SINGAPORE UNIVERSITY OF SOCIAL SCIENCES

ENG335
Machine Learning

Group-based Assignment 02

July 2023 Presentation

GROUP-BASED ASSIGNMENT

GROUP-BASED ASSIGNMENT

This mini-project assignment is worth **15%** of the final mark for **ENG335 Machine Learning**. Total mark assigned to this assignment is 100 marks.

The cut-off date for this assignment is **14 Oct 2023, 23 55 hrs.**

This is a group-based assignment. You should form a group of **maximum** 5 members from your seminar group. Each group is required to upload a single report to Canvas Turnitin via your respective seminar group. Please elect a group leader. The responsibility of the group leader is to upload the report on behalf of the group. In your 1-page cover sheet, please include all project partners' names and student PI numbers.

Note to Students:

You are to submit the GBA assignment i.e. using Canvas in the form of a single MS Word file. It should be saved as ENG335_GBA02_group_number.doc Submission in any other manner like hardcopy or any other means will not be accepted. You are to ensure that the file to be submitted does not exceed 20MB in file size.

Additional Instructions for Submission:

Please follow the submission instructions stated below:

1. Please submit all Program Code / Answers in the form of a Juypter Notebook file (i.e. .ipynb File) for all the programming questions via the additional submission link found under Assignments on ENG335 T01 course site.

2. All Answers for each question should be indicated clearly using the Comments section / markups in the Notebook so that the marker can see clearly which code is for which Question. (e.g. # Answer for Q1a).

3. Submit the file before the submission cut-off date/time via the Canvas T-group course site in **Assignments> GBA02**. You will be then directed to the submission page. Late submission will be subjected to the mark deduction scheme. Please refer to the Student Handbook for details.

Answer all questions (100 marks)

Questions 1

Generative AI is being quickly adopted in diverse fields and there is also a growing concern among people. Discuss at least FIVE (5) ethical concerns of Generative AI.

(10 marks)

Questions 2

Read about the following terms and explain them based on your understanding.

- MLOps
- AGI. Give a real-life example of AGI.
- TinyML and a real-life implementation of TinyML

(9 marks)

Questions 3

Look out for an example for reinforcement learning. Provide snapshots of the working of the reinforcement learning and explain the functioning of the algorithm based on your understanding. Note: Just have a maximum of TWO (2) snapshots and a few lines of text to explain your understanding.

(6 marks)

Questions 4

Load the breast cancer dataset from sklearn package. Perform exploratory data analysis and set up a KNN classifier. Propose an appropriate value for K. Show the relevant performance metrics. Assess whether scaling the data improves the performance. Compare the performance of the best KNN classifier with the Naïve Bayes algorithm.

(25 marks)

Questions 5

Use the wine dataset available in sklearn package. Drop the target variable and apply K-means clustering algorithm. Select appropriate value for K. Discuss any TWO (2) performance metrics.

(25 marks)

Questions 6

Use the diabetes dataset from Kaggle (https://www.kaggle.com/c/diabetes-classification/data). You should use the train.csv dataset for both your training and testing. Explain the attributes and the target. Perform exploratory data analysis and set-up a decision tree. Set up a SVM classifier and compare the performance of the decision tree with the SVM classifier. You are required to perform hyper parameter optimization wherever possible. Compare the performance of the decision tree and SVM classifiers.

(25 marks)

-----END OF GBA ASSIGNMENT-----