Output variable -> yy -> Whether the client has subscribed a term deposit or not

Binomial ("yes" or "no")

#importing the data

data <- read.csv(file.choose())

View(data)

str(data)

attach(data)

##the data we have has been preproceesed with dummy variable##

plot(data$balance,data$y)

plot(default,y)

plot(age,y)

plot(housing,y)

plot(loan,y)

boxplot(loan,y)

plot(campaign,y)

##There is no particular trend dependent on the sigle variable on the out put variable

##model buliding

model <- glm(y~.,data = data,family = "binomial")

summary(model)

##To calculate the odds ratio manually we going r going to take exp of coef(model)

exp(coef(model))

# Confusion matrix table

prob <- predict(model,data,type="response")

summary(model)

# We are going to use NULL and Residual Deviance to compare the between different models

# Confusion matrix and considering the threshold value as 0.5

confusion<-table(prob>0.5,data$y)

confusion

# Model Accuracy

Accuracy<-sum(diag(confusion)/sum(confusion))

Accuracy # .9005

# Creating empty vectors to store predicted classes based on threshold value

pred\_values <- NULL

yes\_no <- NULL

pred\_values <- ifelse(prob>=0.5,1,0)

yes\_no <- ifelse(prob>=0.5,"yes","no")

# Creating new column to store the above values

data[,"prob"] <- prob

data[,"pred\_values"] <- pred\_values

data[,"yes\_no"] <- yes\_no

View(data[,c(32:35)])

table(data$y,data$pred\_values)

# ROC Curve => used to evaluate the betterness of the logistic model

# more area under ROC curve better is the model

# We will use ROC curve for any classification technique not only for logistic

library(ROCR)

rocrpred<-prediction(prob,data$y)

rocrperf<-performance(rocrpred,'tpr','fpr')

str(rocrperf)

plot(rocrperf,colorize=T,text.adj=c(-0.2,1.7))

# More area under the ROC Curve better is the logistic regression model obtained

## Getting cutt off or threshold value along with true positive and false positive rates in a data frame

str(rocrperf)

rocr\_cutoff <- data.frame(cut\_off = rocrperf@alpha.values[[1]],fpr=rocrperf@x.values,tpr=rocrperf@y.values)

colnames(rocr\_cutoff) <- c("cut\_off","FPR","TPR")

View(rocr\_cutoff)

library(dplyr)

rocr\_cutoff$cut\_off <- round(rocr\_cutoff$cut\_off,6)

# Sorting data frame with respect to tpr in decreasing order

rocr\_cutoff <- arrange(rocr\_cutoff,desc(TPR))

View(rocr\_cutoff)

