COURSEWORK SPECIFICATION

ECMM445 & ECM3420 – Learning from Data

Module Leaders: Diogo Pacheco & Chico Camargo

Academic Year: 2024/25

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

Submission deadline: **03 December 2024, 11:59AM**

This assessment contributes **40%** of the total module mark and assesses the following **intended learning outcomes**:

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

This is an individual assessment and you are reminded of the University's regulations on collaboration and plagiarism. You must avoid plagiarism, collusion, and any academic misconduct behaviours. Further details about academic honesty and plagiarism can be found at https://ele.exeter.ac.uk/course/view.php?id=1957.

Use of GenAI tools

The University of Exeter is committed to the ethical and responsible use of Generative AI (GenAI) tools in teaching and learning, in line with our academic integrity policies where the direct copying of AI-generated content is included under plagiarism, misrepresentation and contract cheating under definitions and offences in TQA Manual Chapter 12.3. To support students in their use of GenAI tools as part of their assessments, we have developed a category tool that enables staff to identify where use of Gen AI is integrated, supported or prohibited in each assessment. This assessment falls under the category of ***AI-supported.*** *This is because if they are used wisely, they can improve your understanding of the subject and the quality of your work.*

You can find further guidance on using GenAI critically, and how to use GenAI to enhance your learning, on Study Zone digital.

When submitting your assessment, you must include the following declaration, ticking all that apply:

AI-supported/AI-integrated use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

- ❑ *I have used GenAI tools for developing ideas.*
- ❑ *I have used GenAI tools to assist with research or gathering information.*
- ❑ *I have used GenAI tools to help me understand key theories and concepts.*
- ❑ *I have used GenAI tools to identify trends and themes as part of my data analysis.*
- ❑ *I have used GenAI tools to suggest a plan or structure for my assessment.*
- ❑ *I have used GenAI tools to give me feedback on a draft.*
- ❑ *I have used GenAI tool to generate image, figures or diagrams.*
- ❑ *I have used GenAI tools to proofread and correct grammar or spelling errors.*
- ❑ *I have used GenAI tools to generate citations or references.*
- ❑ *Other: [please specify]*
- ❑ *I have not used any GenAI tools in preparing this assessment.*

I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.

Please note: Submitting your work without an accompanying declaration, or one with no ticked boxes, will be considered a declaration that you have not used generative AI in preparing your work

If a declaration sheet cannot be uploaded as part of an assignment (i.e. at the start of an essay), students understand that by submitting their assessment that are confirming they have followed the assessment brief and guidelines about GenAI use.

Instructions

As part of this term's coursework, you will apply the tools and concepts learnt in class to analyse a dataset. Below, you'll find a detailed breakdown of the coursework requirements. This will offer you the opportunity to apply knowledge and communicate insights.

Objective:

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

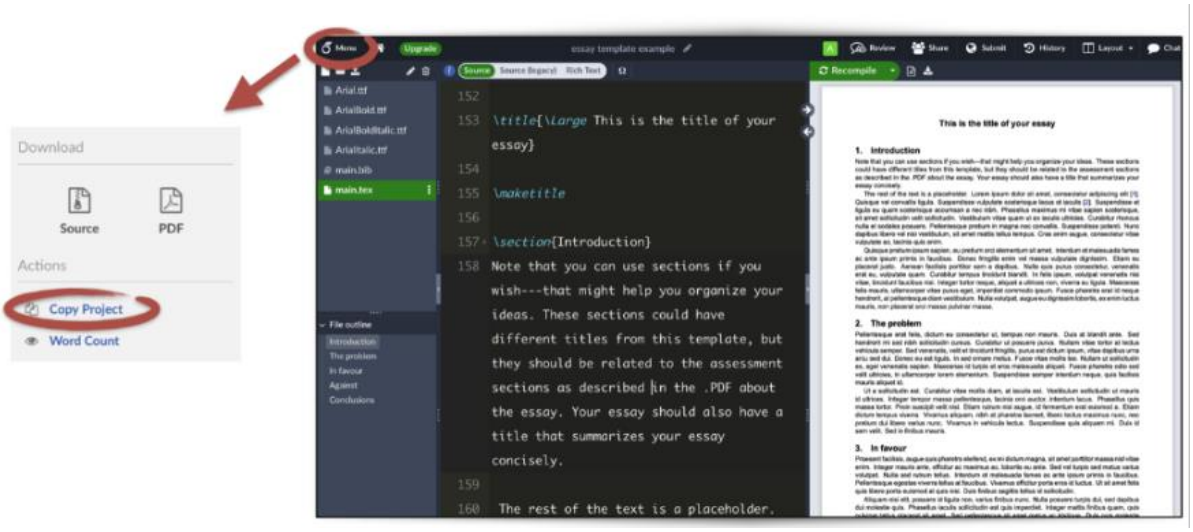
- The research question you're addressing. Why should anyone care about it?
- Introduction to your dataset.
- The type of machine learning technique used (e.g., regression, classification, clustering).
- A description of the machine learning models employed.
- An overview of your data analysis and results.
- 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 decide not to upload the video directly to ELE, please upload a text file with the link to your presentation video. Make sure to grant appropriate access rights to it, meaning that at the very least, anyone within the university should be able to access it.
- If the video is not accessible for markers, it will not receive any marks (i.e., zero marks).

3. Executive Summary:

- Draft a concise, 2-page executive summary, including references, with an additional page dedicated to 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, that the appendix will not be marked.
- Use the LaTeX template **here**, or make something similar in Word.
- To copy the Overleaf template:



4. GenAI Declaration:

Please add a cover page to your executive summary describing if/how GenAI has been used. Note, that even though the declaration will be attached to the executive summary, it should reflect the usage of GenAI in the entire coursework (coding, designing, etc.). Follow the template provided on the second page of this specification.

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 **3rd December 2024, 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!

Some useful material on LaTeX:

- Learn LaTeX in 30 minutes (using Overleaf).

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.