A picture containing icon

Description automatically generated

**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | MSc in Data Analytics | | |
| **Delivery Mode:** | FT/SB+ | | |
| **Cohort Details:** | *MSc in Data Analytics Stage 1 Sem 1* | | |
| **Module Title(s)**: | *ML for Data Analysis*  *Programme schedules are all published on the* [*CCT IQR Provider Profile*](https://irq.ie/providers/cct-college-dublin?id=fec9ea7a-ace4-42c7-9fd5-7fccb6f0a53a&ref=%257B%2522search%2522:%2522cct%2522%257D) | | |
| **Assignment Type:** | *Individual* | **Weighting(s):** | *ML for Data Analysis 100%*  *Capped at 40%* |
| **Assignment Title:** | *MSC\_DA\_ML\_Repeat* | | |
| **Lecturer(s)**: | *Dr. Muhammad Iqbal* | | |
| **Issue Date:** | April 5th, 2025 | | |
| **Submission Deadline Date:** | May 5th, 2025 | | |
| **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**  **Use the submission link on the Data Visualisation and Preparation Module page** | | |
| **Instructions for Submission:** | *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*  *Note that the maximum number of Jupyter Notebooks is 4* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After exam board Jun 2025* | | |

### Assessment Outline

### Description of Assessment Task

## Note: This is an academic exercise and not a hypothetical report and the most important aspect of this report is evaluating and rationalizing your decisions in the domain of Data Analytics NOT the problem domain.

## All Project files MUST be uploaded into MOODLE, this is your responsibility, if any files are not uploaded to MOODLE, even if they are available on GITHUB, they will NOT BE GRADED.

## Criteria of Analysis (ALL EXPERIMENTAL WORK MUST BE CARRIED OUT USING PYTHON IN JUPYTER NOTEBOOK)

***Scenario***

*“Like it or not, every construction company—and solutions provider—is now also in the data business. How well we help our customers transform that data into intelligence that drives better decisions to deliver projects more efficiently and more sustainably, with higher quality, lower costs and fewer risks is what defines the next frontier of construction management. Data is the key to improving the bottom line as well as protecting it. Our ability to break down data silos and transform raw data into action and intelligence is the crux to solving most challenges that rear their head in our industry. Solve the data problem and everything else falls into place.”—Jon Fingland, General Manager, Collaboration Solutions, Trimble*

You have been tasked with analysing Ireland's Construction data and comparing the Irish Construction sector with other countries worldwide. Your Research could include export, import, trade imbalance, house production, material stock, labour/skill pool, etc. (or any other relevant topic EXCEPT Climate change) with Ireland as your base line.

**Note:**

**While topical, Construction impact on Climate Change SHOULD NOT be chosen as an area of research for this assessment.**

You must source appropriate data sets from any available repository to inform your research (all datasets MUST be referenced, and the relevant licence/permissions detailed).

***Criteria of Analysis***

It is Required that you use GitHub Classroom as your version control repository etc with regular commits of code and report versions. You may be called to a Viva to defend your work.

Please find the GitHub Classroom link below:

<https://classroom.github.com/a/4odn-Ai9>

**Machine Learning Tasks**

**You Must Conduct an analysis of your Data based on the problem domain and data you have chosen following which you must compile a report (2000 words +/- 10 % (not including code, code comments, titles, references, or citations)** **) based on the following criteria**

Use of multiple models (at least two) to compare and contrast results and insights gained.

* Describe the rationale and justification for the choice of machine learning models for the above-mentioned scenario. Machine Learning models can be used for Prediction, Classification, Clustering, sentiment analysis, recommendation systems and Time series analysis. You should plan on trying multiple approaches (at least two) with proper selection of hyperparameters using GridSearchCV method. You can choose appropriate features from the datasets and a target feature to answer the question asked in the scenario in the case of supervised learning.

**[0 - 30]**

* Collect and develop a dataset based on the agriculture topic related to Ireland as well as other parts of the world. Perform a sentimental analysis for an appropriate agricultural topic (e.g., product price, feed quality etc…) for producers and consumers point of view in Ireland.

**[0 - 25]**

* You should train and test for Supervised Learning and other appropriate metrics for unsupervised/ semi-supervised machine learning models that you have chosen. Use cross validation to provide authenticity of the modelling outcomes. You can apply dimensionality reduction methods to prepare the dataset based on your machine learning modelling requirements.

**[0 - 30]**

* A Table or graphics should be provided to illustrate the similarities and contrast of the Machine Learning modelling outcomes based on the scoring metric used for the analysis of the above-mentioned scenario. Discuss and elaborate your understanding clearly.

