**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | *Hdip DA for Business* | | |
| **Cohort:** | *FT* | | |
| **Module Title(s)**: | *Data Visualization Techniques, Machine Learning for Business* | | |
| **Assignment Type:** | *Individual* | **Weighting(s)**: | *50%, 50%* |
| **Assignment Title:** | *CA2* | | |
| **Issue Date:** | *16th October 2024* | | |
| **Lecturer(s)**: | *David McQuaid, Muhammad Iqbal* | | |
| **Submission Date:** | *24th November 2024* | | |
| **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 link for Machine Learning for Business Module** | | |
| **Feedback Method:** | ***Please do not ZIP your files. ALL files must be uploaded individually (to a maximum of 20 files)***  *Expected files : Written report (word document only, NO PDF’s) ,Code files (Jupyter notebook (.ipynb) ONLY, NO PYTHON FILES), Data Files, Screencast for practical demonstration. Note that the maximum number of Jupyter Notebooks is 2.* | | |
|  | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After the approval from Exam board* | | |

**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:

**Learning Outcomes Assessed:** Machine Learning for Business

**List the module learning outcomes to be assessed (delete as necessary)**

**MLO 2 -** Perform market basket analysis on representative transactional data. Explore the potential applications of these techniques outside their usual domain.

(Linked to PLO 4, PLO 5)

**MLO 3 -** Implement text categorisation, topic modelling and document summarisation on a range of representative texts.(e.g. twitter, facebook)

(Linked to PLO 3, PLO 5)

**Learning Outcomes Assessed:** Data Visualisation Techniques

**MLO 2 -** Select appropriate data visualisation techniques for a given use case and data

characteristics (Linked to PLO 3 (Stage 4 SLO 3).

**MLO 3 -** Propose, design, develop, and implement data visualisation solutions.

(Linked to PLO 4).

**MLO 4 -** Display effective presentation skills to communicate with peers, team members

and project stakeholders (Linked to PLO 3, PLO 6)

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** | **Level 9 awards** |
| 90% + | Exceptional | Achievement includes that required for a Pass and in **most** respects is significantly and consistently beyond this | 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 | 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 | Attains all the minimum intended programme learning outcomes |
| 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 | 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 | 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 experience 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**

**It is mandatory that you use GitHub Classroom with regular commits (at least 10) of code and report versions. You may be called to a short Viva to defend your work.**

**Please find the GitHub Classroom link below:**

**https://classroom.github.com/a/Q3TzODug**

**You are not allowed to upload a PDF document for your report, It MUST be a word document.**

**Assessment Task**

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

**Assessment details**

**Questions:**

1. Discuss and explain the purpose of a recommendation system for online retail business in machine learning. Briefly compare Content and Collaborative filtering using any dataset of your choice (Datasets used in the class tutorials or exercises are not allowed to use in this CA2). Train and test machine learning models for the user-user and item-item collaborative filtering. Justify your recommendations for the considered scenario by providing a conceptual insight.

In the point 1 I have to use the dataset with clothes and reviews

(60 marks)

2) Perform Market Basket Analysis on the chosen dataset by using Apriori and FP growth algorithms. Can you express major similarities/ divergence between these models? Compare and contrast the machine learning results obtained based on both algorithms.

Market basket Analysis only uses information of the products,there is not dataset yet

(40 marks)

3) Create an interactive Dashboard aimed at younger adults (18 - 35 years) with specific features to summarise the most important aspects of the data and identify through your visualisation why this dataset is suitable for Machine Learning models in an online retail business. Explain how your dashboard is designed with this demographic in mind.

(70 marks)

4) Discuss in detail your rationale and justification for all stages of data preparation for your visualizations.

(30 marks)

**Submission Requirements**

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

* Include the CCT assessment cover page.
* The code and datasets should be provided and uploaded in zip format on Moodle.
* Maximum Number of Words for the report (2500±10% words excluding title page, diagrams, code and HARVARD References).
* Must be clearly specified the number of words used in the report.
* Use Harvard Referencing when citing third party material
* Make sure the dataset should not be used in any previous assessments/ lectures/ tutorials for this CA.
* Be the student’s own work.

