

Will it Rain Tomorrow in Australia?

2023-07-15

Contents

1	Import R Libraries	1
2	Data Preparation	1
2.1	Download the Rain Dataset	1

1 Import R Libraries

```
library(corrplot)
library(ggplot2)
library(caret)
library(magrittr)
library(gridExtra)
library(scales)
library(DMwR2)
library(UBL)
library(caret)
library(MASS)
library(ipred)
library(rsample)
library(mlr)
library(knitr)
library(glmnet)
library(outliers)
library(class)
library(kableExtra)
library(tidyverse)
library(pROC)
library(dplyr)
```

2 Data Preparation

2.1 Download the Rain Dataset

We downloaded the Rain Australia dataset as a CSV from the following link: <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package/code?datasetId=6012&searchQuery=visual>. The dataset is originally composed of 23 columns and 145,460 examples. The aim of the data is to use available information about today's weather, i.e. Temperature, Humidity, Pressure, to predict whether it will rain tomorrow. Therefore, we originally start with 22 features and 1 target variable, RainTomorrow with binary classification (Yes=1, No=0).

```
file_path <- "/Users/Sofia/Desktop/Rain_Australia/weatherAUS.csv"
rain <- read.csv(file_path)
head(rain)
```

```
##      Date Location MinTemp MaxTemp Rainfall Evaporation Sunshine WindGustDir
## 1 2008-12-01  Albury    13.4   22.9     0.6          NA        NA          W
## 2 2008-12-02  Albury     7.4   25.1     0.0          NA        NA        WNW
## 3 2008-12-03  Albury    12.9   25.7     0.0          NA        NA        WSW
## 4 2008-12-04  Albury     9.2   28.0     0.0          NA        NA         NE
## 5 2008-12-05  Albury    17.5   32.3     1.0          NA        NA          W
## 6 2008-12-06  Albury    14.6   29.7     0.2          NA        NA        WNW
##      WindGustSpeed WindDir9am WindDir3pm WindSpeed9am WindSpeed3pm Humidity9am
## 1              44          W        WNW             20             24          71
## 2              44        NNW        WSW              4             22          44
## 3              46          W        WSW             19             26          38
## 4              24          SE         E             11              9          45
## 5              41        ENE        NW              7             20          82
## 6              56          W         W             19             24          55
##      Humidity3pm Pressure9am Pressure3pm Cloud9am Cloud3pm Temp9am Temp3pm
## 1              22      1007.7      1007.1         8        NA      16.9      21.8
## 2              25      1010.6      1007.8        NA        NA      17.2      24.3
## 3              30      1007.6      1008.7        NA         2      21.0      23.2
## 4              16      1017.6      1012.8        NA        NA      18.1      26.5
## 5              33      1010.8      1006.0         7         8      17.8      29.7
## 6              23      1009.2      1005.4        NA        NA      20.6      28.9
##      RainToday RainTomorrow
## 1          No          No
## 2          No          No
## 3          No          No
## 4          No          No
## 5          No          No
## 6          No          No
```

```
summary(rain)
```

```
##      Date      Location      MinTemp      MaxTemp
## Length:145460 Length:145460 Min.   :-8.50 Min.   :-4.80
## Class :character Class :character 1st Qu.: 7.60 1st Qu.:17.90
## Mode  :character Mode  :character Median :12.00 Median :22.60
##                               Mean  :12.19 Mean  :23.22
##                               3rd Qu.:16.90 3rd Qu.:28.20
##                               Max.   :33.90 Max.   :48.10
##                               NA's   :1485  NA's   :1261
##      Rainfall      Evaporation      Sunshine      WindGustDir
## Min.   : 0.000 Min.   : 0.00 Min.   : 0.00 Length:145460
## 1st Qu.: 0.000 1st Qu.: 2.60 1st Qu.: 4.80 Class :character
## Median : 0.000 Median : 4.80 Median : 8.40 Mode  :character
## Mean   : 2.361 Mean   : 5.47 Mean   : 7.61
## 3rd Qu.: 0.800 3rd Qu.: 7.40 3rd Qu.:10.60
## Max.   :371.000 Max.   :145.00 Max.   :14.50
## NA's   :3261 NA's   :62790 NA's   :69835
## WindGustSpeed WindDir9am WindDir3pm WindSpeed9am
## Min.   : 6.00 Length:145460 Length:145460 Min.   : 0.00
## 1st Qu.: 31.00 Class :character Class :character 1st Qu.: 7.00
```

```
## Median : 39.00   Mode :character   Mode :character   Median : 13.00
## Mean   : 40.03                                     Mean   : 14.04
## 3rd Qu.: 48.00                                     3rd Qu.: 19.00
## Max.   :135.00                                     Max.   :130.00
## NA's   :10263                                     NA's   :1767
## WindSpeed3pm   Humidity9am   Humidity3pm   Pressure9am
## Min.   : 0.00   Min.   : 0.00   Min.   : 0.00   Min.   : 980.5
## 1st Qu.:13.00   1st Qu.: 57.00   1st Qu.: 37.00   1st Qu.:1012.9
## Median :19.00   Median : 70.00   Median : 52.00   Median :1017.6
## Mean   :18.66   Mean   : 68.88   Mean   : 51.54   Mean   :1017.6
## 3rd Qu.:24.00   3rd Qu.: 83.00   3rd Qu.: 66.00   3rd Qu.:1022.4
## Max.   :87.00   Max.   :100.00   Max.   :100.00   Max.   :1041.0
## NA's   :3062   NA's   :2654   NA's   :4507   NA's   :15065
## Pressure3pm     Cloud9am     Cloud3pm     Temp9am
## Min.   : 977.1   Min.   :0.00   Min.   :0.00   Min.   : -7.20
## 1st Qu.:1010.4   1st Qu.:1.00   1st Qu.:2.00   1st Qu.:12.30
## Median :1015.2   Median :5.00   Median :5.00   Median :16.70
## Mean   :1015.3   Mean   :4.45   Mean   :4.51   Mean   :16.99
## 3rd Qu.:1020.0   3rd Qu.:7.00   3rd Qu.:7.00   3rd Qu.:21.60
## Max.   :1039.6   Max.   :9.00   Max.   :9.00   Max.   :40.20
## NA's   :15028   NA's   :55888   NA's   :59358   NA's   :1767
## Temp3pm         RainToday         RainTomorrow
## Min.   : -5.40   Length:145460   Length:145460
## 1st Qu.:16.60   Class :character   Class :character
## Median :21.10   Mode  :character   Mode  :character
## Mean   :21.68
## 3rd Qu.:26.40
## Max.   :46.70
## NA's   :3609
```

```
dim(rain)
```

```
## [1] 145460      23
```