



ENG335

Machine Learning

Group-based Assignment (2)

July 2021 Presentation

GROUP-BASED ASSIGNMENT

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

This is a group-based assignment. You should form a group of **minimum** 2 and **maximum** 4 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_yournames.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 Jupyter Notebook file (i.e. .ipynb File) for all the programming questions via the additional submission link found under Assignments on ENG335 T group 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).

The submission deadline for this assignment is announced on L01 course site. **Late submissions carry mark penalty.**

Questions: (Total 100 marks)

- (1) Machine learning algorithms and AI are being quickly adopted in diverse fields and there is also a growing concern among people. Discuss at least **FIVE (5)** ethical concerns of AI.
(10 marks)
- (2) Find out about edge computing and explain this term in your own words. How does Edge ML works?
(10 marks)
- (3) Take any picture using your phone and upload it to Google Vision API (<https://cloud.google.com/vision>). Show the snapshot of your upload and the results. Guess the type of algorithm being used from the results provided by the Google Vision API.
(5 marks)

- (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.
(30 marks)
- (5) Use the iris dataset available in sklearn package. Drop the target variable and apply K-means clustering algorithm. Select appropriate value for K and provide relevant performance metrics.
(15 marks)
- (6) Use the diabetes dataset from Kaggle (<https://www.kaggle.com/c/diabetes-classification/data>). You should use the train.csv dataset for 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 optimization wherever possible. Deploy your trained and tested algorithms on the dataset in test.csv. Compare if the decision tree and SVM classifiers provide the same results for the dataset in test.csv.
(30 marks)

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