Weather Prediction

2022-12-01

```
#Read the data
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plyr)
```

```
## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
```

```
data<-read.csv("weather.csv")
head(data)
```

```
##   MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed
## 1     8.0    24.3      0.0         3.4      6.3          NW            30
## 2    14.0    26.9      3.6         4.4      9.7         ENE            39
## 3    13.7    23.4      3.6         5.8      3.3          NW            85
## 4    13.3    15.5     39.8         7.2      9.1          NW            54
## 5     7.6    16.1      2.8         5.6     10.6         SSE            50
```

```
## 6     6.2    16.9      0.0         5.8      8.2          SE            44
##   WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm
## 1         SW         NW            6           20          68          29
## 2          E          W            4           17          80          36
## 3          N        NNE            6            6          82          69
## 4        WNW          W           30           24          62          56
## 5        SSE        ESE           20           28          68          49
## 6         SE          E           20           24          70          57
##   Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday
## 1      1019.7      1015.0        7        7    14.4    23.6        No
## 2      1012.4      1008.4        5        3    17.5    25.7       Yes
## 3      1009.5      1007.2        8        7    15.4    20.2       Yes
## 4      1005.5      1007.0        2        7    13.5    14.1       Yes
## 5      1018.3      1018.5        7        7    11.1    15.4       Yes
## 6      1023.8      1021.7        7        5    10.9    14.8        No
##   RainTomorrow
## 1          Yes
## 2          Yes
## 3          Yes
## 4          Yes
## 5           No
## 6           No
```

Changing boolean values to 0,1

```
#data$RainToday <- revalue(data$RainToday, c("Yes"=1))
#data$RainTomorrow <- revalue(data$RainTomorrow, c("Yes"=1))
#data$RainToday <- revalue(data$RainToday, c("No"=0))
#data$RainTomorrow <- revalue(data$RainTomorrow, c("No"=0))
data$RainToday=as.factor(data$RainToday)
data$RainTomorrow=as.factor(data$RainTomorrow)
head(data)
```

```
##   MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir WindGustSpeed
## 1     8.0    24.3      0.0         3.4      6.3          NW            30
## 2    14.0    26.9      3.6         4.4      9.7         ENE            39
## 3    13.7    23.4      3.6         5.8      3.3          NW            85
## 4    13.3    15.5     39.8         7.2      9.1          NW            54
## 5     7.6    16.1      2.8         5.6     10.6         SSE            50
## 6     6.2    16.9      0.0         5.8      8.2          SE            44
##   WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am Humidity3pm
## 1         SW         NW            6           20          68          29
## 2          E          W            4           17          80          36
## 3          N        NNE            6            6          82          69
## 4        WNW          W           30           24          62          56
## 5        SSE        ESE           20           28          68          49
## 6         SE          E           20           24          70          57
##   Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm RainToday
## 1      1019.7      1015.0        7        7    14.4    23.6        No
## 2      1012.4      1008.4        5        3    17.5    25.7       Yes
## 3      1009.5      1007.2        8        7    15.4    20.2       Yes
## 4      1005.5      1007.0        2        7    13.5    14.1       Yes
## 5      1018.3      1018.5        7        7    11.1    15.4       Yes
```

```
## 6      1023.8      1021.7        7        5    10.9    14.8        No
##   RainTomorrow
## 1          Yes
## 2          Yes
## 3          Yes
## 4          Yes
## 5           No
## 6           No
```

Remove missing values

```
data=na.omit(data)
```

SPLITTING THE DATA

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
index <- sample(c(TRUE, FALSE), nrow(data), replace=TRUE, prob=c(0.80,0.20))
trainData<-data[index,]
testData<-data[!index,]
dim(trainData)
```

