EXPERIMENT-12

Problem Statement: To implement decision trees using 'reading Skills' dataset.

Theoretical Background

Decision trees:

Decision tree induction is the learning of decision trees from class-labeled training tuples. A decision tree is a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node.

Decision tree is a graph to represent choices and their results in form of a tree. The nodes in the graph represent an event or choice and the edges of the graph represent the decision rules or conditions. It is mostly used in Machine Learning and Data Mining applications using R.

Package:

The R package **"party"** is used to create decision trees.

Installation: Install.packages("party")

Note: R Studio Version should be 4.1.2

The package "party" has the function **ctree()** which is used to create and analyze decison tree.

Syntax

The basic syntax for creating a decision tree in R is −

```
ctree(formula, data)
```

Following is the description of the parameters used −

- formula is a formula describing the predictor and response variable.
- data is the name of the data set used.
- Input Data
- We will use the R in-built data set named readingSkills to create a decision tree. It describes the score of someone's readingSkills if we know the variables "age","shoesize","score" and whether the person is a native speaker or not.
- Here is the sample data.

Program Code:

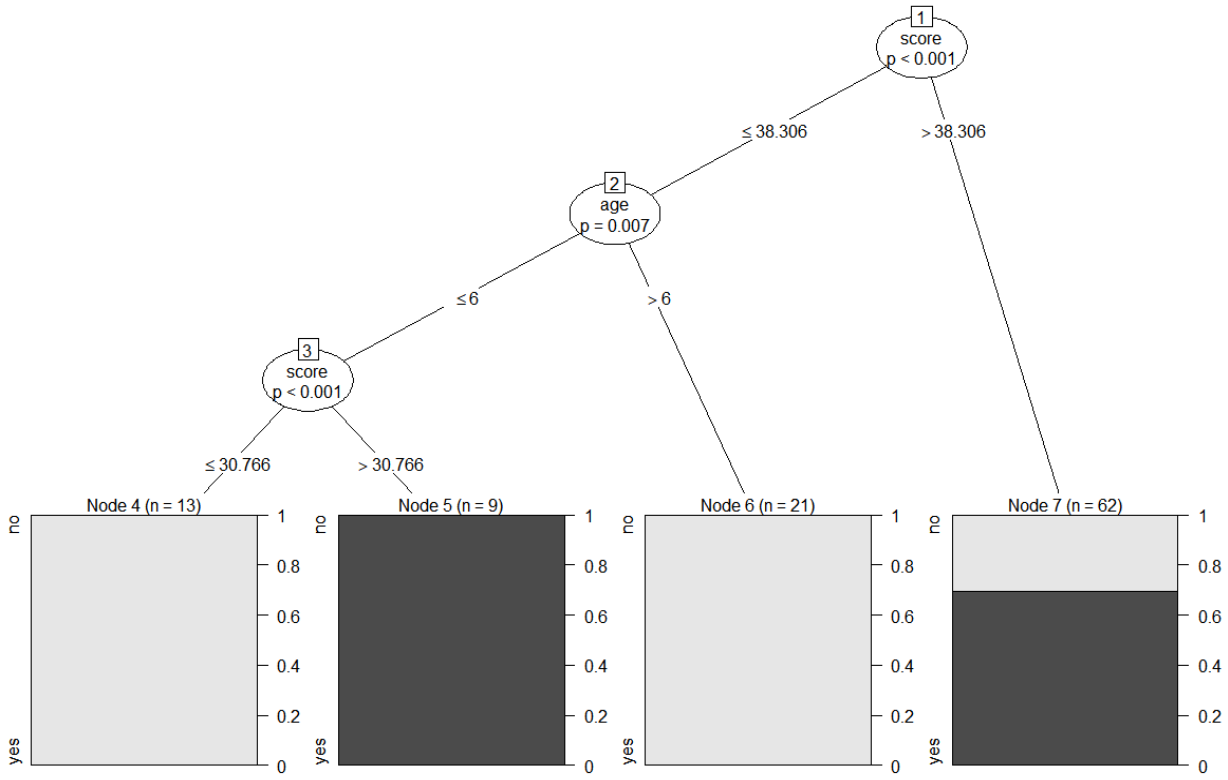
```
install.packages("party")
library(party)
print(head(readingSkills))
input.dat<-readingSkills[c(1:105),]
png(file="decision_tree.png")
output.tree<-ctree(
 nativeSpeaker~ age+ shoesize + core,data = input.dat
)
plot(output.tree)
dev.off()
```

Program Output:

```
> print(head(readingSkills))
   nativeSpeaker age shoeSize    score
  1           yes   5 24.83189 32.29385
  2           yes   6 25.95238 36.63105
  3           no    11 30.42170 49.60593
  4           yes   7 28.66450 40.28456
  5           yes   11 31.88207 55.46085
  6           yes   10 30.07843 52.83124
>
> input.dat<-readingSkills [ c (1:105),]
> png(file="decision_tree.png")
> output.tree<-ctree(
+   nativeSpeaker ~ age + shoesize + score,data = input.dat
```

+)

> plot(output.tree)



> dev.off()

null device

1