CS 232 Final Project

1 Project Overview

For your final project, you will apply supervised machine learning models to an AI task of your choice. You can either choose to **train one model** or to **evaluate three models**.

In addition to implementing your training or evaluation task, you will also conduct an audit of generative AI tools for programming.

You will submit two written reports: a project report describing your supervised machine learning task, and an audit report describing your experiences working with generative AI for programming.

You may choose to work with a partner on your supervised machine learning task. If so, you may write your project report together, but you must write individual audit reports.

If you are working on a thesis or research project, I am also willing to consider a topic that is related. But you **cannot** propose a topic that overlaps with a class project in another course.

Component	Points	Due Date
Project proposal	(part of HW 8)	11/23
Presentation	10 points	12/8
Code	10 points	12/17 **4pm**
Project Report	45 points	12/17 **4pm**
Audit Report	35 points	12/17 **4pm**

All project components must be submitted by 4pm on December 17th. This is a college-wide deadline. I cannot accept any work after this deadline.

2 Project checkpoints

2.1 Project Proposal

As part of HW 8, you will write a brief description and timeline for your project.

2.2 Dataset Exploration

As part of HW 8, you will explore a dataset that you might use in your supervised machine learning task.

2.3 Presentation

We will have short presentations on the final day of class. You will have **4 minutes** to briefly present your project. You should describe your task and what model(s) you are using. You are not required to have results to share, but if you do have preliminary results, you can discuss them.

You should design 1 slide to use in your presentation. Your slide should show an example item from your supervised machine learning task.

3 Final Products

3.1 Codebase

You will submit your code at the end of the semester. Your code should be organized and commented. **You must also submit a README text file that explains how to run your code.**

3.2 Project Report

You will write a short paper about your project, describing your task, the model(s) you used, and your results. The report should be **single-spaced and at least 4 pages**. There is no page limit.

Your report should be structured as follows:

- **Introduction**: introduce and motivate your task. You should discuss and cite related work.
- **Data**: illustrate your task and describe the data that you are using.
- **Model**: explain your model.
- **Metric**: present your evaluation metric(s) and justify why it is appropriate.
- **Results**: present your results. You should analyze any trends or patterns you notice in how the model performs. You should include at least one figure visualizing model performance.
- **Conclusion**: summarize your findings and discuss any threats to their validity.
- **References**: provide citations. This does not count towards the required page length.

3.3 Audit Report

You will also write a short report describing your experiences using generative AI programming tools during the course of your project. This report should reflect on how useful you found the tools; limitations of the tools that you encountered; and your perception of how using these tools affected your learning.

Your audit report should be **single-spaced and at least 3 pages**. It should include the following sections:

Tool Use Describe which tools you used and how you used them. What kinds of tasks did you use them for (programming, code explanation, writing documentation, etc)? Include some example prompts here.

Tool Usability How useful did you find the tools? What were they best at? What did they struggle with? Discuss some example interactions, showing both cases in which the tools were effective and cases where they were less useful.

Interaction Modalities Discuss your experiences with the following features:

- Plan mode: Did you use the tool to generate a natural language plan? If so, was it easy to read?
- Agents asking questions: Did you understand the questions? Did you think they were useful?
- Agents running code: Did you allow the tool to run code autonomously? Why or why not? If you did, how easy was it to understand what it was doing?
- Image understanding: Did you use the tool to interact with images? If so, what did you observe? (Figures count as images.)

Effects on Learning Discuss how you think the use of these tools impacted your learning, considering the following questions:

- Did they help you attempt a more ambitious task?
- How do you think the quality of the code you produced compares with what you would have written by hand?
- Do you think you understand your project code better or worse than if you had written it by hand?

Overall Assessment Based on your experiences, how do you feel about these tools? Do you think you would use them in the future? If so, how? What do you think about their potential impacts on 1) education and 2) the workplace?

4 Generative AI Use

This project involves an assessment of generative AI tools for programming. You are encouraged to use generative AI tools for programming throughout your supervised machine learning project. You will then write a critical assessment of the tools you use in your audit report.

To aid your evaluation of the tools, you will want to log your model interactions so that you can review them later and include examples in your audit report.

