COMP1804	**Applied Machine Learning**	**Faculty Header ID:**	**Contribution 100% of course**
Course Leader Dr. Stef Garasto	**COMP1804 Coursework**		**Deadline Date: 12/04/2023 (23:30 UK time)**

This coursework should take an average student who is up-to-date with the lecturesand the labs approximately 50 hours

Feedback and grades are normally made available within 21 days of the coursework deadline

Learning Outcomes:

1. Rationalise appropriate scenarios for Machine Learning applications and evaluate the choice of machine learning methods for given application requirements.
2. Demonstrate competency in using appropriate libraries/toolkits to solve given real-world Machine Learning problems and develop and evaluate suitable application.
3. Understand and apply the relevant input data preparation and processing required for the Machine Learning models used, and quantitatively evaluate and qualitatively interpret the learning outcome.
4. Recognise and critically address the ethical, legal, social and professional issues that can arise when applying Machine Learning technologies.

Plagiarism is presenting somebody else's work as your own. It includes: copying information directly from the Web or books without referencing the material; submitting joint coursework as an individual effort; copying another student's coursework; stealing coursework from another student and submitting it as your own work. Suspected plagiarism will be investigated and if found to have occurred will be dealt with according to the procedures set down by the University. Please see your student handbook for further details of what is / isn't plagiarism.

All material copied or amended from any source (e.g. internet, books) must be referenced correctly according to the reference style you are using.

Your work will be submitted for plagiarism checking. Any attempt to bypass our plagiarism detection systems will be treated as a severe Assessment Offence.

Coursework Submission Requirements

An electronic copy of your work for this coursework must be fully uploaded on the **Deadline Date of 12th April 2023** using the link on the coursework Moodle page for COMP1804. For this coursework you must submit 2 separate files:

- **A single pdf file named 'report_comp1804_optionX.pdf'** which will be the written report; the written report must have a maximum limit of 3000 words excluding references (if you have any). It is also recommended for the report to have at least 2000 words. The X in the file name should be either 1 or 2 depending on which dataset option you choose (see below). You should include your student ID in the report, but not your name.
- **A single .zip file** containing an ipython notebook file and a pdf (or html) file, showing the machine learning implementation on your dataset. The files should be named 'code_comp1804.ipynb' and 'code_comp1804.pdf'.
    - The pdf file should be generated from the python notebook via "File/Download As/PDF" after executing **all** the cells. It should show all the relevant implementation steps to reproduce the content of your report.
    - If you are submitting a .py file instead of .ipynb, you may skip the pdf/html file.

Please note:

- Any text in the document must not be an image (i.e. must not be scanned) and would normally be generated from other documents (from the latex template in this case).
- There are limits on the file size (see the relevant course Moodle page).
- Make sure that any files you upload are virus-free and not protected by password or corrupted otherwise they will be treated as null submissions.
- You must NOT submit a paper copy of this coursework.
- Lecturers and tutors can *only* accept coursework submitted via the appropriate Moodle link. No other way of submitting the coursework will be considered.

The University website has details of the current Coursework Regulations, including details of penalties for late submission, procedures for Extenuating Circumstances, and penalties for Assessment Offences. See http://www2.gre.ac.uk/current-students/regs

Detailed Coursework Specification

The task is to implement ML solutions and produce a written report **individually**. Two datasets are provided (see below): you should choose one of them and carry out all the associated tasks. Please see below for further details.

Whichever option you choose, you need to use the dataset downloaded from the link listed in this coursework specifications. This is because the datasets provided have been modified to be better tailored to this specific module. While they are in principle available for download in their original form, please **do *not* use data downloaded from anywhere else**. Any coursework submitted with a different dataset will be given 0 marks.

Dataset

Option 1: Predicting the severity of road accidents in the UK.

