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 "Cars 4 You: We buy your car!". 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's overall structure and rationale.

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 inconsistencies 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 a clear and unambiguous strategy for feature selection. Use the 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 problem and select the relevant algorithms.
- Select one model assessment strategy to use throughout your work. **Which metrics are you using to evaluate your model and why?**
- **Train at least 1 model using the train dataset and obtain predictions for the test dataset.**

(Extra 1 point) Be on the Top-5 Best Groups in the Kaggle Competition

PDF file:

5. Describe the overall structure of your pipeline (5 points)

- Provide a schematic representation of your pipeline (stages and what techniques are used in each stage),
- Provide details on preprocessing and feature selection (what is done, to what variables and why).

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: 03.11.2025 - 17:59