**Final project - Card fraud**

Línea horizontal

FECHA

Introduction

The Fraud.csv attachment contains information about many credit and debit card transactions through different channels. For each transaction there is its monetary value and other variables (see the dictionary\_variables.xlsx file). Of particular importance is the FRAUD variable, where 1 appears if the transaction was a fraud or 0 if it was a legitimate transaction.

What are the main objectives in this project?

* Analyze a database.
* Know the data.
* data preparation
* modeling
* final conclusions

1. General analysis

The objective of this practice is to develop a model that allows, based on these data, to predict what the value of the FRAUD variable will be for any transaction.

2. Project organization

Next you will have to create a document where you can **explain in detail** how the current project is **organized**. It is important that it be updated throughout the life of the project.

The document must include at least:

* Exploratory data analysis.
  + Graphics
  + How the variables are distributed
  + Detection of outliers, NA, etc.
  + Others
* Value imputation if necessary (use the missForest library used in an example in class).
* Analysis of dependency/independence of variables
  + Target variable vs independent variables
  + Between independent variables. Is there multicollinearity?
* Modeling
  + Type of model to use
  + Initial interpretation
  + Which variables are significant at first?
  + Stepwise
  + Model evaluation metrics
  + Residue analysis
  + Comparisons between different models
  + Interpretation of coefficients – odds ratio
* Write a paragraph with the conclusions obtained.

3. Requirements

It must be developed in Rmarkdown. Delivering code and pdf with the results.

4. Development

Then you must get down to work and develop a project, it is important that you review the requirements and comply with all of them.

5. Deliverables

Each step that is carried out in practice must have its corresponding written justification, as well as a final summary where the student must write the conclusions obtained from their analysis and modeling.

Explain what lessons you've learned during this project. Conclusions you have obtained from the analyses carried out when necessary

What problems have you encountered when developing this project?

6. Resources

<https://www.cienciadedatos.net/estadistica-con-r.html>

<https://github.com/rvaquerizo02/Estadistica-data-scientist>

<https://www.youtube.com/watch?v=I6FJo8x1wZE>

<https://www.youtube.com/watch?v=KxfC_gzTBVg>