

Final Project Instructions

Unsupervised Learning • STAT GR 5244 • Fall 2025

Project Proposal Due: November 5

Final Project In-Class Presentations: December 3

Final Project Report Due: December 19

The main requirement in this course is a final project. The project is flexible and you can choose to complete a project related to your interests. Projects must be related to the content of the course, broadly defined as unsupervised statistical machine learning. These could include applications of unsupervised learning to analyze a large and complex data set (using best practices from unsupervised workflows). Projects could also include developing software for unsupervised methods, or even novel methodology development. The main requirement is that your work must correctly apply unsupervised techniques learned in class and must contain some novel component (e.g. analyzing a dataset via techniques that have not been applied before). The final project may be completed **individually or in pairs**. If completed as a pair, please delineate each individual's contributions to the project.

Project Proposal:

You should submit a **one-page project proposal by November 5**. The project proposal should include a brief summary of your proposed project, a description of the data to be used, a statement on the project's novelty as well as a very brief overview of related literature, and an outline of your proposed scope of work. Project proposals should be submitted to the instructor and TA via **Ed Discussion** (this can be submitted privately). The instructor will then give you brief feedback on your project proposal.

Project Presentations (In-Class):

Each student (or pair of students) will present their research project in-class on **December 3**. Presentations should be five minutes long, followed by a brief question and answer period. Each presentation should give some background on the area, motivate the problem or data and its significance, briefly outline the existing work on this problem, highlight your contributions, give a very brief overview of your technical approach, and present a summary of your key results or findings. Presentations will count towards 20% of your final project grade and grading rubrics for presentations are delineated below.

Project Report Requirements:

Students should document their work on the final project by submitting a final written report in the style of a peer-reviewed conference or journal article in statistics or machine learning. Reports should be no more than **6 pages** in length for the main report using any Latex format, but you may include additional materials in appendices (note that appendices will not necessarily be graded). (Hint: For a Latex format that allows one to fit much content in 6 pages, check out the IEEE conference template found here: <https://www.ieee.org/conferences/publishing/templates.html>.) The report should be organized with sections typical of peer-reviewed papers in statistical machine learning and should roughly include the following sections: abstract, introduction (with a related works subsection and a contributions subsection), model or methods, theory, empirical studies, discussion, acknowledgments (students working in pairs should delineate contributions here), and references. Note that the report should contain a proper literature review with correctly formatted Bibtex citations in-text and in the reference section.

Project Grading Rubrics:

1. Presentation Communication: 20%

- Are the slides clear and engaging? Are visuals and figures used to aid in communication? Is the presentation well-organized and structured in a manner that is easy to follow? Is the motivation, context, and contribution of the work clearly outlined? Is the technical approach and results presented in a manner that makes the main contributions clear? Is the oral communication clear and engaging?

2. Report Communication: 20%

- Is the report well-written? Are best practices in writing employed and all spelling and grammar correct? Is the content organized in a coherent manner? Is the technical approach outlined in sufficient detail and with proper notation? Are all empirical studies explained in enough detail to permit results to be reproduced? Are mathematics, tables, or figures clearly explained?

3. Technical Approach: 35%

- Is the technical approach sound and using best practices in unsupervised learning? Are findings and conclusions correct and justified?

4. Novelty: 10%

- Is the project novel, presenting a new angle given the existing literature in unsupervised learning or analyzing a new data set or an existing data set in a new way? Is a thorough literature review conducted and the work discussed in the context of this literature?

5. Significance and Impact: 15%

- What is the significance or impact of the project? Is the project well-motivated with sufficient background given to understand its significance?

Note on AI Usage:

Students are permitted to employ AI for assistance on this project. However, work that can be produced solely using generative AI with minimal prompting will receive at least a 35% reduction in grading.