SCHOOL OF HUMAN SCIENCES & TECHNOLOGY

MACHINE LEARNING I

Big Data Blended
2020

SUPERVISED GROUP ASSIGNMENT
Instructions

Main data set to be used is the file called "**Houses for rent in madrid_ assignment_2020.xlsx**" which contains information about +2,000 ads of houses set up for rent in Madrid. The information comes from "idealista.com" web site.

Groups must choose between one of the following two assignments:

1. Linear Regression

The task consists of the specification and estimation of a LINEAR REGRESSION MODEL to explain the house rental price, based on a set of explanatory variables related with 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 under their theoretical estimated price.

2. Classification

In this case, a model will have to be created to analyze variables that influence that a house has a rent above 2.000€ a month.

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

Work in groups and create the appropriated model. Prepare a report with **3-4 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 included
- A short description of the process and technical details of your study
- Conclusions and recommendations, if any.
- If classification is chosen, an optional simulation analysis on balancing the errors: how would you do it?

Submit your work before the deadline **using only Blackboard assignment tool** (assignments will not be accepted by email).

Deadline: End of July, 4th

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

- Executive Summary,
- Professional Look,
- Technical approach,
- Data audit,
- Data cleaning/manipulation,
- Model evaluation,
- Conclusions and Recommendations,
- Report clarity and readiness,
- Annex.