

Project Phase – 1

Proposal and Data Exploration/Preprocessing

Due date Monday, May 1st, 2023 (11:59pm).

The course project aims to producing high-quality work involving machine learning (ML). The project will have two phases, the details of which will be given in due times. All students are required to work in groups of three and produce comprehensive technical reports by the end of the project.

In this phase, you need to form a group, propose a project idea, perform data exploration/preprocessing, and initial modeling. There are no restrictions on project ideas as long you have data to train ML models for specific objective(s). Specifically, you should:

- Brainstorm project ideas with your team members. You can consider competing in Kaggle, make use of publicly available datasets, or private data that you have permission to access.
- Produce various insightful summary statistics and visualizations of the data.
- Remove any outliers and suggest plans for missing values and duplicates, if any. Investigate relationships, such as “correlations”, between attributes.
- Establish a baseline performance by training initial models.
- Report your work: write an introduction that includes motivations and objectives of the project, show results, justify choices, highlight interesting findings, and draw conclusions. Submit all this as a notebook. Ensure that the notebook is readable.
- Be **meticulous** and **innovative**!