

MATH 156: MACHINE LEARNING PROJECT

Fall 2025

1 Project Description

The idea of this project is that you are to pick a real data set for which you believe there are interesting questions to answer. You will then try out at least two machine learning approaches that we have covered/will cover in this course to try to find the best way to answer these questions (you are welcome to choose additional approaches either from class or your own reading). An important aspect of your results will be appropriate error assessments for the results you provide.

2 Deliverables

1. **Members' names for each team.** Due Wednesday, October 29, 2025.
2. **Project proposal.** 1-2 pages long in LaTeX. Due Wednesday, November 12, 2025.
 - (a) Members' names;
 - (b) Description of the problem;
 - (c) Description of the dataset (dimensions, names of variables with their description);
 - (d) Statement about whether the problem you want to solve is a regression problem or classification problem;
 - (e) The methods you plan to try; The error metrics you plan to use and the algorithm for assessing them;
 - (f) Comments and/or concerns.
3. **Slides** for a 10 minutes presentation. Due Wednesday, November 26, 2025.
 - (a) All presentation slides will be due on the same day regardless of the presentation day;
 - (b) Description of the data and the questions that you attempted to answer;
 - (c) Review of the approaches that you tried or thought about trying. It is interesting and useful to discuss both successes and failures!;
 - (d) Summary of the final approach you thought worked best and why you chose that approach;
 - (e) Summary of the results;
 - (f) Conclusions.
4. **Project report.** 5-15 pages long in LaTeX. Due Wednesday, November 26, 2025.
 - (a) Full detail of the material covered in the presentation;
 - (b) Project report LaTeX template available on BruinLearn.

3 Evaluation

Points will be allocated for the explanation of the question of interest, the descriptions of the data and approach you used, the reasons you chose your final approach, and the conclusions you were able to draw, both positive and negative.

4 Data Repositories

Data repositories you might consider:

- [UCI Machine Learning Repository](#);
- [Kaggle](#);
- [Statlib Datasets](#).