IN-COURSE ASSESSMENT (ICA) SPECIFICATION

Module Title: **Machine Learning**	Module Leader: **Dr. Alessandro Di Stefano**
	Module Code: **CIS4035-N**
Assignment Title: **Machine Learning Application and Report**	Deadline Date: **20/05/2022**
	Deadline Time: **4:00pm**
	Submission Method: Online (Blackboard) ☑ Middlesbrough Tower ☐

Online Submission Notes:

- Please follow carefully the instructions given on the Assignment Specification
- When Extenuating Circumstances (e.g. extension) has been granted, a fully completed and signed Extenuating Circumstances form must be submitted to the School Reception or emailed to scdt-assessments@tees.ac.uk.

Central Assignments Office (Middlesbrough Tower M2.08) Notes:

- All work (including DVDs etc) needs to be secured in a plastic envelope or a folder and clearly marked with the student's name, number and module title.
- An Assignment Front Sheet should be fully completed before the work is submitted.
- When Extenuating Circumstances (e.g., extension) has been granted, a fully completed and signed Extenuating Circumstances form must be submitted to the School Reception or emailed to scdt-assessments@tees.ac.uk.

FULL DETAILS OF THE ASSIGNMENT ARE ATTACHED INCLUDING MARKING & GRADING CRITERIA

Machine Learning

CIS4035-N

Teesside University
Inspiring success

In-course Assessment

Overview of Requirements

Assessment for **Machine Learning** (CIS4035-N) requires you to develop machine learning applications and make predictions about unseen data. The summative assessment for this module is via **in-course assessment (100%)** which will evaluate all learning outcomes (see below).

The assessment will emulate the "shared task" framework appearing in several machine learning venues, in which participants are supplied with a task description and annotated data and must develop a machine learning solution that makes predictions about an unannotated data set.

The assessment for this module is individual and it contains **two elements**:

1. The **first element** (50%) consists of developing a Machine learning application and its predictions on the labelled or unlabelled dataset (50%) to assess Learning Outcomes 2, 3, 4, 5 and 6. The student will produce a voice over brief walk-through video (2-3 minutes), showing and demonstrating what has been done in the distinct parts of the project, from the design choices to the implementation stages of the solution. It is expected that the student will introduce the entire project and its objectives as well as briefly discuss what has been achieved. Throughout the walk-through video, reference should be made to the student's report (second element) to ensure that the application presented is consistent with the description in the report. Explanations should be provided when things did not go according to plan. It is also recommended that the student highlights the issues and limitations encountered during implementation. **[50 points].**

2. The **second element** (50%) consists of a conference-style paper of approximately 2,000 words that reports on the development of their system (50%) and it will assess Learning Outcomes 1, 2 and 7 **[50 points].**

The purpose of this assessment is to is to demonstrate achievement of the module learning outcomes (see Learning Outcome section).

Further details are given below and there will be a supporting briefing session on the ICA.

Submission of materials must be made via Backboard to the link provided. The submission date is specified in the submission schedule.

Learning Outcomes

Personal and Transferable Skills

1. Select, apply, and defend the selection and application of machine learning methodologies and experiments in academic reports.

2. Demonstrate a systematic understanding of machine learning algorithms and their selection for solving a specific problem.

Research, Knowledge and Cognitive skills

3. Investigate state-of-the-art machine learning algorithms.

4. Design appropriate representations of machine learning problems for input into machine learning packages and critically evaluate their effectiveness.

5. Design and evaluate neural network configurations and learning mechanisms for sample problems.

6. Analyse empirical results of the selected machine learning algorithms and justify the performance.

Professional skills

7. Autonomously implement and evaluate appropriate machine learning technique for particular learning tasks, taking into consideration professional, ethical and legal issues.

Task Description

Problems in machine learning vary from domain to another. In this coursework, you will select a dataset related to a real-world problem that best suits your area of interest. There are abundant of websites that provide publicly available datasets. A categorised list of datasets from GitHub can be found at https://github.com/caesar0301/awesome-public-datasets. The UCI Machine Learning Repository at https://archive.ics.uci.edu/ml/index.php is another long-standing source of benchmark datasets for data mining and machine learning research. Kaggle (https://www.kaggle.com/datasets) has interesting real-world problems and datasets.

