

Data & Web Mining Lab Project

Claudio Lucchese claudio.lucchese@unive.it

Lab Project



- **Prediction Task: Home Value Prediction**

<https://www.openml.org/search?type=data&status=active&id=43926&sort=runs>

- **To be delivered:**

- One or more notebooks (and possibly other python files) and a 5-page pdf report discussing your critical choices. No more than 5 pages! The pdf report can be the export of a notebook.
- Project is discussed at the oral exam, about one week after the written exam.
- Deadline: a couple of days before the oral exam via Moodle.

Lab Project

- **Work to be done:**

- Add custom features to your dataset (e.g., avg. prices in the same geographical area) and evaluate their contribution
- Compare at least two prediction methods, after fine-tuning their parameters
- Investigate instances with the most correct predictions and with the most wrong predictions. Do they have some peculiarities? Any strange feature distribution?
- Experiment beyond scikit-learn, e.g., with LightGBM or XGBoost or CatBoost
- Can be done in groups, multiply the work by the size of the group

- **Evaluation:**

- Quality of the report
- Quality of the code (I'll ask about every single line of code)
- Depth of the analysis
- Project can be insufficient (the exam is not passed),
or it may add up to 3 points to the score of the written exam

- **Bonus:**

- rewrite one of my notebooks (mail me first)