Kennesaw State University
Department of Computer Science

CS 8267: Advanced Machine Learning
Project

Projects can be implemented by individuals or as well as groups of up to 4 members. The expectations from the project are the same regardless of the number of students implementing the project. Therefore, group projects are encouraged. A maximum of 8 projects will be allowed.

Group Project Policy:

You are responsible for choosing group members. While choosing your group, consider the following:

    a. Schedules of students in the group,
    b. Interest in the project topic,
    c. Programming languages known by group members.

2. Each group must choose a leader. The leader is responsible for the management of the group.
3. Each group needs to maintain how many times they met, group members who attended the meeting, dates of the meetings, and problems solved. If a group member misses a meeting, he or she should let me know as soon as possible if he or she has any excuse.
4. In the first meeting, assign the roles of each group member as accurately as possible.
5. I expect all group members to participate at all levels of the project.
6. If a group member does not contribute to the project at a reasonable level, please let me know ASAP. If a group member misses 20% of group meetings, the student will get 0 grade on the project.
7. Assignment of responsibilities in the following ways is <u>not acceptable</u>:
    a. Student 1: read the papers and wrote the report; Student 2: implement the project.
    b. Student 1 and Student 2 have worked on all parts equally.
8. In the report, each group member must state **at least one major contribution** for the project (e.g., proposing a solution for a problem that is met while implementing or designing).
9. The **group report** will also include meeting info and responsibilities of group members. There should be a table showing the percentage of meetings attended by each group member along with the total number of meetings held.

Design & Implementation

Design and implement a machine learning application:
1. *Topic:* Determine a machine learning application.
2. *Papers:* Select 3 papers on your application using similar or the same datasets.
    a. Ideally select the latest papers in this area in the last 5 years.
3. *Data source:* Choose your data source. Make sure the data is large enough.
4. *Description:* State a detailed description of the machine learning problem. Consider the machine learning libraries and programming languages to be used and describe the

responsibilities of each group member. You may determine the libraries and languages from the papers you read.

5. *Report:* You are going to prepare a report that compares these three papers. The report length should be around 8 pages using 12pt Time New Roman font and single spacing. Your report should be prepared in the form of a paper. It will have the following parts.

   a. Title
   b. Abstract
   c. Introduction
   d. Related Work: This should provide a summary of 10 papers including 3 papers you have selected.
   e. Datasets: Description of datasets (size, schema, format, examples, classes, etc.)
   f. Methods
      i. **Option Improvement:** Provide a brief description of 3 methods and a detailed description of the one you have chosen. Implement their algorithm and run their method on the same dataset. Propose suggestions to improve the algorithm and explain why those suggestions may work. Implement one of those suggestions.
      ii. **Option Comparison:** Explain and compare 3 methods. Provide the performance of these methods on another dataset.
   g. Experiments
   h. Responsibilities of Group Members
   i. Major Contributions by Each Group Member
   j. Conclusion
   k. References
   l. Appendix
      i. Group report (not part of 8-page limit)

6. Provide your results in the experiments section.

7. Bonus (10% on Paper/Report):
   a. Identify a candidate conference for possible paper submission. This could be one of the conference papers where your selected papers were published. Find out the submission guidelines and prepare your report according to those guidelines. The bonus part should clearly be indicated as a section in your report stating the target conference venue.
   b. Optional:
      i. Submit your paper to this conference.
      ii. *Novel Idea.* You should have at least one novel idea that shows significant improvement.
      iii. *Public Datasets:* If you use a public dataset, most likely the experimental results of other papers should be available in those papers.
      iv. *Compare.* Try your ideas on several datasets by comparing them with at least two other techniques.

8. Make sure that large datasets are utilized. An account may need to be created on the HPC cluster of the university if you plan to use KSU's HPC. The HPC access information is available at https://research.kennesaw.edu/computing/index.php. Please let me know

ASAP so that I could request accounts for you. In addition, CCSE has a DGX server with A100 GPUs. Please let me know if you need to access this server.

9. Extra Bonus (**3% on the course**): Present your work on C-Day or Analytics Day of CCSE. Please follow the deadlines.

Submitting Materials

- Submit your Jupyter notebook with outputs of intermediate stages. If you use another IDE, provide your source code and outputs.
- Provide a Readme file on how to execute your project if you are not using Jupyter notebook.
- Provide your project report (including major contributions of each group member).
- Include group report (meeting info, responsibilities, attendance) as Appendix to your project report.
- Your presentation slides.
- (Optional) Your presentation video.

General Guidelines for Presentation
- The duration of a demo will be based on the number of presentations. You may assume that each demo will take around 15-20 minutes including Q&A after your presentation.
- Provide a brief presentation of your paper.
- Title slide (Names of students, Project Title, Course Name, Semester or Date) (1 slide)
- Summarize the research topic (1 slide)
- Explain the methods in 3 papers (1 slide per paper: 3 slides)
- Comparison of 3 techniques based on their results in the paper (1 slide)
- Option Improvement:
    - Explain your method (1-2 slides)
    - Go over your code and explain it briefly.
    - Present your demo.
- Option Comparison:
    1. Explain the new dataset (1 slide)
    2. Go over the code you have used/developed (especially ML modeling parts) and explain it.
    3. Present your demo on a sample dataset. You may run it on a partial dataset.
- Conclusion (1 slide)

Important Dates

These intermediate dates are very important.

LATE SUBMISSION: **You will lose 1% per week (from report and presentation) if you miss these intermediate deadlines.**

Topic submission with the names of group members: August 27th, 2025
Finding Papers: September 15th, 2025
Title submission: September 22nd, 2025

Data Acquisition: September 29th, 2025
Description of Datasets: October 1st, 2025
Related Work & References: October 29th, 2025
Report: December 1st, 2025
Presentations: December 1st and 3rd, 2025