

# Generazione di modelli per predire l'interesse dei clienti per l'acquisto di una polizza

Francesco Gozzoli

29/10/2021

## INTRODUZIONE

Il dataset contiene informazioni sull'interesse di clienti di una compagnia di assicurazioni per l'acquisto di polizze auto. Inizialmente le colonne che compongono il dataset si trovano nella situazione mostrata dalla funzione `summary`.

```
insurance <- read.csv("insurance.csv")
summary(insurance)
```

```
##      id      Gender      Age      Driving_License
## Min.   :    1  Length:381109  Min.   :20.00  Min.   :0.0000
## 1st Qu.: 95278  Class :character  1st Qu.:25.00  1st Qu.:1.0000
## Median :190555  Mode  :character  Median :36.00  Median :1.0000
## Mean   :190555              Mean   :38.82  Mean   :0.9979
## 3rd Qu.:285832              3rd Qu.:49.00  3rd Qu.:1.0000
## Max.   :381109              Max.   :85.00  Max.   :1.0000
## Region_Code  Previously_Insured  Vehicle_Age      Vehicle_Damage
## Min.   : 0.00  Min.   :0.0000  Length:381109  Length:381109
## 1st Qu.:15.00  1st Qu.:0.0000  Class :character  Class :character
## Median :28.00  Median :0.0000  Mode  :character  Mode  :character
## Mean   :26.39  Mean   :0.4582
## 3rd Qu.:35.00  3rd Qu.:1.0000
## Max.   :52.00  Max.   :1.0000
## Annual_Premium  Policy_Sales_Channel  Vintage      Response
## Min.   : 2630  Min.   : 1      Min.   : 10.0  Min.   :0.0000
## 1st Qu.: 24405  1st Qu.: 29      1st Qu.: 82.0  1st Qu.:0.0000
## Median : 31669  Median :133      Median :154.0  Median :0.0000
## Mean   : 30564  Mean   :112      Mean   :154.3  Mean   :0.1226
## 3rd Qu.: 39400  3rd Qu.:152      3rd Qu.:227.0  3rd Qu.:0.0000
## Max.   :540165  Max.   :163      Max.   :299.0  Max.   :1.0000
```

Per poter essere utilizzate efficientemente, alcune colonne hanno bisogno di essere manipolate. In particolare:

- La colonna `Gender`, viene trasformata nella colonna `Male`, che ammette valori binari 0 e 1 dove 1 rappresenta il sesso maschile e 0 quello femminile.

```
insurance[which(insurance$Gender == "Male"),]$Gender = 1
insurance[which(insurance$Gender == "Female"),]$Gender = 0
colnames(insurance)[2] = "Male"
```

- La colonna `Vehicle_Age` viene convertita da categorica ordinata a numerica. I valori utilizzati sono:
  - -1 per le auto immatricolate da meno di un anno.
  - 0 per le auto immatricolate tra gli 1 e i 2 anni precedenti.
  - 1 per le auto immatricolate da più di 2 anni.

```
insurance[which(insurance$Vehicle_Age == "< 1 Year"),]$Vehicle_Age <- -1  
insurance[which(insurance$Vehicle_Age == "1-2 Year"),]$Vehicle_Age <- 0  
insurance[which(insurance$Vehicle_Age == "> 2 Years"),]$Vehicle_Age <- 1
```