**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | *Hdip DA for Business* | | |
| **Cohort:** | *PT* | | |
| **Module Title(s)**: | *Machine Learning for Business* | | |
| **Assignment Type:** | *Integrated* | **Weighting(s)**: | *50%* |
| **Assignment Title:** | *CA2* | | |
| **Issue Date:** | *2nd November 2023* | | |
| **Lecturer(s)**: | *Sam Weiss/James Garza* | | |
| **Submission Date:** | *24th 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**  Upload Word document (2000 words max), Jupyter notebooks, dataset and any supporting information. Be sure to include a link to your GitHub repository at the top of the report! | | |
| **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:**

**MLOs Machine Learning**

**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)

**MLO 4 -** Apply modelling to time series data to facilitate business intelligence needs

(Linked to PLO 1, PLO 2, PLO 3)

**MLOs Data Visualisation**

**MLO 1** - Discuss the concepts, techniques and processes underlying data visualisation (Linked to PLO 1)

**MLO 3** - Select appropriate data visualisation techniques for a given use case, data characteristics and multiple transmission media (Linked to PLO 3, PLO 4)

**MLO 5** - Present a detailed visualisation of a data analysis to 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** | **QQI Description of Attainment** | |
| **Level 6, 7 & 8 awards** | **Level 9 awards** |
| 70% + | 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 |
| 60 – 69% | 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% | 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% | Attains all the minimum intended programme learning outcomes |
| 35 – 39% | Nearly (but not quite) attains the relevant minimum intended learning outcomes | Nearly (but not quite) attains the relevant minimum intended learning outcomes |
| 0 – 34% | Does not attain some or all of the minimum intended learning outcomes | Does not attain some or all of the minimum intended learning outcomes |

The CCT Grade Descriptor describes the standard of work for grade boundaries summarised below. The full descriptor is available on Moodle. Understanding grading in the Irish Higher Education context is available at:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Grade** | 90-100% | 80-89% | 70-79% | 60-69% | 50-59% | 40-49% | 35-39% | <35% |
| **Performance** | Exceptional | Outstanding | Excellent | Very Good | Good | Acceptable | Fail | Fail |

**Assessment Task**

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

**Note that ALL Students are required to use Git for any Assignments that they are working on.**

**Assessment details**

Question 1:

Discuss the concept and application of Time series analysis using machine learning modelling by providing a real-world data set. What is the purpose of The Augmented Dickey-Fuller test in time series?

a) Apply an appropriate Box-Jenkins model to the chosen dataset (ARMA, ARIMA, SARIMA etc). Check for the model adequacy.

b) Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

c) Make a time series plot of the data, and further calculate and illustrate the sample autocorrelation and partial autocorrelation (PA). Is there significant autocorrelation in the chosen time series?

(40 Marks)

Question 2:

Discuss the concept and application of Text Analytics (eg text categorisation, topic modelling and document summarisation) using machine learning modelling by providing a real-world data set of social media posts.

Apply appropriate text analytics tools to the chosen dataset.

(40 marks)

Observations about the assessment:

In your report, please include:

* **Introduction:** Briefly explanation of the topics/ questions and the steps that were followed to write the report.
* **Argument:** Justification of process you followed and the machine learning models performed to gain insights about the topics you are working on. Variety, research, and rationality explanation is expected.
* **Conclusions:** Here you should lead to some final comments regarding the topics you worked with and the relation between the raw data and machine learning.

(20 marks for report)

**Data Visualisation**

Create an interactive dashboard using your chosen dataset. Create a wireframe proposing the design of the dashboard before implementation. The dashboard will include at least three rows and two columns for a total of six sections. There should be a minimum of four plots in total. The remaining sections can be made up of text, tables or any other relevant information you deem necessary that can give critical insights to the dashboard viewer.  The dashboard will include a range of visualisations that effectively communicate the key insights derived from the exploratory data analysis.

**The dashboard could include any of the following visualisations:**

A heatmap showing the correlation matrix between all continuous variables. A heatmap could help the viewers understand the strength and direction of the relationships between variables.

A scatter plot matrix will show pairwise relationships between all continuous variables. A scatter plot matrix could enable the viewer to visually identify potential outliers or non-linear relationships between the variables.

A bar chart showing the distribution of the target variable could help the viewers understand the range and distribution of the target variable.

A histogram for each continuous variable could help the viewers understand the distribution of the continuous variables.

A stacked bar chart showing the distribution of each categorical variable could help viewers understand the distribution of the categorical variables and how they relate to the target variable.

A box plot showing the distribution of the target variable for each category of each categorical variable could enable viewers to understand how the target variable varies across the different categories of categorical variables.

A line chart showing any patterns or trends that change over time across a continuous variable.

A scatter plot showing the relationship or association between two continuous variables. Particularly useful for identifying patterns, trends, clusters or correlations in the data.

The dashboard will be designed to be interactive, allowing the audience to filter and explore the data in more detail. For example, the audience can filter the categorical variables by category or select specific data points in the scatter plots to explore and understand the underlying data. The dashboard will also briefly describe the dataset, its provenance and domain, and a summary of the key insights derived from the exploratory data analysis. There will need to be at least four interactive plots in the dashboard. You can use any Python visualisation libraries, such as Plotly Dash or Altair, to create the interactive dashboard.

**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 to Moodle.

● Maximum Number of Words for the report (2000 +- 10% words excluding title page, diagrams, code and HARVARD References).

● Must be clearly specified the number of words used in the report.

● Include the **GitHub** repository **link** in the report.

● The rubric is provided for the detailed breakdown of marks at the end of this CA.

● 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.

**Marking Schedule Data Visualisation:**

|  |  |
| --- | --- |
| **Description** | **Weighting** |
| Explain what was done with the data and the pre-processing and cleaning. | 10% |
| Creativity and originality of the dashboard. Explain the colour selection and Python visualisation library use. Effectiveness of the dashboard in communicating key insights. Explain the selection of the plots and key insights found. | 30% |
| Suitability of visualisation techniques for the given use case and data characteristics. Explain the plots and why you chose the specific visualisations. | 40% |
| Clarity of dashboard design and user interface. An initial wireframe of the dashboard and rationale on the six sections giving key insights to the viewer of the dashboard. The plots are interactive and easy to use. | 20% |
| Poor referencing, spelling, grammar and layout will incur marking penalties. |  |
| **Total** | **100%** |

**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. |