School of Computing, Engineering and Built Environment

Department of Computing

Artificial Intelligence and Machine Learning

Module Code: MMI226824

Coursework 2-1st diet

Issued: 27th November 2023

This coursework comprises 50% of the overall mark for the module.

This coursework is to be submitted electronically via GCULearn, no later than:

Hand-in date: Wednesday 10th January 2024 (end of day)

An average student should be able to complete this assignment in about 20 hours of work.

Attention is drawn to the university regulations on plagiarism. Whilst discussion of the coursework between individual students is encouraged, the actual work has to be undertaken individually. Collusion may result in a zero mark being recorded for the coursework for all concerned and may result in further action being taken.

COURSEWORK 2

This module has two coursework components, this is Coursework 2. The pass mark for each coursework element is 50%.
This coursework 2 (CW2) consists of two main components:

1. **Practical Component (50%):** A natural language processing (NLP) classification task
2. **Written Component (50%):** Value and Ethical Issues of AI in Fraud Detection

The practical component must be carried out using the provided *Jupyter notebook*. The written component must be completed on the provided template and submitted via Turnitin.

1. PRACTICAL COMPONENT: improving news classification using NLP

Please use template (***CW2_NLP_template.ipynb***) for this task. It contains a summary of lab 8 (news classification with NLP) already setting up a baseline classifier.

You can reuse the *Conda* environment provided for lab 8 (*lab8_NLP_news.yml*) to easily get the necessary libraries.

Locate the different TODO entries along the notebook. You can implement one of them at a time and check how much of a improvement (if any) they did:

- TODO 1: Remove the occurrences of residual html code (“http", "href", "https", "www", “quot”, “`lt`”, …) from the datasets which indicate weblinks. Also, remove the news agencies names (“Reuters”, “AP”, …).
- TODO 2: Apply *stemming* or *lemmatization* to simplify grammatical word variations.
- TODO 3: Apply the model to some *recent* pieces of news to test how they are classified. One fresh example for each category is enough. Also, find one example of a piece of news incorrectly predicted, and briefly discuss why it might have been misclassified.
- TODO 4: Try using a different model architecture. For example, increase the number of layers, or change the number of neurons in each layer. Or a 1D CNN.
- TODO 5: Discussion. Discuss how your steps 1 to 4 are improving (or not) the performance, and why it might be this way.

The implementation in the Jupyter notebook should contain Markdown (text) cells or in-code comment explaining the steps that are being carried out.

The maximum word limit for all Markdown cells (excluding instructions already written in the template, inline code comments and the “Sources” Markdown cell) is **1000 Words**.

2. WRITTEN COMPONENT: Value and Ethical Issues of AI in Fraud Detection

The second part of this coursework consists of a short report of 1000 words. Please use template (***CW2_written_template.docx***) for this task.

The application of Artificial Intelligence (AI) in fraud detection is being adopted across various sectors. Financial institutions, insurance companies and e-commerce platforms now utilise AI to scrutinize financial transactions for potential fraud, encompassing activities such as credit card usage, insurance claims, and online purchases.

Additionally, telecommunications companies are employing AI to detect phone fraud through the analysis of call patterns. This extensive use of AI underscores its significance in identifying fraudulent activities across a wide range of financial operations.

Your report should be organized into two sections:

Section 1: How Does AI in Fraud Detection Add Value to Business and Society? *(500 Words)*

- Discuss how AI-powered fraud detection systems analyse **large volumes** of activities in **real time** to identify patterns and anomalies that suggest fraudulent behaviour.
- Explain the role of AI in enhancing the **accuracy and efficiency** of fraud detection compared to traditional methods, thus saving costs and resources for businesses.
- Highlight the **benefits to society** and the overall reduction in fraudulent losses.

Section 2: Ethical Considerations in AI-Powered Fraud Detection *(500 Words)*

Plagiarism

Glasgow Caledonian University has accepted and adopted the Russell Group Principles on the use of generative AI tools in Education.

It is not appropriate for students to submit work which passes knowledge, argument, or content of another off as their own work. This is akin to the use of a ghost writer or an essay mill, as articulated within the Student Code of Conduct. Beyond this, the following guidance applies for students:

Appropriate	***Not appropriate***
Using AI as a resource, like a search engine, to explore a topic or question.	**Trusting AI to provide factual content. Always cross check with other search tools to establish validity. Including AI generated content without appropriate acknowledgement and contextualisation.**
Using AI as a critical friend to discuss ideas.	**Not acknowledging the use of AI as a co-creator / collaborator.**
Using AI to augment knowledge and information from class.	**Using AI as the main source of information.**

You should also pay attention to the university's codes and practices[1] as well as their plagiarism regulations[2]. Any kind of content (images, text, code) that was copied from any source and used in your coursework **without acknowledging of the source** is bad academic practice and could fall under plagiarism.

The discussion of coursework between students is encouraged, but **the work must be undertaken individually**. Collusion (copying work between students) may result in a zero mark being recorded for everyone involved and further action being taken.

