CCT College Dublin Continuous Assessment

Programme Title:	HDip in Data Analytics for Business		
Cohort:	FT		
Module Title(s):	Machine Learning for Business		
Assignment Type:	Group (Max 2 members)	**Weighting(s)**:	50% (60% Group work and 40% Individual)
Assignment Title:			
Lecturer(s):	Dr. Muhammad Iqbal		
Issue Date:	14th September 2023		
Submission Deadline Date:	5th November 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 or multiple files as 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. Critically evaluate and implement appropriate clustering algorithms and interpret and document their results. (Linked to PLO 1, PLO 5)
2. Apply modelling to time series data to facilitate business intelligence needs (Linked to PLO 1, PLO 2, 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	

80 – 89%	Outstanding	Achievement includes that required for a Pass and in **most** respects is significantly and consistently beyond this
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

Students are advised to review and adhere to the submission requirements documented after the assessment task.

This is a group-based project (Max 2 students) using the PYTHON programming language. Develop and deploy machine learning models in any one of the following areas only and analyse the results.

- Covid-19 datasets
- Transport datasets
- Energy
- Stock market dataset from only website: https://data.world/datasets/finance

The dataset should have at least 8000 rows and 10 columns (for example, type of variables may be categorical, continuous, and discrete) 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 domain of the dataset that your group is considering.

Project questions could be: (this is a small, suggested, sample of questions, other questions may be more appropriate to your project)

- How to measure similarity or dissimilarity between different clusters?
- Which clustering solution do you prefer, and why?
- How to analyse and investigate an inflation rate for a specific product?

Your group may start with a simple approach to initiate your project work based on project objectives and enhance your work using distinct approaches. 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 (supervised/ Unsupervised) for the chosen problem and dataset (s). Justify the rationale for using the project management framework/ activities (CRISP-DM, KDD, or SEMMA).
b) Machine Learning models can be used for Prediction, Classification, Clustering and time series analysis. Your group should plan on trying multiple approaches (at least two), with proper parameter-selection using hyperparameter tuning and a comparison between the chosen modelling approaches if essential.
c) You/ Your group should train the Machine learning models, test and further validate the models. Perform a comparison of two or more ML modelling outcomes using a Table or graph visualization. Your group may employ dimensionality reduction methods to prepare the dataset based on your project requirements.
d) Depending on the complexity of the problem, develop the clustering profiles that clearly describe the characteristics of the specific data within the cluster.

Your group will present their findings and defend the results in the report (MS Doc). Your report should capture the following aspects that are relevant to your project investigations.

i) Motivation, description of problem domain, justification of project objectives in the above-mentioned areas.

(15 marks group)

ii) Characterization and normalization of data if required, train and test supervised ML models based on three different splits in the case of supervised learning and discuss the variation in accuracy/ score obtained from the models. Use appropriate metrics to justify your results in the case of unsupervised learning.

(25 marks group)

iii) Interpret and justify the results based on the problem specification or project objectives. Comments and description of Python code, conclusions of the project should be specified in the report as well as jupyter notebook. Citations and references should be in the Harvard Style.

(20 marks group)

iv) 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)

v) 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.
- Number of Words for the report (4000 words +/-10%) excluding title page, diagrams, code and HARVARD References). Number of words used to express individual contributions is part of the mentioned words.
- Clearly detail the number of words used in the report.
- 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. Use version control like Github or any other tool to show the progress of both team members in CA1.
- 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 5th November 2023 (23:59).
- Must be clearly specified the number of words used after each section in the report.

GRADING RUBRIC – Machine Learning for Business - 2022								
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 (15%)	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 adequate 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 (25%)	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 adequate 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 based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An exceptional justification is provided. An exceptional description of code using logical comments. The	An outstanding interpretation and explanation of the results based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An outstanding advocacy is provided. An outstanding description of code using rational	An excellent interpretation and explanation of the results based on problem specification and objectives. The results clearly state that the models are neither overfitted nor underfitted. An excellent defence is provided. An excellent description of code using comments. The comments are detailed and provide an	A very good interpretation and explanation of the results based on problem specification and objectives. The results state that the models are neither overfitted nor underfitted. A very good justification is provided. A very good description of code using comments. The comments are brief and provide a clear	A good interpretation and explanation of the results based on problem specification and objectives. The results state that the models are overfitted but not under fitted. A good justification is provided. A good description of code using comments. The comments are very brief and provide an understanding of the	An adequate interpretation and explanation of the results based on problem specification and objectives. The results state that the models are adequate. An adequate justification is provided. An adequate description of code using comments. The comments are not satisfactory and provide a partial understanding	A poor interpretation and explanation of the results based on problem specification and objectives. No clear results obtained. A poor description of code using comments. The comments are not satisfactory. A poor demonstration of conclusions or no conclusions. A report along with errors.	An impecunious interpretation of the results. No clear results obtained. An impecunious description of code using unsatisfactory comments. An impecunious demonstration of conclusions or no conclusions. An inadequate report.

	comments are detailed and provide a remarkable understanding of the functionality of the code. An exceptional presentation of conclusions. An exceptional report along with proper citations and references in all sections.	comments. The comments are detailed and provide an impeccable understanding of the functionality of the code. An outstanding manifestation of conclusions. An outstanding report along with appropriate citations and references in all sections.	explicit understanding of the functionality of the code. An excellent demonstration of conclusions. An excellent report along with proper citations and references in all sections.	understanding of the functionality of the code. A very good demonstration of conclusions. A very good report along with proper citations and references in all sections.	functionality of the code. A good demonstration of conclusions. A good report along with citations and references in some sections.	of the functionality of the code. An adequate demonstration of conclusions. An adequate report along with incomplete citations and references.		
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.