



MACHINE LEARNING II

Big Data – 2025

Instructions for Group Assignment

Every group **must choose its own database** to complete this assignment. Kaggle (or any other website) databases are allowed, only if they do not include a classification model solved on them. In case of doubt, ask the professor for validation.

The goal is to build a **classification** model using a **selected target** variable, too. You need to take all decisions about the work on your own, using the resources we have seen throughout the course.

EXPECTED OUTCOME

Work in the groups you were assigned to and upload the assignment in zip format to blackboard section created, with your group name (i.e. Group Z.zip) including:

1. A **summary report** of **no more than 4 pages** in which you explain:
 - An executive summary of your main findings
 - A brief description of the technical approach to your analysis along with some details on your best solution with conclusions.
 - Recommendations for managers
 - Any other information that you think is relevant
2. A **technical annex** supporting your conclusions
3. The **database** that was used, with some explanation about the variables and their coding
4. The final **notebook**
5. The final ppt **presentation** (or any other format chosen) that will be used.

Please note this is not a report addressed to your professor.

One of the group members must submit the file via blackboard, **before 11:59 PM on March 18th**. **Please note it will not be allowed to use a different presentation file than the one submitted the day before.**

- **50% of the final score** for this assignment will be based on the deliveries. Main points that will be evaluated here are:
 - Executive Summary with main findings,
 - Professional Look (Index, Table of Contents, document structure, etc), report clarity and readiness (highlighting relevant ideas, concise language, etc),
 - Exploration Data Analysis,
 - Data processing, transformations, variable creations, feature selection, etc
 - Technical approach,
 - Validation strategies,
 - Model evaluation,
 - Conclusions & Recommendations,
 - Technical Annex.
- **The other 50% of the final grade** will consider:
 - **Group presentation**
 - **Individual presentation** and **individual ability to answer** the questions posed by the professor.
 - **Peer review** (evaluated based on similarity to the teacher's evaluation).

Please note that the professor might ask any question related to the contents seen in ML2, not only related to the work group done.

- Every group will have a maximum of 9 minutes to present their work, plus 4 minutes to answer questions.
- The **order of the presentations** will be randomly assigned at 9:02am on the presentation day and first group will start at 9:05am sharp (Quickly attendance this day will finish at 09:02am).

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