“**Experiment 1.5”**

**Student Name: SUMIT KUMAR UID: 20BCS8226**

**Branch: CSE Section/Group: DM-720 A**

**Semester: 6th Date of Performance: 03-04-23**

**Subject Name: Data Mining Lab Subject Code: 20CSP-376**

**Aim:**

To perform the classification by decision tree induction using WEKA tools.

**Objective:**

**Decision tree algorithm** falls under the category of supervised learning. They can be used to solve both **regression** and **classification problems**. Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of the tree. We can represent any boolean function on discrete attributes using the decision tree.

**Code:**

library(RWeka)

library(partykit)

library(caTools)

iris\_data= iris

str(iris\_data)

summary(iris\_data)

spl= sample.split(iris\_data,SplitRatio=0.5)

dataTrain= subset(iris\_data,spl==TRUE)

dataTest= subset(iris\_data,spl==FALSE)

m1<-J48(Species~.,dataTrain)

summary(m1)

dataTestPred <- predict(m1,newdata=dataTest)

table\_matrix <- table(dataTest$Species, dataTestPred)

print(table\_matrix)

accuracy\_Test <- sum(diag(table\_matrix))/sum(table\_matrix)

cat("Test Accuracy is: ", accuracy\_Test)

#Initiate PDF File

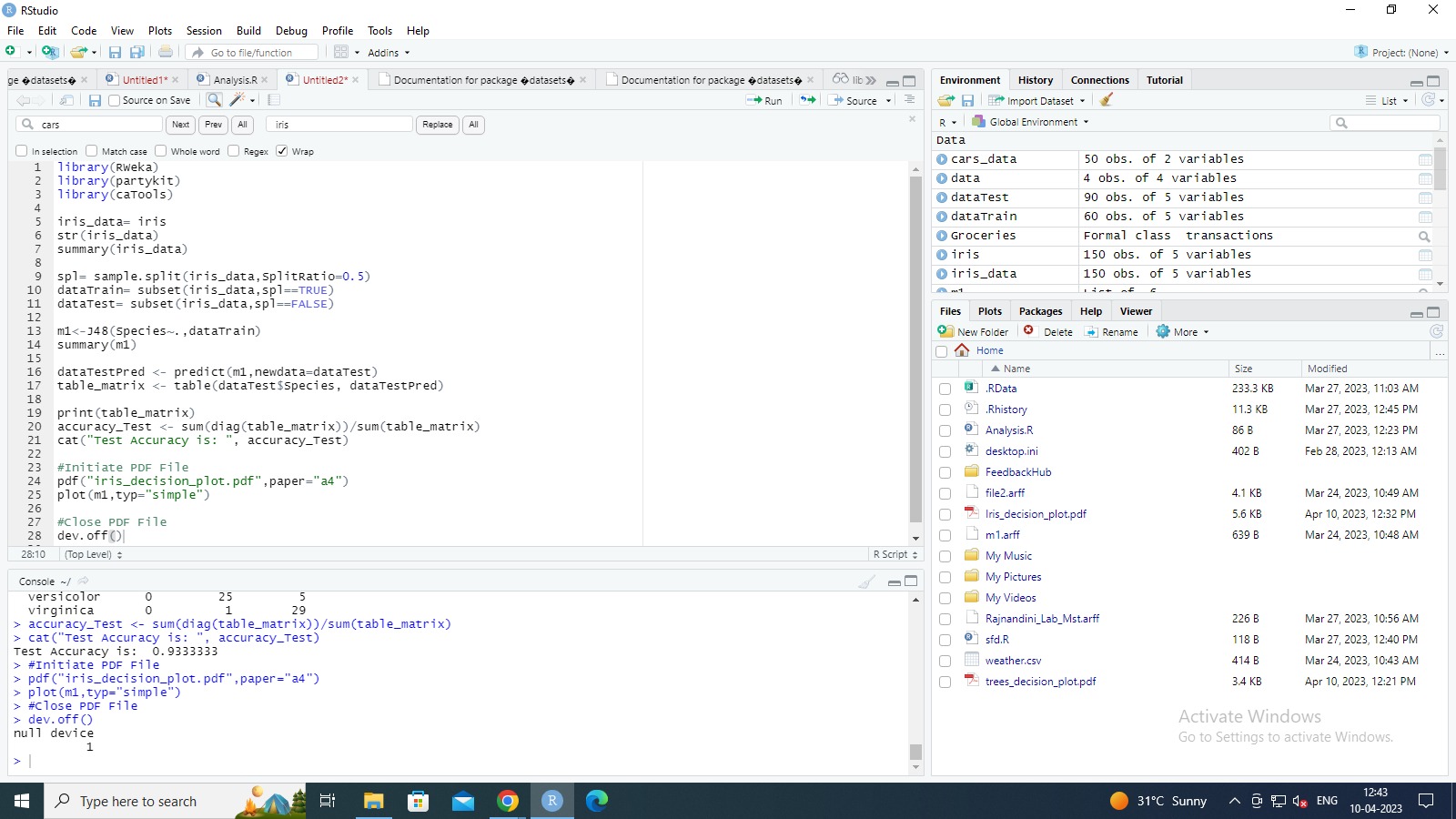
pdf("iris\_decision\_plot.pdf",paper="a4")

