**Random Forest:**

1. **Fraud Data:**

#Building random forest on fraud data

#install.packages(“caret”, dependencies = True)

#install.packages(“randomForest”)

library(randomForest)

Fraud\_check$Urban <- as.numeric(Fraud\_check$Urban) – 1

Fraud\_check$Undergrad <- as.numeric(Fraud\_check$Undergrad) – 1

Tax.inc <- cut(Fraud\_check$Taxable.Income, breaks = c(0,30000,99700), labels = c(‘Risky’,’Good’))

View(fc\_rf)

Fraud\_check <- dummy\_cols(Fraud\_check, select\_columns = ‘Marital.Status’, remove\_selected\_columns = TRUE)

fc\_rf <- cbind(Tax.inc, Fraud\_check)

fc\_rf <- fc\_rf[,-3]

#Building the model

model\_rf <- randomForest(fc\_rf$Tax.inc ~. , data = fc\_rf, ntree = 1000)

print(model\_rf)

##Importance of the variable – Lower Gini

print(importance(model\_rf))

#prediction

pred\_rf <- predict(model\_rf, fc\_rf[,-1])

table(pred\_rf, fc\_rf$Tax.inc)

#”City Population” and “Work Experience” play a major role in determining whether an individual is likely to

#committ an act of fraud or not.

1. **Company data:**

#Building random forest with target variable sales

View(Company\_Data)

Company\_Data$Urban <- as.numeric(Company\_Data$Urban) - 1

Company\_Data$US <- as.numeric(Company\_Data$US) - 1

Company\_Data <- dummy\_cols(Company\_Data, select\_columns = 'ShelveLoc', remove\_selected\_columns = T)

Sales <- cut(Company\_Data$Sales, breaks = c(0,10,17), labels = c('<=10','>10'))

Company\_Data <- Company\_Data[,-1]

sales\_rf <- cbind(Sales,Company\_Data)

library(gam)

sales\_rf1 <- na.omit(sales\_rf)

View(sales\_rf1)

#building the model

mod\_rf <- randomForest(sales\_rf1$Sales ~., data = sales\_rf1, ntree = 1000)

print(mod\_rf)

##Importance of the variable - Lower Gini

print(importance(mod\_rf))

#prediction

pred\_sales\_rf <- predict(mod\_rf, sales\_rf1[,-1])

table(pred\_sales\_rf, sales\_rf1$Sales)

#"Price", Good Shelf Location" , "Age", "Income", "Competitor's Price", "Advertising" and "Population" are

#the attributes that determine high sales of the company.