**CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning for Business |
| **Assessment Title:** | CA1 Project |
| **Lecturer Name:** | Dr. Muhammad Iqbal |
| **Student Full Name:** | Daniela Mariano Barreto |
| **Student Number:** | 2023278 |
| **Assessment Due Date:** | 21/04/2024 |
| **Date of Submission:** | 21/04/2024 |

**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. |

Table of Contents

[**Introduction** 3](#_Toc162462504)

[Motivation 3](#_Toc162462505)

[Problem Domain and objectives 3](#_Toc162462506)

[Why this dataset? 3](#_Toc162462507)

[Word count 3](#_Toc162462508)

[**Clustering Algorithms** 4](#_Toc162462509)

[Ttttt1 4](#_Toc162462510)

[Tyyy2 4](#_Toc162462511)

[**ARMA / ARIMA model – Time Series Data** 4](#_Toc162462512)

[**Assessment** 5](#_Toc162462513)

[**Conclusion** 5](#_Toc162462514)

[**References** 5](#_Toc162462515)

# **Introduction**

A global trend in pedagogical approach, known as Massive Open Online Courses (MOOCs), serves as an alternative and supplement to traditional models of learning by utilizing online platforms (Sharma, 2013, p.19). It provides high-quality educational content that students worldwide can access more quickly and easily (Thuy et al., 2023, p.1-2).

Since 2012, numerous platforms have emerged for online education, among which Coursera stands out as the most popular due to its variety of courses and strong partnerships with prestigious higher educational institutions like MIT, Harvard, and Stanford (Sharma, 2013, p.19; Zotova et al., 2021, p.167).

## Motivation

During the pandemic period, MOOCs experienced a significant increase (Serravallo, 2020, p.1). This market is estimated to reach USD 22.8 billion this year, with an expectation of USD 119 billion by 2029 (www.mordorintelligence.com, n.d.), making it a lucrative investment opportunity. According to this source, the Coursera platform experienced a 640% increase during the pandemic compared to the previous period.

## Problem Domain and objectives

With the modernization of the educational system with MOOCs, platforms such as Coursera developed and sought to serve the most diverse areas of knowledge, signalling promising investment prospects.

In this context, this project seeks to compare two clustering algorithms (DBSCAN and OPTICS) to evaluate Coursera's course area, focusing on student satisfaction rates, which could enable recommendations on areas of study to enrich online learning. Analysis of stock market for Coursera using the ARIMA algorithm will also be covered.

## Data description

The dataset used is from the Kaggle repository (Elvin, 2024, p.1) and contains records of various courses available on the Coursera platform with attributes like 'rating' and 'keywords', which are aligned with the project purpose (Figure 1). The time series data is from the stock market of Coursera and was gathered from Yahoo Finance. It can also provide information about the actual financial situation of this educational platform.

Time series data: [Coursera, Inc. (COUR) Stock Price, News, Quote & History - Yahoo Finance](https://finance.yahoo.com/quote/COUR/)

**Figure 1:** Data dictionary based on the repository information.

A close-up of a document

Description automatically generated

## Word count

# **Clustering Algorithms**

Unsupervised learning algorithms can be characterized by their ability to make predictions when the target label is unknown; the learning process depends exclusively on resources/features (Müller and Guido, 2017, p.131). Within this framework, clustering algorithms separate objects (data points) based on specific criteria/algorithms, typically by similarity. Furthermore, this technique is useful in various business areas, including recommendations and segmentation (Avinash Navlani, Fandango and Idris, 2021, p.325).

The Silhouette score and Davies-Bouldin index (BDI) are internal performance metrics used to evaluate the quality of clustering results (Avinash Navlani, Fandango, and Idris, 2021, p.350). In clustering methods, a high score does not necessarily indicate effective clustering results, as is the case with BDI, which evaluates the compactness and separation of clusters based on the ratio of intra-cluster distance to inter-cluster distance. Better clusters are indicated by a lower BDI, while the Silhouette score measures how well-separated the clusters are, and high values suggest better clustering results (Avinash Navlani, Fandango, and Idris, 2021, p.351).



* 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.

OPTICS + DBSCAN

## Ttttt1

## Tyyy2

# **ARMA / ARIMA model – Time Series Data**

What insights can you derive from the initial exploration of the time series data based on the provided topics? Describe any trends, seasonality, or anomalies observed. How did you determine the appropriate parameters (p, d, q) for the ARIMA model. Evaluate the performance of the ARIMA model in forecasting future values, highlighting any strengths and limitations based on your chosen dataset.

# **Assessment**

# **Conclusion**

# **References**

Thuy, T., Thanh Tu Tran, Gia Khuong An and Nguyen, P. (2023). Students’ Perception Towards Learning Massive Open Online Courses on Coursera Platform: Benefits and Barriers. *International Journal of Emerging Technologies in Learning (ijet)*, 18(14), pp.4–23. doi:https://doi.org/10.3991/ijet.v18i14.39903.

Sharma, R.C. (2013). How to Offer a Massive Open Online Course (page 19-20). *EduComm Asia*. [online] Available at: https://www.academia.edu/5505040/How\_to\_Offer\_a\_Massive\_Open\_Online\_Course\_page\_19\_20\_. [Accessed 16 Mar. 2024].

Zotova, M., Likhouzova, T., Shegai, L. and Korobeynikova, E. (2021). The Use of MOOCS in Online Engineering Education. International Journal of Engineering Pedagogy (iJEP), [online] 11(3), p.157. doi:https://doi.org/10.3991/ijep.v11i3.20411. [Accessed 16 Mar. 2024].

Serravallo, J. (2020). Connecting with students online : strategies for remote teaching & learning. Portsmouth, New Hampshire: Heinemann.

www.mordorintelligence.com. (n.d.). Massive Open Online Course (MOOC) Market | Growth, Trends, and Forecasts (2020 - 2025). [online] Available at: <https://www.mordorintelligence.com/industry-reports/massive-open-online-course-mooc-market>. [Accessed 17 Mar. 2024].

Avinash Navlani, Fandango, A. and Idris, I. (2021). Python Data Analysis. Packt Publishing Ltd.

Müller, A. C. and Guido, S. (2017). Introduction to machine learning with Python: a guide for data scientists. 1st ed. United States of America. O’reilly Media.

Elvin, R. (2024). Coursera Dataset. [online] Available at: https://www.kaggle.com/datasets/elvinrustam/coursera-dataset. [Accessed 12 Feb. 2024].