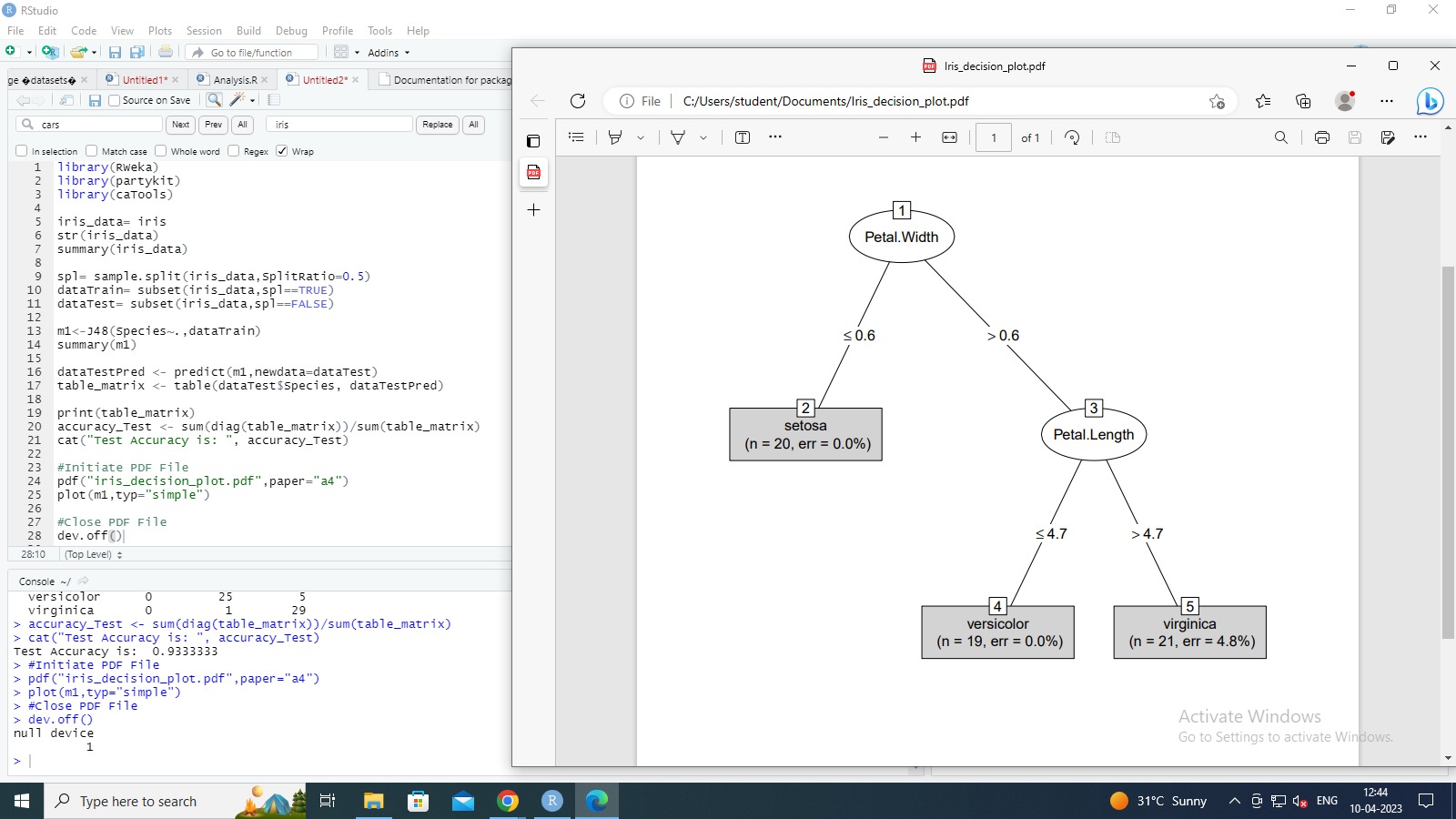
plot(m1,typ="simple")

#Close PDF File

dev.off()

**Output:**





## Observations & Conclusion:

While building Decision Tree using Information Gain, the essentials are to start with all training instances associated with the root node, Use info gain to choose which attribute to label each node with, and no root-to-leaf path should contain the same discrete attribute twice. Also, recursively construct each subtree on the subset of training instances that would be classified down that path in the tree. If all positive or all negative training instances remain, the label that node “yes” or “no” accordingly. If no attributes remain, label with a majority vote of training instances left at that node. If no instances remain, label with a majority vote of the parent’s training instances.

## Learning outcomes (What I have learnt):

1. Learn that feature values are preferred to be categorical.
2. Leant that if the values are continuous then they are discretized prior to building the model.
3. Learn that on the basis of attribute values, records are distributed recursively.
4. Learnt that we use statistical methods for ordering attributes as root or the internal node.