

# Final Project (s23)

**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. A repository containing all data, code, and results
2. A 10-minute in-class presentation summarizing your work
3. A 3-5 page scientific report detailing your methods and results

**Goals:** There are two major goals for the project. The first is for you to demonstrate a high degree of competency in applying machine learning methods covered during the semester to a novel application/dataset. The second is for you to extend the knowledge base you developed during the semester in at least one tangible way. This might entail learning an entirely new model/method, or an extension of a familiar model, or concepts related to a familiar model that weren't discussed during the semester.

## Project Timeline

- Monday 4/10 - One person must email Professor Miller the names of each member in your group. Anyone who does not belong to a group will be randomly assigned on Tuesday 4/11.
- Friday 4/14 - One person must email Professor Miller a **1-paragraph proposal** describing your data source and project plans.
- Friday 4/28 - Your group must have a cleaned version of your data (and a data cleaning script, if applicable) available in your **project repository**.
- Tuesday 5/2 and Thursday 5/4 - **Progress briefings** during class
- Tuesday 5/9 and Thursday 5/11 - **Final presentations** during class
- Wednesday 5/17 - **Final reports** are due at noon (the end of the final exam period assigned to our course)

## Details

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

1. You have a repository containing all data, code, and results
2. You consider and properly compares multiple machine learning methods. For deep learning applications, you may include one or more simpler methods as “straw man” competitors (perhaps using a flattened version of your data).
3. Your work incorporates at least one new method, extension, or new concept.

Under special circumstances requirement #2 may be waived.

**Identifying a New Method:** As part of the project you are expected to study and utilize one or more methods/topics that were not covered in our lectures, labs, and assignments. Substantial extensions of a method that was introduced in class qualify as new methods (for example, multi-task Lasso qualifies as an extension of Lasso regression).

To begin learning about suitable new methods and extensions, I encourage you to consider 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 Tuesday 5/2 or Thursday 5/4 each group will share a 3-5 minute progress briefing with the class. This presentation will briefly describe the project's data source, goals, and current status. Each group will receive completion credit for adequately delivering this presentation. There are no set requirements for the amount of progress that must be achieved by the time of the presentation; however, if you aren't able to present anything substantive you will not receive credit for this step.

**Presentation (10 %):** On Tuesday 5/9 and Thursday 5/11 each group will deliver a 10-minute formal 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 it were given at a scientific conference (in any scientific discipline), meaning you may assume some familiarity with machine learning methods and concepts. Your presentation will be scored on the following criteria:

1. Comprehensiveness - Did you cover everything necessary for someone to understand your work? Or were there important pieces of your work that are not covered?
2. Coherence - Did the presentation flow logically and smoothly? Or was it hard for the audience to follow?
3. Professionalism - Were your presentation materials appropriately prepared? Did each group member contribute to the presentation in a manner that didn't interrupt its flow?

**Report (30 %):** Final reports are to be submitted via P-web on Wednesday 5/17 by noon (the ending of our assigned final exam time). Your paper should be 3-5 pages in length (not including figures, tables, references, and supplemental material). The report should contain a link to the repository containing your data and code. Your report will be scored according to the following criteria:

1. Structure - Does your report follow the proper scientific structure (Intro, Methods, Results, Discussion, References)? Does the report include figures and tables? Information is placed 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 are there errors, misstatements, or inaccuracies?
3. Level of Detail - Are your methods described in sufficient detail? Are your results thorough and comprehensive?
4. Coherence - How well could a reader of your paper understand and describe your work? Are there gaps, awkward transitions, or missing information?
5. Professionalism - Are visualizations and tables professional in appearance? Do you cite scientific sources? Is the report free of grammatical and spelling errors?

For additional details on scientific writing, I recommend this article, which discusses general scientific articles.