```
## [1] 257  21
```

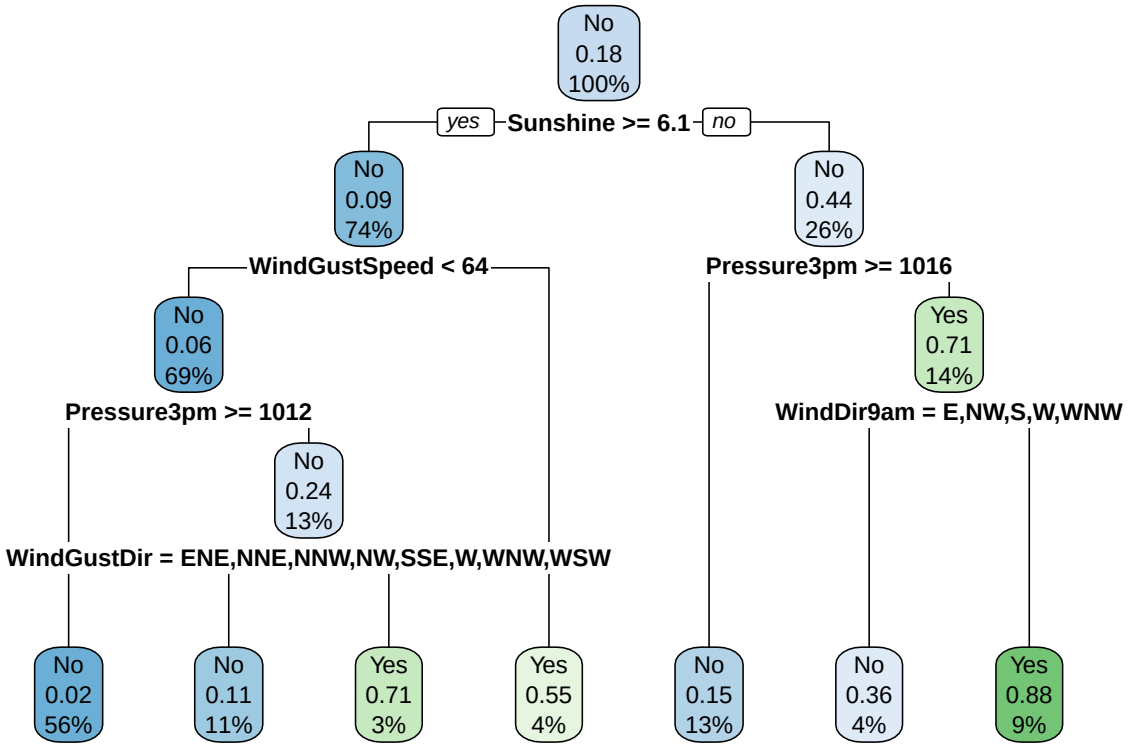
```
dim(testData)
```

```
## [1] 71 21
```

BUILDING THE DECISION TREE MODEL

```
library(rpart)
library(rpart.plot)
```

```
model=rpart(RainTomorrow~.,trainData)
rpart.plot(model)
```




```
Prediction <- predict(model, testData, type= "class")
confusionMatrix(Prediction, testData$RainTomorrow, positive= "Yes")
```

```
## Confusion Matrix and Statistics
##
##          Reference
## Prediction No Yes
##        No  56   8
##        Yes  2   5
##
##               Accuracy : 0.8592
##                 95% CI : (0.7562, 0.9303)
##    No Information Rate : 0.8169
##    P-Value [Acc > NIR] : 0.2256
##
##                  Kappa : 0.4265
##
##  Mcnemar's Test P-Value : 0.1138
##
##             Sensitivity : 0.38462
##             Specificity : 0.96552
##          Pos Pred Value : 0.71429
##          Neg Pred Value : 0.87500
##              Prevalence : 0.18310
##          Detection Rate : 0.07042
```

```
##    Detection Prevalence : 0.09859
##       Balanced Accuracy : 0.67507
##
##        'Positive' Class : Yes
##
```

The Accuracy of the model is 87.7%