CCT College Dublin Continuous Assessment

Programme Title:	HDip in Science in Data Analytics for Business/ AI		
Cohort:	PT		
Module Title(s):	Machine Learning (10 ETCS)		
Assignment Type:	Group (Max 3 members)	**Weighting(s)**:	50% (60% Group work and 40% Individual)
Assignment Title:			
Lecturer(s):	Dr. Muhammad Iqbal		
Issue Date:	13th March 2023		
Submission Deadline Date:	30th April 2023		
Late Submission Penalty:	Late submissions will be accepted up to **5** calendar days after the deadline. All late submissions are subject to a penalty of **10%** of the mark awarded. Submissions received more than 5 calendar days after the deadline above **will not** be accepted and a mark of 0% will be awarded.		
Method of Submission:	**Moodle**		
Instructions for Submission:	Upload one zip file composed of pdf/ word file, jupyter notebook, dataset and any supporting information.		
Feedback Method:	**Results posted in Moodle gradebook**		
Feedback Date:	Three weeks after submission		

Learning Outcomes:
Please note this is not the assessment task. The task to be completed is detailed on the next page.
This CA will assess student attainment of the following minimum intended learning outcomes:

1. Develop a machine learning strategy for a given domain, communicate this strategy effectively to team members, peers and project stakeholders (CRISP-DM)
   (Linked to PLO 1, PLO 4, PLO 6)
2. Implement a range of classification and regression techniques and detail /document their suitability for a variety of problem domains.
   (Linked to PLO 5)
3. 5. Critically evaluate and optimise the performance of Machine Learning models.
   (Linked to PLO 3)

Attainment of the learning outcomes is the minimum requirement to achieve a Pass mark (40%). Higher marks are awarded where there is evidence of achievement beyond this, in accordance with QQI *Assessment and Standards, Revised 2013*, and summarised in the following table:

Percentage Range	CCT Performance Description	QQI Description of Attainment
		Level 6, 7 & 8 awards
90% +	Exceptional	Achievement includes that required for a Pass and in **most** respects is significantly and consistently beyond this
80 – 89%	Outstanding	
70 – 79%	Excellent	
60 – 69%	Very Good	Achievement includes that required for a Pass and in **many** respects is significantly beyond this
50 – 59%	Good	Achievement includes that required for a Pass and in **some** respects is significantly beyond this
40 – 49%	Acceptable	Attains all the minimum intended programme learning outcomes
35 – 39%	Fail	Nearly (but not quite) attains the relevant minimum intended learning outcomes
0 – 34%	Fail	Does not attain some or all of the minimum intended learning outcomes

Please review the CCT Grade Descriptor available on the module Moodle page for a detailed description of the standard of work required for each grade band.

The grading system in CCT is the QQI percentage grading system and is in common use in higher education institutions in Ireland. The pass mark and thresholds for different grade bands may be different from what you have experienced in the higher education system in other countries. CCT grades must be considered in the context of the grading system in Irish higher education and not assumed to represent the same standard the percentage grade reflects when awarded in an international context.

Assessment Task

This is a group-based project (Max 3 students) using the PYTHON programming language. Develop and deploy machine learning models in any one of the following areas only and analyse the results.

- Population and Society
- Energy & Environment
- Economy and Finance

You can find any public dataset from an authentic resource repository and the dataset should have at least 3000 rows and 10 columns after cleaning and there is not any upper bound.

The type of question(s) that you should formulate for the project will depend on the chosen area of the dataset that your group is considering for the machine learning project.

Suggested possible analysis / project questions are mentioned below (this is a small, suggested, sample of questions, other questions may be more appropriate to your project)

- What are the most important features for predicting X as a target variable?
- Which classification approach do you prefer for the prediction of X as a target variable, and why?
- How to classify the loyal and churn customers using Support Vector Machines?
- Why is dimensionality reduction important in machine learning?

The group would need to consider the following instructions (a - d) during the development of this group project.

a) Logical justification based on the reasoning for the specific choice of machine learning approaches.
b) Multiple machine learning approaches (at least two) using hyperparameters and a comparison between the chosen modelling approaches.

c) Visualise your comparison of ML modelling outcomes. You may use a statistical approach to argue that one feature is more important than other features (for example, using PCA).
d) Cross-validation methods should be used to justify the authenticity of your ML results.

Your group will present their findings and defend the results in the report (MS Doc/ pdf or any other readable format) by highlighting their individual contribution. Your report should capture the following aspects that are relevant to your project investigations.

1. Motivation, a description of the problem domain, and an explanation of how the project's goals are justified using Prediction / Classification / Clustering Rules / Dimensionality Reduction etc..

(10 marks)

2. Characterization of data, explanation and description of techniques used for the variation in the accuracy across three training splits (10% / 20%/ 30%) using cross validation techniques.

(30 marks)

3. Interpret and explain the results obtained, discuss overfitting / underfitting / generalisation, provide a rationale for the chosen model and use visualisations to support your findings. Comments in Python code, conclusions of the project should be specified at the end of the report. Harvard Style must be used for citations and references.

(20 marks)

4. Each team member presents a PowerPoint presentation of their work (maximum 5 slides) to emphasize their distinctive contributions based on their involvement in the project's conceptual understanding, code development, and deployment.

(20 marks individual)

5. Each team member fully described their individual contributions to the project in a reflective journal, using at least 500 to 700 words as well as images, diagrams, figures, and visualizations to elaborate his/ her work.

(20 marks individual)

Submission Requirements

All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the marks awarded.

