CS 4342 Term Project

For the term project, we focus on applying multiple classification techniques on sample datasets. For the term project, you can work on one of the following datasets:

Project 1: Wine Quality Dataset, Number of classes: 10
The dataset for this project can be found on the course website (Canvas). The goal is to model wine quality based on physicochemical tests. The dataset description can be found at https://archive.ics.uci.edu/ml/datasets/Wine+Quality

Project 2: Default of Credit Card Clients Dataset, Number of classes: 2
The dataset for this project can be found on the course website. The goal of this project is to build an accurate classifier to predict the credit default. The dataset description can be found at https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

For each project, there will be a training and a test dataset. The test dataset does not have the target or class labels.

Please note that there are multiple scientific articles published on these datasets. You can search in Google Scholar or other search engines for the papers linked to these datasets. There are also many ML projects being developed for these datasets. You can find these projects in the Kaggle website, or other open-source platforms like GitHub. You can use these codes in your algorithm development; however, you need to reference any online source code you used in your project. Please note we have altered the datasets and the result on our test dataset might be different from what you get using the online datasets.

Term Project Rules:
1. You need to build your team – maximum 3 people can be in a team - and assign a name to your team by April 14th.
2. You are free to pick your classification algorithm(s). You can pick algorithms we studied in the class, or any other algorithms being suggested for these datasets.
3. You will get both the training and the test data. For the test data, the target or class labels are being omitted.
4. You will report your classification result on the test dataset in a CSV file.

Grading:
1. For each project, the team with the highest score will get the complete point plus 5 bonus points (15 points for each member of the team) – the score will be defined based on the prediction performance.
2. For each project, the team with the second-highest score will get the complete point plus 2 bonus points (12 points for each member of the team)
3. Teams based on their ranking in the score list will get from 10 to 5 points.
4. Grades are conditioned on providing the report and source code. Teams without the report or source code will lose the points for the term project.

Test result data format:
You need to provide a CSV file with only one column in it, where the column provides the classification result.
For the first dataset, the number is from 1 to 10 corresponding to the wine quality. A value 0 per sample will be assigned randomly to either of 10 classes. A file with any number beyond 10 will receive score 0.
For the second problem, the number is 1 or 2 corresponding to the credit default. A value 0 per sample will be assigned randomly to either of the two classes. A file with any number beyond 2 will receive score 0.

Report file:
The report file should provide a concise description of your algorithm. Most of Machine Learning 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 classifier models you examined and what was the result of each classification method – we expect you will check at least three different methods. You need to describe how you picked the best classifier 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 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: Project final deadline is May 5th, 11:59 pm. The project report and result being submitted after the deadline will not be graded.