

Análisis y modelado de datos

Pablo Domínguez

2022-06-23

```
library(tidyverse)
```

Planteamiento del problema a abordar

Nos encontramos con un conjunto de datos obtenidos a partir de mediciones meteorológicas realizadas por el gobierno de Australia¹. Estos datos, recogidos en distintas localidades, se han capturado realizando mediciones diarias de temperatura, lluvia, evaporación, sol, viento, humedad etc.

En la referencia mencionada advierten que el control de calidad aplicado a la captura de estos datos ha sido limitado, por lo que es posible que existan imprecisiones debidas a datos faltantes, valores acumulados tras varios datos faltantes o errores de varios tipos. Es por este motivo que empezaremos nuestro estudio realizando una revisión de la calidad y estructura del dato. Tras este proceso, construiremos una serie de variables que transformarán el problema y la estructura de datos para que puedan aplicarse los modelos de clasificación supervisada planteados.

Partiendo de la base de datos procesada, la segmentaremos para aplicar varios modelados diferentes por zonas (siguiendo cierto criterio). Finalmente, compararemos los modelos, los ensamblaremos y presentaremos unos resultados de la precisión del modelo final.

Con esta aplicación práctica de los modelos teóricos abordados en el capítulo anterior buscamos reflejar la capacidad de herramientas matemáticas abstractas a la hora de resolver situaciones que pueden tener un gran beneficio en varios ámbitos, tales como sociales, económicos o medioambientales.

Origen de los datos y variable objetivo

El buró de meteorología australiano coordina una serie de estaciones meteorológicas repartidas a lo largo del territorio. De esta manera, recopila y reporta datos sobre mediciones meteorológicas. En nuestro caso,

```
db <- read.csv("../db/weatherAUS.csv", stringsAsFactors = TRUE)
db$Date <- as.Date(db$Date, format="%Y-%m-%d")
attach(db)
db %>% head()
```

##	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

¹Notes about Daily Weather Observations - Australian Government (2004)

```
## 6 2008-12-06   Albany   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
```

```
# Voy a descomponer en:
# - trend
# - seasonality
# - noise
#
# SMA(n):moving average of last n days --> smoothing
# decompose()
```

```
str(db)
```

```
## 'data.frame':   145460 obs. of  23 variables:
## $ Date       : Date, format: "2008-12-01" "2008-12-02" ...
## $ Location    : Factor w/ 49 levels "Adelaide","Albany",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ MinTemp     : num  13.4 7.4 12.9 9.2 17.5 14.6 14.3 7.7 9.7 13.1 ...
## $ MaxTemp     : num  22.9 25.1 25.7 28 32.3 29.7 25 26.7 31.9 30.1 ...
## $ Rainfall    : num  0.6 0 0 0 1 0.2 0 0 0 1.4 ...
## $ Evaporation : num  NA NA NA NA NA NA NA NA NA NA ...
## $ Sunshine    : num  NA NA NA NA NA NA NA NA NA NA ...
## $ WindGustDir  : Factor w/ 16 levels "E","ENE","ESE",...: 14 15 16 5 14 15 14 14 7 14 ...
## $ WindGustSpeed: int   44 44 46 24 41 56 50 35 80 28 ...
## $ WindDir9am   : Factor w/ 16 levels "E","ENE","ESE",...: 14 7 14 10 2 14 13 11 10 9 ...
## $ WindDir3pm   : Factor w/ 16 levels "E","ENE","ESE",...: 15 16 16 1 8 14 14 14 8 11 ...
## $ WindSpeed9am : int   20 4 19 11 7 19 20 6 7 15 ...
## $ WindSpeed3pm : int   24 22 26 9 20 24 24 17 28 11 ...
## $ Humidity9am  : int   71 44 38 45 82 55 49 48 42 58 ...
## $ Humidity3pm  : int   22 25 30 16 33 23 19 19 9 27 ...
## $ Pressure9am  : num  1008 1011 1008 1018 1011 ...
## $ Pressure3pm  : num  1007 1008 1009 1013 1006 ...
## $ Cloud9am     : int    8 NA NA NA 7 NA 1 NA NA NA ...
## $ Cloud3pm     : int   NA NA 2 NA 8 NA NA NA NA NA ...
```

```
## $ Temp9am      : num  16.9 17.2 21 18.1 17.8 20.6 18.1 16.3 18.3 20.1 ...
## $ Temp3pm      : num  21.8 24.3 23.2 26.5 29.7 28.9 24.6 25.5 30.2 28.2 ...
## $ RainToday    : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 2 ...
## $ RainTomorrow : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 2 1 ...
```

```
for(city in levels(db$Location)){
}
```

Australian Government. 2004. "Notes about Daily Weather Observations." Bureau of Meteorology. <http://www.bom.gov.au/climate/dwo/IDCJDW0000.pdf>.