

number = 10,

repeats = 3)

#logistic regression

seed <-7

metric <- 'Accuracy'

set.seed(seed)

lg\_reg <- train(Churn~.,

data = train,

method = 'glm',

metric = metric,

trControl = control,

maxit=100)

#predictions

p1<-predict(lg\_reg ,train)

head(p1)

#accuracy

tab1<-table(p1,train$Churn)

accuracy1<-sum(diag(tab1))/sum(tab1)

#prediction for test

p2<-predict(lg\_reg,test)

head(p2)

#accuracy

tab2<-table(p2,test$Churn)

accuracy2<-sum(diag(tab2))/sum(tab2)

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

library(nnet)

seed <-7

metric <- 'Accuracy'

set.seed(seed)

my\_modle<- train(Churn~.,

data = train,

method = 'multinom',

metric = metric,

trControl = control,

maxit=100)

p1<-predict(my\_modle ,train)

head(p1)

tab1<-table(p1,train$Churn)

accuracy1<-sum(diag(tab1))/sum(tab1)

table(train$Churn)

#prediction for test

p2<-predict(lg\_reg,test)

head(p2)

#accuracy

tab2<-table(p2,test$Churn)

accuracy2<-sum(diag(tab2))/sum(tab2)

#model performance evaluatiion

library(ROCR)

pred<-predict(my\_modle ,train,type='prob')

head(pred$`1`)

head(train$Churn)

hist(pred$`0`)

hist(pred$`1`)

pred1<-prediction(pred$`1`,train$Churn)

eval<-performance(pred1,"acc")

plot(eval)

abline(h=0.87,v=0.38)

#identify bestvalues

max<-which.max(slot(eval,"y.values")[[1]])

acc<-slot(eval,"y.values")[[1]][max]

cut<-slot(eval,"x.values")[[1]][max]

#ROC (reciver operating characteristic) curve

pred1<-prediction(pred$`1`,train$Churn)

roc<-performance(pred1,"tpr","fpr")

plot(roc)

abline(a=0,b=1)

#can add some more things

plot(roc,

colorize=T,

main="ROC curve",

ylab="sensitivity",

xlab="1-specficity")

abline(a=0,b=1)

#AUC (area under curve)

auc<-performance(pred1,"auc")

auc<-unlist(slot(auc,"y.values"))

auc<-round(auc,4)

legend(.8,.2,auc,title = "AUC",cex = 1.2)

#model performance evaluatiion for test

library(ROCR)

pred<-predict(my\_modle ,test,type='prob')

head(pred$`1`)

head(test$Churn)

hist(pred$`0`)

hist(pred$`1`)

pred1<-prediction(pred$`1`,test$Churn)

eval<-performance(pred1,"acc")

plot(eval)

abline(h=0.87,v=0.38)

#identify bestvalues

max<-which.max(slot(eval,"y.values")[[1]])

acc<-slot(eval,"y.values")[[1]][max]

cut<-slot(eval,"x.values")[[1]][max]

#ROC (reciver operating characteristic) curve

pred1<-prediction(pred$`1`,test$Churn)

roc<-performance(pred1,"tpr","fpr")

plot(roc)

abline(a=0,b=1)

#can add some more things

plot(roc,

colorize=T,

main="ROC curve",

ylab="sensitivity",

xlab="1-specficity")

abline(a=0,b=1)

#AUC (area under curve)

auc<-performance(pred1,"auc")

auc<-unlist(slot(auc,"y.values"))

auc<-round(auc,4)

legend(.8,.2,auc,title = "AUC",cex = 1.2)

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

#random forest

control <- trainControl(method = 'repeatedcv',

number = 5,

repeats = 3)

library(randomForest)

set.seed(7)

mtry <- sqrt(ncol(data))

rf\_random<- train(Churn~.,

data = train,

method = 'rf',

metric = metric,

trControl = control,

maxit=100)

print(rf\_random)

plot(rf\_random)

#predictions for train

predictions<- predict(rf\_random,train)

#accuracy

tab1<-table(predictions,train$Churn)

accuracy<-sum(diag(tab1))/sum(tab1)

#or

confusionMatrix<- confusionMatrix(predictions,train$Churn)

confusionMatrix

#for test data

predictions1<- predict(rf\_random,test)

#accuracy

tab2<-table(predictions1,test$Churn)

accuracy2<-sum(diag(tab2))/sum(tab2)

#model performance evaluatiion

library(ROCR)

pred<-predict(rf\_random ,train,type='prob')

head(pred$`1`)

head(train$Churn)

hist(pred$`0`)

hist(pred$`1`)

pred1<-prediction(pred$`1`,train$Churn)

eval<-performance(pred1,"acc")

plot(eval)

abline(h=0.99,v=0.4)

#identify bestvalues

max<-which.max(slot(eval,"y.values")[[1]])

acc<-slot(eval,"y.values")[[1]][max]

cut<-slot(eval,"x.values")[[1]][max]

#ROC (reciver operating characteristic) curve

pred1<-prediction(pred$`1`,train$Churn)

roc<-performance(pred1,"tpr","fpr")

plot(roc)

abline(a=0,b=1)

#can add some more things

plot(roc,

colorize=T,

main="ROC curve",

ylab="sensitivity",

xlab="1-specficity")

abline(a=0,b=1)

#AUC (area under curve)

auc<-performance(pred1,"auc")

auc<-unlist(slot(auc,"y.values"))

auc<-round(auc,4)

legend(.8,.2,auc,title = "AUC",cex = 1.2)

#model performance evaluatiion for test

library(ROCR)

pred<-predict(rf\_random ,test,type='prob')

head(pred$`1`)

head(test$Churn)

hist(pred$`0`)

hist(pred$`1`)

pred1<-prediction(pred$`1`,test$Churn)

eval<-performance(pred1,"acc")

plot(eval)

abline(h=0.97,v=0.55)

#identify bestvalues

max<-which.max(slot(eval,"y.values")[[1]])

acc<-slot(eval,"y.values")[[1]][max]

cut<-slot(eval,"x.values")[[1]][max]

#ROC (reciver operating characteristic) curve

pred1<-prediction(pred$`1`,test$Churn)

roc<-performance(pred1,"tpr","fpr")

plot(roc)

abline(a=0,b=1)

#can add some more things

plot(roc,

colorize=T,

main="ROC curve",

ylab="sensitivity",

xlab="1-specficity")

abline(a=0,b=1)

#AUC (area under curve)

auc<-performance(pred1,"auc")

auc<-unlist(slot(auc,"y.values"))

auc<-round(auc,4)

legend(.8,.2,auc,title = "AUC",cex = 1.2)