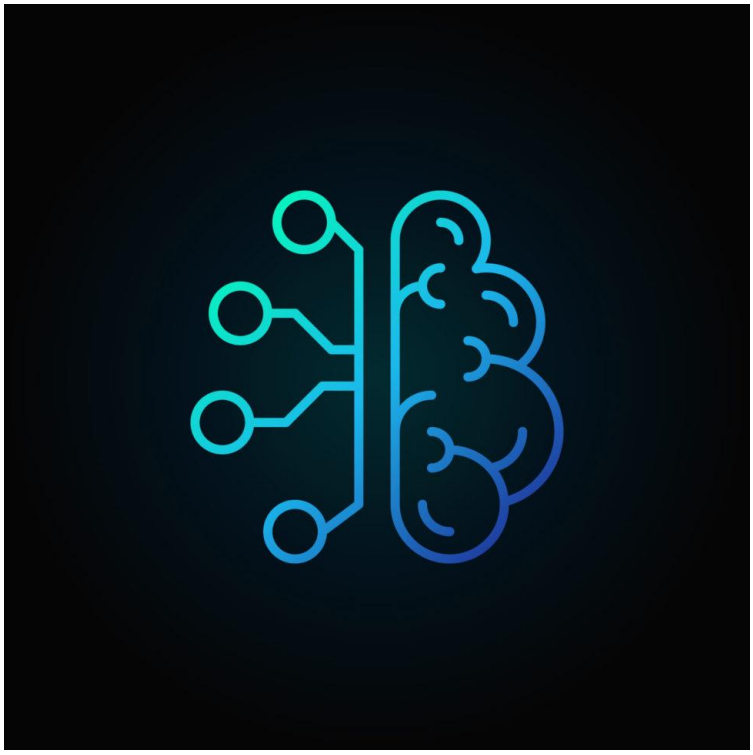

School Name	School of Computing
Semester	AY2324 Semester 2
Course Name	DAAA
Module Code	ST1511
Module Name	AI & Machine Learning

Assignment 2 (CA2: 40%)

The objective of the assignment is to help you gain a better understanding of machine learning tasks of time series forecasting and unsupervised learning.

Guidelines

1. You are to work on the problem set individually.

2. In this assignment, you will solve typical machine learning tasks.

3. Write a Jupyter notebook including your code, data visualization and comments. Create a presentation PowerPoint slides for your work.

4. You are also required to submit the “Declaration of Academic Integrity” form.

5. Submit your Jupyter notebook, PowerPoint slides and Declaration of Academic Integrity form in a zip file. Please name the zip file using this format: **p1234567_WilsonQiu_CA1.zip**.

6. The normal SP’s academic policies on Copyright and Plagiarism applies. Please note that you are to cite all sources. You may refer to the SP academic policy at: https://www.sp.edu.sg/docs/default-source/as-exams/sp_policy-on-use-of-ai-tools-for-academic-work4272ccebf2744e1e9945401cef8f0d2b.pdf?sfvrsn=f00682de_0#:~:text=Academic%20Integrity%20is%20a%20central,copying%20and%20using%20plagiarised%20material

Submission Details

Deadline: Feb 9, 2024, 23:59
Submit through: BrightSpace

Late Submission

50% of the marks will be deducted for assignments that are received within ONE (1) calendar day after the submission deadline. No marks will be given thereafter.
Exceptions to this policy will be given to students with valid LOA on medical or compassionate grounds. Students in such cases will need to inform the lecturer as soon as reasonably possible. Students are not to assume on their own that their deadline has been extended.

PART A: Stock Price Forecasting (50 marks)

Background

Using the Stock Price Dataset to train time series models and forecast the stock price of Apple, Amazon and DBS for the next 60 days.

Dataset

You are to use the dataset: **CA2-Stock-Price-Data.csv**.

Tasks

1. Write the code to solve the prediction task. You should use Statsmodels library to build the forecasting models (do not use other 3rd party libraries such as autoML).

2. Tune the hyperparameters of the time series model to maximize the accuracy for training data and testing data prediction.

3. Write a short report detailing your implementation, your experiments and analysis in the Jupyter notebook (along with your python code and comments).

4. Create a set of slides with the highlights of your Jupyter notebook report. Explain the entire machine learning process you went through, data exploration, data cleaning, feature engineering, model building and evaluation, and model improvement. Write your conclusions. The slides should not exceed 20 pages

Evaluation criteria:

Background Research & Data Exploration	10 marks
Feature Engineering	10 marks
Modelling and Evaluation	10 marks
Model Improvement	10 marks
Demo/Presentation and Quality of report (Jupyter)	10 marks

PART B: Student Segmentation (50 marks)

Background

We collected the information of 1000 students including their age, gender and subject scores. We want to group the students into multiple clusters so that we can cater needs for different students.

By the end of this study, you would be able to answer below questions.

- How to perform student segmentation using unsupervised machine learning algorithm in Python?
- Describe the characteristics of each student cluster.
- Which group of students needs to be more attention to?

Dataset

You are going to use the dataset: **Student_Performance_dataset.csv**.

Tasks

1. Write the code to solve the clustering task. Use scikit-learn library to build the model (do not use other autoML libraries).

2. **In the Jupyter notebook**, write your report detailing your implementation, your experiments and analysis (along with your python code and comments).

3. Test your clustering with different possible values of k.

4. Determine the best possible value of k. And show how you can determine that this is the best value for k.

5. Use more than just one clustering (k-means) algorithm.

6. Create a set of slides with the highlights of your Jupyter notebook report. Explain the entire machine learning process you went through, data exploration, data cleaning, feature engineering, model building and evaluation, and model improvement. Write your conclusions. The slides should not exceed 20 pages.

Evaluation criteria:

Background Research & Data Exploration	10 marks
Feature Engineering	10 marks
Modelling and Evaluation	10 marks
Model Improvement	10 marks
Demo/Presentation and Quality of report (Jupyter)	10 marks

— End of Assignment —