ML PROJECT Part 2:
Final project

CSCI 4371/5371/6397

Due in Blackboard by 11:59 pm on December 5

This project phase is worth **8 points**.

Instructions:

Submit your project code, which must be a .py file (more than one file is ok).

Submit a report describing your project. The report should include the following sections, which must be numbered and labeled:

1) **Data**. Describe your dataset.
2) **Methods**. Describe how you designed your experiment(s).
3) **Results**. Report the results of your experiments and interpret them. Use well-labeled tables and figures, rather than screenshots of program outputs, to help communicate your results clearly.
4) **Discussion**. Briefly reflect on (a) how the project went (e.g., obstacles you encountered), (b) what you would do differently in the future if you were to try this again, and (c) what you learned by doing the project.
5) **References.** If you used an LLM while working on your project, describe how you used it and give a few example prompts that you used. For any other resources you used, you can just list them.

Note that the number of modeling decisions you consider would meet the following *minimums* based on what section of the class you are enrolled in:

- Students in **CSCI 4371** should use at least **two different model types** and should investigate at least **four sets of modeling decisions**

- Students in **CSCI 5371** should use at least **three different model types** and should investigate at least **five distinct sets of modeling decisions**. At least one model type must be some form of artificial neural network.
- Students in **CSCI 6397** should use at least **three different model types** and should investigate at least **six distinct sets of modeling decisions**. At least one model type must be some form of artificial neural network.

For example, a student in CSCI 4371 could, at the minimum, choose to investigate logistic regression and kNN as two different model types, and the four sets of modeling decisions might be (1) logistic regression w/ default parameters, (2) kNN with k=3, (3) kNN with k=13, and (4) kNN with k=33.

Note that the goal of the project is not to satisfy the minimum requirements above, but to spend a large amount of time playing around with a modeling problem. The more you immerse yourself in the problem, try different things, and think through what is happening, the more you will learn.

Reports that include AI-generated language will receive a zero, will result in an academic integrity violation, and may result in a zero for the entire course.