

StatWars: models predictius

Arnau Garcia

2024-10-14

StatWars: els meus primers models predictius

En aquesta activitat, us volem presentar una eina molt poderosa que utilitzem en estadística per entendre millor el món que ens envolta: els **models predictius**. Aquests models ens ajuden a predir què passarà en el futur a partir de dades que ja coneixem. A través de conceptes senzills com les **regressions lineals** i les **regressions lineals generalitzades**, veurem com podem utilitzar les matemàtiques per establir relacions entre diferents variables i fer prediccions amb aquestes. Descobrireu com la combinació d'estadística i matemàtiques ens pot ajudar a prendre decisions basades en dades, d'una manera més informada i precisa.

Aquest document serveix com a exemple de com ajustar dos models predictius fent servir el llenguatge de programació **R**. **R**, [https://es.wikipedia.org/wiki/R/_\(lenguaje_de_programaci%C3%B3n\)](https://es.wikipedia.org/wiki/R/_(lenguaje_de_programaci%C3%B3n)), és el principal llenguatge de programació pels estadístics. Es tracta d'un software lliure, gratuït que es troba a l'abast de tothom.

Veurem dos models predictius:

- Predicció de preus d'immobles.
- Predicció d'espècies de pingüins.

Predicció de preus d'immobles

Farem servir la següent taula de dades:

```
cases <- read.csv("House Price Prediction Dataset.csv", header = TRUE)
head(cases)
```

##		Id	Area	Bedrooms	Bathrooms	Floors	YearBuilt	Location	Condition	Garage	Price
## 1	1	1360		5	4	3	1970	Downtown	Excellent	No	149919
## 2	2	4272		5	4	3	1958	Downtown	Excellent	No	424998
## 3	3	3592		2	2	3	1938	Downtown	Good	No	266746
## 4	4	966		4	2	2	1902	Suburban	Fair	Yes	244020
## 5	5	4926		1	4	2	1975	Downtown	Fair	Yes	636056
## 6	6	3944		1	2	1	1906	Urban	Poor	No	93262

En el codi d'R d'amunt, podeu veure com llegim la taula de dades, i el nom que li posem (**cases**). També veiem com fer servir la funció **head()** per ensenyar les primeres observacions de la taula de dades.

Podem observar que en la taula de dades tenim les següents variables:

- **Id**: identificador de l'immoble.
- **Area**: superfície de la casa.
- **Bedrooms**: número d'habitacions.
- **Bathrooms**: número de lavabos.
- **Floors**: número de plantes.
- **YearBuilt**: any en que es va construir l'immoble.

- **Location:** localització, pot ser **Downtown** (centre), **Urban** (urbana), **Suburban** (afores), **Rural** (rural).
- **Condition:** estat en que es troba l'immoble. Pot ser **Excellent** (excel·lent), **Fair** (acceptable), **Good** (bo), **Poor** (pobre).
- **Garage:** garatge a l'immoble. Pot ser **Yes** o **No**.
- **Price:** preu de l'immoble.