Rubric for Machine Learning Final Project

Your final project will primarily be graded based on the questions you choose to answer about your data, the models you create to analyze the data, and your evaluation of the results. You will also be graded on your presentation on the last day of class, your final writeup, and the oral quiz.

Deliverables:

- You will turn in all the code for the final project that you used to perform your analysis. This should include all code to process your data (if needed), the code to create and run your models, and the code to perform the evaluation.
- You will give a "lightning presentation" on the last day of class (no more than 5 minutes) where you summarize what you did for the project.
- You will turn in a final report detailing your data, your models, and your evaluation.
- You will have an oral quiz with me where I will ask you about your code. This is mostly to ensure that if you have used LLMs to write any parts of your code (which is fine, and even encouraged!) that you have taken steps to figure out what the code does.

Category	Points	Criteria
Dataset Appropriateness and Description	10	Dataset is relevant, sufficiently large (>=100 examples unless justified), and well-described. Includes number/type of features and real-world source or motivation.
Modeling Effort	20	Two questions posed; two different ML models per question (4 total). Models are appropriate and reasonably learned. Shows effort beyond "default settings."
Evaluation and Analysis	20	Training/testing split is done correctly. Results are reported with appropriate metrics (e.g., accuracy, precision/recall, F1, MSE, ROC curve, confusion matrix, etc.). Models are compared, and conclusions are drawn with insight. Strengths, limitations, and unexpected findings are discussed.
Code Quality	10	Code is clean, modular, and documented.
Lightning Presentation	10	4 minute presentation on the last day of class. Covers: data origin, questions, models used, results, and main takeaways. Clear slides or other visuals; each group member speaks or contributions are stated. Emphasis on communication and clarity over specific technical details.
Oral Quiz	15	Each group meets with instructor for a short 1-on-1 code walkthrough. Must explain the logic behind major code sections, choice of algorithms, and how to interpret results. Demonstrates solid understanding, even if AI tools were used.
Final Report	15	3–5 page (including tables/figures) write-up summarizing the entire project: background, dataset, questions, models, results, and conclusions. Includes thoughtful personal/group reflections: what was learned, challenges faced, surprises, and ideas for future work. Written clearly with appropriate tables and/or graphs.

(More details on next page.)

Lightning Demo Guidelines

- 5 minutes per group
- Content:
    - What was your dataset, and why did you pick it?
    - What were your two questions/predictions?
    - What models did you use?
    - What did you find out?
    - What was hard, surprising, or rewarding?
- No need to show code unless there's something in particular you want to emphasize.

Oral Quiz Grade Breakdown

- 13–15 points: Demonstrates deep understanding of models, code, and design choices. Can answer most questions clearly.
- 10–12: Solid overall understanding but some gaps in reasoning or explanation.
- 7–9: Understands surface-level functionality but unsure of key ideas or code mechanisms.
- 0–6: Cannot explain large portions of code or decisions

Final Report Guidelines

Include the following points. You don't have to structure the document this way, but you can use this as a guide.

1. Introduction and Motivation – Why this dataset and topic?
2. Dataset Overview – Size, structure, source, and interesting attributes.
3. Predictive Questions – What are you trying to predict? Why?
4. Modeling Process – Which models did you try? Why those ones? How did you choose your hyperparameters?
5. Results – Evaluation on training vs. test sets, comparisons between models.
6. Takeaways and Reflection – What worked? What didn't? What was surprising? What would you do differently? What did you learn about ML?

You should include figures throughout the document as appropriate – this would include relevant tables, charts, or graphs to aid explanation.