DS3000: Introduction to Machine Learning

Course Project Information

Fall 2025

Objective:

This course project provides hands-on experience in applying machine learning (ML) techniques to solve real-world problems. In this project, groups are expected to select a project topic, implement/analyze, and optimize ML models on a dataset to investigate an engineering problem, and finally submit a final report along with the code (in Python).

Important Notes:

- The project should be conducted in groups. All group members can be selected from different lab sections. Groups of **four students** are responsible for working equally and making sure that each member contributes equally to the project.
- The instructor/TA asks questions from the group members individually in the lab, and accordingly, members may not receive the same grades.
- The project has these deliverables:
  - List of group members, project description and the link to the dataset to work with. Due date: **Sep 26, 2025 at 11:59 PM.**
  - Progress Report (1 page). Due date: **Nov 14, 2025 at 11:59 PM.**
  - Project Presentation (in-person in the lab). The last week of labs.
  - CodeDemo (in-person in the lab). The last week of labs.
  - GitHub repository submission. Due date: **Dec 1, 2025 at 11:59 PM.**
  - Final Project Report (4 pages). Due date: **Dec 9, 2025 at 11:59 PM.**

Project Phases and Marking Scheme:

- The course project is worth 35% of the total final grade.

Phase #	Weight
Phase I - Group, Description & Dataset Link	**0%**
Phase II- Progress Report	**5%**
Phase III - Project Presentation (In-person group presentation)	**13%**
Phase IV - Code Demo (In-person group demo)	**1%**
Phase V- Code repository submission	**1%**
Phase VI - Final Report	**15%**
Total:	**35%**

Phase I – Group, Dataset and Description Submission

The students need to pick a group of 4 students and a dataset, either own collected data or an open-source dataset (*e.g.,* Kaggle, UCI, Data.gov, Google Dataset Search, Awesome Public Datasets) to work with. The dataset should be relevant, suitable for machine learning, and not being used for years by many users, which doesn't leave any gaps to work on. The students need to submit a text within the allocated space on OWL which includes the names of the group members, the link to their dataset, and a short paragraph describing the problem statement and the proposed project by the deadline. Once the dataset is validated by the

TA/Instructor, they can start working on that project. This phase has no mark but is mandatory to be completed.

In case the submission is not approved (e.g., improper dataset, wrong/unclear problem statement or broken link to the dataset), students can resubmit by the second deadline (Oct 3, at 11:59 PM) and consult with the Instructor/TAs to get the approval in the second attempt.

Phase II – Progress Report Submission

A 1-page report (text submission on OWL) of the progress on the project will be due approximately midway through the term (Nov 14). The purpose of the report is to allow the instructor/TAs to provide feedback on the direction of the project and its correctness before the final submission of the project in December.

Phase III – Group Presentation

Each group should present their project's slides (in-person, in the lab) to the Instructor/TAs the last week of labs. This will include Introduction, problem statement and objectives, dataset description/visualization, methodology and implementation details, and finally, showing the results and conclusion and the main remarks). The total time for each group is up to **8 minutes**. All group members should contribute equally to the talk. Participation of all group members and their presence is mandatory.

Phase IV – Code Demonstration

Each group should demonstrate their project code (in-person, in the lab) to the Instructor/TAs the last week of labs. They need to explain the code details, run it and show the results. Participation of all group members and their presence is mandatory. All members will be asked random questions about the code content and functionality.

Phase V –Code Repository Submission

A link to a GitHub repository with the code for the project is required to be submitted and shared with the Instructor/TAs. Please look at the GitHub website for more details.

Phase VI –Final Project Report Submission

Groups should submit a final project report of 4-page double-column in any IEEE format. The templates for MS Word/LaTeX markups can be found in this link: Report Template. Attached to the report, please add an appendix that describes what each group member worked on and contributed to the project.

Project Evaluation:

Each deliverable of your project will be evaluated based on several factors:

- The extensiveness of the study and experiments. A project that produces a more intelligent system by combining/comparing several Machine Learning techniques, or a project that involves well-designed experiments and thorough analysis of the experimental results. The projects that nicely incorporate more complex models and real-world applications will receive bonus marks.
- The writing style and the clarity of the written paper & code.

If you have any questions, please contact the TA or course instructor.

Good Luck,

Soodeh Nikan