A picture containing icon

Description automatically generated

**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | HDip Data Analytics for Business | | |
| **Delivery Mode:** | SB+ | | |
| **Cohort Details:** | *HDIPDAB Sep24 SB+ Stage1 Semester 2]* | | |
| **Module Title(s)**: | *Machine Learning for Business, Data Visualisation Techniques* | | |
| **Assignment Type:** | *Individual* | **Weighting(s):** | *50%, 50%* |
| **Assignment Title:** | *CA2* | | |
| **Lecturer(s)**: | *Sam Weiss, David McQuaid* | | |
| **Issue Date:** | *April 2025* | | |
| **Submission Deadline Date:** | *18th May 2025 @ 23:55* | | |
| **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:** | **This assignment is submitted via Moodle.** | | |
| **Instructions for Submission:** | *A report of approx. 2500 words as a Word document (.docx)*  *Your code as a Jypyter Notebook (.ipynb, 2 max)* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After approval from Exam Board* | | |

CA Git Repository:

**https://classroom.github.com/a/GLPMXw7l**

# Assessment Outline

## Description of Assessment Task

This is an individual continuous assignment using Python programming language in a Jupyter Notebook.

You have been provided with data to complete the following tasks.

**Machine Learning for Business**

**Clustering**

Apply and evaluate various clustering techniques with the aim of generating actionable insights from the data.

* Explain your analysis objectives (eg data exploration, insight generation etc).
* Select and justify the features you will be using.
* Apply appropriate clustering algorithms to the dataset.
* Evaluate the performance of the algorithms and make a recommendation as to which gives the “best” results.
* Include in your report your own interpretation of the results.

**[60 marks]**

Train and test machine learning models for collaborative filtering, in the context of a recommender system.

* Select and justify the features you will be using.
* Apply appropriate collaborative filtering algorithms to the dataset.
* Evaluate the performance of the algorithms and make a recommendation as to which gives the “best” results.
* Discuss whether you think that including elements of/replacing your algorithm with a content filtering approach may be beneficial to this recommender system.
* Include in your report your own interpretation of the results.

**[40 marks]**

**Data Visualisation**

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)

**All code and writing progress should be kept in the CA Git repository with a minimum of 8 commits each**

<https://classroom.github.com/a/GLPMXw7l>

## Assessment Requirements

All assessment submissions must meet the following minimum requirements:

* Your report as a Word document (.docx) approx. 2500 words.
* Your Python code as a maximum of 2 Jupyter Notebooks (.ipynb)
* Be submitted by the deadline date specified or be subject to late submission penalties.
* Be submitted via Moodle upload.
* Use [Harvard Referencing](http://40.115.124.2/sp/subjects/guide.php?subject=harvardref) when citing third party material.
* Be the student’s own work.
* Include the CCT assessment cover page.

## Learning Outcomes:

This assessment addresses the following module learning outcomes for this module:

**Machine Learning for Business**

|  |  |
| --- | --- |
| **MLO 1** | Critically evaluate and implement appropriate clustering algorithms and interpret  and document their results. |
| **MLO 3** | Implement text categorisation, topic modelling and document summarisation on a range of representative texts.(e.g. twitter, facebook) |

**Machine Learning for Business**

|  |  |
| --- | --- |
| **MLO 2** | Select appropriate data visualisation techniques for a given use case and data  characteristics |
| **MLO 3** | Propose, design, develop, and implement data visualisation solutions. |
| **ML) 4** | Display effective presentation skills to communicate with peers, team members  and project stakeholders |

## Statement of Acceptable Use of Artificial Intelligence

|  |
| --- |
| **Acceptable and Unacceptable Use of AI**  *This statement is useful when you are allowing the use of AI tools for certain purposes, but not for others.*  *Adjust this statement to reflect your particular parameters of acceptable use, and your discipline context.* |
| * The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:   + Brainstorming and refining your ideas;   + Fine tuning your research questions;   + Finding information on your topic;   + Drafting an outline to organise your thoughts; and   + Checking grammar and style.   + Generating code fragments (appropriately labelled) but not entire code sections. * The use of generative AI tools is not permitted in this course for the following activities:   + Impersonating you in classroom context   + Completing group work that your group has assigned to you   + Generating entire code sections for your assignment   + Writing a draft of a writing assignment   + Writing entire sentences, paragraphs or papers 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. |

## Grading Criteria

**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. |

**Machine Learning for Business**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Marking Criteria** | **Weighting** | **Excellent (+70%)** | **Very Good (60-69%)** | **Good (50-59%)** | **Acceptable (40-49%)** | **Fail (<39%)** |
| **Clustering Research Objectives** | 10 marks | Clear and thorough explanation of the purpose and motivation behind the clustering algorithms, 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. |
| **Clustering implementation** | 50 marks | Comprehensive and insightful comparison between various diverse clustering algorithms, demonstrating deep conceptual understanding; accurate and a well-documented implementation, with meaningful results and well-justified recommendations. | Strong comparison with relevant insights and a very good implementation of various diverse clustering algorithms, 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. |
| **Recommender 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. |

**s**

**The Irish Grading System**

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.

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, and review the marking criteria outlined in this assignment brief for a breakdown of the marking criteria for this specific assignment.

**Additional Information**

* Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer’s discretion.
* In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
* Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
* Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
* Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
* Students are advised that disagreement with an academic judgement is not grounds for review.
* For additional support with academic writing and referencing students are advised to contact the CCT Library Service.
* For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](https://moodle.cct.ie/course/view.php?id=827)
* For additional support with IT subject content, students are advised to access the [CCT Support Hub](https://moodle.cct.ie/course/view.php?id=1861).