

Objective:

The main objective of this coursework is to provide you with a hands-on experience of posing a research question, identifying a dataset, and employing machine learning techniques to answer that question.

1. Dataset Identification:

Source a dataset from reliable platforms such as **GitHub**, **kaggle.com**, **data-is-plural.com**, or any other reputable data archive.

2. Research Question:

Formulate a research question that can be addressed using machine learning techniques based on the dataset you've chosen. By machine learning methods, we mean something like regression, classification, or clustering.

3. Data Analysis:

Download and, if necessary, preprocess the dataset to make it suitable for analysis. Use the machine learning methods best suited to your research question and dataset to analyse the data.

Deliverables

You are expected to submit 3 files:

1. Presentation Slides:

Prepare a 7-minute presentation encompassing:

- The research question you're addressing. Why should anyone care about it?
- Introduction to your dataset.
- The type of machine learning technique used (e.g., regression, classification, clustering).
- A description of the machine learning models employed.
- An overview of your data analysis and results.
- Concluding remarks on what you have learned.
- A concise discussion on the limitations of your data and chosen methods.

2. Presentation Video:

- Make sure to upload your presentation to either ELE, Panopto, OneDrive or YouTube.

- If you choose to not upload the video directly on ELE, please upload a txt file with the link to your presentation video. Make sure you grant the appropriate right access to it.
- If the video is not accessible for markers, it will not receive any marks (i.e., zero mark).

3. Executive Summary:

- Draft a concise, 2-page executive summary, including references, with an additional page dedicated for a figure, which can have multiple subfigures/panels).
- This document should encapsulate all aspects discussed in the presentation.
- An unlimited appendix can be attached containing any supplementary figures, tables, or relevant material. Note, the appendix will not be marked.
- Use the LaTeX template [here](#), also attached below, or make something similar in Word.
- To copy the Overleaf template:

Submission:

- Upload the presentation slides (as a PDF), the video or the txt with a link to your video, and the executive summary (as a PDF) to ELE by the deadline.
- The university will release further instructions about the new submission system via ELE.

Deadline:

All submissions must be made by the **6th December 2023, 11:59am**.

Please adhere to the requirements mentioned above. This coursework offers an invaluable opportunity to not only understand the nuances of machine learning but also to communicate your findings to an audience. We hope you find this experience both challenging and rewarding. Additionally, you may use this opportunity to anticipate analyses and/or explorations related to your final project (outside this module).

Best wishes on your research and analysis!

Marking Criteria

These assessment criteria are designed to evaluate students' performance in key areas that align with the Intended Learning Outcomes (ILOs) of the module. They ensure that students are not only capable of applying knowledge and skills but also of thinking critically, formulating research questions, and effectively communicating their findings.

1. Problem Formulation and Research Question (20%):

Criteria: To what extent did the student formulate a clear research question answerable with the chosen dataset? Did the research question demonstrate an understanding of the principles of machine learning techniques?

Rationale: This criterion assesses the students' ability to apply principles of machine learning to formulate a research question. It also evaluates their capacity to analyse novel problems and establish a well-defined question given the data availability.

2. Utilisation of Machine Learning Techniques (50%):

Criteria: How effectively did the student apply a range of supervised and unsupervised machine learning techniques to address the research question and analyse the dataset? Did they justify the selection of these techniques based on the problem at hand? Did they prepare the data accordingly?

Rationale: This criterion evaluates the students' ability to utilize machine learning techniques to solve real-world problems, demonstrating their proficiency in addressing novel challenges.

3. Critical Thinking and Reflection (30%):

Criteria: To what extent did the student discuss the advantages and disadvantages of every choice that was made in the process of the model? Did they effectively identify the compromises and trade-offs that must be made when translating theory into practice? Were the limitations of the research clearly acknowledged?

Rationale: This criterion assesses the students' ability to engage in critical thinking, demonstrating an understanding of the importance of each choice made in their machine learning pipeline, and their ability to identify and reflect on the compromises and trade-offs involved in practical applications.