**Random Forest**

**Uploading dataset**

**data = read.csv("Dataset\_spine.csv")**

**getwd()**

**View(data)**

**summary(data)**

**#Removing Outliers**

**bx=boxplot(data$pelvic\_incidence)**

**bx$stats**

**quantile(data$pelvic\_incidence,seq(0,1,0.02))**

**data$pelvic\_incidence<-ifelse(data$pelvic\_incidence>=94,94,data$pelvic\_incidence)**

**bx=boxplot(data$pelvic\_incidence)**

**bx=boxplot(data$pelvic\_tilt)**

**bx$stats**

**quantile(data$pelvic\_tilt,seq(0,1,0.02))**

**data$pelvic\_tilt<-ifelse(data$pelvic\_tilt>=39,39,data$pelvic\_tilt)**

**bx=boxplot(data$pelvic\_tilt)**

**bx=boxplot(data$lumbar\_lordosis\_angle)**

**bx$stats**

**quantile(data$lumbar\_lordosis\_angle,seq(0,1,0.02))**

**data$lumbar\_lordosis\_angle<-ifelse(data$lumbar\_lordosis\_angle>=94,94,data$lumbar\_lordosis\_angle)**

**bx=boxplot(data$lumbar\_lordosis\_angle)**

**bx=boxplot(data$sacral\_slope)**

**bx$stats**

**quantile(data$sacral\_slope,seq(0,1,0.02))**

**data$sacral\_slope<-ifelse(data$sacral\_slope>=70,70,data$sacral\_slope)**

**bx=boxplot(data$lumbar\_lordosis\_angle)**

**bx=boxplot(data$pelvic\_radius)**

**bx$stats**

**quantile(data$pelvic\_radius,seq(0,1,0.02))**

**data$pelvic\_radius<-ifelse(data$pelvic\_radius>=146,146,data$pelvic\_radius)**

**data$pelvic\_radius<-ifelse(data$pelvic\_radius<=89,89,data$pelvic\_radius)**

**bx=boxplot(data$pelvic\_radius)**

**bx=boxplot(data$degree\_spondylolisthesis)**

**bx$stats**

**quantile(data$degree\_spondylolisthesis,seq(0,1,0.02))**

**data$degree\_spondylolisthesis<-ifelse(data$degree\_spondylolisthesis>=91,91,data$degree\_spondylolisthesis)**

**bx=boxplot(data$degree\_spondylolisthesis)**

**#------------------------------------------------------------**

**#Checking correlation between Variables**

**library(car)**

**scatterplot(data$pelvic\_tilt,data$Status)**

**mosaicplot(data$Status~data$pelvic\_incidence, color='skyblue')**

**library(corrplot)**

**cor(data[,1:8])**

**corrplot(cor(data[,1:8]), method="circle")**

**#Creating Train and Test data.**

**set.seed(1234)**

**splitIndex <- createDataPartition(data$Status, p = .70,list = FALSE, times = 1)**

**trainSplit <- data[ splitIndex,]**

**testSplit <- data[-splitIndex,]**

**print(table(trainSplit$Status))**

**print(table(testSplit$Status))**

**prop.table(table(trainSplit$Status))**

**prop.table(table(testSplit$Status))**

**#Random Forest model**

**library(randomForest)**

**modelrf <- randomForest(as.factor(Status) ~ . , data = trainSplit, do.trace=T)**

**modelrf**

Call:

randomForest(formula = as.factor(Status) ~ ., data = trainSplit, do.trace = T)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 16.59%

Confusion matrix:

