Project Guidelines: Machine Learning & Data Mining (OSUPR)

Overview

To complete the OSUPR course, students must undertake a project that encompasses knowledge of data mining and machine learning. There are two distinct stages to the project:

1. **Phase 1:** Proposal Selection & Approval (Deadline: Jan 18th)

2. **Phase 2:** Project Execution & Defense (Deadline: August)

   The project can be carried out **individually or in pairs**.

1 Phase 1: Project Selection & Proposal

The topic is up to the student to choose based on their interests. You may choose between three types of projects:

1. **Applied Data Mining/ML:** Select a dataset (e.g., from Kaggle) and apply algorithms to solve a specific problem.

2. **Algorithm Implementation:** Implement a specific data mining or machine learning algorithm from scratch in a programming language of your choice.

3. **Custom Proposal:** If you have a unique idea that meets the course requirements, you may propose it.

The Proposal Document

By the end of the semester, we must agree on your topic to ensure the difficulty is fair and balanced. You must submit a PDF proposal to domen.vake@famnit.upr.si.

 Submission Deadline: The start of the exam period (**January 18th**). You may submit earlier.

The proposal should be approximately **one page** long. **Note: The proposal itself is grounds for grading.** It must include:

- **Student Info:** Name(s) of participating student(s).

- **Problem Description:** A clear description of the problem you wish to tackle.

- **Data:** The dataset(s) that will be used and their descriptions (include links).

- **Methodology:** Modeling algorithms and technologies you intend to use.

- **Work Division:** (If working in a pair) A clear division of tasks between students.

Process: The assistant will assess all proposals to adjust for difficulty. You will receive feedback to adjust the scope if necessary. You must have an accepted topic by the end of the exam period.

2 Phase 2: Development & Defense

Timeline

Once the proposal is approved, you have until **August** to complete the project (see e-classroom submission for exact date).

- You can complete the project earlier.

- **Warning:** Do not wait until the last minute. After the deadline, no corrections are allowed.

Development Requirements (Code)

The development process must be transparent.

- Students must use **GitHub** or **GitLab** for code development (if using Kaggle, you can version your code directly on Kaggle).

- The history of commits must show the development process; submitting only the final code without version history is not acceptable.

- If you wish to use a different platform, you must get approval during Phase 1.

Submission & Defense

1. **Report:** The final submission must include a report detailing all steps in the development of the algorithm/model.

2. **Submission:** Submit the report and a link to your code repository to the e-classroom.

3. **Defense:** Once submitted, you must defend your project.

   - **All defenses are in person.**
   - The specific date/time will be arranged after submission.