Emergency services in the UK (non-commercial entities) are looking into developing a system to provide a more effective response in case of road accidents. They want to know if it would be possible to predict the severity of an accident using some variables that can potentially be gathered on the place of the accident. They provide you with a dataset of all the accidents that occurred in 2019 in the UK which contain their variable of interest.

They want you to use machine learning techniques to predict the accident severity in details (that is, whether an accident is "fatal", "serious" or "slight" – 3 classes in total), from all the other features in the dataset. For the avoidance of doubt, the information about accident severity is contained in the column titled "accident_severity".

They want to compare traditional machine learning algorithms with neural networks to see if the latter offer significantly higher performance.

They want you to write the results of your analysis and implementation in a report. More details about what to include in the report are provided below.

The dataset can be downloaded from this Moodle link. The dataset has been adapted to the requirements of this module; the original data are taken from UK's statistics on Road Safety Data[1] and have been released under the Open Government Licence v3.0.

Option 2: Predicting the topic of customers' banking questions.

You are consulting for a bank (a commercial entity) on how to make their online customer service more effective. They want to trial an automated first level of filtering for questions asked by customers in an online chat. Specifically, they want to know whether it is possible to identify the topic a question relates to, for some specific topics of interest. You are provided with a dataset with sample questions and their associated topic.

They want you to use machine learning to predict the question topic in details (that is, whether a question is about "card queries or issues", "needs troubleshooting", "top up queries or issues" or "other" – 4 classes in total). For the avoidance of doubt, the question topic is contained in the column titled "label".

They want to compare traditional machine learning algorithms with neural networks to see if the latter offer significantly higher performance.

They want you to write the results of your analysis and implementation in a report. More details about what to include in the report are provided below.

The dataset can be downloaded from this Moodle link. The dataset has been adapted to the requirements of this module; the original data were released under the terms of the CC-BY-4.0 licence by Casanueva et al.[2]

---

[1] Road Safety Data by the Department for Transport: Vehicles and Accidents data from 2019. Contains public sector information licensed under the Open Government Licence v3.0.

[2] Casanueva et al. (2020) *Efficient Intent Detection with Dual Sentence Encoders*. Proceedings of the 2nd Workshop on NLP for ConvAI - ACL 2020. Data available at https://github.com/PolyAI-LDN/task-specific-datasets.

Report

You should submit a report detailing the work you did to solve the chosen machine learning task, including any experiment performed in search of the “best” model.

The report should contain the following sections:

0. Executive summary

Briefly summarize what the report contains. That is: the task you are solving and why it is important; the outline of the ML methods you implemented and any experiments performed; the summary of your results and your conclusions. The executive summary should be between 100 and 200 words.

1. Exploratory data analysis

Describe the exploratory data analysis performed and comment on what its implications are for the machine learning task. As part of the exploratory data analysis, you should use dimensionality reduction techniques to show the dataset (including the target labels) in a 2-dimensional plot.

2. Data preprocessing

Describe the steps performed for data cleaning, splitting (training/validation/test) and pre-processing (where appropriate: normalization/standardization, imputation of missing values, feature encoding, over/under-sampling, text processing). Provide justifications, based on theory and/or experiments, for your design choices.

3. Classification using traditional machine learning

Describe your solution to the classification task (accident severity for option 1 and question topic for option 2) using traditional machine learning techniques. You should describe the final model hyper-parameters in details, ideally in a table, and give a brief explanation of how the algorithm works.

Describe the experiments you did to optimize your model (hyper-parameters optimization and comparison with other models) – these experiment should be rigorous and follow best practice. Provide justifications, based on theory and/or experiments, for your design choices.

Evaluate the model performance using a) a confusion matrix; b) two performance metrics (explain what each metric compute, why it is an appropriate metric to use and what are the implications of the results for the task you are solving); c) a comparison with one “trivial” baseline (for example, random guess or majority class).

Results should be presented in well-formatted figures and tables.

