Test Accuracy:

Confusion Matrix and Statistics

Reference

Prediction 0 1

0 62391 5862

1 4870 16952

Accuracy : 0.8809

95% CI : (0.8787, 0.883)

No Information Rate : 0.7467

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6804

Mcnemar's Test P-Value : < 2.2e-16

Sensitivity : 0.9276

Specificity : 0.7431

Pos Pred Value : 0.9141

Neg Pred Value : 0.7768

Prevalence : 0.7467

Detection Rate : 0.6927

Detection Prevalence : 0.7577

Balanced Accuracy : 0.8353

'Positive' Class : 0

Model plot

Chart

Description automatically generated with medium confidence

Code

#input data

InputData.df <- read.csv("C:/Users/skb24/Downloads/Study/Input Data.csv", header=TRUE, stringsAsFactors=TRUE)

TrainLabels.df <- read.csv("C:/Users/skb24/Downloads/Study/train\_labels.csv", header=TRUE, stringsAsFactors=TRUE)

merged.df <- merge.default(TrainLabels.df, InputData.df)

#convert target to factors with 2 levels 0,1

merged.df$target <- as.factor(merged.df$target)

#convert date column from factor to date type

#merged.df$S\_2 <- as.Date(merged.df$S\_2)

#remove cust id

merged.df <- merged.df[-c(1)]

#remove date

merged.df <- merged.df[-c(3)]

str(merged.df)

#replace na values with column mean -- working method

data2 <- merged.df

for(i in 1:ncol(merged.df)) {

data2[ , i][is.na(data2[ , i])] <- mean(data2[ , i], na.rm = TRUE)

}

#random forest

library(randomForest)

library(datasets)

library(caret)

set.seed(222)

ind <- sample(2, nrow(data2), replace = TRUE, prob = c(0.1, 0.9))

train <- data2[ind==1,]

test <- data2[ind==2,]

rf.model <- randomForest(target~., data=train, proximity=TRUE, na.action=na.roughfix)

print(rf.model)

#prediction and confusion matrix train data

p1 <- predict(rf.model,train)

confusionMatrix(p1, train$target)

#prediction and confusion matrix test data

p2 <- predict(rf.model,test)

confusionMatrix(p2, test$target)

#plot model on graph

plot(rf.model)