**[0 - 15]**

**Total Mark = 30+25+30+15=100:(100%)**

## Assessment Requirements

Note ALL Students are required to use Git Classroom for any Assignments that they are working on. This assignments Git Classroom link is <https://classroom.github.com/a/4odn-Ai9>

This means that ALL changes must be committed to the assignments Git classroom during your assignment. (Not just a single commit at the end!) This is to allow you to display your incremental progress throughout the assessments, give you practice for your capstone/thesis, allows you to create an online portfolio that can be used to showcase your work and to ensure that there are no problems with final uploads (as all your work will be available on GitHub). It is expected that there will be a minimum of 10 commits (with many of you making very many more).

You may Only use your CCT email for your git account, private/work email-based accounts will not be accepted. You DO NOT NEED TO include your lecturer's CCT email as a collaborator on your submission as they have automatic access.

**NOTE As well as committing to the Git Classroom you must also upload your work to Moodle as usual for grading. Failure to do so (even if the files are available on Git) will result in no grades for those files .**

All assessment submissions must meet the following minimum requirements:

* Be submitted in the format outlined in the assignment summary table.
* 2000 (+/- 10%) words in report (not including code, code comments, titles, references, or citations)
* Report submission MUST be a word document only (No PDF’s!).
* Code in a Jupyter Notebook file (.ipynb) only but may be referenced in the word document.
* 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 Data Analysis**

A white text on a white background

AI-generated content may be incorrect.

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.* |
| * 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. * The use of generative AI tools is not permitted in this course for the following activities:   + Impersonating you in classroom context   + Generating code for your assignment   + Writing a draft of a writing assignment   + Writing entire sentences, code, 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. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | **Criteria 1** **Rationale and Justification for Model Choice** | **Criteria 2:**  **Dataset Development and Sentiment Analysis** | **Criteria 3:**  **Training and Testing with Machine Learning Models** | **Criteria 4: Interpretation and Explanation of ML Results** |
| **Weighting per Criteria** | **30 marks** | **25 marks** | **30 marks** | **15 marks** |
| **Excellent (+70%)** | Comprehensive and insightful justification, exceptional alignment to scenario. Sophisticated use of GridSearchCV and strong reasoning behind feature and target selection. | Comprehensive and well-curated dataset with insightful sentiment analysis that considers producer and consumer perspectives. | Thorough training/testing with robust results and metrics. Cross-validation and dimensionality reduction are applied expertly to meet modelling requirements. | Comprehensive and visually compelling table/graphics. In-depth analysis and clear understanding of modelling outcomes using scoring metrics. |
| **Very Good (60 - 69%)** | Clear rationale, strong alignment with scenario needs. GridSearchCV used effectively for hyperparameter tuning. Features and target chosen thoughtfully. | Dataset relevant and well-constructed. Sentiment analysis addresses agricultural topic with good interpretation. | Effective training/testing with detailed results. Cross-validation ensures authenticity. Dimensionality reduction used to refine dataset. | Clear and well-designed table/graphics showing meaningful comparisons. Good discussion of scoring metrics and outcomes. |
| **Good (50 - 59%)** | Moderate justification with some alignment to the scenario. Attempts multiple approaches with basic GridSearchCV implementation. | Dataset somewhat relevant and includes basic features. Sentiment analysis performed but lacks depth or focus. | Models are trained and tested with acceptable results. Some use of cross-validation and basic dimensionality reduction applied. | Table/graphics illustrate some comparisons, with basic discussion of outcomes and scoring metrics. |
| **Acceptable (40 - 49%)** | Basic rationale provided but lacks depth. Limited explanation of why models are suitable. Minimal attempt at hyperparameter tuning. | Dataset minimally relevant or poorly developed. Sentiment analysis is incomplete or poorly executed. | Minimal training and testing with poorly explained results. Limited cross-validation or dimensionality reduction. | Offers a basic interpretation of results with limited explanations. The report presents the findings but would benefit from a clearer connection to the dataset and more detailed interpretation. |
| **Fail (< 39%)** | No table or graphics provided. No discussion of outcomes. | Basic table or graphic provided but lacks clarity or insight. Minimal discussion of outcomes. | No training or testing conducted, or results are irrelevant. No use of cross-validation or dimensionality reduction. | Fails to interpret or explain the results. The report is disorganized, with no connection to the dataset or meaningful conclusions. |

**Grading Criteria Machine Learning for Data Analysis**

**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).