



# Term Project

## *Design, Implement, and Evaluate ML Models*

### 1. Overview:

This project forms 20% of your total marks, and it is expected to be an overall implementation framework for what you learn during the course. The objective of this assessment is to enable the students to design, implement, and evaluate various machine learning models (ML) in a range of real-world applications. Moreover, the project aims to assess how the students work effectively in teams to accomplish a common goal.

### 2. Instructions:

- You should team up into 2-3 students maximum. The students group must be in the same section.
- Refer to Section 3 of this document, to select a domain of interest, i.e., the domain of problems to be adapted as project idea.
- There are three expected outcomes of this project: technical report, documented source codes, and presentation. Our assessment will consider these outcomes with their related criteria. More details of these requirements, refer to Section 4.
- You will be required to present your work and discuss implementation details.
- **There will be two deadlines**, one is an optional deadline to submit a follow-up draft of your report, and the other is a hard deadline to submit all the project requirements.
  - (Optional) October 7, 2023, at 11:59 p.m -- submitting **Project Overview** to describe the problem to be addressed, your choice of dataset, introduce your teamwork, and a Gantt Chart timeline for completing the project tasks.
  - (Mandatory) October 26, 2023, at 11:59 p.m – Submitting **all the project requirements (codes, report, and slides)** in one zipped folder.
- The submission box shall be available on Blackboard.

### 3. Project Scope/Domain:

You will have the choice to select a dataset to work on, and we recommend using dataset repositories such as Kaggle or Mendeley Data. Another choice is selecting a dataset from `sklearn.datasets` from SciKit package in Python which embeds various, readily processed, datasets to start with. The following references might be useful for you:

- <https://www.kaggle.com/datasets>
- <https://data.mendeley.com/>
- <https://scikit-learn.org/stable/datasets.html>

## 4. Submission Requirements:

There are three main requirements for submitting this project

- Source coded and documentation.
- Technical report
- Presentation slides

### 4.1 Source Code Documentation

You have the choice to choose the implementation environment, however, we highly recommend you using Python and online collaborative Jupyter-based notebook such as: DeepNote or Google CoLab. Documentation in the source code file is expected, explaining your implementation steps and necessary packages to run your code. We are expecting the following parts:

1. Importing needed packages and explain their uses in your code.
2. Importing the selected dataset, and visualizing the dataset contents.
3. Splitting the dataset into train, test, and perhaps “validate” dataset.
4. Setting up the selected ML model and with the default or set of parameters.
5. Training the ML model with your dataset.
6. Testing and evaluating the trained ML model(s) against the test dataset, and report the result visually.

### 4.2 Technical Reporting

A technical report, not more than 20 pages, of your project must be submitted, and it must contain the following parts:

- **Project Overview:** describe the problem to be addressed in general, and introduce your teamwork, and Gannt Chart timeline for completing the project tasks.
- **Dataset Description:** describe the dataset you select, its contents, attributes/features types, and the target concept. Using visual figure and tables is expected.
- **Machine Learning Model Selection:** describe the ML model selected for solving the proposed problem and the rationale of your selection.
- **Implementation:** describe the technical framework you use for training and testing the models.
- **Evaluation & Results Discussion:** discuss the performance of your trained models against unseen (test) dataset, and explain the metrics you have used for evaluation
- **Conclusion:** short summary of your project where you highlight the main results you obtained and any challenges you face that could be addressed in the future work.

### 4.3 Presentation

For the presentation, all team members must be there during their presentation dates. The presentations timetable will be announced on Blackboard. It is expected to be a short presentation, not more than 10 minutes (and 10 for QA).