- The code and datasets should be provided and uploaded in zip format on Moodle.
- Clearly detail the number of words used in the report.
- Number of Words in the report (3000 words +/-10%) excluding diagrams, code, references and titles. Number of words used to express individual contributions is part of the mentioned words.
- Describe the contribution of each team member in the project clearly and use a bar chart or pie chart to represent the effort and time spent during this project.
- The rubric is provided for the detailed breakdown of marks at the end of this CA.
- Use Harvard Referencing when citing third party material
- Be the student's own work.
- Include the CCT assessment cover page.
- Be submitted by the deadline date specified or be subject to late submission penalties
- Note: The names of group members must be uploaded on the link provided on Moodle until 17th March 2023 (23:59).
- Must be clearly specified the number of words used after each section in the report.

GRADING RUBRIC – Machine Learning - 2022/2023								
GRADE	**90-100%**	**80-90%**	**70-79%**	**60-69%**	**50-59%**	**40-49%**	**35-39%**	**<35%**
Performance	**Exceptional**	**Outstanding**	**Excellent**	**Very Good**	**Good**	**Acceptable**	**Fail**	**Fail**
Introduction to problem Description, Motivation and Objectives (10%)	An exceptional introduction to problem description and motivation that provide a concise and clear case for the proposed Machine Learning project. An exceptional specification of objectives concisely.	An outstanding introduction to problem description and motivation that provide a compact and clear case for the proposed Machine Learning project. An outstanding specification of objectives precisely.	An excellent introduction to problem description and motivation that provide a precise and clear case for the proposed Machine Learning project. An excellent specification of objectives succinctly.	A very good introduction to problem description and motivation that provides a very convincing case for the proposed Machine Learning project. A very good specification of objectives.	A good introduction to problem description and motivation that furnishes a largely convincing case for the proposed Machine Learning Project. A good specification of objectives.	An acceptable introduction to problem description and motivation that offers a somewhat weak case for the proposed Machine Learning Project. An adequate specification of objectives.	A poor introduction to problem description and motivation that fails to motivate the problem or provide a case for the proposed Machine Learning Project. A poor specification of objectives.	An impecunious introduction to problem description that fails entirely to motivate the problem. An impecunious specification of objectives.
Characterization and cleaning of Dataset, Training and Testing of Models (30%)	An exceptional characterization and cleaning of a dataset that abstracts all details from source to fields. An exceptional accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the generalizability of the model and it should justify the results in an exceptional way.	An outstanding characterization and cleaning of dataset that highlights all details from source to fields. An outstanding accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the generalizability of the model and it should justify the results in an outstanding way.	An excellent characterization and cleaning of the dataset that summarizes all details from source to fields. An excellent accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the generalizability of the model and it should justify the results in an excellent way.	A very good characterization and cleaning of the dataset that summarizes all details from source to fields. A very good accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the partial generalizability of the model and it should justify the results.	A good characterization and cleaning of the dataset that summarizes all details from source to fields. A good accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used to test the partial generalizability of the model.	An acceptable characterization and cleaning of the dataset that summarizes all details from source to fields. An adequate accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is used.	A poor characterization and cleaning of the dataset that summarizes all details from source to fields. A poor accuracy obtained based on the training and testing of ML models using three logical splits. Cross-validation is not used.	An impecunious characterization and cleaning of the dataset. An impecunious obtained based on the training and testing of ML models using three logical splits. Cross-validation is not used.
Interpretation of results, Code description and comments, Conclusions, citations, and references (20%)	An exceptional interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An exceptional justification is provided.	An outstanding interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An	An excellent interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An excellent defence is provided.	A very good interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are neither overfitted nor underfitted. A very good justification is provided.	A good interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are overfitted but not under fitted. A good justification is provided.	An acceptable interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. The results state that the models are adequate. An adequate justification is provided.	A poor interpretation and explanation of the results, code description, comments, conclusions, citations, and references based on problem specification and objectives. No clear results obtained.	An impecunious interpretation of the results. No clear results obtained.

		outstanding advocacy is provided.						
Powerpoint presentation (20%) - Individual	The presentation is delivered in an exceptional manner, is well-organized and visually appealing, and successfully explains the topic's essential concepts, ideas, and code.	The presentation is delivered in an outstanding manner, is well-organized and visually appealing, and successfully explains the topic's essential concepts, ideas and code.	The presentation is delivered in an excellent manner, is well-organized and visually appealing, and successfully explains the topic's essential concepts, ideas, and code.	The presentation is delivered in a very good manner, is nicely organized and visually appealing, and decently explains the topic's essential concepts, ideas and code.	The presentation is delivered in a good manner, is organized and visually appealing, and explains the topic's essential concepts, ideas, and code.	The presentation is delivered in an acceptable manner, is organized, and explains the topic's essential concepts, ideas, and code to some extent.	The presentation is delivered in a poor manner, is not organized, and has an unsuccessful explanation of the topic's concepts, ideas, and code.	The presentation is not delivered according to the guidelines.
Reflective journal for individual group member (20%) - Individual	Reflection demonstrates an exceptional level of engagement and understanding of the group project material, and shows exceptional evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates an outstanding level of engagement and understanding of the group project material, and shows outstanding evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates an excellent level of engagement and understanding of the group project material, and shows excellent evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates a very good level of engagement and understanding of the group project material, and shows very good evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates a good level of engagement and understanding of the group project material, and shows good evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates an acceptable level of engagement and understanding of the group project material, and shows some evidence of critical thinking, self-reflection, and collaboration.	Reflection demonstrates a poor level of engagement and understanding of the group project material, and shows incomplete evidence of critical thinking, self-reflection, and collaboration.	Reflection does not demonstrate any engagement and understanding of the group project material, and shows no evidence of critical thinking, self-reflection, and collaboration.