4. Classification using neural networks

Describe your solution to the classification task (accident severity for option 1 and question topic for option 2) using neural networks. You should describe the final model hyper-parameters in details, ideally in a table, and give a brief explanation of how the algorithm works.

Describe the experiments you did to optimize your model (hyper-parameters optimization and comparison with other neural networks) – these experiment should be rigorous and

follow best practice. Provide justifications, based on theory and/or experiments, for your design choices.

Evaluate the model performance using a) a confusion matrix; b) two performance metrics (explain what are the implications of the results for the task you are solving); c) a comparison with one “trivial” baseline (for example, random guess or majority class). Unless you are using performance metrics that are different from the previous section, you do not need to explain again what each metric computes. However, you should compare the results with those from the previous section.

Results should be presented in well-formatted figures and tables.

5. Ethical discussion

Identify and discuss some of the social and ethical implications of your chosen task, from data collection and processing to the ML prediction. It is highly recommended that you structure the discussion using either Data Hazard Labels or the Ethical OS Toolkit. The discussion should take into account communities and people that may be affected by the ML system.

6. Recommendations

You should provide three bullet points detailing the following:

- Which of your machine learning model is the best candidate for the task and why.
- Whether the final model is good enough to be used in practice and why (or why not).
- Your top suggestion for future improvements and why.

7. Retrospective

The last section in the report is a reflection on the work you have done for this coursework. You should write a maximum of 50 words answering the following question: if you were to start the coursework all over again, what aspect of it would you want to investigate more in depth and why?

8. References

If needed, cite references and sources used at the end. These can be academic papers, blogs, code repositories, and more. Remember to give credit if someone else’s work has helped you complete your coursework.

Reproducibility and report presentation

The report should be written in a professional style, in English, with clear language and a logical flow (report presentation). Double check your spelling, punctuation and grammar.

In principle, the reader should be able to recreate your final machine learning solution based on the report alone, without needing to look at the code (reproducibility). To achieve this, all the necessary implementation details must be included in the report.

The report must include only and all the sections detailed above. Marks may be removed if different sections are used. You are allowed to add sub-sections. To be safe, you are highly recommended to use one of the template provided. You can download the coursework templates (Latex and Word) from Moodle.

Make sure that all figures have legible text and an appropriate label – use the same label to reference the figure in the text.

Do not include any screenshots of your code.

Code

The code must be developed in python and submitted as part of the coursework. It should show the implementation leading to all the results described in the coursework and should be well documented.

In principle, tutors should be able to run the code without errors, if needed. Tutors may review your code to check that the implementation is consistent with what is written in the report. Inconsistencies between the report and the code may be penalized. However, the code itself is not marked. Only what is described in your report counts towards your final mark.

Deliverables

For the submission to be admissible, all coursework submission requirements as specified in the Coursework Submission Requirements section above should be uploaded by the Deadline Date using the link on the coursework Moodle page for COMP 1804.

Assessment Criteria

Marks breakdown

The mark breakdown is given below. Generally, both the solution implemented and the quality of explanation in the report will be taken into account.

- Executive summary (5 marks)
- Data exploration (8 marks)
- Data preparation (15 marks)
- ML implementation – traditional machine learning techniques (22 marks)
- ML implementation – neural networks (22 marks)
- Ethical discussion (10 marks)
- Retrospective (3 marks)
- Recommendations (5 marks)
- Reproducibility and report presentation (10 marks)

Rubric

Generally, marks will be given for:

- Features implemented. That is, the extent to which a comprehensive, rigorous, optimized, clear and properly justified machine learning solution was implemented. The complexity and appropriateness of the implementation are also taken into account. Note that rigorous, clear and justified can be achieved in many ways: it could be based on theoretical or empirical reasoning, it can refer to knowledge acquired before, during and after the (often iterative) implementation. Also, optimized means showing how you tried to improve the model and why.
- Note that there is not a direct relationship between the achieved accuracy and the final mark. You are expected to do your best to obtain a good prediction accuracy, but it is recognized that the available tasks are challenging. Your mark will depend much more strongly on the soundness of your approach, evaluation and analysis than on the final accuracy.
- Critical understanding of relevant concepts, appropriate explanation, recommendations and discussion that are backed up by the results.
- Quality of the report: Are all the required sections included and completed properly? Is the executive summary clear and concise? Is the report clear, well formatted and easy to read? Does it have a logical structure? Does it have a discussion on design decisions? Is the evaluation realistic, does it show that you have really thought about your system and how you went about developing it?
- An ethical discussion that is not generic but tailored to the specific task, that identifies potential risks (with reasoning given) and that includes a wide range of perspectives and stakeholders.

Please, see the table below for further details on how marks will be allocated:

	0-29% Fail	30-49% Fail	50-59% Good	60-69% Very Good	70-79% Excellent	80-100% Exceptional
Executive summary (5 marks).	No summary or highly unsatisfactory summary that does not adequately represent the work done, is unclear and irrelevant.	A poor summary that does not adequately represent the work done, is unclear and irrelevant.	A good summary that mostly represent the work done in an adequate manner, is clear enough and relevant.	A very good summary that represents well the work done, is very clear and relevant.	An excellent summary that represents the work done optimally and clearly, with only minor imprecisions.	An exceptional summary that represents the work done optimally and clearly.
Exploratory data analysis (8 marks)	No or extremely little exploratory data analysis has been performed.	Some exploratory data analysis has been performed, but is largely incomplete and with mistakes. Implications are scarcely mentioned.	A good exploratory data analysis, but may be incomplete and with mistakes. Some implications are mentioned, but are superficial or have mistakes.	A very good exploratory data analysis, which is mostly complete, but may have mistakes. Implications are assessed clearly, but may be superficial or have mistakes.	An excellent and comprehensive exploratory data analysis, but may have minor mistakes. Implications are assessed mostly in a critical manner, but may have minor mistakes.	An exceptional and comprehensive exploratory data analysis, with very little left to improve. There is a critical assessment of the implications.
Data preparation (15 marks)	Highly unsatisfactory data preparation steps. Data cleaning and pre-processing is missing or incorrect. No or extremely little reasoning/evidence is given. Data splitting is missing or incorrect.	Poor data preparation steps. There is little or mostly incorrect data cleaning and pre-processing. No or extremely little reasoning/evidence is given. Data splitting may have been performed incorrectly.	Good data preparation steps. Attempts at cleaning and pre-processing the data are satisfactory but are incomplete and/or with mistakes. Some reasoning/evidence behind the design choices given, but is mostly superficial and may have mistakes. Data splitting is performed correctly.	Very good data preparation steps. Attempts at cleaning and pre-processing the data are mostly complete but may have some mistakes. There is clear reasoning/evidence behind the design choices, but may have some mistakes. Data splitting is performed correctly.	Excellent data preparation steps. Comprehensive data cleaning and pre-processing, may have some minor mistakes. Reasoning/evidence behind the design choices is presented largely in a critical manner, but may have some imprecisions. Data splitting is performed correctly.	Exceptional data preparation steps, with very little that can be improved. Data cleaning and pre-processing is correct and comprehensive. There is critical reasoning/evidence behind the design choices. Data splitting is performed correctly.