You can select a dataset from the above sources, or another one that is available online. The dataset should be publicly available. The chosen dataset should have a minimum of 1,000 instances (rows) and a minimum of 5 attributes (columns). You have to complete the following stages in this assignment:

1. Define the problem for the selected data set and identify the machine learning algorithms that are applicable to this problem.
2. Data exploration and preparation: The nature of the dataset may dictate some data exploration and preparation that can help inform the solutions. For example, higher dimensional datasets (those with too many attributes/columns) may require applying a data reduction method like Principal Component Analysis (PCA).
3. Propose solutions: In this step, you will propose three machine learning algorithms that are applicable to the selected data set/problem.
4. Design, implementation, modelling and evaluation: design, model and implement the proposed solutions and critically evaluate the solutions. Use appropriate visualisation for the results.
5. Reflect on professional, ethical, and legal issues in relation to the problem and the data set.

Moreover, each student is required to create a voice over brief **walk-through demo video** (2-3 minutes) of the application, showing and demonstrating what has been done in the distinct parts of the project, from the design choices to the implementation stages of the solution. Thus, the student will discuss the design of the steps (pipeline) and all the data pre-processing and exploratory data analysis conducted, the selection of Machine Learning algorithms for the given task, and implementation details showing experiments and the code (software architecture and implementation) and the obtained results though a quantitative performance of application.

The purpose of the video is to allow the assessors to see the product and identify its standard. This gives the student the opportunity to explain their project, demonstrate their work, explain how they have solved (or otherwise) challenges but also defend the decisions that they took during the project. It is therefore expected that the student will introduce the entire project and its objectives as well as briefly discuss what has been achieved. Throughout the walk-through, reference should be made to the student's report (second element) to ensure that the application presented is consistent with the description in the report. Explanations should be provided when things did not go according to plan. It is also recommended that the student highlights the issues and limitations encountered during implementation.

Element 1 Deliverable – Contribute 50% of the Module Mark

Element 1 will assess learning outcomes LO 2, 3, 4, 5 and 6.

Deadline: 20/05/2022

What to Hand-In

Submission method is online on Blackboard. You are required to submit a file in a pdf format via Blackboard that includes all source code and screenshots from your experiments appropriately labelled and commented.

You are required to submit copy of the source code and screenshots from your experiments appropriately labelled and commented via Blackboard. Moreover, you are required to hand-in a produce a voice over brief walk-through video (2-3 minutes) as discussed above.

The code and experiments will be assessed on:

- Appropriateness of machine learning algorithm selected for the given task.
- Quality of software architecture and implementation.
- Quantitative performance of application.

Element 2 – Contribute 50% of the Module Mark

Element 2 will assess learning outcomes LO 1, 2 and 7.

Deadline: 20/05/2022

What to Hand-In

- A case study report maximum of 2000 words that documents the process of the entire case study, including data set, problem, data preparation and exploration, selected algorithms, critical evaluation and justification of the algorithms and findings.
- Submission method is online on Blackboard. You are required to submit a file in a pdf format via Turnitin on Blackboard.

The hand-in is electronically via Blackboard, all deliverables shall be labelled with project name, your student's name, and university number (or student ID).

The report will be assessed on:

- understanding of machine learning task
- review of relevant literature
- development methodology
- justification of design decisions
- consideration of professional, ethical, and legal issues

The report could broadly include the following sections:

- Abstract
- Introduction (introduce the problem and its significance, write short literature review of related work)
- Data exploration and features selection
- Experiments
- Results
- Discussion, Conclusions and Future Work
- References

These are generic section titles, which you may adapt appropriately to the application/problem that is investigated. You may include sections describing modifications of algorithms or developments that are novel and specific to your work.

Outline Marking Scheme

Your submission will be assessed according to the following criteria:

1. Machine Learning application **[50 points]**.
2. Report (conference-style paper) **[50 points].**

Below is a provisional indication of the criteria applied to determine points for each element.

