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: **15/05/2024**
	Deadline Time: **4:00pm**
	Submission Method: TuOnline (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.

Task Description

Problems in machine learning vary from one 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 websites that provide publicly available datasets. A categorized 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 visualization for the results.
5. Reflect on professional, ethical, and legal issues in relation to the problem and the data set.

Moreover, as part of Element 1 of the ICA, 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 through a quantitative performance of the 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 and 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: 08/05/2024

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 the application.

Element 2 – Contribute 50% of the Module Mark

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

Deadline: 08/05/2024

What to Hand-In

- A case study report maximum of 2,000 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 electronic via Blackboard, all deliverables shall be labelled with the 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.

Knowledge, Skills, and Behaviours (KSB) – Degree Apprenticeship

This module supports the delivery and assessment of the following Knowledge Skills and Behaviours (KSBs)

Knowledge

- Understanding the fundamental concepts and principles of machine learning
- Knowledge of different types of machine learning algorithms and their applications
- Familiarity with common machine learning terms and techniques, such as overfitting and feature selection.

Skills

- Ability to preprocess and analyze a chosen dataset, including data cleaning, feature engineering, and data visualization.

- Proficiency in implementing machine learning algorithms using appropriate programming languages and libraries, such as Python and scikit-learn.
- Skill in evaluating and selecting the most suitable machine learning model for a given problem, considering factors like accuracy, interpretability, and computational efficiency.

Behaviours

- Demonstrating critical thinking and problem-solving skills in designing and developing a machine learning application.
- Showing attention to detail and accuracy in preprocessing data, selecting features, and tuning model parameters.
- Displaying effective communication skills in presenting the machine learning application and discussing the project in the report

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.

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	Demonstrates an exceptional understanding of the chosen machine learning algorithms, including their strengths, weaknesses, and suitability for the given task, providing insightful analysis of their suitability for the task and potential optimizations. Clear evidence of running the experiments with code that is excellently organized and commented, also shown through an excellently produced walk-through demo video. Provides clear justification for the selection of specific machine learning models or techniques, including consideration of data characteristics, scalability, and interpretability. Effectively addresses potential challenges such as overfitting, underfitting, and data imbalance, demonstrating robustness and reliability in real-world applications. Utilizes sophisticated and appropriate techniques for data preprocessing, feature engineering, and model evaluation, enhancing the overall quality and reliability of predictions. Achieves high performance compared to baseline models or industry standards, showcasing the effectiveness of the chosen approach. Presents insightful analysis and interpretation of model outputs, demonstrating an exceptional grasp of the underlying principles and implications, and how the machine learning algorithms contribute to solving the problem at hand. Conducts thorough evaluations of the model's performance, demonstrating an excellent understanding of the models' generalizability and robustness through various methods (cross-validation, feature importance analysis, etc.) Deep understanding is shown.

Very Good 60% - 69%	Demonstrates a solid understanding of the chosen machine learning algorithms, providing thoughtful analysis of their strengths and limitations in the context of the task. Very good evidence of running the experiments with code that is well-organized and commented, also shown through a well-produced walk-through demo video. Offers sound justifications for the selection of machine learning models, considering relevant factors such as data characteristics and scalability. Attains commendable predictive accuracy and performance metrics, meeting, or exceeding expectations with well-executed methodologies. Addresses common challenges like overfitting, underfitting, and data imbalance effectively, with practical solutions demonstrating proficiency. Implements reliable data preprocessing, feature engineering, and model evaluation techniques, contributing to the overall effectiveness of predictions. Provides insightful interpretations of model outputs, showcasing a good understanding of their implications for solving the problem. Very good understanding is shown.
Satisfactory 50% - 59%	Demonstrates a basic understanding of the chosen machine learning algorithms, outlining their relevance to the task without significant depth. Satisfactory evidence of running the experiments with code that is organized and commented on and a satisfactory walk-through demo video. Provides adequate justifications for the selection of machine learning models, considering basic factors such as data characteristics. Achieves satisfactory predictive accuracy and performance metrics, meeting minimum requirements with acceptable methodologies. Addresses common challenges like overfitting, underfitting, and data imbalance with standard solutions, albeit with limited innovation. Implements fundamental data preprocessing, feature engineering, and model evaluation techniques, ensuring basic reliability of predictions. Offers straightforward interpretations of model outputs, capturing key insights without significant depth or elaboration. Satisfactory understanding is shown.
Fail Less than 50%	Demonstrates a lack of understanding or misunderstanding of the chosen machine learning algorithms, with minimal analysis of their relevance to the task. Little evidence of running the experiments with code that is not well-organized and commented on, as well as the walk-through demo video. Provides insufficient or incorrect justifications for the selection of machine learning models, showing little consideration of relevant factors, and the selected machine learning algorithms are not appropriate for the given task. Fails to achieve adequate predictive accuracy and performance metrics, falling significantly short of expectations with flawed methodologies.

	Struggles to address common challenges like overfitting, underfitting, and data imbalance, with ineffective or erroneous solutions. Implements rudimentary or inappropriate data preprocessing, feature engineering, and model evaluation techniques, resulting in unreliable predictions. Offers superficial or incorrect interpretations of model outputs, lacking insight into their implications for solving the problem. Conducts minimal or inadequate evaluations of the model's performance. Poor understanding is shown.
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 the 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 are 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 a 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 the narrative. Appropriate references and citations. Very good evaluation and discussion of the significance of the results. Legal, social, ethical, security, and professional issues are fully considered.

Satisfactory 50% - 59%	Satisfactory technical quality. Produced and demonstrated good quality solutions 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 are 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 is 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 artifacts (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 by including your student ID and full name.

Make sure your student ID and full name are 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 no 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 acknowledgment.