**PRACTICAL-3**

**Implement Decision Tree using ID3 Algorithm**

**CODE:**

library(rpart)

> library(rpart.plot)

> data("ptitanic")

> str(ptitanic)

'data.frame': 1309 obs. of 6 variables:

$ pclass : Factor w/ 3 levels "1st","2nd","3rd": 1 1 1 1 1 1 1 1 1 1 ...

$ survived: Factor w/ 2 levels "died","survived": 2 2 1 1 1 2 2 1 2 1 ...

$ sex : Factor w/ 2 levels "female","male": 1 2 1 2 1 2 1 2 1 2 ...

$ age :Class 'labelled' atomic [1:1309] 29 0.917 2 30 25 ...

.. ..- attr(\*, "units")= chr "Year"

.. ..- attr(\*, "label")= chr "Age"

$ sibsp :Class 'labelled' atomic [1:1309] 0 1 1 1 1 0 1 0 2 0 ...

.. ..- attr(\*, "label")= chr "Number of Siblings/Spouses Aboard"

$ parch :Class 'labelled' atomic [1:1309] 0 2 2 2 2 0 0 0 0 0 ...

.. ..- attr(\*, "label")= chr "Number of Parents/Children Aboard"

> set.seed(123)

> tree <- rpart(survived ~ ., data = ptitanic, control = rpart.control(cp = 0.0001) )

>

> printcp(tree)

Classification tree:

rpart(formula = survived ~ ., data = ptitanic, control = rpart.control(cp = 1e-04))

Variables actually used in tree construction:

[1] age parch pclass sex sibsp

Root node error: 500/1309 = 0.38197

n= 1309

CP nsplit rel error xerror xstd

1 0.4240000 0 1.000 1.000 0.035158

2 0.0210000 1 0.576 0.576 0.029976

3 0.0150000 3 0.534 0.570 0.029863

4 0.0113333 5 0.504 0.566 0.029787

5 0.0025714 9 0.458 0.530 0.029076

6 0.0020000 16 0.440 0.530 0.029076

7 0.0001000 18 0.436 0.534 0.029157

> bestcp <- tree$cptable[which.min(tree$cptable[,"xerror"]),"CP"]

> tree.pruned <-prune(tree, cp = bestcp)

> conf.matrix <- table(ptitanic$survived,predict(tree.pruned,type = "class"))

> rownames(conf.matrix) <- paste("Actual", rownames(conf.matrix), sep = ":")

> colnames(conf.matrix) <- paste("pred", colnames(conf.matrix), sep = ":")

> print(conf.matrix)

pred:died pred:survived

Actual:died 749 60

Actual:survived 169 331

> plot(tree.pruned)

> text(tree.pruned, cex = 0.8, use.n = TRUE, xpd = TRUE)

Error in text.default(xy$x, xy$y + 0.5 \* cxy[2L], rows[left.child], ...) :

plot.new has not been called yet

> prp(tree.pruned, faclen = 0, cex = 0.8, extra = 1)

> tot\_count <- function(x, labs, digits, varlen)

+ {

+ paste(labs,"\n\nn =", x$frame$n)

+ }

> prp(tree.pruned,faclen = 0, cex = 0.8, node.fun = tot\_count)

> only\_count <- function(x, labs, digits, varlen)

+ {

+ paste(x$frame$n)

+ }

> boxcols <- c("pink","palegreen3")[tree.pruned$frame$yval]

> par(xpd = TRUE)

> prp(tree.pruned, faclen = 0, cex = 0.8, node.fun = only\_count, box.col = b oxcols)

> legend("bottomleft", legend = c("died","survived"), fill = c("pink","palegreen3"), title = "Group")

**GRAPH:**