ML implementation – traditional machine learning techniques (22 marks)	A highly unsatisfactory implementation of a traditional machine learning method for classification. It is incorrect or absent, has no justification for design choices, no experiments at all, an inappropriate or absent evaluation procedure. Explanations are incorrect.	A poor implementation of a traditional machine learning method for classification, which has major mistakes, little experimentation and justification for design choices. It has an inappropriate or absent evaluation procedure. Explanations are mostly incorrect.	A good implementation of a traditional machine learning method for classification, with some experimentation and justification for the design choices. There is some evaluation and some explanations are given. May have mistakes, be superficial and/or limited.	A very good implementation of a traditional machine learning method for classification, with clear experimentation and justification for the design choices. Comprehensive evaluation is performed and explanations are given. May have some mistakes.	An excellent implementation of a traditional machine learning method for classification. Experimentation and justification for the design choices are rigorous and mostly critical. Comprehensive evaluation is performed and explanations are given. May have some minor mistakes.	An exceptional implementation of a traditional machine learning method for classification. Experimentation and justification for the design choices are rigorous and critical. Comprehensive evaluation is performed and in-depth explanations are given.
ML implementation – neural networks (22 marks)	A highly unsatisfactory implementation of a neural network for classification. It is incorrect or absent, has no justification for design choices, no experiments at all, an inappropriate or absent evaluation procedure. No or little comparison is performed.	A poor implementation of a neural network for classification, which has major mistakes, little experimentation and justification for design choices. It has an inappropriate or absent evaluation procedure. Comparison is mostly incorrect or absent.	A good implementation of a neural network for classification, with some experimentation and justification for the design choices. There is some evaluation and some comparison is performed. May have mistakes, be superficial and/or limited.	A very good implementation of a neural network for classification, with clear experimentation and justification for the design choices. Comprehensive evaluation and comparison are performed. May have some mistakes.	An excellent implementation of a neural network for classification. Experimentation and justification for the design choices are rigorous and mostly critical. Comprehensive and mostly critical evaluation and comparison are performed. May have some minor mistakes.	An exceptional implementation of a neural network for classification. Experimentation and justification for the design choices are rigorous and critical. Comprehensive and critical evaluation and comparison are performed.
Ethical discussion (10 marks)	Missing or highly unsatisfactory ethical discussion, which covers almost no relevant point.	Poor ethical discussion, which barely cover some relevant points, but is largely generic.	Good ethical discussion, which covers some relevant points, and is at least partially specific to	Very good ethical discussion, which covers many relevant points, is justified and mostly specific to the task implemented.	Very good ethical discussion, which covers most relevant points, is well justified and fully specific to the task implemented.	Exceptional ethical discussion, with very little left to improve. It is highly specific, critically justified and covers

			the task implemented.			all the relevant points.
Recommendations (5 marks)	Highly unsatisfactory recommendations are given, which fail to address the work done or give any reasoning.	Poor recommendations are given, that are only loosely related to the work done and with very little reasoning.	Good recommendations are given. They are coherent with the work done and some reasoning is attempted. There may be mistakes or superficial reasoning.	Very good recommendations are given. They are coherent with the work done and backed up by good reasoning. There may be imprecision or partially superficial reasoning.	Excellent recommendation are given. They are coherent with the work done, assessed correctly and mostly in a critical manner.	Exceptional recommendation are given, with very little left to improve. They are coherent with the work done and critically and correctly assessed.
Reproducibility and report presentation (10 marks)	Unsatisfactory clarity and logicality, almost always hard to follow. Layout and section structure are not acceptable. No details for replication are provided.	Poor clarity and logic, with many mistakes and often hard to follow. Layout and section structure are not acceptable. Some, but not all essential details for replication are provided.	Overall clear and logical, but with many mistakes. Layout is acceptable. A suitable section structure is followed. The essential details for replication are provided, but some may be missing.	Overall clear and logical, but there may be minor mistakes. Layout is good. A suitable section structure is followed. The essential details for replication are provided, but some may be missing.	Writing is mostly clear, concise and logical. Layout is correct, with minor mistakes. The appropriate section structure is followed. Most of the necessary details for replication are given, with only minor omissions.	Writing is clear, concise and logical, with very little that can be improved. Layout is correct. The appropriate section structure is followed. All necessary details for replication are given.
Retrospective (3 marks)	No retrospective.		A superficial or very generic retrospective.	A good retrospective, at least partially specific to the work done.	A good retrospective, that is specific to the work done.	