Use Random Forest to prepare a model on fraud data

treating those who have taxable\_income <= 30000 as "Risky" and others are "Good"

**Ans :**

**R Code :**

## Random Forest

########## Fraud Check Data Set #########

install.packages("randomForest")

library(randomForest)

Fraud\_check\_r <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Random Forests\\Fraud\_check\_r.csv')

model<-randomForest(Fraud\_check\_r$Status~.,data=Fraud\_check\_r,ntree=1000)

# View the forest results.

print(model)

#Importance of the variable - Lower Gini

print(importance(model))

#Prediction

pred<- predict(model,Fraud\_check\_r[,-7])

table(pred,Fraud\_check\_r$Status)

**Results :**

model<-randomForest(Fraud\_check\_r$Status~.,data=Fraud\_check\_r,ntree=1000)

> # View the forest results.

> print(model)

Call:

randomForest(formula = Fraud\_check\_r$Status ~ ., data = Fraud\_check\_r, ntree = 1000)

Type of random forest: classification

Number of trees: 1000

No. of variables tried at each split: 2

OOB estimate of error rate: 0.17%

Confusion matrix:

Good Risky class.error

Good 475 1 0.00210084

Risky 0 124 0.00000000

> #Importance of the variable - Lower Gini

> print(importance(model))

MeanDecreaseGini

Undergrad 0.5537986

Marital.Status 1.0480055

Taxable.Income 181.2023333

City.Population 6.1666890

Work.Experience 3.8105818

Urban 0.5398872

> #Prediction

> pred<- predict(model,Fraud\_check\_r[,-7])

> table(pred,Fraud\_check\_r$Status)

pred Good Risky

Good 476 0

Risky 0 124

**Inference :**

Taxable Income was found as the important feature as it was having the highest index.