IEOR 242A
Fall 2025

Project Description, Information, and Guidelines

Dates:
Report and Presentation: 11:59pm Friday Dec 19
Project proposals (optional, ungraded) if you want feedback: 11:59pm Friday, Nov 21

Submission:
Each team submits a single file on Gradescope, including
1. PDF of the report, Link to code, Source of your data (links only, do not submit actual data).
2. YouTube video link of the presentation (set to unlisted).
3. PDF slides.

Goal: This project is intended to be an additional learning experience to use or build on what we have learned throughout the class through working with a real dataset in a topic that you are interested in. You may work in **groups of up to 5** students. You will submit a report and give a short team presentation on what you did. See details below.

Description:
We have seen lots of different ML/analytics methods and techniques throughout the course. In this project, you will find or build a dataset in an area that you are interested in, conceptualize a use-case / application of a learning problem on that data, and then use ML / DA techniques to meet that described use-case / application. We have seen instances of this in class and on homeworks: Wine quality predictions, loan decision making, interpretable decision making in criminal justice, car pricing, medical decision making / patient risk identification… so there are applications in pretty much any industry or discipline.

I encourage you to pursue something that you care about and would be interested in outside of the course, or to think about something that could be part of a product/app/tool that could be useful.

I encourage you to take risks with this. Grading will be done holistically, and so something that is an interesting idea can absolutely receive full credit even if it doesn't really pan out in the end. This also goes for if you end up procuring your own dataset (whether scraping from the web or even gathering data "by hand").

Finally, I encourage you to look at this project as not something you are doing primarily for a grade or even for this class, but as something you can use to show to other people to showcase your skills.

Grading:

Grading will be done semi-holistically for this project; there is no exact “checklist” of a rubric to be followed because I want you to be able to focus on the aspects that you find the most interesting and the aspects that are simply most relevant for the topic you choose. However, here are the guidelines that will be part of the project:

- Project description and motivation /use-case: Is there a clearly identified, well-described use-case in the project? Is it well-motivated?
- Data and learning problem: Is the dataset to be used clearly described and explored, and is it clear what you are trying to learn from the data? If there are any particular challenges with the data, are they discussed? What kind of relationships might you expect to see? If you gathered data yourself, how did you go about this?
    - An excellent resource to find datasets is Google Dataset Search. I encourage you to try searching for different topics you may be interested in. This is a great resource to find sources of data for the project.
- Methodology: Is the methodology of applying ML / DA tools sound? Are multiple models considered and compared if relevant? Are the used models tuned / modified appropriately to try to improve performance?
- Discussion of results: Are the results presented in a clear way with meaningful metrics? If applicable, what kind of improvements in the actual decision making process does the method produce? Do you think the method has good enough performance to be useful in practice, and how would you judge its success? Note that “we tried, did this and this thing, but because of xyz reasons this is not good enough to be useful” may be completely valid!
- Presentation: You will give a short (~5 minutes) presentation discussing your project. This should give an overview of all of the above (of course there’s not enough time for too much detail, but that is not needed – you can keep things high-level).

These will each be ranked as “insufficient, lacking, satisfactory, great, exceptional”. Any project that is deemed satisfactory in all these aspects will receive a 90% mark. Each deviation above satisfactory is +5%, and any deviation below satisfactory is -5% (so in any category a “great” is +5%, an “exceptional” is +10%, a “lacking” is -5%, and an “insufficient” is -10% to the total project grade, but capped at 100%).

Bonus points possibilities (up to +5% bonus, past 100% possible)

- Build a dashboard (check out plotly/dash for example) / tool of some sorts / “deploy” your method and make it available online, allowing others (incl the instructor and GSIs) to interact with it or explore your application, method, results, and/or data. +2.5% for a great rating, +5% for an exceptional rating. Include a link in your report to the tool (preferred) or to a recording of a demo of it.

Report format:
Your project report will be up to 5 pages, aiming to capture the above points. The only addition is that your report should have an “Abstract” that gives a short summary of what you did. Include code as an appendix. You may also include additional figures in the appendix that you can reference in the main report if you need extra space for more figures.

On the Presentation:

- The presentations will be at most 5 minutes in length.
- Ideally, all project members would present, though we understand this may be difficult to coordinate in some cases so is not strictly necessary.
- Consider the presentation as a five minute “pitch” of your project idea and what you have done so far.
- The presentation should entice us to read your report to see how the “story” ends.

Finding a team:
We will make a megathread in Ed for finding teams. If you have an area / idea / application (or multiple) you are interested in, please make a post there mentioning what you are potentially interested in. Even if it’s vague (maybe something like “I’d like to do a project in sports / marketing / financial …. analytics”) that would probably be helpful.

Proposals:
You may optionally submit a proposal on Gradescope for feedback. This will be ungraded, only for feedback – you may format it however you wish. Please also feel free to come to office hours to discuss project ideas (if this ends up making OH too full, we can schedule). We will also devote discussion sessions on Dec 5 for proposal feedbacks.

Github:
I highly encourage you to use Github for this project to help you with coding collaboratively on the project as a team. Here is a guide for integrating Github in VSCode, though you can use it with any IDE. This way you can make a repository for your project code that all team members can access and contribute to. You will also almost certainly use Git in any ML / DA type job, so if you are not familiar it is a great tool to learn.