



University
of Exeter

COURSEWORK SPECIFICATION

ECMM445 – Learning from Data

Module Leader: Diogo Pacheco

Academic Year: 2025/26

Title: **Find your data and learn from it**

Submission deadline: **18 November 2025, 11:59AM**

This assessment contributes **30%** of the total module mark and assesses the following **intended learning outcomes**:

1. Apply principles for statistical and neural pattern recognition to novel data.
2. Analyse novel pattern recognition and classification problems, establish models for them and write software to solve them.
3. Utilise a range of supervised and unsupervised pattern recognition and machine learning techniques to solve a wide range of problems.
4. State the importance and difficulty of establishing principled models for pattern recognition.
5. Use Python or other programming languages for scientific analysis and simulation.
6. Identify the compromises and trade-offs that must be made when translating theory into practice.
7. Critically read and report on research papers.

Plagiarism

This is an **individual assessment**. Plagiarism is interpreted by the university as the act of presenting the work of others as one's own work, without acknowledgement. It is considered academically fraudulent and an offence against university discipline. Your attention is drawn to the [university's regulations on plagiarism](#). Your work will be scrutinised for detection of plagiarism.

Generative AI

This assessment has been categorised as **AI-Assisted**. You can find further information in the [University's policies around using AI in assessed work](#). For this assessment, you are allowed to use GenAI tools to support different stages of your work. Acceptable uses include developing ideas, assisting with research or gathering information, and helping you to understand key theories and concepts. You may also use GenAI tools to suggest a plan or structure for your assessment and to provide feedback on a draft. In addition, you are permitted to use them to generate code for images, figures, or diagrams (but not for the underlying data shown), as well as to proofread and correct grammar or spelling errors. Finally, you may use GenAI tools to identify potential citations or references; however, all references included in your work must be properly accessed and read by you.

Coursework Specification

Instructions

As part of this term's coursework, you will apply the tools and concepts covered in class to analyse a dataset. This assignment is designed to give you practical experience in formulating a research question, applying machine learning methods, and communicating your findings effectively through both written and oral formats.

Objective

The main objective of this coursework is to provide you with hands-on experience in:

- Posing a research question.
 - Identifying and preparing a dataset.
 - Employing machine learning techniques to address the question.
 - Communicating research findings clearly in both written and presentation formats.
-

Requirements

1. Dataset Identification

Select a dataset from a reliable source such as GitHub, Kaggle, Data Is Plural, or another data archive.

2. Research Question

Formulate a research question that can be addressed using machine learning techniques (e.g., regression, classification, clustering). Ensure that your question is meaningful and relevant.

3. Data Analysis

Download and, if necessary, pre-process the dataset to make it suitable for analysis. Apply appropriate machine learning methods to address your research question.

Deliverables

You are required to submit two deliverables: **an executive summary report** and **a presentation**. These are assessed together as a single submission, with the marking criteria applied across both.

1. Executive Summary Report (Primary Deliverable)

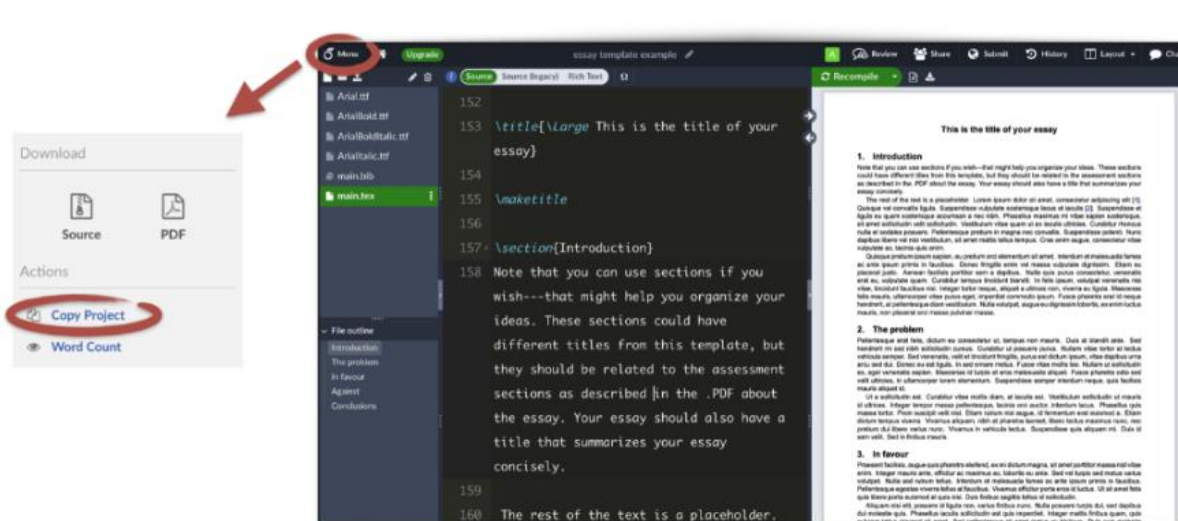
The executive summary report is the core of your submission and must be **4 pages** in length, structured as follows:

