CW2: Machine Learning Case Study - 40%

Start Assignment

- Due 8 Apr by 12:00
- Points 40
- Submitting a website url or a file upload
- Available 13 Feb at 0:00 - 15 Apr at 23:59

Weighting %:	40	Submission deadline (for students):	12 April 2024, 12:00
Authorship:	Individual	Target date for returning marked coursework:	22 April 2024
Tutor setting the work:	Rafael S. de Souza	Number of hours you are expected to work on this assignment:	25-30

This Assignment assesses the following module Learning Outcomes (from Definitive Module Document):

1. Demonstrate a knowledge and understanding of advanced machine learning methods, and their effectiveness for the analysis of numerical data
2. Be able to critically evaluate research papers and methodologies
3. Demonstrate knowledge and understanding of a range of neural networks as models of neural computation
4. Be able to work as part of a team to achieve computer data analysis tasks at a professional or equivalent level.

Assignment Tasks:

The objective of this assignment is to harness machine learning techniques for analyzing a dataset of your choosing, aiming to conduct a study focused on either regression or classification. Utilize either scikit-learn or PyCaret. You are free to choose any model that suits your analysis, including but not limited to Random Forests, Multilayer Perceptrons (MLPs), or Support Vector Machines (SVMs). You may

select your dataset from Kaggle or any other reliable source; however, ensure you provide a link to the original data in your presentation slides.

Your Submission Should:

- Describe at least one specific ML mode used in your analysis.
- Highlight any preprocessing techniques applied to the data before feeding it into the model.

For Your Submission, You Should Create:

1. **Slide Presentation:**

   - Construct a 10-page slide presentation detailing your findings.
   - The slides should cover:
     - Introduction to the chosen dataset.
     - Preprocessing techniques employed.
     - The ML architecture and parameters chosen.
     - Results and critical evaluation of your model.
     - Limitations and potential areas of improvement.

2. **Colab Notebook:**

   - Create a **Google Colab** ▱ **(https://colab.research.google.com/)** notebook that contains all the code used for your analysis.
   - The notebook should be well-commented, ensuring that steps are clear and replicable.
   - Include sections for data loading, preprocessing, model definition, training, prediction, and evaluation.
   - Ensure the notebook can be executed seamlessly from start to finish.

Upon completion, each student should **submit both one slide presentation (in pdf format) and the shareable link to your Colab notebook**.

Marks awarded for:

- You will be marked on the breadth and depth of your analysis, your understanding of the methodology, how well you present the context and conclusions, the quality of references, and the overall quality of your presentation.
- You will receive zero marks if you do not submit a shareable link to your Colab notebook or if the link does not work.
- You will receive a maximum of 50% marks if no slide presentation is submitted.

Additional information:

- Regulations governing assessment offences including Plagiarism and Collusion are available from **https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct.pdf** ▣→ **(https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct.pdf)** (UPR AS14).
- Guidance on avoiding plagiarism can be found here: **https://herts.instructure.com/courses/61421 (https://herts.instructure.com/courses/61421)** (see the **Referencing** section)
- For **postgraduate modules**:
  - a score of 50% or above represents a pass mark.
  - late submission of any item of coursework for each day or part thereof (or for hard copy submission only, working day or part thereof) for up to five days after the published deadline, coursework relating to modules at Level 7 submitted late (including deferred coursework, but with the exception of referred coursework), will have the numeric grade reduced by 10 grade points until or unless the numeric grade reaches or is 50. Where the numeric grade awarded for the assessment is less than 50, no lateness penalty will be applied.

Coursework - Image Classification

Criteria	Ratings				Pts
Problem statement	**10 Pts** **Full marks** You provided a comprehensive overview of the subject. Both your slides and Colab have a professional look.	**5 Pts** **Good** Generic text. Either slides or Colab lack high quality.	**2.5 Pts** **Poor** Poorly written text	**0 Pts** **No marks**	10 pts
Exploratory Data Analysis	**10 Pts** **Full marks** Publication quality plots.	**5 Pts** **Good** Basic visualization	**2.5 Pts** **Poor** Poor visualization	**0 Pts** **No marks**	10 pts
Model implementation	**10 Pts** **Full marks** The model is efficient and clearly tailored toward the task. You explored few approaches to reach the final model.	**5 Pts** **Good** The model performs well on the data, but you did not provide any further insight in terms of model choice and data-preprocessing.	**2.5 Pts** **Poor** Off-the-shelf approach.	**0 Pts** **No marks**	10 pts
Performance Evaluation	**10 Pts** **Full marks** You have provided a critical and logically sound evaluation of your results.	**5 Pts** **Good** You have provided an average discussion.	**2.5 Pts** **Poor** Generic discussion or lack thereof.	**0 Pts** **No marks**	10 pts
					Total points: 40