

Dataset

El dataset seleccionado para realizar los experimentos consiste en datos relevantes para el estudio del sueño. Los datos se obtuvieron a través de Kaggle, en el siguiente link. El dataset consiste de 374 observaciones con 13 variables:

- Numéricas:
 - Age
 - Sleep duration (hrs)
 - Quality of sleep (1-10)
 - Physical activity level (min/día)
 - Stress level (1-10)
 - Blood pressure
 - Heart rate
- Categóricas
 - Gender
 - Occupation
 - BMI Category
 - Sleep disorder

El objetivo es construir un modelo para predecir si una persona posee algún sleep disorder en base a variables de salud, ocupación, duración y calidad de sueño, etc.

Se debe transformar la columna de “Blood pressure” que viene dada en el formato yyy/xx a 2 columnas diferentes una para cada valor: yyy, xx. Por otro lado, se debe transformar la columna “Sleep disorder” en etiquetas binarias: posee o no algún sleep disorder.

```
sleep_data <- read.csv("data/sleep_health.csv")
```

Transformamos también la variable a predecir, dado que el original viene con los valores None o el nombre del disorder.

```
sleep_data$Sleep.Disorder[sleep_data$Sleep.Disorder != "None"] <- "Yes"
sleep_data$Sleep.Disorder[sleep_data$Sleep.Disorder == "None"] <- "No"
```

```
# Cast column and split measurements
sleep_data$Blood.Pressure <- as.character(sleep_data$Blood.Pressure)
split_bp <- strsplit(sleep_data$Blood.Pressure, "/")

# Create new columns for systolic and diastolic
sleep_data$systolic <- as.numeric(sapply(split_bp, `[`, 1))
sleep_data$diastolic <- as.numeric(sapply(split_bp, `[`, 2))

# Remove the original Blood Pressure column if desired
sleep_data$Blood.Pressure <- NULL
```

```
# Write the modified DataFrame to a CSV file with custom options
write.table(sleep_data, "data/sleep_health_proc.csv", sep = ",", col.names = TRUE, row.names = FALSE)
```

```
sleep_data <- read.csv("./data/sleep_health_proc.csv", sep = ",")
head(sleep_data, n=5)
```

```
##  Person.ID Gender Age      Occupation Sleep.Duration Quality.of.Sleep
## 1         1   Male  27   Software Engineer         6.1             6
## 2         2   Male  28         Doctor         6.2             6
## 3         3   Male  28         Doctor         6.2             6
## 4         4   Male  28 Sales Representative     5.9             4
## 5         5   Male  28 Sales Representative     5.9             4
##  Physical.Activity.Level Stress.Level BMI.Category Heart.Rate Daily.Steps
## 1                42          6   Overweight       77       4200
## 2                60          8     Normal       75      10000
## 3                60          8     Normal       75      10000
## 4                30          8       Obese       85       3000
## 5                30          8       Obese       85       3000
##  Sleep.Disorder systolic diastolic
## 1             No      126       83
## 2             No      125       80
## 3             No      125       80
## 4            Yes      140       90
## 5            Yes      140       90
```