

# Course Project

## Overview

---

The project is your opportunity to apply data science techniques and machine learning algorithms to an interesting problem of your choice! You can choose the problem that fits your work domain or your research interests. Address a data-related problem in your professional field or a field you're interested in.

Want to enjoy this project?

Make sure you pick a subject that you're passionate about. You won't feel it is a project! Rather, it will be lots of fun moments exploring this data and I assure you that your performance and grade will be great!

Some public data sources can be found in a separate document.

## Project Team

---

You will work closely with other classmates in a 3-person project team. Please submit only one submission per team!

## Project Deliverables

---

### 1. **Project paper:** technical

That covers:

- Introduction and motivation
  - Clearly presenting the Problem statement and hypothesis
  - Contribution
  - Novelty: summarize how your work compares to existing approaches
- Dataset description
- Preprocessing (if any)
- Exploratory Data Analysis (EDA)
- Modeling process
  - Algorithm selection
  - Algorithm validation

- Challenges you faced
  - Conclusion and future work
2. **Project presentation:** pitch, storytelling, maybe a business application
  3. **Code:** well-narrated and documented

## Milestones

---

### Week 3: Proposal (Milestone 0)

What is the question you hope to answer? Why did you choose this topic? Who is your team?

Your proposal should be a PDF document, including:

- title of the project
- the full names of all of your team members and their ID numbers
- 350 word description of what you plan to do
  - Motivation
  - What is the problem?
  - Where will you get the data from?
  - Include at least one reference to prior research (related work).

### Week 7: Milestone 1

Just to make sure you are on track and working on your project (not leaving it till the last minute!!!).

I would recommend writing this milestone as if it is an early draft of the final report so that so that you can re-use most of it in your final report.

For your milestone, I expect you to have acquired, cleaned, and explored your dataset. You should also explain in more detail what will go into your final analysis.

Make sure you answer the following questions:

What data have you gathered and how?

How did you explore the data?

How did you clean your data?

What insights have you gained from your exploration?

What visualizations have you done?

What is your detailed plan going forwards?

**Format:** At most 3 pages that include:

- Introduction and Motivation: Describe again your problem in a style of report or research paper (~0.5 pages)
- Related work (~0.5 pages)
- Dataset and exploration (~0.5 pages)
- Preliminary experiments (~0.5 pages)
- Future plan: Given your preliminary results, what are the next steps that you're considering? (~0.5 pages)

## **Week 12 (Last Lecture): Milestone 2**

Final paper following the predefined structure!

Final presentation

Put all the deliverables in a folder and send it as a zip folder including: paper, slides, code, and data.

## **Publishing (Optional)**

---

If you feel that your work has a novel idea that can be submitted to a conference, please discuss this with me and I will give you some probable publication venues suitable to your work.