

Final Project: Submission Guidelines and Grading Rubrics

95828 Machine Learning for Problem Solving

1 Submission Guidelines

- In this final project, you will study how data analytics and machine learning can be used to guide CDC in designing vaccination policy. To help you initialize the project, we recommend you go through the following three phases in a timely manner:
 1. Understanding the Data and the Problem
 2. Data Cleaning, Preparation, and Exploration
 3. Predictive Modeling and Decision-Making
- You are expected to submit two files via Gradescope: An assignment write-up should be submitted as **.pdf**, and a **.zip** file that contains all your code. **Your written report should follow a 8-page limit.**
- **Only one submission per group is required.** Please make sure to mark your group members on your submission and Gradescope while submitting.
- The assignment is due at **11:59 PM on Friday Apr 26, 2025.**
- If you have any questions, please use Piazza or visit course staff during office hours and recitations.
- Do not copy from other sources, share your work with others, or search for solutions on the web. Plagiarism will be penalized according to the university rules.

2 Report Formatting and Grading Rubrics

2.1 Report Formatting

Your report should follow a **8-page limit** and include the following sections:

1. **Title:** Concise and informative.
2. **Introduction:** Clearly state the chosen task. Explain why this task is important or relevant in practice. Briefly describe the provided dataset, including its dimensions, key features, and target variables.
3. **Data Analysis:** In this section, you should demonstrate the steps you take in your exploratory analysis. This should include: data cleaning, feature selection/transformation, correlation analysis, and any other steps you take in your analysis.

4. **Baseline Model:** Briefly describe your chosen baseline model. Include any relevant details, such as specific techniques or regularizations applied (e.g., kernel methods, Ridge or Lasso regularization for the regression models).
5. **Methods:** Clearly describe the models you implemented beyond the baseline. Discuss your motivations for choosing these specific models. You're strongly encouraged to introduce original ideas or explore creative applications beyond simply applying existing machine-learning methods. Additionally, briefly discuss any models or approaches you attempted but discarded due to poor performance or other practical limitations (if necessary).
6. **Results:** Present clear and insightful experimental results. Include plots, tables, or figures to compare and visualize model performance. Clearly label all results and thoroughly interpret them. Discuss whether the outcomes aligned with your expectations and why. Explicitly state the metrics you used for evaluation.
7. **Analysis and Policy Recommendation:** In this section, critically analyze your methods and results. Clearly discuss limitations or assumptions inherent to your approach (e.g., simplifying assumptions, constraints, data quality, computational issues). Identify conditions or scenarios where your chosen methods were ineffective, if necessary. Provide insights about the behavior of your models or about the problem environment based on your results. You should also include your answers for the decision-making task. You should clearly demonstrate and justify how your answers align with your observation in the data analysis step and the results of your prediction task.
8. **References and Citations:** Include correctly formatted citations and references for all external resources, papers, or tools used in your project.

2.2 Grading Rubrics

We will evaluate your final project based on the following criteria:

Results Evaluation (10 pts) We will evaluate your results based on the quantitative evaluation on your prediction models and the qualitative evaluation of your answers for the decision-making and policy recommendation task.

- **Prediction Task Performance (15 pts):** Evaluated using appropriate metrics (e.g., Mean Squared Error, Mean Absolute Error, accuracy), clearly reported for each model. You should include your analysis and the empirical results for in-sample and out-sample error, cross-validation for the model and hyperparameter selection (if any).
- **Decision-Making/Policy-Recommendation (5 pts):** Evaluated based on your answers for the decision-making task. You should justify how your answers align with your observation in the data analysis step and the results of your prediction task. Please refer to Section 3.2 in the Project Guidelines for the recommended questions you may want to address.

Written Report Evaluation (90 pts) Your written report will be assessed qualitatively as follows:

1. **Introduction [10 pts]**: Clear motivation, problem definition, and dataset overview. Effectively communicates the importance and context of the chosen task.
2. **Data Exploratory Analysis [10 pts]**: We will evaluate the completeness and the validity of your data analysis. This includes data cleaning, feature selection/transformation, correlation analysis, and any other steps you take in your analysis.
3. **Methods and Experimental Setup [25 pts]**: Clearly describe all the models you choose and the rationale for their selection. Proper explanation of experimental details (hyperparameters, training/test splits, evaluation setup) is expected. Originality or creative application of ML techniques is strongly rewarded.
4. **Results [25 pts]**: Clarity, correctness, and completeness of presented results. Results should include thorough comparisons, clear visualizations (plots, tables), interpretation, and meaningful analysis of performance across tasks.
5. **Discussion and Analysis [25 pts]**: Depth and quality of analysis, clearly articulating limitations, assumptions, and practical constraints. Insightful interpretation of results, including thoughtful reflections and meaningful suggestions for improvement or future directions.
6. **Formatting [5 pts]**: Professionalism, clarity, readability, and organization of the document. Proper formatting of equations, tables, figures, citations, and references.