You may use generative AI tools for any part of your supervised machine learning project, including generating or debugging code; computing evaluation metrics; generating data visualizations; and writing documentation.

4.1 Tool Use in Writing

Your **audit report must be entirely written by you**: you may not use generative AI tools to help you plan, write, or edit your audit report.

However, **you may choose to use generative AI tools when writing your project report**. If you choose to do this, you must include an additional section entitled Use in Writing in your audit report, where you discuss how you used generative AI tools in your writing process, and how useful you felt they were.

4.2 AI Tools

For this project, you will use an *agentic* programming tool. This means a tool that can read your files, generate code, and run code (if you choose).

My class demo used Google's Antigravity tool. This is a good option because you have access through your Wellesley Google account to basic Google models for free.

Another option is to redeem Github's Student Developer Pack, which comes with basic access to Copilot for free. This requires verifying your student status with Github.

If you have a different tool that you prefer, that is okay. Here are some other platforms, most of which require some monthly fee:

- OpenAI Codex
- ClaudeCode
- Cursor

5 Rubrics

Presentation Rubric (10pt)

- Is the task explained well?
- Is it clear what model(s) are being used?
- Is it clear how model performance will be measured?
- Does the talk make good use of the slide, without merely reading off of it?
- Does the slide contain an example to illustrate the task?
- Are figures captioned and sources cited?

Codebase Rubric (10pt)

- Does the code successfully run the model(s)?
- Is the evaluation metric appropriate?
- Does the code evaluate model performance?
- Is the code commented and organized?
- Is there a README that describes how to run the code?

Project Report Rubric (45pt)

- **Introduction (10pt)**
  - Is the research question clearly explained?

  - Is the research situated with respect to previous work?
  - Is previous work cited properly?
  - Is the task clearly articulated?
- **Data (5pt)**
  - Is the data clearly explained?
  - Was care taken to select high-quality data?
  - Is it clear how the data is going to be used?
  - Are there examples of the task?
- **Model (5pt)**
  - Is the model design clearly explained?
  - Is the model appropriate to the task?
  - Are the design decisions related to the model(s) explained clearly and thoroughly?
- **Metric (5pt)**
  - Is the evaluation paradigm clear?
  - Is it clear how model success or failure will be measured, for each model?
  - Is the evaluation metric(s) used to assess model performance clearly explained?
  - Is the proposed evaluation metric appropriate?
- **Results (10pt)**
  - Is the discussion of model performance clear and thorough?
  - Are trends in the model performance highlighted and discussed?
  - Is there at least one visualization?
- **Conclusion (5pt)**
  - Are the findings summarized in a concise and clear way?
  - Are the claims about model performance made clear?
  - Are threats to the validity of the findings discussed?
- **General (5pt)**
  - Is the report well-organized?
  - Is it easy for a reader to follow?
  - Has it been proofread?

Audit Report Rubric (35pt)

- **Tool Use (6pt)**
  - Is it clear how the tool was used?
  - Are the tasks that the tool was used for described clearly?
  - Are some example prompts included and discussed?
  - Is there some discussion of how you decided when to use the tool?
- **Tool Usability (6pt)**
  - Are tool strengths discussed?
  - Are tool failures discussed?
  - For tool failures, is there a discussion of what caused the interaction to fail?
  - Are example interactions included and discussed?
- **Interaction Modalities (6pt)**
  - If Plan Mode was used, did you discuss this?
  - Are agent-asked questions discussed?

  - Is there a discussion of how you decided whether or not to let the tool run code?
  - If you used the tool to interact with images, did you discuss this?
  - If AI tools were used in writing, was this discussed?
- **Effects on Learning (6pt)**
  - Is there a discussion of the effects tool use had on your learning?
  - Is there a discussion of the effect tool use had on the scope of your project or what you accomplished?
  - Is there a discussion of the effect tool use had on your understanding of your code?
- **Overall Assessment (6pt)**
  - Is there a thoughtful discussion of the overall experience?
  - Is there a reflection on individual choices about tool use in the future?
  - Is there a discussion of potential impacts on education?
  - Is there a discussion of potential impacts on the workplace?
- **General (5pt)**
  - Is the report well-organized?
  - Is it easy for a reader to follow?
  - Has it been proofread?