Instructions for CS-GA 2565 Final Course Project

Student Name 1
Affiliation
Address
`email`

Student Name 2
Affiliation
Address
`email`

Student Name 3
Affiliation
Address
`email`

Abstract

This document outlines the instructions for the final course project of CS-GA 2565 Fall 2024. Your project proposal and final report should use this space for abstract.

1 Goals

The final course project constitutes 30% of your overall grade. The objective is to apply the machine learning concepts acquired during this course to a real-world problem. Choose a pertinent and applicable issue, identify an appropriate data source for your machine learning solution, and if no suitable data source exists, propose methods to gather the required data efficiently.

2 Key Dates and Details

Please mark your calendars with these key dates:

- **Oct 8, 2024**: Form groups of three students. Registration link is `https://forms.gle/nNEHn1k1dQv5rJ3W7`. Groups will be assigned a number. Students not part of a group will be assigned a group by the instructor.
- **Oct 31, 2024**: Submit project proposals on Gradescope by 12PM. While this proposal is mandatory, it will not be graded. Its intent is to establish a checkpoint. With the submission, book a mandatory consultation with the instructor to discuss/approve your selected topic and proposed methodologies. This consultation can be outside of regular office hours. Calendar booking link: `https://calendar.app.google/iNPzAVEHR6tnTSWf9` (Link will be activated between Oct 16 - Nov 15).
- **Nov 15, 2024**: Complete at least one consultation with the instructor.
- **Dec 9, 2024**: Submit the project presentation slides by 11:59PM. The submission link is `https://forms.gle/Dw8b9JCeJuBxCy8r9`.
- **Dec 10, 2024**: Course project presentation from 4:55PM to 6:55PM.
- **Dec 13, 2024**: Final report submission on Gradescope by 12PM.
- **Dec 14, 2024**: Complete the self and peer evaluation via Google Form. This is crucial for determining your contribution to the project. Evaluation link: `https://forms.gle/geMc7ZKxmyostPQe9`.

3 Submission Guidelines

Page format. Adhere to the LaTeX template provided with **this document** (`template.sty`). Avoid altering font type, size, line spacing, margins, or heading arrangements. Use standard LaTeX structures. For guidance on LaTeX, refer to `https://www.latex-project.org/get/` or utilize the online editor Overleaf at `https://www.overleaf.com/`. Note: Not using the provided template (including fonts, headings, margins, etc.) will result in an automatic fail on the project.

Citation format. Employ BibTeX for citations. An illustrative reference is [1]. The references for this template reside in `template.bib`.

For the project proposal:

- **Page limit:** 2 pages, excluding references. Note: Exceeding the page limit is not acceptable. Please stay under the page limit.
- **References:** No page limit.

For the final project report:

- **Page limit:** 8 pages, excluding references, including tables and figures. Note: Exceeding the page limit is not acceptable. Please stay under the page limit. It is ok to have less than 8 pages, but if the report is significantly below the limit, then it may also be graded unfavorably due to a lack of content.
- **References:** No page limit.
- **Code:** Submit your code as a zip file.

4 Report Structure

Your proposal should contain the following sections:

- **Abstract:** Brief report summary.
- **Introduction:** Define the problem, its significance, and relevance.
- **Related Work:** Summarize previous attempts or similar problems, along with relevant data or approaches.
- **Proposed Dataset and Approach:** Detail the dataset, data metrics, your proposed solution, and potential alternatives.
- **Expected Results and Milestones:** Enumerate anticipated outcomes and timelines. If working in a group, specify responsibilities for each member.

The final report should contain the following sections:

- **Abstract:** Brief report summary.
- **Introduction:** Define the problem, its importance, historical solutions, dataset origins, whether you gathered your own data. Briefly summarize the experiments and the significance of your findings.
- **Related Work:** A deeper dive into prior solutions, drawing parallels and differentiating your approach.
- **Method:** Elucidate your methodology, data collection processes, and the machine learning techniques employed.
- **Dataset:** Describe your dataset in detail, including its structure, size, and presentation. Incorporate visuals and samples.
- **Experiments:** Lay out your experimental framework, benchmark comparisons, and hyper-parameter settings. Incorporate results and comprehensive analysis.
- **Discussion/Conclusion:** Reflect on your method's efficacy, its implications, and overarching insights derived.

5 Presentation

Presentations will be held during the course's final lecture. Prepare a slide deck and submit the slides as a PDF a day prior to the presentation. Aim for a duration of 5 minutes (around 5 slides), followed by a 1-minute Q&A session. The precise time allocation may vary based on the total number of groups. The exact time allocation will be announced.

6 Grading

Your project will be graded out of 30 points, distributed in Table 1.

Table 1: Grading guideline

Area	Points	Criteria
Topic Selection	3	Relevance to ML, applicability, and novelty.
Literature Review	3	Comprehensive survey, proper citation, and insightful connections.
Dataset Selection	5	Suitability for ML, quality, preprocessing efforts, and data collection processes.
Modeling	5	Appropriateness of the proposed method, clarity of the learning objective and optimization methods, and mathematical correctness.
Experiments	5	Thoroughness, clarity of results, and effective analyses.
Writing	3	Clarity, structure, and motivation.
Presentation	4	Clarity, organization, time management, and content depth.
Participation	2	Attendance during presentations and participation during Q&A.

References

[1] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.