

CS 5393 – Introduction to Machine Learning

Fall 2025

PROJECT PROGRESS REPORT GUIDELINES

0. Project Title and Author Names

1. Progress Update Summary

- Provide a brief description of the work you have completed so far.
- Clearly state whether you are **on schedule**, **ahead of schedule**, or **behind schedule**, and explain the reasons for your current status.

2. Key Project Milestones

Ensure that your report provides clear, detailed explanations for each of the points below.

2.1 Dataset Selection and Exploration

- Identify the real-world dataset(s) used for building machine learning models.
- Provide summary statistics for each variable or attribute in the dataset(s).
- Visualize the frequency distribution of each attribute using appropriate plots.
- Use boxplots to identify potential outliers.
- Other interesting finding about the dataset(s)

2.2 Machine Learning Tasks

- Clearly define the machine learning tasks being performed.
- Explain why these tasks are necessary for your project.

2.3. Algorithm Selection

- List the machine learning algorithms chosen for each task. You may compare multiple machine learning algorithms to find the best model for your project.
- Justify your selection by explaining why these algorithms are suitable for your problem.

2.4 Data Preprocessing

- Describe the data preprocessing steps required before running the selected algorithms.
- This may include:
 - Handling missing data (e.g., imputation techniques)
 - Encoding categorical variables
 - Standardizing or normalizing numerical data
 - Any other necessary preprocessing techniques

2.5 Performance Evaluation

- Explain how you will assess the performance of the models.
- Specify the evaluation metrics and criteria used to compare different models.
- Justify how you will determine the best model(s) for your project.

3. Work to be done

Describe in detail the tasks you still need to perform in order to conclude the project. Then provide a timetable for the remaining tasks. This timetable must specify clearly the remaining tasks to be accomplished, the deliverables of EACH remaining task, the starting and ending dates of EACH remaining task, and if you work in team, the name of the team member responsible for EACH remaining task.

4) References

List all sources in proper citation format. Sort references alphabetically by the last name of the first author. Each reference should include:

- Author(s)
- Title of the article/book
- Journal/conference/book name
- Publisher (for books)
- Month and year of publication
- Page numbers (if applicable)
- For websites, include the **access date**

SUBMISSION & GRADING

Submission Deadline: Submit all deliverables to the class website on Canvas by **11:59 PM** on the due dates listed below.

This project accounts for **25% of your final grade**. You may complete it individually or in a team of **two**.

Items	Due Date	Grade
Project Proposal	Saturday, Oct 11, 2025	5%
Project Progress Report	Saturday, Nov 8, 2025	5 %
Project Presentation	Tuesday, Dec 9, 2025	5 %
Project Final Report	Tuesday, Dec 9, 2025	10%

Submit your progress report as a **ZIP file** containing the following:

- **Jupyter Notebook (.ipynb) file**
- **Rendered HTML version** of the notebook
- **Additional files** (e.g., figures or datasets used in your notebook)

USEFUL LINKS

MACHINE LEARNING DATASET REPOSITORY

- <https://archive.ics.uci.edu>
- <https://www.kaggle.com/datasets>
- <https://openml.org/search?type=data&sort=runs&status=active>

WRITING REPORT USING MARKDOWN IN JUPYTER NOTEBOOK

- <https://www.datacamp.com/tutorial/markdown-in-jupyter-notebook>

- <https://www.ibm.com/docs/en/watson-studio-local/1.2.3?topic=notebooks-markdown-jupyter-cheatsheet>