School Name	School of Computing
Semester	AY2022-2023 Semester 2
Course Name	DAAA
Module Code	ST1508
Module Name	Practical AI

CA2 (100%)

The objective of CA2 is to help you gain a better understanding of the data science project workflow, including applying advanced data processing techniques on the dataset, developing and optimizing machine-learning models, tracking and monitoring machine learning models, and deploying the machine learning model as a graphical user interface.

Guidelines

1. You are to work on the data science project on a group basis.

2. In this assignment, you will build the second part of a data science project and write a report that describes your solution to the tasks.

3. Write a Jupyter Notebook including your code, comments.

4. Create a nice presentation slide for your project.

5. Students are required to submit their assignment using the assignment link under the CA2 folder. Please remember to include student names and student numbers in the notebook and slides (using zip file, name of file, a student name+class, e.g. John-2B02).

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 citation guide available at: http://eliser.lib.sp.edu.sg/elsr_website/Html/citation.pdf

Submission Details:

Deadline: 2023-02-10 23:59H
Submit through Blackboard

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.

Task:

1. Apply **advanced techniques** on the dataset (you can use, but not limited to, feature selection, feature engineering, high-dimensional outlier detection, etc.)

2. Select, build, optimize and scale **machine-learning models** on this huge dataset.

3. **Track and monitor machine learning** cycles and experiments (you can use MLflow, or other third party tools).

4. Deploy the prediction models as a **graphical user interface** (you are recommended to use Tkinter, or other third-party libraries approved by your lecturer).

CA2 deliverables:

1. A well-written **project report** (word document, < 20 pages), describing the whole process in phase two, from advanced data processing, machine learning modelling, to machine learning experiment monitoring and deployment..
2. A **Jupyter Notebook** (.ipynb)**,** with advanced data processing, machine learning modelling and optimization.
3. An **executable graphical user interface** with the prediction models.
4. **PowerPoint slides** (< 30 pages) for 20-minute presentation.
5. **Scrum Document**: Product Backlog, Bi-Weekly Scrum Reports..
6. **Peer Evaluation Form** (every individual student needs to submit the form)

CA2 Evaluation Criteria:

Advanced data pre-processing	10 %
Machine learning modelling and optimization	30 %
Track and monitor machine learning cycles	10 %
Phase 2 Report & Scrum Documentation	20 %
Final presentation and demonstration	15 %
Build graphical user interface	15 %