**SCCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | *MSc in Data Analytics* | | |
| **Cohort:** | *MSc in Data Analytics FT* | | |
| **Module Title(s)**: | *Advanced Data Analytics* | | |
| **Assignment Type:** | *Individual* | **Weighting(s)**: | *Advanced Data Analytics – 100%*  *Capped at 40%* |
| **Assignment Title:** | *MSC\_DA\_ ADAv5 Repeat* | | |
| **Issue Date:** | *April 2025* | | |
| **Due Date:** | *5th January 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 Advanced Data Analytics 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, Screencast for practical demonstration. Note that the maximum number of Jupyter Notebooks is 2* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After exam board May 2025* | | |

**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:

**Advanced Data Analytics**

3. Analyse a set of requirements to determine the type of Advanced Data Analysis for a particular problem set. Document and justify choices made to stakeholders and peers through insight gained from the process.(linked to PLO 4, PLO 5)

4. Develop a solution, reliant on temporal data (e.g., social media feed, sensor data) to solve a given problem set.(linked to PLO 1, PLO 2)

5. Critically assess the existing state of the art in Natural Language Processing and propose a strategy toward optimisation.(linked to PLO 1, PLO 2, PLO 4)

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 9 awards** |
| 70% + | 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 |
| 40 – 59% | Attains all the minimum intended programme learning outcomes |
|
| 35 – 39% | 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 |

The CCT Grade Descriptor describes the standard of work for grade boundaries summarised below. The full descriptor is available on Moodle.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |

**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:   + 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   + Completing group work that your group has assigned to you   + Writing a draft of a writing assignment   + 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. |

**Assessment Task**

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

**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/qemICDnr>

**You may not upload a PDF document for your report, It MUST be a word document. This is Not a Scientific Paper and must be formatted in standard report format including index, appendix, etc.**

In this continuous assessment, You are required to identify and carry out an analysis of a large dataset gleaned from the twitter API and Yahoo Finance and is available on Moodle as “stock-tweet-and-price.zip”

you are required to analyse the data and make a CLOSE price forecast of 1 day, 3 days, and 7 days for at least 5 of the companies using the tweets AND the financial price data.

**Context**

**Data in file “stocktweet.csv”**

Data collection period : Jan 2020 - Dec 2020

Number of Tweets : 10,000

It contains the following 4 fields:

* ids: The id of the tweet (eg. 100001)
* date: the date of the tweet (eg. 01/01/2020)
* ticker: The ticker value for the company (eg AMZN)
* tweet: the text of the tweet (eg. $AMZN Dow futures up by 100 points already ðŸ¥³)

**Data in folder “stockprice”**

Data collection period : Jan 2020 - Dec 2020

38 companies historical price data in csv format.

- Tickers:

'AAPL', 'ABNB', 'AMT', 'AMZN', 'BA', 'BABA', 'BAC', 'BKNG', 'BRK.A', 'BRK.B', 'CCL', 'CVX',

'DIS', 'FB', 'GOOG', 'GOOGL', 'HD', 'JNJ', 'JPM', 'KO', 'LOW', 'MA', 'MCD', 'MSFT', 'NFLX',

'NKE', 'NVDA', 'PFE', 'PG', 'PYPL', 'SBUX', 'TM', 'TSLA', 'TSM', 'UNH', 'UPS', 'V', 'WMT', 'XOM'

Each CSV file contains the following 7 fields:

* Date : the date of the stock price (eg. 01/01/2020)
* Open : the opening value of the stock price that day (eg. 123.33)
* High : the highest value of the stock price that day (eg. 125.45)
* Low : the lowest value of the stock price that day (eg. 121.54)
* Close : the closing value of the stock price that day (eg. 122.49)
* Adj Close : the adjusted closing value of the stock price that day (eg. 122.49)
* Volume : the number of stocks traded that day (eg. 100805600)

Following your analysis, you are then required to make a time series forecast of the CLOSE price **for at least 5 of the companies using the tweets AND the financial price data** at 1 day, 3 days and 7 days going forward. This forecast must be displayed as a dynamic dashboard.

