MACHINE LEARNING I
MBDS

Instructions for Group Assignment

The file "**Houses for rent in madrid.xlsx**" contains a set of which contains information about 2,100 ads of houses set up for rent in Madrid. The information comes from a Spanish web site, called "idealista.com" and was extracted several years ago.

The assignment has two phases:

1. Segmentation

This task consists of the execution of a segmentation of the data file to find groups of similar houses. You have total freedom to choose the variables that you consider.

With some of the segments found (chose one or two that you think are interesting), proceed with the following task.

2. Groups must choose between one of the following two models:

A. Linear Regression

The task consists of the specification and estimation of a LINEAR REGRESSION MODEL to explain the rental price of the group/s chosen in the previous phase, based on a set of explanatory variables related to their characteristics.

The model may help us:
1. to estimate the rental price if we were a Real Estate agency.
2. to find good opportunities in the market looking for flats that may be below their theoretically estimated price.

B. Classification with Logit Regression

If you choose this alternative, a classification model will have to be created to analyze variables that influence a house that has a rent equal or above 1.800€ a month.

The model may help us:
1. to understand market factors.
2. to allocate new houses fairly to the correct group based on their details

Work in groups. Prepare a report with **4 pages maximum** + **technical Annex** supporting your conclusions, addressed to the Real State agency manager, presenting:

- A clear exposition of your results, defending the final model chosen.
- A short description of the process and technical details of your study
- Conclusions and recommendations.

Please be aware AI tools to write text in this assignment are NOT allowed.

One of the group members must submit your work before the deadline **using only Blackboard assignment tool** (assignments will not be accepted by email).

Deadline: End of February 28th (Madrid time)

Main points that **will be evaluated** are:

- Executive Summary with main findings on first page,
- Data audit, cleaning and data manipulation,
- Technical approach explained,
- Business explanation of segmentation solution,
- Labels and profile description of groups found,
- Correct validation of the supervised model chosen
- Full model evaluation,
- Professional Look, Report clarity and readiness,
- Conclusions and Recommendations,
- Annex,
- Extras, if any.

Remember: 2+ members of any group might report (by sending a private email) over a student who did not participate in the assignment.