CS 4342 Term Project

For the term project, we focus on applying multiple ML techniques on sample datasets. In the past, students worked on the datasets like:

Wine Quality Dataset, Number of classes: 10
https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Default of Credit Card Clients Dataset, Number of classes: 2
https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

however, **you can work on your choice of dataset**. Some of topics which can be of interest to our community are: 1) COVID-related datasets, 2) datasets addressing election results and/or predictions, and/or 3) mental health and social interactions.

Please note there are many online data repositories/websites that you can search to find ML-ready datasets. For instance, Kaggle website provides many interesting datasets that you can use. For most of these datasets, there scientific articles (and source code) examining different ML methods and their performance. You can search in Google Scholar or other search engines for the papers linked to the datasets. You can also use the available source code in your algorithm development; however, you need to reference any online source code you used in your project.

Term Project Rules:

1. You need to build your team – a maximum of 4 people can be in a team - and assign a name to your team by Nov 11th.
2. You are free to pick your ML algorithm(s). You can pick any algorithm(s), we studied in the class, or any other algorithms being suggested for these datasets.
3. You need to demonstrate your result through test/train dataset or cross-validation.
4. You need to provide a report that clearly describes your dataset, the questions that you want to answer, and demonstrate why you picked a specific approach (check **Report File**).

Grading:

1. The maximum score for each project is 8 points. This point will be assigned to each member of the team.
2. Grades are conditioned on providing a concise report, dataset, and source code. Teams without the report or source code will lose points for the term project.

Report file:

The report file should provide a concise description of your algorithm. Most of the Machine Learning (ML) problem starts with the data visualization and data cleaning steps. For your report, you need to provide these steps and explain what you did. You

need to describe what sort of ML techniques or models you examined and what was the result of each method – we expect you will check **at least three different methods**. You need to describe how you picked the best model. As a part of the report, you need to describe how you validated your result and what metrics you used for your model assessment. You need to reference the papers and online resources like the source code you used in your research.

Please upload the result to Canvas. You can also send your report to my email address, with CS 4342 and your team's name in the subject line.

Ethics of research: We will follow the following guideline in our research. https://libguides.library.cityu.edu.hk/researchmethods/ethics

Deadline: The project final deadline is Friday, Dec 16th, 9 pm. The project report and result being submitted after the deadline will not be graded.