**Acceptable and Unacceptable Use of AI**

|  |  |
| --- | --- |
| **Acceptable and Unacceptable Use of AI** | ● The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:  o Brainstorming and refining your ideas;  o Fine tuning your research questions;  o Finding information on your topic;  o Drafting an outline to organise your thoughts; and  o Checking grammar and style.  ● The use of generative AI tools is not permitted in this course for the following activities:  o Impersonating you in classroom context  o Completing group work that your group has assigned to you  o Writing a draft of a writing assignment  o Writing entire sentences, paragraphs, papers, code fragments, functions, scripts to complete class assignments.    ● You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited.  ● Any assignment that is found to have used generative AI tools in an unauthorised way will be subject to college disciplinary procedures as outlined in the QA Manual.  ● When in doubt about permitted usage, please ask for clarification. |

**Rubric for MLB Module**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Marking Criteria** | **Weighting** | **Excellent (+70%)** | **Very Good (60-69%)** | **Good (50-59%)** | **Acceptable (40-49%)** | **Fail (<39%)** |
| **Problem Understanding and Motivation** | 20 marks | Clear and thorough explanation of the purpose and motivation behind recommendation systems for online retail, with relevant context and specific details. | Strong understanding with good motivation and context; explanation is convincing and appropriate. | Adequate explanation with acceptable understanding but lacking some depth or context. | Basic understanding, with some missing details or unclear purpose. | Inadequate explanation, with little or no connection to recommendation systems or motivation. |
| **Comparison and Implementation of Algorithms (Content vs Collaborative Filtering)** | 40 marks | Comprehensive and insightful comparison between content-based and collaborative filtering, demonstrating deep conceptual understanding; accurate and well-documented implementation of user-user and item-item collaborative filtering, with meaningful results and well-justified recommendations. | Strong comparison with relevant insights and a very good implementation of collaborative filtering, with clear documentation and well-supported recommendations. | Adequate comparison, lacking depth, with an acceptable implementation that has some gaps in execution or explanation of results. | Basic comparison with missing details; implementation is partial with unclear recommendations and some logical gaps. | Missing or poor comparison with little to no conceptual understanding of filtering methods; inadequate or missing implementation with no clear recommendations. |
| **Implementation of Apriori and FP Growth Algorithms** | 30 marks | Accurate and well-documented implementation of both Apriori and FP Growth algorithms, with comprehensive results and clear understanding of both. | Very good implementation of both algorithms, well supported by clear results and insights. | Adequate implementation, but some steps are lacking clarity or full results. | Basic implementation with some missing steps or weak understanding of algorithms. | Inadequate or missing implementation with little understanding or poor results. |
| **Comparison of Results and Conceptual Insight (Apriori vs FP Growth)** | 10 marks | Thorough comparison of results from Apriori and FP Growth, with clear conceptual insight into the major similarities/divergences between the two. | Strong comparison with good insights into both algorithms' results and differences. | Adequate comparison, though lacking depth or complete insight into similarities and differences. | Basic comparison with weak understanding or missing insights into differences. | Poor or missing comparison with little or no conceptual insight into the algorithms or results. |

**Rubric for DVT Module**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Marking Criteria** | **Weighting** | **Excellent (+70%)** | **Very Good (60-69%)** | **Good (50-59%)** | **Acceptable (40-49%)** | **Fail (<39%)** |
| **Dashboard Design and User-Centric Features** | 35 marks | Exceptionally designed dashboard tailored for younger adults (18-35) with engaging visual elements and interactive features; meets the target audience’s preferences effectively. | Strong design with relevant features for younger adults; good use of visual elements and interactivity. | Satisfactory design that meets some user preferences but lacks depth in interactivity or visual appeal. | Basic dashboard design with limited user engagement; some features may not cater to the target demographic. | Poorly designed dashboard with little regard for the target audience; lacks interactivity and visual appeal. |
| **Data Visualization and Summary of Key Aspects** | 35 marks | Comprehensive visualizations that clearly summarize important aspects of the dataset, effectively supporting insights suitable for Machine Learning. | Strong visualizations that summarize key aspects; insights are relevant and understandable for the audience. | Adequate visualizations but may lack clarity or completeness in summarizing key aspects. | Basic visualizations that may not effectively summarize important aspects or insights. | Poor or missing visualizations that do not summarize key aspects or are unclear for the audience. |
| **Rationale and Coherence of Data Preparation** | 30 marks | Comprehensive and detailed discussion of all stages of data preparation, with clear rationale and strong justification for each step; exceptionally clear and coherent explanation. | Good discussion of data preparation stages, with relevant rationale and justification; clarity is generally strong. | Adequate discussion with some rationale but may lack depth; clarity could be improved. | Basic discussion with limited rationale; explanations may confuse the reader. | Poor or missing discussion, with little to no rationale for data preparation stages. |

**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** |  |
| **Assessment Title:** |  |
| **Lecturer Name:** |  |
| **Student Full Name:** |  |
| **Student Number:** |  |
| **Assessment Due Date:** |  |
| **Date of Submission:** |  |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |