CSCI 475/675 Final Project/Term Paper Information

CSCI 475/675 Intro to Machine Learning course's main goal is to involve the participating students to apply machine learning algorithms to tackle real-world tasks or to conduct research in their research fields. The course has both undergraduate (CSCI 475) and graduate students (CSCI 675).

Undergraduate students (enrolled in CSCI 475) are expected to work on a machine learning project, where they will either build an application involving machine learning (ML) algorithm(s) or solve a problem using ML. Students can use a publicly available data set or collect their own data set for their final project. However, it's recommended to use a (popular) publicly available data set for such data sets are usually balanced, require less data cleaning. Undergraduate students can also work on the final project in a group with a maximum of 2 people in each group.

Graduate students (enrolled in CSCI 675) are expected to work on a term paper, where they will use machine learning algorithm(s) for a research problem. The students are recommended to pick a research problem in their research domains that can be tackled using ML. Graduate students should aim to produce a term paper that is publishable. They can either work on a **survey paper** or a **technical paper** with at least preliminary results. Either way, the term paper should be **publishable**. Survey paper must present a coherent and extensive compilation of machine learning algorithms being used to solve research problems in a particular research domain. On the other hand, a technical paper must propose a research problem statement, and a proposed solution involving machine learning algorithm(s). Preliminary results to the proposed research statement are acceptable, which can be further extended to make the paper publishable.

While the term paper is expected from graduate students, undergraduate students can also present a term paper if they are interested in writing one.

Project/Term Paper Topics

First of all, you will have to pick a project topic. If you are looking for project ideas, please come to the office hours, where we can brainstorm together and come up with meaningful project ideas.

CSCI 475
Some potential project ideas for undergraduate students:

1. Application project: Students can pick an application (web application, mobile application, or merely a computer program) that they are interested in developing/writing, and explore how they can integrate machine learning algorithm(s) to solve any problem.

2. Develop a machine learning pipeline which takes a (publicly available) data set, and

conduct a classification/regression/clustering depending on what the students are trying to solve.

A well-written application can always be published/shipped. More importantly, students can showcase this work as part of their portfolio or can be extended as a long term project where the students can further add new features to make their applications more functional and richer.

CSCI 675
Term paper ideas for graduate students:

1. Pick a research problem statement in your research domain, and propose a potential solution to it that involves machine learning algorithm(s).
2. In case your planning to produce a survey paper, compile the machine learning algorithm(s) that are being actively used in your research domain to tackle research problems. A survey paper needs to extensive and coherent to be publishable. Writing a good survey paper can be a good practice to explore research problems and start working towards one.

A good term paper will be a publishable or nearly-publishable piece of work. **Graduate students can take this course as an opportunity to extend their current research using machine learning**, and potentially publish their research findings (or their survey paper) in a conference or a journal.

For any machine learning algorithm, it requires one or several datasets. Not all datasets are balanced, pre-processed that might suit your task. Therefore, one needs to be mindful about this aspect while designing their project/term paper because pre-processing a dataset can be extremely time consuming.

Project/Term Paper Parts: Proposal, Milestone, Presentation, and Final Report

In this section, I breakdown different parts of your project/term paper.

#	**Parts**	**Weight**	**Points**
1	Proposal	20%	100
2	Milestone	20%	100
3	Presentation	10%	100
4	Final Writeup	50%	100

*Note: This point breakdown is for **project/term paper only**. Considering a student receives 100% in the project/term paper, it will constitute 30% of total grade in case of an undergraduate student, and 35% of total grade in case of a graduate student, as stated in the syllabus.*

Submission

You will submit all four parts of your project/term paper (**proposal, milestone, presentation slides, final report**) on Canvas.

Evaluation

Your project/term paper will be evaluated based on:

- **Technical quality of the work**: The technical elements of your work should make sense, for instance, the correctness of your algorithm, how suitable the algorithm is to solve your problem statement.
- **Originality**: It should be your original work regarding all components of your project/term paper.
- **Communication**: You should be able to clearly and effectively explain your work in the final writeup.

Project/Term Paper Proposal [Due: Friday, Oct 11]

In your project/term paper proposal, you will pick a project/paper idea. You will clearly explain what you are going to work on as a problem statement.

Format: Your proposal should a PDF document, including a title for your project/term paper, full name, email, and student ID.

Your project/term paper proposal should include the following information:

- Introduction/Motivation: What kind of problem you are trying to solve? Are you developing an application, or writing a technical paper, or a survey paper?
- Method: What kind of machine learning techniques are you planning to apply or improve upon? In case of survey paper, you may still run some experiments to showcase preliminary results to illustrate the results on the usage of machine learning algorithms.
- Intended Experiments: What are the experiments you are planning to run? How are you going to evaluate your machine learning algorithm? For instance, what are the metrics you will be using measure the performance of your trained ML model?

Grading: As long as these elements are outlined clearly on your proposal, you will receive full points on your proposal.

Milestone [Due: Friday, Nov 08]

The milestone is to make sure that the students are on track, and update the instructor about your progress. With the milestone, you will share your "early draft" which will develop into the final draft. Your milestone should look like a progressing paper or project with some portion of it completed, and some portion yet to be completed. Your milestone

should include your full name, student ID, and the project/paper title. If it's a group project (in case of undergraduate students), it will be a group submission, where contribution from each member will be described/noted in the submission.

Contribution: With a group project (in case of students enrolled in CSCI 475), please make sure that the tasks are divided equally among the group members, and their contributions is noted accordingly in the milestone.

Grading: You will receive full points, as long as the milestone has some goals, features, and/or components completed as mentioned in your proposal. A group can have a maximum of 3 students.

Format: It will be a PDF of your "early draft" writeup/report. Similar to your proposal, it should have the following sections:

- Introduction/Motivation: What kind of problem you are trying to solve? Are you developing an application, or writing a technical paper, or survey paper?
- Method: Method: What kind of machine learning techniques are you planning to apply or improve upon? In case of survey paper, you may still run some experiments to showcase preliminary results to illustrate the results on the usage of machine learning algorithms.
- Preliminary results/experiments: Describe the experiments you have run as part of your baseline experiment, their results, and error you might have run into, and how you are planning to tackle it.
- Term paper (CSCI 675: Graduate students):
  - In case of term paper, your milestone should have **Introduction/Motivation**, **Related Work**, **Methodology** sections completed, and you should be in the phase where you are running your experiments to collect the results. You should have a few preliminary results by this point.
  - If it's a survey paper, it should have a complete outline of the paper with half of its elements should be written.
- Group Project (CSCI 475: Undergraduate students):
  - In case of a group project, if you are writing an application (web app, mobile app, or simply a computer program), you should have implemented roughly have of the features/components of your application. Similarly, for the machine learning algorithm used in the project, you should be running experiments to collect baseline results, so that your ML model is in the way to be trained, and deployed via your application/computer program.
  - In case of a machine learning pipeline, you should be running your at least your baseline experiments, and collecting results.

Presentation [Nov 27 & 29]

You will present your work, either group project or term paper, during the last week of November. You are also required to present and submit your slides to receive your points for presentation.

Final Writeup [Due: Nov 29]

CSCI 475: Group Project

Format: It will be a PDF of your writeup (preferably written in LaTeX). You may use the proposal template itself to complete the writeup or you can use some other template. The final writeup/report should have (tentatively) following section:

1. Abstract [$\approx$ 1 paragraph]
2. Introduction/Motivation [$\approx$ 0.5 pages]
3. Related Work [$\approx 0.5 - 1$ pages]
4. Dataset and Features [$\approx 0.5 - 1$ pages]
5. Methods [$\approx 0.5 - 1$ pages]
6. Experiments/Results/Discussion [$\approx 1 - 2$ pages]
7. Conclusion/Future Work [$\approx$ 0.5 pages]
8. Appendices (if any)
9. References/Bibilography

CSCI 675: Term Paper

Format: If it's a technical paper, it can be up to 6 pages using IEEE Conference Template, including figures. The final report should have (tentatively) following sections:

1. Abstract [$\approx$ 1 paragraph]
2. Introduction/Motivation [$\approx$ 0.5 pages]
3. Related Work [$\approx 0.5 - 1$ pages]
4. Dataset and Features [$\approx 0.5 - 1$ pages]
5. Methods [$\approx 1 - 2$ pages]
6. Experiments/Results/Discussion [$\approx 1 - 3$ pages]
7. Conclusion/Future Work [$\approx$ 0.5 pages]

8. Appendices (if any)

9. References/Bibilography (at least 20 references)

If you have any questions regarding the project/term paper, please come visit me during my office hours (Tuesdays 1pm – 2pm & Wednesdays 2pm-3pm) or feel free to email me to setup a (virtual or in-person) meeting.