



School Name	School of Computing
Semester	AY2324 Semester I
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 regression 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 and write a report that describes your solution to the tasks.
3. Write a Jupyter notebook including your code and comments and visualizations. Create a short presentation file (about 10 slides) for your project. Submit your Jupyter notebook, data and the slides in a compressed package (zip file).
4. Students are required to submit their assignment using the assignment link under the Assignment folder. Please remember to include your student name and student admission number on the first page of your assignment report.
5. 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 citation guide available at: http://eliser.lib.sp.edu.sg/elsr_website/Html/citation.pdf

Submission Details

Deadline: August 11, 2023, 23:59H

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: TIME SERIES (50 marks)

Background

- a) Using the Energy Consumption Dataset to train time series models and forecast the gas consumption, electricity consumption and water consumption in the future.
- b) You will be given a csv dataset to build your time series model.

Tasks

1. Write the code to solve the time series prediction. For the time series model, use Statsmodels only (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. Explain the time series prediction process, model building and evaluation. Write your conclusions.

Submission requirements

1. Submit a zip file containing all the project files (Jupyter notebook), all data sets used, and the slides (PPTX or pdf).
2. Submit online via the Assignment link.

Evaluation criteria:

Background Research & Data Exploration	25%
Modelling and Evaluation	25%
Model Improvement	25%
Demo/Presentation and Quality of report (Jupyter)	25%

PART B: UNSUPERVISED LEARNING (50 marks)

Background

You are running a shopping mall, and you have some data about your customers like Age, Gender, Income and Spending.

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

- How to perform customer segmentation using unsupervised machine learning algorithm in Python?
- Describe the characteristics of each customer cluster.
- Which group of customer is the most valuable to the shopping mall so that the shopping mall should do something to retain them.

Dataset

Use the Customer_dataset.csv

Tasks

1. Write the code to solve the clustering task. Use scikit-learn only (do not use other 3rd party libraries).
2. Write a short report detailing your implementation, your experiments and analysis in the Jupyter notebook (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 are able to determine that this is the best value for k.
5. Use more than just one clustering (k-means) algorithm.
6. Create a set slides with the highlights of your Jupyter notebook. Explain the unsupervised machine learning process, model building and evaluation. Write your conclusions.

Submission requirements

1. Submit a zip file containing all the project files (Jupyter notebook), all data sets used, and the slides (PPTX or pdf).
2. Submit online via the Assignment link.

Evaluation criteria:

Background Research & Data Exploration	20%
Feature Engineering	20%
Modelling and Evaluation	20%
Model Improvement	20%
Demo/Presentation and Quality of report (Jupyter)	20%

— End of Assignment —