**Decsion Tree**

**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))**

**# DECISION TREE**

**library(rpart)**

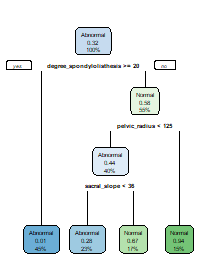
**library(rpart.plot)**

**library(caret)**

**fit = rpart(Status ~ ., data = trainSplit,method = "class", control = rpart.control(minsplit = 65,cp = 0.01))**

*#Plotting the Decision Tree rpart.plot(fit)*

**rpart.plot(fit)**

****

**summary(fit)**

Call:

rpart(formula = Status ~ ., data = trainSplit, method = "class",

control = rpart.control(minsplit = 65, cp = 0.01))

n= 217

CP nsplit rel error xerror xstd

1 0.2714286 0 1.0000000 1.0000000 0.09837388

2 0.1571429 1 0.7285714 0.8285714 0.09312894

3 0.0100000 3 0.4142857 0.5714286 0.08159973

Variable importance

degree\_spondylolisthesis sacral\_slope lumbar\_lordosis\_angle pelvic\_radius

26 18 17 16

pelvic\_incidence pelvic\_tilt Direct\_tilt sacrum\_angle

15 7 1 1

**print(fit)**

n= 217

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 217 70 Abnormal (0.67741935 0.32258065)

2) degree\_spondylolisthesis>=20.08504 98 1 Abnormal (0.98979592 0.01020408) \*

3) degree\_spondylolisthesis< 20.08504 119 50 Normal (0.42016807 0.57983193)

6) pelvic\_radius< 125.4417 86 38 Abnormal (0.55813953 0.44186047)

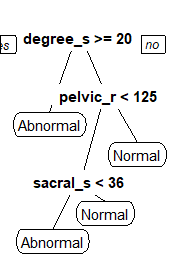
12) sacral\_slope< 35.55356 50 14 Abnormal (0.72000000 0.28000000) \*

13) sacral\_slope>=35.55356 36 12 Normal (0.33333333 0.66666667) \*

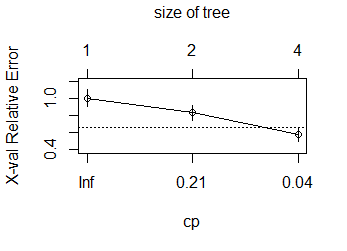
7) pelvic\_radius>=125.4417 33 2 Normal (0.06060606 0.93939394) \*

***#ANOTHER METHOD For Plotting the Decision Tree***

**prp(fit)**

****

**plotcp(fit)**



**printcp(fit)**

Classification tree:

rpart(formula = Status ~ ., data = trainSplit, method = "class",

control = rpart.control(minsplit = 65, cp = 0.01))

Variables actually used in tree construction:

[1] degree\_spondylolisthesis pelvic\_radius sacral\_slope

Root node error: 70/217 = 0.32258

n= 217

CP nsplit rel error xerror xstd

1 0.27143 0 1.00000 1.00000 0.098374

2 0.15714 1 0.72857 0.82857 0.093129

3 0.01000 3 0.41429 0.57143 0.081600

***#Checking confusion matrix on Train data***

**predtr <- predict(fit,trainSplit,type = "class" )**

**confusionMatrix(predtr,trainSplit$Status)**

Confusion Matrix and Statistics

Reference

Prediction Abnormal Normal

Abnormal 133 15

Normal 14 55

Accuracy : 0.8664

95% CI : (0.8137, 0.9086)

No Information Rate : 0.6774

P-Value [Acc > NIR] : 1.194e-10

Kappa : 0.6931

Mcnemar's Test P-Value : 1

Sensitivity : 0.9048

Specificity : 0.7857

Pos Pred Value : 0.8986

Neg Pred Value : 0.7971

Prevalence : 0.6774

Detection Rate : 0.6129

Detection Prevalence : 0.6820

Balanced Accuracy : 0.8452

'Positive' Class : Abnormal

***#Checking confusion matrix on Train data***

**predtest <- predict(fit,testSplit, type = "class" )**

**confusionMatrix(predtest,testSplit$Status)**

Confusion Matrix and Statistics

Reference

Prediction Abnormal Normal

Abnormal 51 4

Normal 12 26

Accuracy : 0.828

95% CI : (0.7357, 0.8983)

No Information Rate : 0.6774

P-Value [Acc > NIR] : 0.0008401

Kappa : 0.632

Mcnemar's Test P-Value : 0.0801183

Sensitivity : 0.8095

Specificity : 0.8667

Pos Pred Value : 0.9273

Neg Pred Value : 0.6842

Prevalence : 0.6774

Detection Rate : 0.5484

Detection Prevalence : 0.5914

Balanced Accuracy : 0.8381

'Positive' Class : Abnormal