Decision Tree Tutorial Part 3 - Reading Skills

John Kelly

12 February 2019

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.

Step 1: Load required packages and libraries

```
# Load the party package. It will automatically load other
# dependent packages.
install.packages("party", dependencies = TRUE)
```

```
## Installing package into 'C:/Users/jpkel/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
```

```
## package 'party' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\jpkel\AppData\Local\Temp\RtmpG8viI0\downloaded_packages
```

```
install.packages("TH.data", dependencies = TRUE)
```

```
## Installing package into 'C:/Users/jpkel/OneDrive/Documents/R/win-library/3.5'
## (as 'lib' is unspecified)
```

```
## package 'TH.data' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\jpkel\AppData\Local\Temp\RtmpG8viI0\downloaded_packages
```

```
library(party)
```

```
## Loading required package: grid
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: modeltools
```

```
## Loading required package: stats4
```

```
## Loading required package: strucchange
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
## Loading required package: sandwich
```

```
# Print some records from data set readingSkills.
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
```

```
# Load the party package. It will automatically load other
# dependent packages.
library(party)
```

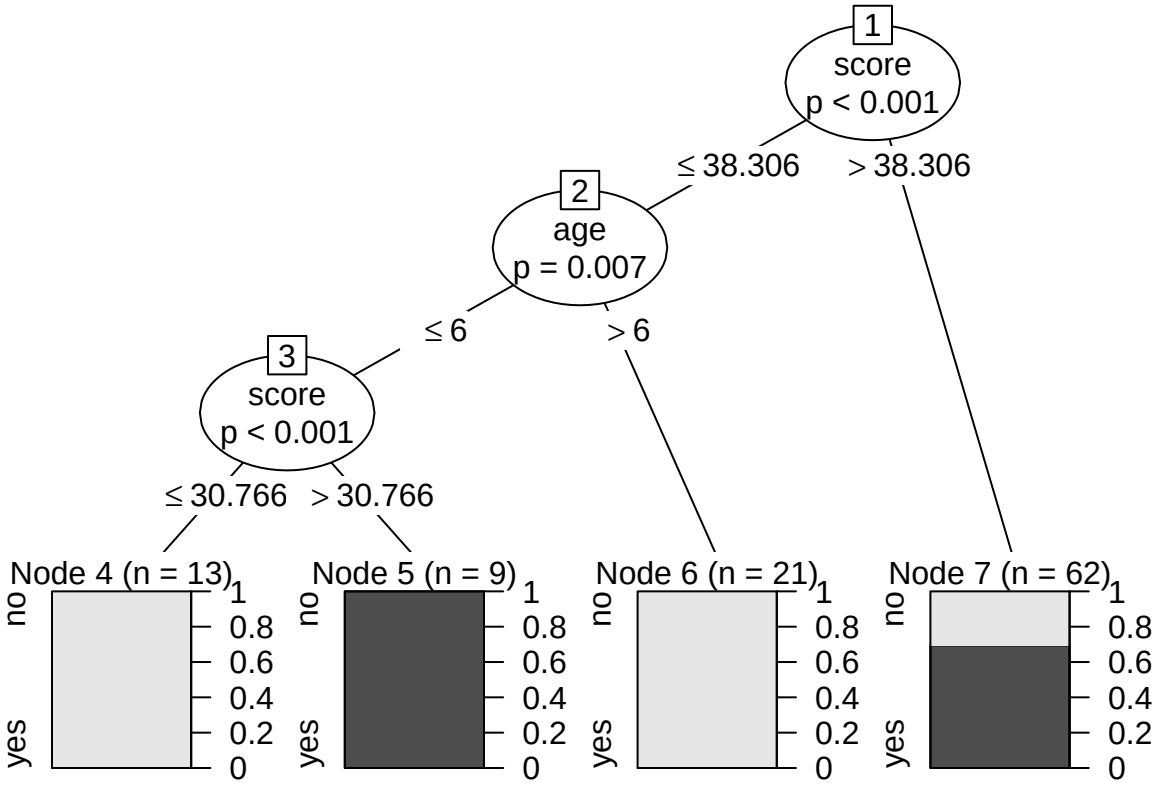
Step 2: Use the ctree algorithm to create the tree and view output in saved png file.

```
# Create the input data frame.
input.dat <- readingSkills[c(1:105),]

# Give the chart file a name.
#png(file = "decision_tree.png")

# Create the tree.
output.tree <- ctree(
  nativeSpeaker ~ age + shoeSize + score,
  data = input.dat)

# Plot the tree.
plot(output.tree)
```




```
# Save the file.
dev.off()
```

```
## null device
##           1
```

Step 3: Conclusion

From the decision tree shown above we can conclude that anyone whose readingSkills score is less than 38.3 and age is more than 6 is not a native Speaker.