

Digging into Data: Project

The capstone of this course is a project to show off the skills that you have learned in the course. There are many different avenues to create a successful project:

1. Analyze a dataset
2. Create annotations for a dataset
3. Create a classifier for a problem that interests you
4. Create a regression for a problem that interests you
5. Create an unsupervised analysis of a dataset

There are many other possible topics you could tackle. Talk to me or discuss possible ideas on Piazza. Make sure that it represents a reasonable amount of work (I expect more from larger groups), but don't select something that is dissertation-sized.

Find Teammates

I strongly encourage you to form groups for this project. If you really want to work by yourself, you can, but I do not recommend it. Groups should have no more than four members, and most groups should have 2-3 people.

Write a Project Proposal

On April 7, you'll need to turn in a project proposal. This 2-3 paragraph post should describe:

- Who is in your group
- What you will do
- Where you will get the data
- Who will do what
- How do you know that what you did worked
- What you'll turn in for your midterm deliverable (more later)

I will create this as a separate Piazza thread. Reply with your proposal there.

Turn in a Project Deliverable

Before April 28, you'll need to turn in a project deliverable. This should be a super simple milestone that you need to complete before moving on to bigger and better things but that shows that you're working on your project.

For example:

- If you're doing classification, then you may want to run experiments with a basic set of features that you'll expand later for your project.
- If you're doing annotation, then you may want to annotate five data points and give definitions for each of your annotation classes.

This will not be graded separately, but will help me judge the group's progress thus far.

Present Your Project in Class

On May 12, you'll present your project in class. Each team will have 10 minutes to present their work. You can do a live demo, a slide presentation, or use the blackboard. Choose a medium that will best highlight your work.

Turn in a Project Writeup

By 23:59 May 12, have the person in your group whose last name is alphabetically first submit their project writeup explaining what you did and what results you achieved. This document should make it clear:

- Why this is a good idea
- What you did
- Who did what
- Whether your technique worked or not

Please do not go over 3000 words unless you have a really good reason. Images are a much better use of space than words, usually (there's no limit on including images, but use judgement and be selective).

Grading

The final project will make up a large part of your grade.

- *10 points: Presentation* - For your oral presentation, do you highlight what you did and make people care?
- *10 points: Clarity of writeup* - Does the writeup explain what you did in a way that is clear and effective?
- *10 points: Technical Soundness* - Did you use the right tools for the job, and did you use them correctly?
- *10 points: Effort* - Did you do what you say you would, and was it the right amount of work?