MSc in Data Science – Data Analytics and Algorithms
Continuous Assessment

Provisional Date Due	:	Friday, April 11th, 2025
Time	:	11:59pm
Value	:	100%

INSTRUCTIONS TO CANDIDATES:

- Include your name and login-id as a comment at the top of all of YOUR submissions and/or code.
- You must submit all program, document and other files for this assessment. You may use any method for submission including Jupyter notebooks, GitHub, etc., however you **must** reference all sources used correctly.
- **NOTE** It is your responsibility to keep a backup(s) of all solution(s) and other files for this assessment.
- You may use any programming language(s) or other technologies you choose; however, you must provide a detailed outline of your reasoning behind your choice(s).
- ***This is an INDIVIDUAL assessment***; students will be assessed and marked individually.
- ***Any attempt at plagiarism will be referred to the office of the Registrar & VP for Academic Affairs and dealt with accordingly***.

Project Description

You are required to create a professional data science & machine learning portfolio. You are required to add a Jupyter notebook implementation for each machine learning algorithm covered in the module. You may use the Jupyter notebooks provided in the module notes or you may choose to create your own notebooks, or alternative implementation; however, alternative implementations must be discussed and agreed with the module lecturer.

For each notebook, you are required to add the notebook to your portfolio, and you **must make significant** adjustments/alterations/additions to the notebook. The purpose of these adjustments/alterations/additions is to gain significant theoretical and practical understanding of the workings of the underlying machine learning algorithm(s).

You are required to maintain a consistent and **detailed** log of the work that you undertake for each algorithm; this can be in any format you choose, however it will most likely form part of your revised notebook for each algorithm. This log should be logically structured and must show clearly show the adjustments/alterations/additions that you have made to each notebook, the practical impact of the adjustments/alterations/additions (e.g. changes in the accuracy measures for the algorithm) and your analysis of the adjustments/alterations/additions. You may use a process model to help you format the structure of this log and to help track and identify the work that you complete on each of the notebooks that you add as projects to your portfolio. For example, CRISP-DM has six phases and you could complete work on your models in line with each of these phases and document this work in your log under these six phases.

While all aspects of the assessment are important, the following are considered the most important aspects of this continuous assessment 1) using new/substantially altered data set(s) and making the algorithm(s) work on this new/altered data set(s), 2) changing the algorithms (i.e. using different versions of the algorithms) and tuning the parameters/hyperparameters of the algorithm(s) and 3) creating a minimal deployment (e.g. app, webapp etc.) of a serialised version of the model.

By the end of the reading week of this semester (11:59pm, Friday 21st February, 2025), you are required to submit your data science & machine learning portfolio containing the algorithm(s) completed to date, worth 30%.

The final completed version of your portfolio must be submitted by the end of week 10 of this semester (11:59pm, Friday March 28th, 2025), worth 40%.

You are required to present and demonstrate your portfolio to the class group in weeks 11 or 12, the last two weeks of the semester (Thursday April 3rd & 10th, 2025) – this is worth 30%. A presentation schedule will be circulated closer to the time, but you should be prepared to present anytime from April 3rd, 2025. In this presentation you should focus on the technical work that you have completed on the notebooks and in particular on the 3 main elements identified above. No marks will be awarded for your overall work if you fail to present and demonstrate it – please note that this is your opportunity to demonstrate the quality of your portfolio.

A professional standard of work is expected for all elements of the assessment. Include comments in your code as you deem necessary. Marks will be awarded for presentation, adhering to best programming practices (good use of comments, naming conventions, using functions etc.).

Your final submission must include your completed portfolio, all code, notes and sources of information and anything else that you feel is appropriate to be included.

If you have any queries regarding this assessment, please do not hesitate to contact the lecturer at any time.

Deliverables

1) Portfolio completed to date (30%) 11:59pm, Friday, February 21st, 2025

   This is the **final** version of the notebooks that should be completed at this time.

2) Final Portfolio (40%) 11:59pm, Friday, March 28th, 2025

   This is the **final** version of the additional notebooks that should be completed at this time.

3) In–class presentation and demonstration (30%) Thursday April 3rd & 10th, 2025

   You must present and demonstrate or no marks will be awarded for your portfolio

*All dates are provisional and are subject to change

E&OE