Data Science Project "Navigating Missed Hospital Visits"

Problem: many patients miss their scheduled appointments.
Goal: Develop a machine learning model that predicts if a patient will miss a future appointment.

You will develop, train and evaluate a machine learning model with Python. The project will solve a classification problem with structured data. Teams should work independently and may not share solutions or ideas.

If you use Python templates or build on other solutions described in books or on the Internet, for example, you must cite them as sources.

The final report should be a runnable Jupyter Notebook. The notebook will mix Python code, results, and your text (as Markdown cells). Detailed explanations of your approach within your notebook are mandatory.

Each member of the team must have an active role in the presentation and be able to explain all aspects of the project results. The presentation will be through slides.

The notebook should at least cover the following parts:

1. General information:
    - Name of the used dataset
    - List of all participating students (group has max. 3 students)
2. Business Understanding
    - Short description of the business problem that you want to solve (it is ok to use your own assumptions)
3. Data Understanding (Exploratory Data Analysis)
    - Description of the data set
    - Data visualization
    - Analysis of missing data
4. Data Preparation
    - How and why did your select the features?
    - Did you derive new features?
    - How did you deal with missing data?
    - How and why did you transform the data?
    - How did you split the data?
        - A split into test and training data is required.
        - A cross-evaluation or validation set can also be used to optimize the parameters (train, validation, test set).
5. Model Building
    - Which machine learning models did you use and why?
    - How did you adjust the parameters of your machine learning model and why?
    - What are the resulting machine learning models?
    - A plus is if you build more than one machine learning model.

    - You should add an understandable description of the types of machine learning models you used
6. Model Evaluation
    - Evaluation of the selected models: How will your models perform in practice?
    - Did you compare different models? If yes, which one is better? For what metric?
    - If you have a classification task, try to find a model and parameters with a high accuracy in the test set. Are both errors (false positive and false negative) equally important? How could you adjust for different costs of the errors?
    - Try to find a model and parameters with a high accuracy in the test set.
    - Description of the approach chosen to optimize the parameters
    - Discuss evaluation results (e.g., are false positive and false negative errors equally important? How could you balance the different costs of the errors)?
7. Possible Business Use
    - How would your model help the business? How is it solving the business problem?
    - What is the meaning of your model and your solution from a business perspective?

The presentation will be a recorded video of your team presenting a slide deck.
The duration of your presentation should be limited to 15min.

You should 1) upload your presentation (video and PDF) and 2) upload your notebook (as HTML and PDF) to Moodle.

Grading Criteria:

- Comprehensiveness and correctness of the description of the approach.
- Technical sophistication and creativity of your solution
- Connection of the technical solution to the business problem or benefit
- Correct and transparent citation of all aids and sources
- Discussion and understanding
- Presentation and quality of slides