

COSC 425: Intro to Machine Learning

Final Project Rubric

Dataset(s) Selection Due: November 3, 2023

In-Class Presentations: November 30/December 5, 2023

Full Submission: December 8, 2023, 11:59 PM

Introduction

In your final project, you will be selecting a dataset on which you will perform machine learning. Your project will include: the analysis and visualization of this dataset, the application of machine learning approaches to the dataset, and a discussion of hyperparameters required for this dataset.

The goal for the final project is to give you an opportunity to apply the topics you have learned over the course of the semester to a real-world dataset.

Project Structure

You will first select a dataset or datasets. You will be doing some data analysis on the dataset before you apply machine learning to it. As part of that analysis, you should understand and report the broader characteristics of the data, clean the data, visualize the data, and prepare the data, perhaps by normalizing, scaling, or transforming and/or down-selecting. You will do an examination of the impact of hyperparameters on the machine learning approach(es).

You will then apply one or more machine learning approach that we have learned in class. **NOTE:** Your ML approach cannot be deep learning. We're covering that way too late and there is an entire class devoted to deep learning in this department.

Project Report

Your final project report should be structured as an academic paper for a conference or journal submission is structured. Your final project report should use the IEEE 2-column conference format with the **sigconf** format. There are both LaTeX and Word templates available here: <https://www.ieee.org/conferences/publishing/templates.html>

The final project report should be at least 6 pages and no more than 8 pages long. The report length includes any references, plots, and figures that you include.

Your final report will be composed of the following components:

Abstract:

150-250 words to briefly introduce your dataset, the key question/problem you're addressing with ML applied to the dataset, and a brief summary of your results. Someone should be able to read just the abstract of your report and have an idea of what you did and what the results were.

Introduction and Motivation

You should address the following questions in the introduction and motivation section of your report:

- What is the dataset or datasets you chose for your project, and why did you choose it/them?
- What is the overarching goal of you're trying to achieve with ML on this dataset?
- (Briefly) Which ML approaches are you using, and why did you choose those?

Your introduction should also include a discussion of what the goals are of the project and how you're determining whether you are successfully addressing those goals. Finally, your introduction should include a summary of the rest of the paper.

Dataset:

In this section, you should describe in detail the dataset that you're using. You should use visualizations to illustrate characteristics of the dataset. You should also describe which features of the dataset you're using and why. If you omitted any features that are usually part of this dataset, describe why you omitted them. You should also describe in this section any transformations or pre-processing you performed on the data and why you chose to use those.

Important Note: You must include at least two plots describing aspects of the data in this section.

Machine Learning Approaches and Methodology:

In this section, you should describe the ML approaches you chose to apply to this dataset and why you chose those ML approaches. You should describe what hyperparameters you will be investigating of those approaches. Here you will also describe which metrics you will be reporting and how those metrics are calculated. You should describe **in detail** the experimental setup you have defined:

- Are you comparing two different algorithms?
- Are you comparing performance across hyperparameter values?

- How are you defining whether the project is successful?

Results:

In this section, you will describe the results of your approach. You should depict the results visually through plots, and you should provide a discussion of each plot and the results you obtained. If something unexpected happens in the experiments, it may be worthwhile to probe into that further to try to explain why it happened. Where appropriate, you should visualize the machine learning approach (through decision boundaries or some other visualization technique). If you can interpret the machine learning approach results, you should include a discussion of what you learned about the dataset from this interpretation (which features are most important, etc.)

Discussion, Conclusion, and Future Work

Here, you will provide a discussion of the results and any major conclusions you obtained. You will also provide at least a paragraph of “future work” discussion:

- If you had more time, what would you do next?
- Did this project open up any new research questions?
- Was there something else you would have liked to have done and didn’t get a chance to do?

Contribution of Team Members

This is a short section that is NOT included in the page count. Here, you should describe what each team member did in the project. There should be clear contributions of each member of the team. **Note:** If you’re working alone, you can omit this section.

In-Class Presentation

You will make a three-minute in-class presentation on either November 30 or December 5. You will have four TOTAL slides in your in-class presentation:

- Title Slide: Includes the title of the project and the names of the project team members (you should spend very little time showing this slide)
- Dataset slide: Include a description and visualizations of the dataset
- Machine learning methods slide: Include a description of the machine learning approach(es) you are applying and which hyperparameters you’re examining
- Results: Include an overview of the key results you obtained

You should keep text at a minimum and include visuals instead. You will have a strict time limit (3-5 minutes) to deliver the presentation, so you should practice presenting to ensure that you will stay within the time limit. All team members will stand at the front, but it is fine for one person to take the lead on presenting.

Recorded Presentation

You will create a video presentation of 8-10 minutes. This should be a narration over a slide show. However, if you want to do it differently in some way, I'm open to that, but you need to get it approved first. Every team member should present at least one part of the presentation. You should submit this presentation as an mp4 or mov video file, and you will also submit the associated slides you presented as a PDF.

Note: I recommend recording this via Zoom. Setup a Zoom session with the full group, have one person in charge of advancing slides and do a Zoom recording.

Recorded Presentation

- **Report (70 points):**
 - Abstract: 5 points
 - Introduction and Motivation: 10 points
 - Dataset: 10 points
 - ML Approaches and Methodology: 10 points
 - Results: 25 points
 - Discussion, Conclusions, and Future Work: 10 points
- **In-Class Presentation (10 points):**
 - Content (5 points)
 - Presentation quality (5 points)
- **Recorded Presentation (20 points):**
 - Slide content (10 points)
 - Presentation quality (10 points)

Submission

You will submit a zip file that includes:

- PDF of your report
- PDF of the slides presented in class
- PDF of slides from pre-recorded presentation
- Video file of pre-recorded presentation
- Dataset (if the dataset is not publicly available)
- Code files generated as a result of this project

Each member of the team should submit the SAME zip file.

Late Penalty: Your final project can be submitted no later than December 12. There will be 10 points deducted for each day late. If your project is missing any of the required components, you will also received a penalty if you submit those late.