

Predictive Modeling

Master in Big Data Analytics

First Homework: Topics 1 and 2

Deadline: **December 14, 2025 at 23:59**

Upload to Aula Global: data used (or link), notebook (Rmd or Jupyter), and compiled version (html)

General Objectives

- HW1 will consist of two parts: Topic 1 (60%) and Topic 2 (40%).
- You should use open datasets, maybe two different databases for each part, or just one, but remember it is always advisable to use new databases to gain experience and skills.
- The bigger and more diverse the dataset, the better
- Originality is valued: e.g., considering interactions between variables, reducing the dimensionality of the problem, feature extraction, etc.
- For all the considered models, you should provide interpretation, conclusions, and predictive performance

Evaluation

- ① Linear Regression data preparation: 1 point
- ② Linear Regression explanatory analysis: 2 points
- ③ Linear Regression predictive analysis (with interpretation and prediction intervals): 2 points
- ④ Linear Regression originality and report: 1 point
- ⑤ GLM data preparation: 0.5 points
- ⑥ GLM model development and interpretation: 2 points
- ⑦ GLM predictive performance (including prediction intervals): 0.5 points
- ⑧ GLM originality and report: 1 point

Important: It is allowed to take inspiration from other analyses, always citing the source

Remember the deadline: December 14, 2025 at 23:59