MACHINE LEARNING II

Big Data PT
2025

Instructions for Group Assignment

The file "Spotify_songs.xlsx" contains a set of songs with information about their popularity, album, artist and some audio features. Variables are explained in the "legend" sheet.

The goal is to build a classification model that tries to predict the **future success (in popularity)** of the songs, based on the provided data.

To do this, each group must create its own discrete version of the popularity variable (since it is metric), which will become the target one, and build a model with the supervised algorithms studied (including optionally dimension reduction), choosing only one final model.

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 final **notebook** explained.
4. 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 23:59 on Nov 23rd. Please note the file (Powerpoint or similar) you submit in the assignment will be the one to be used in the group presentation (no changes will be allowed).**

- **50% of the group assignment score** 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 remaining 50%** will consider:
    - **Group presentation**
    - **Individual presentation** and **individual ability to answer** the questions posed by the professor.

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 15 minutes to present their work, plus 10 minutes to answer questions.
- Please contact the professor to arrange a convenient time and hour for the presentation on the final week (starting Nov 24th).

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.