

# Final Project (Spring 24)

**Overview:** The final project is a capstone experience where you will work collaboratively in a group of 2-4 students on a self-selected application of machine learning. There are three major components on which your project will be evaluated:

1. The clarity, organization, and structure of your approach including your code, data preparation, choice of methods, and degree of professionalism.
2. A 7-10 minute in-class presentation summarizing your work
3. A 3-5 page scientific report detailing your methods and results

You are free to pursue any topic that allows you to demonstrate your knowledge of machine learning methods including (but not limited to): model training and evaluation, data preparation and pre-processing, comparing and choosing between different algorithms.

You may work on a Kaggle competition, reanalyze data from a scientific paper, or pursue another application that interests you.

## Project Timeline

- Monday 3/11 - One person in your group must indicate the members of your project group via email. Anyone who does not belong to a group will be randomly assigned on Tuesday 3/12
- Thursday 3/14 - You may use class-time to get started on your project. You must submit an official proposal of your plans by Friday 3/15.
- Thursday 4/25 - **Progress briefings** during class.
- Tuesday 5/7 and Thursday 5/9 - **Final presentations** during class
- Friday 5/17 - **Final reports** are due at 5pm. This is a hard deadline that is college-wide.

## Details

**Essential Components:** At minimum, your project is expected to satisfy the following:

1. You must submit all code and data in a manner that allows for the results in your presentation and final report to be easily replicated. Your code should be organized, appropriately commented, and reasonably efficient.
2. You must compare multiple machine learning approaches. If your application lends itself to a specific type of deep learning approach you may fulfill this requirement using a “straw man” (for example an SVM trained on flattened data).
3. Your project must involve a higher degree of difficulty than in-class labs and homework assignments. There are two primary ways for you to fulfill this requirement:
  - Working with unstructured data that require substantial feature engineering or preparation.
  - Researching and using one or more new methods that were not covered in class.

**Identifying a New Method:** To help you learn about new methods not covered I encourage you to begin by using the resources below:

- List of Scikit Learn’s Supervised Learning Methods
- PyTorch Tutorials - see the expandable navigation bar on the left side of the page

**Progress Briefings:** On Thursday 4/25 each group will present a short introduction to their project topic and its current status. There are no set requirements for the amount of progress that must be achieved by the time of this presentation; however, if you aren't able to present anything substantive you will not receive credit for this step.

**Presentation:** On Tuesday 5/7 and/or Thursday 5/9, each group will deliver a 7 to 10 minute presentation to the class. This presentation should summarize the key components of your group's work, including the project's goals, data, methods, and results. You should treat the presentation as if you were at a scientific conference, meaning you may assume some familiarity with basic machine learning methods and concepts, but you should thoroughly cover more advanced methods or non-standard methods. Your presentation will be scored on the following criteria:

1. *Comprehensiveness* - Did you communicate all of the necessary information that someone would need to understand your work? Were any important details not adequately explained? Could an audience member provide an accurate overview of everything you did?
2. *Coherence* - Was the presentation organized in a logical order? Did it flow smoothly? Was it easy for the audience to follow along?
3. *Professionalism* - Were your presentation materials appropriately prepared? Did you adhere to the 7-10 minute time frame and speak using appropriate volume and pacing? Did each group member contribute to the presentation in a manner that didn't interrupt its flow?

**Report:** Final reports, code, and data are to be submitted via P-web no later than 5pm on Friday at the end of final's week. Your paper should be 3-5 pages in length (not including figures, tables, references, and any supplemental material). You are encouraged to embed a link to a repository containing your code and data (rather than submitting them directly). Your report will be scored according to the following criteria:

1. *Structure* - Does your report follow a proper scientific structure (Intro, Methods, Results, Discussion, References)? Does the report include appropriate figures and tables that help communicate your main results? Does information appear in the proper section? (ie: results are not given in the Methods section, etc.)
2. *Technical Correctness* - Are your methods and results described correctly? Or there any errors, misleading statements, or important information that is omitted?
3. *Level of Detail* - Are your methods described in sufficient detail? Are your results thorough and comprehensive?
4. *Coherence* - Can a reader of your paper fully understand your work? How easily could a reader summarize the main steps of your project? Are there any gaps, awkward transitions, or omitted sections?
5. *Professionalism* - Are your visualizations and tables professional in appearance? Do you cite scientific sources when appropriate? Is the report written in a scientific tone and free of grammatical/spelling errors?

For additional details on scientific writing, I recommend reading this article, which provides some general guidance on scientific articles.