MATH637, Spring 2023
Final project

You will form groups of 5-6 students and work on a machine learning project. This project will contribute to **40% (+5% bonus)** of the total course grade.

Recommended Datasets

Below are some recommended datasets (with links). You are free to choose your own project (not included in this list) if you have data available. Some datasets require a Kaggle account and joining the corresponding competition to download.

1. Fraud detection
2. Predict survival on the Titanic
3. Predict air pollution
4. Predict corporate credit rating
5. Predict next-day rain in Australia
6. Sign language MNIST
7. Job change prediction
8. House price prediction
9. Healthcare analytics
10. Predict water quality

You can use the task associated with the above datasets, or you can also define your own task using a dataset. However, your project should focus on only **one task** (such as predicting a target). Note that you are **not** graded based on your model's accuracy nor whether you can successfully solve the task.

Deliverables

Deliverables for the project include: a presentation (10%) and a project report (30%). You will be graded based on how you apply the knowledge in the course to approach the problem, whether your experiment setups are reasonable to evaluate your methods, and whether your conclusions are supported by your experiments. You will also be graded based on the clarity of your report.

Bonus Points

You can score a maximum 5% bonus points toward the final course grade by: (1) using models that are not covered in the lectures, or (2) adding novelty to your project (such as proposing a new method). You should **explicitly specify such models or your novel method** in the report.

Project Report

Your project report should be in the IEEE conference format. You can download the LaTeX templates for the format here:

> https://www.ieee.org/conferences/publishing/templates.html

The maximum length of your report should be 4 pages, excluding the references. You can use at most 1 additional page for the references. Please strictly follow the above templates. You are not allowed to modify the template (such as changing the margin or font size).

Your report should include the followings:

- An abstract (short paragraph summarizing your work)
- An introduction (giving an overview of your work)
- Related work (discussing briefly previous work on the problem)
- Data and Methods (discussing the problem, the dataset, and your approaches to the problem, etc.)
- Experiments (detailing your experiment setups to evaluate your methods, presenting and discussing the results of your experiments; include any tables or figures to show your results)
- Discussions and conclusions (any discussions and conclusions that you can draw from your work)

Your report should contain enough details for a reasonably knowledgeable reader (such as your classmates) to understand and possibly reproduce your work. However, you do not need to include very basic knowledge, unless it is important for your arguments/discussions.