Your project must incorporate the following elements:

* Explore at least 2 methods of time-series forecasting including at least 1 Neural Network and 1 autoregressive model (ARIMA, SARIMA etc…) . (Hint: that this is a Short time series, How are you going to handle this?)
* Use both the financial AND tweet data
* Evidence and justify your choices for your final analysis and include your forecasts at 1 day, 3 days and 7 days going forward.
* Your dashboard must be dynamic and interactive. Include your design rationale expressing Tufts principles.

**Deliverables:**

The results of the analysis must be presented in the form of a project report. This report should discuss the storage and processing of big data using advanced data analytics techniques. The report should be 2500 ± 10% words in length (excluding references, titles, and code) and must follow the Harvard styles format in addition to employing appropriate referencing methods and academic writing style. The report should include the following:

**Advanced Data Analytics**

1. A discussion of the rationale, evaluation, and justification for the choices you have made in terms of EDA, data wrangling, machine learning models and algorithms that you have implemented (This is particularly important as you must use both the tweets AND the Stock Price)**.[0-40]**
2. **E**valuation and justification of the hyperparameter tuning techniques that you have used **[0-20]**
3. Your analysis of the data and your forecast of the CLOSE price for at least 5 of the companies using the tweets AND the financial price data at 1 day, 3 days and 7 days going forward**[0-20]**
4. Presentation of results by making appropriate use of figures along with caption, tables, etc and your dashboard for your forecast, Discuss Tufts Principles in relation to your Dashboard **.[0-20]**

**SUBMISSION:**

**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:

* 2500 words +- 10% (excluding references, titles, citations and quotes)
* Word Document for report (No PDF’s), Jupyter notebook for code, Screencast for practical demonstration.
* 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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criteria** | **70+ (Excellent)** | **60-69 (Very Good)** | **50-59 (Good)** | **40-49 (Pass)** | **<=39(Fail)** |
| **1. Rationale, Evaluation & Justification for EDA, Data Wrangling, ML Models using Tweets & Stock Price [0-40]** | Comprehensive and insightful discussion of all stages; strong justification for EDA, wrangling, and model choices; excellent integration of tweets and stock price data. | Clear and well-structured discussion of processes; good justification for methods; effectively uses both data sources. | Basic discussion with limited justification; some use of tweets and stock data but lacks full integration or depth. | Minimal rationale or unclear explanation; weak or incomplete use of both data types. | No or poor discussion; fails to use both data sources meaningfully. |
| **2. Evaluation & Justification of Hyperparameter Tuning Techniques [0-20]** | Detailed evaluation with strong rationale for chosen techniques; demonstrates clear understanding and optimization strategy. | Good explanation of techniques with relevant justification; shows understanding of tuning importance. | Basic explanation of tuning methods; justification lacks depth or clarity. | Limited discussion; unclear or unjustified tuning strategy. | No hyperparameter tuning or explanation provided. |
| **3. Analysis & Forecast of CLOSE Price (5 Companies, 1/3/7 Days Using Tweets & Financial Data) [0-20]** | Thorough and accurate forecast analysis with strong integration of both data types; clear and insightful multi-day predictions for all companies. | Good analysis with accurate forecasts; uses both data types; minor gaps in prediction depth or explanation. | Basic analysis and forecast; may lack clarity or consistency; moderate use of tweets and financial data. | Limited forecast or weak use of data; unclear predictions. | No meaningful forecast or fails to use data sources effectively. |
| **4. Presentation of Results: Visuals, Dashboard, Tuft Principles Discussion [0-20]** | Highly effective visuals and dashboard; excellent use of figures/tables with captions; thoughtful discussion of Tuft’s principles with clear application to dashboard design. | Good use of visuals and dashboard; mostly clear and relevant; sound discussion of Tuft’s principles. | Basic visuals and dashboard; Tuft’s principles discussed but not well-integrated; minor clarity issues. | Limited visuals or dashboard clarity; minimal reference to Tuft’s principles. | No clear visuals or dashboard; no discussion of Tuft’s principles. |

**Additional Information**

* Lecturers are not required to review draft assessment submissions.
* 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 requested by contacting your lecturer AFTER the publication of results,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 or access the [CCT Learning Space](http://learningspace.cct.ie/subjects/index.php).
* For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](https://moodle.cct.ie/mod/forum/view.php?id=55148)
* 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).

**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. |