CMSC-327

Machine Learning

Final Project Guidelines

Introduction

This document provides information on the final project. It lists some datasets and possible directions to explore when formulating a problem statement for your project.

How to Start

1. **Form a project group**: Maximum number of members-3. If you prefer to work individually, you may do so. However, the minimum project requirement will be the same as the other teams.

2. **Approach 1**- You can adapt an idea from an existing paper and re-run the experiments using the stated algorithm(s) or your own choice of algorithm(s). You should check if the dataset used in the paper is publicly available. This link may help you explore datasets with cited papers. You can also choose one of the papers discussed in the class for this (if the data is publicly available).

3. **Approach 2**- Use the options below to explore various datasets. Based on your interest, choose the dataset, visualize/understand it and decide on the problem you want to explore.

Datasets

Here are some datasets that you could use for the project.

1. https://data.gov/
2. OpenML
3. Kaggle
4. UC Irvine Machine Learning Repository
5. TensorFlow Datasets
6. Wikipedia's List
7. Political Science Dataset (shared by Dr. Sandra F. Joireman, Global Studies Program, and Dr. Mmachi Obiorah, Department of Computer Science). Please reach out to me if you are interested in reviewing this dataset.

You may also choose some other publicly available dataset if you prefer.

Project Timeline

Task	Due on	Description	Deliverable
Project proposal	October 16, 2025	• On a high level, describe the overall plan (goal, plan, etc.). • An assignment link will be created on Blackboard Learn, and a template will be shared for the proposal.	Proposal document Referred paper (if using approach 1).
Project progress	November 21, 2025 (tentative).	I will discuss the progress with each team.	
Project Files Submission	December 05, 2025	All members are required to submit the code files.	➢ Well-documented Jupyter Notebook project files (code) ➢ Team Evaluation Form
Final project presentation	December 15, 2025 2-5 PM, Jepson-G06	• Each team will present the findings to the class. • A template will be shared for the presentation and report.	➢ Presentation slides ➢ A project report (template will be shared).

1. All submissions should be made individually by all team members.
2. Please note that the final presentation should be based on the files submitted by the deadline—December 5, 2025. No further changes should be made to the project code/results after the deadline has passed.

3. If you start working on the deliverables on time, they will be feasible within the given timeline. NO project extensions will be provided. As stated in the course syllabus, a two-day late submission policy will be in effect.

Minimum Requirement

While you can include several tasks in your analysis, the project MUST include the following-

1. If using approach 1- Replicate the experiments used in the paper. If you are replacing an algorithm with another one, that's fine. The overall direction should be similar to what was discussed in the paper. Once your proposal is ready, I will evaluate if it meets the project requirements.

2. If using approach 2-
    a. Data cleaning: The dataset should include substantial cleaning and preprocessing efforts. Often, some websites have already preprocessed datasets. Please do not use them for our class project.
    b. Data visualization
    c. Feature extraction
    d. Dimensionality reduction (if applicable)
    e. Correlations
    f. Supervised learning
    g. Unsupervised learning
    Steps e, f, and g may vary depending on the dataset being used.

 Ultimately, the project proposal will decide if the planned approach meets the project requirements.

IMPORTANT

Project Authenticity and Use of AI Tools

The project must be original and demonstrate both collaborative and individual contributions. While it is acceptable to use generative AI tools to look up specific machine learning/visualization functions or clarify technical concepts, the overall project should not be created or generated by such tools. The primary objective is to apply the conceptual knowledge gained in the classroom to a hands-on implementation, showcasing genuine understanding and effort.

Grading Rubric

Project proposal	**20%**
Presentation	**20%**
Report& code	**20%**
Analysis and quality of work	**40%**
Going above and beyond	**20% (BONUS)**

The bonus section is for exceptional work, which will be assessed based on various factors, including the idea, conceptualization, and notable findings. If using approach 1, you can build upon the existing work done by the authors.

Grading

Equal project grades will be assigned to all team members if the work distribution is equal. Otherwise, the scores will be assigned based on the responsibility shared by each member while working on the project. Please be honest in your team evaluation form about the individual contributions (yourself and members).