Final Machine Learning Project

In a group of no more than four people, students must analyze a set of real data of their choosing, write a report and make a presentation.

The data set must be in tabular form and have a complexity (number of features and instances) proportional to the size of the group (the more students the more complex the problem must be). **Students must consult with the teacher if the data is adequate for the goals of the course before starting the analysis**. The analysis must necessarily involve one or more prediction tasks.

The dataset must belong to the next database: https://archive.ics.uci.edu/

You must include in your report the link (website) where you get your dataset

The analysis should include preprocessing of data, including the calculation of descriptive statistical values of the features, assessment of feature normality and feature correlations, identification of possible outliers and missing values, visual exploration of the data, and analysis of what features provide more information to solve the problem. If appropriate, it should also include feature transformations.

Several predictive models should be properly trained and validated. The evaluation of the models must include (1) their interpretability, (2) their interest, (3) their novelty, (4) their accuracy and (5) the estimated cost of applying the model. It should also include a comparative assessment of the performance of the best models. Keep in mind that a model is more interpretable the fewer features it uses. It is interesting to find precise models, but also interpretable ones. The best model in terms of precision does not have to be the best in interpretability.

The report should explain where the data comes from, what the data to be analyzed is, what analytical tasks are to be performed, and why those tasks are relevant to the problem. The report must be defended individually by each member of the team. The maximum grade a student may get in the defense is the grade assigned to the report, although depending on the student's defense the grade can be lower. **All team members must know the details of how all the contents of the report have been**

derived. The final date for the submission of the project is **May 21st**. **Failure to submit by that date implies not passing the course**.

Students will also have to make a **presentation of the report.** The presentation should last approximately four minutes for each member of the team plus 10 min of teacher questions about the job, with a total of 20-25 min per group. In the event of being part of a group of three students the presentation should last approximately five minutes for each member of the team plus 10 min of teacher's questions