

```
#Solutions for DATA230, Decision Tree and CART Assignment_ptitanic (Fall 2021)

#install.packages("naniar")
#install.packages("rpart")
#install.packages("rpart.plot")

#step 1
library(naniar)
library(rpart)
library(rpart.plot)
data("ptitanic")

#step 2
any_na(ptitanic)

#step 3
n <- nrow(ptitanic)
t_idx <- sample(seq_len(n), size = round(0.7 * n)) #random sample 70% data as training data,
the rest 30% is the test data
traindata <- ptitanic[t_idx,]
testdata <- ptitanic[ - t_idx,]

#step 4
tree <- rpart(survived ~ ., data = traindata,
              method = "class") #change to anova for numerical

#step 5
printcp(tree) #find the best/optimal stopping point (CP, complexity paramater)
tree.pruned <- prune(tree,cp = tree$cptable[which.min(tree$cptable[, "xerror"]), "CP"])

#step 6
rpart.plot(tree.pruned)

#step 7
future <- predict(tree.pruned, testdata, type="class")
future <- as.data.frame(future)

#step 8
final <- cbind(future, testdata)
confusion <- table(final$survived, final$future, dnn = c("truth", "predicted"))
confusion

#step 9
accuracy <- sum(diag(confusion)) / sum(confusion)
accuracy
```