1. Use the below given data set

DataSet

2. Perform the below given activities:

a. Create classification model using different classifiers

b. Verify model goodness of fit

c. Apply all the model validation techniques.

Solution:

library(C50)

data(churn)

head(churnTrain)

head(churnTest)

#churnTrain = churnTrain[1:500,]

#churnTest = churnTest[1:500,]

# logistic regression model:

fit <- glm(churn~.,data = churnTrain,family = binomial(link='logit'))

summary(fit)

install.packages('ResourceSelection')

library(ResourceSelection)

hoslem.test(churnTrain$churn, fitted(fit))

#plot the fitted model

plot(fit$fitted.values)

x <- 1:20

y <- 21:40

library(MASS)

boxcox(y~x)

plot(1/y^2~x)

#check for multicollinearity

library(car)

vif(fit)

vif(step\_fit)

pred <- predict(fit,newdata = churnTest,type = 'response')

#check the AUC curve

library(pROC)

g <- roc(churn ~ pred, data = churnTest)

g

plot(g)

library(caret)

#with default prob cut 0.50

churnTest$pred\_churn <- ifelse(pred<0.7,'yes','no')

table(churnTest$pred\_churn,churnTest$churn)

#training split of churn classes

round(table(churnTrain$churn)/nrow(churnTrain),2)\*100

# test split of churn classes

round(table(churnTest$churn)/nrow(churnTest),2)\*100

#predicted split of churn classes

round(table(churnTest$pred\_churn)/nrow(churnTest),2)\*100

#create confusion matrix

confusionMatrix(churnTest$churn,churnTest$pred\_churn)

#------------

# load libraries

library(caret)

library(rpart)

# define training control

train\_control<- trainControl(method="cv", number=10)

# train the model

model<- train(churn~., data=churnTrain, trControl=train\_control, method="glm")

# make predictions

predictions<- predict(model,churnTest)

# append predictions

pred<- cbind(churnTest,predictions)

# summarize results

confusionMatrix<- confusionMatrix(pred$predictions,pred$churn)

confusionMatrix