Please note:
Exceptionally, whilst points are allocated to specific parts, outstanding work in one area may be used to trade-off points against poorer work in another area.

Machine Learning application [50 points]	*Source Code Documentation and Demo*
Excellent 70% and above	Clear evidence of running the experiments with code that is excellently organized and commented, also showed through an excellently produced walk-through demo video. Machine learning algorithms selected are appropriate for the given task. Excellent quality of software architecture and implementation. Excellent quantitative performance of application. Deep understanding shown. Excellent walk-through demo video.
Very Good 60% - 69%	Very good evidence of running the experiments with code that is well-organized and commented, also showed through a well-produced walk-through demo video. Machine learning algorithms selected are appropriate for the given task. Very good quality of software architecture and implementation. Very good quantitative performance of application.

	Very good understanding.
Satisfactory 50% - 59%	Satisfactory evidence of running the experiments with code that is organized and commented, and satisfactory walk-through demo video. Machine learning algorithms selected are appropriate for the given task. Satisfactory quality of software architecture and implementation. Satisfactory quantitative performance of application. Satisfactory understanding.
Fail Less than 50%	Little evidence of running the experiments with code that is not well-organized and commented, as well as the walk-through demo video. Machine learning algorithms selected are not appropriate for the given task. Poor quality of software architecture and implementation. Poor quantitative performance of application. Poor understanding.
NS NON-SUBMISSION	N/A

Report **[50 points]**	*Academic Quality of the Paper*
Excellent 70% and above	Excellent technical quality (rigor of the experiments, data preparation, justification and correct application of the selected algorithms and suitability of the selection). Produced and demonstrated a comprehensive, high quality solution to the problem. Sufficient information for the reader is provided to reproduce the results. Outstanding evidence of systematic review using multiple high quality academic sources. Logical, clear development of narrative. High quality references and citations. Outstanding evaluation and discussion of the significance of the results (Why the results are important? How does the paper advance the state of the art? How would the results be useful to other researchers or practitioners? Is this a “real” problem or a small “toy” problem?). Legal, social, ethical, security and professional issues fully considered. A paper, which could be, with minor modifications, suitable for a publication – or form the basis for a postgraduate project. There is some element of a novel approach to the problem or novel use of techniques.

Very Good 60% - 69%	Very good technical quality. Produced and demonstrated very good quality solution to the problem. Sufficient information for the reader is provided to reproduce the results. Very good evidence of systematic review using multiple high quality academic sources. Logical, clear development of narrative. Appropriate references and citations. Very good evaluation and discussion of the significance of the results. Legal, social, ethical, security and professional issues fully considered.
Satisfactory 50% - 59%	Satisfactory technical quality. Produced and demonstrated good quality solution to the problem. Good evidence of reviewing multiple academic sources. Some references and citations. Good evaluation and discussion of the significance of the results. Legal, social, ethical, security and professional issues fully considered.
Fail Below 50%	Not adequate technical quality. Produced and demonstrated a solution to the problem, which is flawed, despite some effort. Poor evidence of reviewing academic sources. Little evaluation and discussion of the results. Little consideration of legal, social, ethical, security and professional issues. Narrative difficult to follow. Poor quality of references and citations.
NS NON-SUBMISSION	N/A

Deliverables & Submission

You are required to submit your work to Blackboard via the assessments link by the due date. You may use a zip file to package your submission artefacts (i.e., the fprg files and your reflective report). All the submitted files should be labelled as follows for identification purposes:

studentID_lastname_firstname.zip (e.g. *x1234567_smith_jane.zip*)

Your reflective report should also be labelled in a similar manner with your student ID.
Make sure your student ID and name is present on all documentation you submit.

Logistics

After the ICA briefing has been given, you will be provided with opportunities to progress your in-course work during some timetabled sessions. Feedback – but not points – will be given on your work in progress to assist you in submitting a considered and well-developed ICA submission.

Academic Misconduct and Plagiarism

Please note that the University takes the issue of academic misconduct and plagiarism very seriously. You should not copy anyone else's work or use copyright materials without due acknowledgement.