**CCT College Dublin**

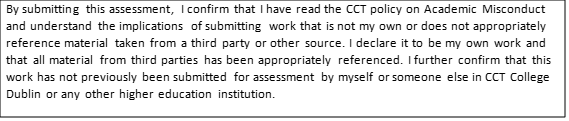
**Assessment Cover Page**

Form

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning for Business |
| **Assessment Title:** | CA1 Project |
| **Lecturer Name:** | Dr. Muhammad Iqbal |
| **Student Full Name:** | Rone Soares |
| **Student Number:** | 2023322 |
|  |  |
|  |  |
|  |  |
|  |  |
| **Assessment Due Date:** | 24-04-2024 |
| **Date of Submission:** | 29-04-2024 |

Form

**Declaration**



[Introduction: 2](#_Toc647214344)

[Objectives: 3](#_Toc1651477923)

[General Goal: 3](#_Toc1647198617)

[Success Criteria/Indicators: 3](#_Toc1086756879)

[Technologies Used: 3](#_Toc1425489756)

[Libraries: 3](#_Toc2050993090)

[Ethical Considerations: 4](#_Toc826538904)

[Part 2- 4](#_Toc1411902443)

[Which clustering algorithms would you consider for segmentation, and why? Explain the differences between silhouette score and Davies-Bouldin index in the context of clustering. Compare the results obtained from any two clustering algorithms from the chosen dataset. 4](#_Toc796760746)

[Kmeans: 4](#_Toc746755168)

[Hierarchy or clustering: 4](#_Toc193281493)

[Explain the differences between silhouette score and Davies-Bouldin index in the context of clustering. 4](#_Toc1483025749)

[Conclusion: 4](#_Toc1852161007)

Introduction:

Analysis and forecasting of Apple's stock prices represent a complex challenge of great relevance for investors and financial analysts. This project aims to explore and apply advanced data analytics and machine learning techniques to address key questions related to Apple's stock prices over time.

The financial market is dynamic and highly volatile, with stock prices reflecting a complex interplay of economic, political, and social factors. Understanding and predicting the patterns of change in Apple's stock prices is critical to making informed investment decisions and mitigating financial risks. This project arises from the need to develop robust and effective methods for analyzing and predicting Apple stock prices based on historical data and relevant market indicators.

Description of the problem domain. The problem domain encompasses financial time series analysis and the application of machine learning models to predict Apple's stock prices. This includes identifying trends, seasonal patterns, and anomalies in historical data, as well as developing accurate and reliable forecasting models.

Objectives:

The main objectives of this project are:

· Identify patterns of change in Apple stock prices over time.

· Develop machine learning models to predict future movements of Apple stock prices, evaluate the performance of prediction models, and highlight their strengths and limitations.

# General Goal:

The main objective of the final work is to establish a standard that identifies with a high degree of assertiveness the forecast price of Apple’s stock market, using the best machine learning model.(ARIMA)

# Success Criteria/Indicators:

Be able to achieve the proposed objective 1, 2 and and evaluate the degree of accuracy of the chosen machine learning technique.

Technologies Used:

All the work was carried out in the Python language, using the jupyter platform for the creation and execution of codes and commands. The original database has the CSV format, "comma separeted value".

# Libraries:

For this step I will only use many libraries and the ARIMA machine learning model.

# Ethical Considerations:

My dataset of Stock Market was extracted from [Apple Inc. (AAPL) Stock Price, News, Quote & History - Yahoo Finance](https://finance.yahoo.com/quote/AAPL)

## Part 2-

### Which clustering algorithms would you consider for segmentation, and why? Explain the differences between silhouette score and Davies-Bouldin index in the context of clustering. Compare the results obtained from any two clustering algorithms from the chosen dataset.

I did analysis using the Kmeans and hierarchical clustering algorithms, where both indicated the presence of 4 clusters as being an appropriate choice, we can consider these 2 algorithms as options for segmentation. The reasons why I considered these algorithms were as follows:

### Kmeans:

* Simplicity: Kmeans is a simple and easy-to-understand algorithm, which can make it easier to interpret the results.

Computational efficiency: It is computationally efficient and works well with large data sets, such as Apple's stock data.

Interpretable results, the clusters formed by Kmeans tend to have a convex shape, which can make it easier to interpret and explain the results.

### Hierarchy or clustering:

* Flexibility: Hierarchical clustering allows for the creation of dendrograms, which can be useful for visualizing the hierarchical structure of clusters and the relationship between them.

No need to specify the number of clusters, unlike Kmeans, you don't need to specify the number of clusters in advance, which can be an advantage if you don't know a priori how many clusters to expect.

* Robustness: Hierarchical clustering can be more robust with respect to the choice of different distance metrics and connection criteria, allowing for a more detailed analysis of the data structure.

### Explain the differences between silhouette score and Davies-Bouldin index in the context of clustering.

In summary, while the silhouette coefficient measures how well each data point fits into its assigned cluster compared to other clusters, the Davis Bouldin index measures the average of the similarities between clusters, where similarity is the ratio of the distance within the cluster and the distance between clusters. So while both are used.

# Conclusion:

Based on the results obtained from this comprehensive analysis of Apple's stock prices, I was able to draw several important conclusions about the patterns of variation, trends, and seasonality over time.

Firstly, exploratory analysis of the data revealed upward trends in Apple's stock prices. Indicating an overall increase in the value of the shares of 218% over the analyzed period. In addition, seasonal patterns and market cycles were identified that influenced price movements, reflecting the product launch cycle and seasonal fluctuations in sales. Anomaly detection also provided valuable insights into what may have impacted stock prices, such as unexpected product announcements and financial reports.

By applying machine learning techniques like clustering, we can segment the data into distinct groups and develop accurate prediction models. Clustering algorithms have identified natural groupings in the data, providing a deeper understanding of the underlying structure of stock prices. In turn, the ARIMA model demonstrated its ability to capture patterns of variation and predict future movements of the stock price, highlighting its effectiveness as a forecasting tool.

Importantly, this study contributes significantly to the understanding of the stock market and informed decision-making in the field of investing. The analyses carried out offer valuable insights for those who operate in the financial market, allowing a better understanding of the patterns of variation in stock prices and facilitating the formulation of more effective investment strategies.

It is important to recognize the limitations of this study, while I have explored a variety of analytical techniques and forecasting models, there is always room for improvement and refinement, as well as developing more advanced machine learning models for stock price prediction.

Ultimately, this work represented a significant step forward in understanding Apple's stock prices and highlights the potential of advanced analytical techniques to inform investment decisions and drive financial success.