Sources and citation

To avoid plagiarism or poor academic practice, you need to ensure that you specify **where** you obtained any material you use, and how you have interpreted or modified it. This MUST include

[1] https://www.gcu.ac.uk/academicquality/regulationsandpolicies/universityassessmentregulationsandpolicies/
[2] https://www.gcu.ac.uk/library/smile/plagiarismandreferencing/

specific web addresses (not just *google.co.uk* or similar) that can be followed to reach the original content.

Sources must be **listed** at the end of the document (under a section called ***Sources*** or ***References***), and **cited** (using your preferred citation method) at the point of the document where you used them. Just adding sources into the final list, without citing them from the main document, is bad academic practice. **THIS WILL BE EXPLICITLY CHECKED BY THE MARKING TEAM**.

A template for this is included on at the very bottom of the supplied templates (notebook and Word document). The notebook version is shown in Figure 1.

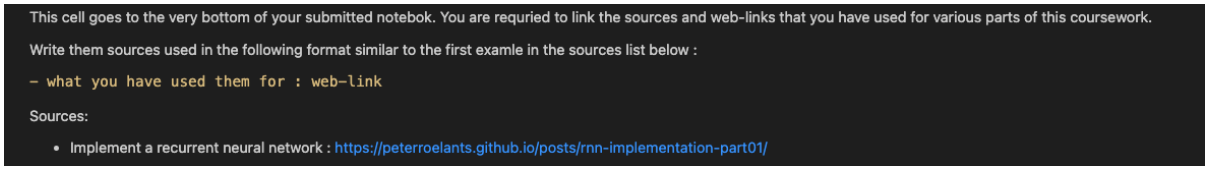

Figure 1: "Sources" markdown cell to be included at the very end of the submission

Coursework Submission

You are expected to submit your Jupyter Notebook with your code and markdown cells, using the following **naming** scheme: ***CW2_LastName_FirstName_StudentID.ipynb***

For instance, if your name is Nicola Sturgeon with the StudentID S1234567, name the file: *CW2_sturgeon_nicola_S1234567.ipynb*

The written report, in the provided Word template, will be submitted via Turnitin.

Both coursework 2 files are submitted using GCU Learn.

- Ensure that the practical submission (notebook) contains the **output** of the cells. Fewer marks will be awarded if the notebook does not contain the outputs and the code does not run on the markers' computer.
- Double-check your files before submission to ensure you are submitting the correct ones.
- Omitting either of the submission files will lead to less marks being awarded.

Marking Scheme

The marking scheme which will be used to assess the coursework is appended at the end of this document.

Artificial Intelligence & Machine Learning
Coursework 2 – 1st diet 23/24

Assessment Rubric – PRACTICAL PART

	80-100	70-79	65-69	60-64	50-59	40-49	1-39	0	Mark
	Exceptional	Excellent	Very Good	Good	Satisfactory	Marginal Fail	Clear Fail	N/S	
Criteria	Demonstrates **exceptional or consistently excellent** ability, skills and behaviours across specified items	Demonstrates **mostly excellent** ability, skills and behaviours across specified items.	Demonstrates **overall very good** ability, skills and behaviours across specified items.	Demonstrates **overall good** ability, skills and behaviours across specified items	Demonstrates **overall satisfactory** ability, skills and behaviours across specified items	Demonstrates **overall poor** ability, skills and behaviours across specified items with **some satisfactory** elements	Demonstrates **overall poor** ability, skills and behaviours across specified items with **no satisfactory** elements		
TODO 1 (5%) *Remove html residues and news agencies names.*					Several agency names, or several html code residual words, have been removed.				
TODO 2 (10%) *Apply stemming or lemmatization.*					Basic stemming or lemmatization applied.				
TODO 3 (5%) Apply the model to some recent pieces of news. One example of a piece of news incorrectly predicted, discuss why.					At least 2 suitable examples provided, with all pre-processing steps applied.				
TODO 4 (20%) *Improved model*					Trivial modifications over the baseline RNN or just changing some hyperparameters.				
TODO 5 (10%) *Discussion*					Summary of results with little critical discussion.				
Total (out of 50%)									

Artificial Intelligence & Machine Learning
Coursework 2 – 1st diet 23/24

Assessment Rubric – Written PART

	80-100	60-79	50-59	1-49	0	Mark
	Excellent	Good	Satisfactory	Fail	N/S	
Criteria	Outstanding understanding and presentation. Comprehensive and insightful content, clear arguments, professional writing.	Clear understanding with minor gaps. Well-researched and organized content, clear arguments, mostly error-free writing.	General understanding and presentation with gaps. Adequately researched, some developed arguments, acceptable writing with some errors.	Poor understanding and execution. Limited or insufficient content, poorly developed arguments, unclear writing with numerous errors.		
Discuss How AI-Powered Fraud Detection Systems Work. **(5%)**			Provides a general introduction with some context and rationale for AI in fraud detection.			
How Does AI in Fraud Detection Add Value to Business? **(10%)**			Offers a general explanation with examples of how AI contributes value to businesses.			
How Does AI in Fraud Detection Add Value to Society? **(10%)**			Provides a general discussion with examples of societal benefits.			
Ethical Considerations Discussion **(20%)**			Offers a general discussion of at least two ethical issues with examples.			
Presentation of information **(5%)**			Information presented for most of the materials; formatting, structure and references mostly fine including headings.			
Total (out of 50%)						