I. INTRODUCTION

This homework is designed to help you with the final project, covering some steps of your future work. In this task, you will work with the dataset from your Project "To Grant or Not to Grant". You must explore and prepare data, perform feature selection, build a simple model, and assess the performance.

II. DELIVERABLES

The deliverables for this task are:

- Jupyter Notebook (or a zip of multiple notebooks) with your code and markdown comments inside.
- A 2-page PDF file describing your project report's expected structure and contents. For each section, you should detail what information you intend to include in your Project Report.

The file naming convention should follow Homework_GroupXX, where GroupXX should be your group number.

III. TASK & EVALUATION

Notebook:

1. Import the dataset and explore the data (3 points)

- Check data contents, provide descriptive statistics and check for incoherencies in the data.
- Explore data visually and extract **relevant insights.** Explain your rationale and findings. **Do not forget to analyse multivariate relationships.**

2. Clean and pre-process the dataset (5 points)

- Are there any missing values? Take action to handle them.
- Check the dataset for outliers and pre-process them. Justify your decisions.
- Deal with categorical variables.
- Review current features and create extra features if needed. Explain your steps.
- Perform data scaling. Explain the reasoning behind your choices.

3. Feature selection (3 points)

- Define and implement an unambiguous strategy for feature selection. Use methods discussed in the course. **Present and justify your final selection.**

4. Build a simple model and assess the performance (4 points)

- Identify the type of the problem and select the relevant algorithms.
- Use cross-validation to assess performance. **Which metrics did you use and why?**
- **Train at least 1 model using the train dataset and obtain predictions for the test dataset.**

(Extra 1 point) Submit a solution to Kaggle with a macro F1-score of more than 0.6.

PDF file:

5. Describe the overall structure of your final report (5 points)

- Detail your intended contents for each section of your final report.
- Present and explain the rationale of your open-ended section.

Readability and consistency will be considered during grading, so include sufficient comments and maintain a well-defined structure throughout your work.

Final grade = min(20, your points)
Deadline: 11.11.2024 - 11:59 AM