Abnormal Normal class.error

Abnormal 131 16 0.1088435

Normal 20 50 0.2857143

P N

P Tp fn

N Fp tn

**#Checking variable importance in Random Forest**

**importance(modelrf)**

MeanDecreaseGini

pelvic\_incidence 7.302535

pelvic\_tilt 8.579979

lumbar\_lordosis\_angle 7.940512

sacral\_slope 7.437386

pelvic\_radius 12.991508

degree\_spondylolisthesis 23.748114

pelvic\_slope 3.446864

Direct\_tilt 3.580621

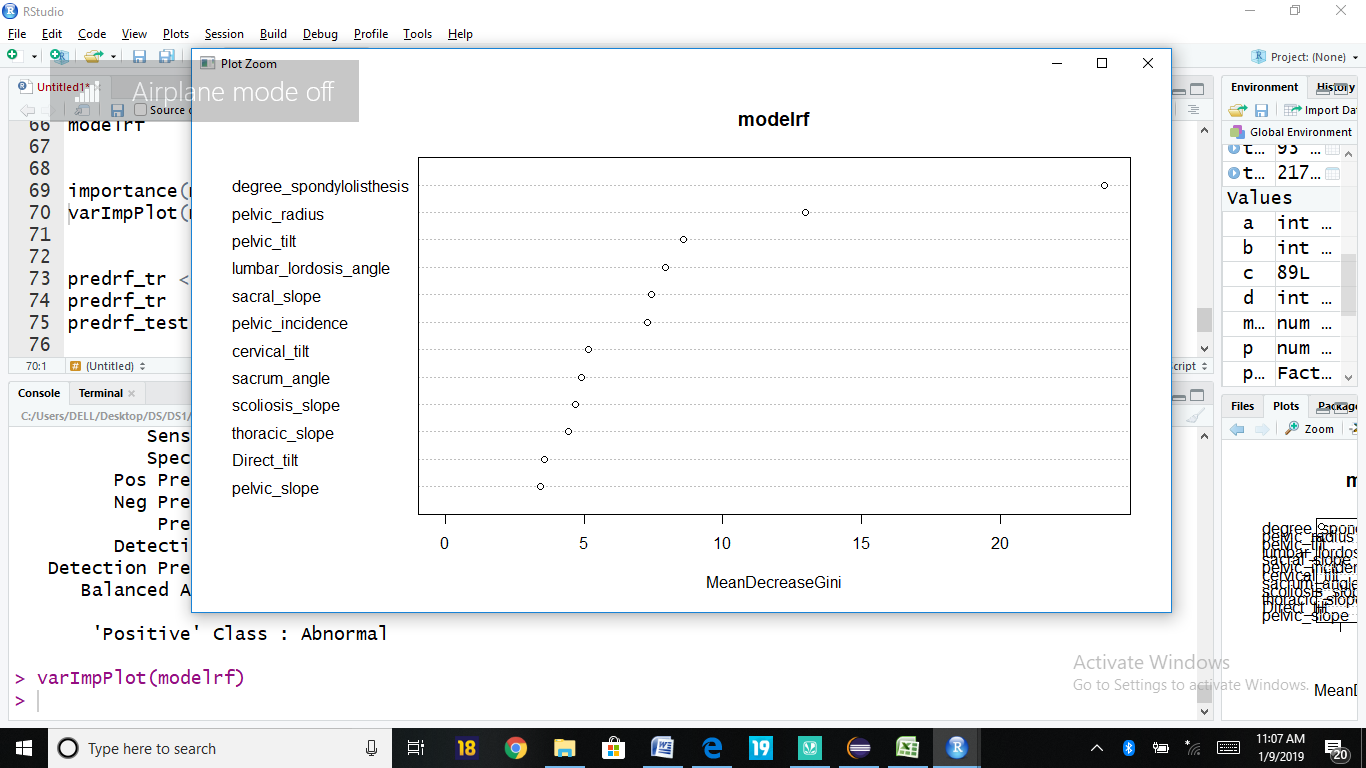
thoracic\_slope 4.456488

cervical\_tilt 5.164046

sacrum\_angle 4.898668

scoliosis\_slope 4.682147

**varImpPlot(modelrf)**

****

**#Prediction and Model Evaluation using Confusion Matrix**

**predrf\_tr <- predict(modelrf, trainSplit)**

**predrf\_test <- predict(modelrf, testSplit)**

**confusionMatrix(predrf\_tr,trainSplit$Status)**

Confusion Matrix and Statistics

Reference

Prediction Abnormal Normal

Abnormal 147 0

Normal 0 70

Accuracy : 1

95% CI : (0.9831, 1)

No Information Rate : 0.6774

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

Kappa : 1

Mcnemar's Test P-Value : NA

Sensitivity : 1.0000

Specificity : 1.0000

Pos Pred Value : 1.0000

Neg Pred Value : 1.0000

Prevalence : 0.6774

Detection Rate : 0.6774

Detection Prevalence : 0.6774

Balanced Accuracy : 1.0000

'Positive' Class : Abnormal

**confusionMatrix(predrf\_test,testSplit$Status)**

Confusion Matrix and Statistics

Reference

Prediction Abnormal Normal

Abnormal 56 7

Normal 7 23

Accuracy : 0.8495

95% CI : (0.7603, 0.9152)

No Information Rate : 0.6774

P-Value [Acc > NIR] : 0.0001313

Kappa : 0.6556

Mcnemar's Test P-Value : 1.0000000

Sensitivity : 0.8889

Specificity : 0.7667

Pos Pred Value : 0.8889

Neg Pred Value : 0.7667

Prevalence : 0.6774

Detection Rate : 0.6022

Detection Prevalence : 0.6774

Balanced Accuracy : 0.8278

'Positive' Class : Abnormal