Statistical & Machine Learning

Group Project Description

The project is a group task (max. 3 students/group) where each group has to (i) produce a complete machine learning benchmark experiment, (ii) optimize and debug the machine learning experiment; then (ii) summarize and present the results in a structured format.

Purpose

Your group will work on an **InClass Kaggle competition** and will follow different phases of a predictive modeling project pipeline to implement the machine learning methods discussed in this course. The goals of this project are (1) practicing the full predictive modeling pipeline; and (2) interpreting and explaining the results.

The output of this project should be presented based on the modeling pipeline as follows:

- Problem definition (e.g. business, data science)
- Data (e.g. data exploration, data processing, error correction, feature engineering)
- Modeling (e.g. models, hyper-parameter)
- Experimental setup (e.g. variable selection, resampling, cross-validation, evaluation metrics, debugging)
- Result (e.g. model scoring, model interpretation, variable importance)

Submission

- A data science presentation: 10min presentation + 5min Q&A
- PPT + Jupyter Notebook R/Python script (*.ipynb) + basetable (*.csv)
- Note: Upload the project to GitHub

Evaluation

The project will be evaluated on:

- The correctness of the benchmark experiment setup (correctly set up the main parts of the machine learning pipeline) [40%].
- The degree of depth of the machine learning pipeline (implement and compare different techniques to optimize the result) [30%].
- The quality of the presentation and result interpretation [20%].
- The quality of the R/Python script [10%].
- Bonus: Top 1 group on Kaggle competition = +2 for final grade.
- Note: Using the work (text or programming scripts) of other people without citing the source is considered as plagiarism, and is strictly prohibited.

Timeline information

- Exam Section, 11 April 2022
- Student can use the Data Science Presentation Template (Hackathon) as a reference.