- **Pages 1–2:** Main text (research question, methods, analysis, results, conclusions).
- **Page 3:** Figures (plots, tables, visualisations).
- **Page 4:** References (plus video link if your presentation is pre-recorded).

An **unlimited appendix** may be attached containing code, supplementary figures, tables, or additional material. The appendix will not be marked, but it can provide supporting detail.

Format guidelines:

- Use the [provided LaTeX template](#) (or create a similar format in Word). The template includes additional instructions and guidance on how to prepare your report, including detailed notes on what content to include in each section.
- To copy the Overleaf template:



2. Presentation (Supporting Deliverable)

The presentation is a concise demonstration of your report. It should be **no longer than 7 minutes** and must use slides.

Emphasis should be placed on **visual communication** to make your findings clear and engaging. This includes figures, plots, and tables that illustrate what was learned from the data.

The presentation should cover:

- Your research question and why it matters.
- An introduction to the dataset.

- The machine learning technique(s) applied.
- A description of the models used.
- Key findings supported by visuals.
- Reflections on limitations and what was learned.

The presentation format will be confirmed in advance (live or pre-recorded).

If the presentation is **pre-recorded**:

- Upload your video to ELE, Panopto, OneDrive, or YouTube.
 - Provide the video URL at the end of your report.
 - Ensure the video is accessible (at a minimum, within the university).
 - Note: if the video cannot be accessed by the markers, it will be considered as not submitted.
-

Marking Criteria

Your work will be assessed using the following criteria, applied across both the executive summary report and the presentation.

1. Problem Formulation and Research Question (20%)

Criteria:

- Did you formulate a clear and well-defined research question that can be answered using the chosen dataset?
- Does the research question demonstrate an understanding of machine learning principles?

Rationale:

Assesses your ability to apply machine learning principles in framing a research question and to identify novel problems based on available data.

2. Application of Machine Learning Techniques (50%)

Criteria:

- How effectively did you apply supervised and/or unsupervised machine learning techniques to address your research question?
- Did you justify your choice of techniques in relation to the dataset and problem?
- Was the dataset suitably prepared and pre-processed for analysis?

Rationale:

Assesses your ability to apply machine learning methods to real-world data, demonstrating technical proficiency, appropriate methodological choices, and practical handling of challenges.

3. Critical Thinking and Reflection (30%)

Criteria:

- Did you discuss the advantages and disadvantages of your methodological choices?
- Did you identify and reflect on compromises and trade-offs made when translating theory into practice?
- Were the limitations of the dataset, methods, and results clearly acknowledged?

Rationale:

Assesses your ability to critically evaluate your work, demonstrating awareness of trade-offs, practical limitations, and the implications of your methodological decisions.

Important Note on Presentation Submission

Although the report and presentation are assessed together, a **penalty of 30% of the total marks** will be applied if the presentation is **not submitted** or if the submitted video **cannot be accessed** by the markers.

Submission

- Upload your executive summary report as a **PDF** to ELE by the deadline.
 - If your presentation is pre-recorded, include the video URL in your report.
 - Further submission system details will be released via ELE.
-

Deadline

18 November 2025, 11:59 AM

Late submissions will follow the university's standard late submission policy.

Final Notes

- The **executive summary report** is the central deliverable. The **presentation** is a demonstration of your report, designed to highlight key insights with effective visuals.
- You are **MSc or 3rd year undergraduate students**. The **quality and complexity** of your research question and dataset should reflect this.
- Avoid simple datasets commonly used for teaching demonstrations. If you choose such a dataset, you must extend your analysis to go **beyond the ordinary**.

We hope you find this coursework both challenging and rewarding. Best wishes on your research and analysis!