

WQD7006 Machine Learning for Data Science

Group Assignment (4-5 per pax)

30%

Deadline 1 Jan 2025 11.59pm

Submission Link: <https://forms.gle/yAmq98U1wrBtnzLH6>

Group Link:

<https://docs.google.com/spreadsheets/d/1vZPmcSUXVLoICr6lAxtqRBhp9lvq2MecsuOliusriSY/edit?usp=sharing>

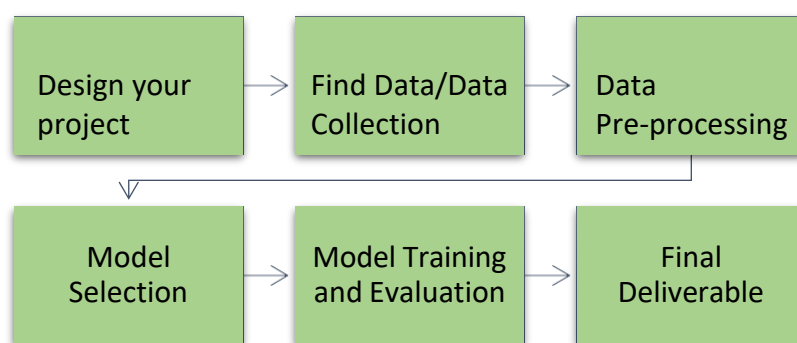
This project is designed for Master of Data Science students who are taking the WQD7006 Machine Learning for Data Science.

For the theme of the project, your group can choose from the themes below. Choose only ONE theme.

1. Machine Learning related to your current work (if you are working)
2. Machine Learning in Finance
3. Machine Learning in Engineering
4. Machine Learning in Environment

Note: Make sure your project is not something popular or that can easily be found on Kaggle.

In this project, you will be working in a group of 5 to explore and analyze a specific theme using machine learning techniques. The goal is to train and optimize several machine learning models and compare their performance in solving real-world problems.



1. Design your project: As a group, you should first brainstorm and come up with a specific theme that you are interested in exploring. This could be anything from predicting the weather (Renewable Energy related) to predicting when a machine will fail (Predictive Maintenance).
2. Data Collection: Once you have decided on a theme, you will need to collect/find data that

is relevant to your project. You can use publicly available datasets or collect data on your own. Ensure that your data is relevant, reliable, and sufficient for your project.

3. **Data Pre-processing:** After collecting your data, you will need to preprocess it to make it suitable for machine learning. This may include cleaning, normalization, feature extraction, and transformation.
4. **Model Selection:** Once your data is ready, you will need to choose an appropriate machine learning algorithm for your project. You may choose from supervised or unsupervised learning, and select the model based on the problem you are trying to solve.
5. **Model Training and Evaluation:** Once you have selected the model, you will need to train it on your preprocessed data and evaluate its performance using appropriate metrics. You may need to fine-tune your model to improve its performance.
6. **Final Deliverable:** Finally, you will need to present your findings and results in a clear and concise manner. Document it using Google Colab/Kaggle/Jupyter Notebook. There are 3 deliverables that you must submit
 - (i) The link to your notebook. Please make grant access for me to view and assess your work.
 - (ii) A poster presentation (Just one page). It must contain (1) The problem (2) The dataset (3) The Insights (4) The Machine Learning Models and Methods (5) Results and Conclusion
 - (iii) A Video presentation of 5 minutes explaining your poster. Everyone needs to present, so if 5 people per group, that'll be around 1 minute per person. As you can see, the poster has 5 parts as above.