CA1 Machine Learning for Business

**Performing Clustering Algorithm for Education Student Performance and Investigate how Covid 19 pandemic affect the stock market of hotel in 2020 and 2021**

*By*

*Rachel Mae Lademora – 2023396*

Higher Diploma in Science in Data Analytics for Business

Machine Learning for Business

**Dr. Muhammad Iqbal**

**CCT College**

**Dublin, Ireland**

**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Machine Learning for Business |
| **Assessment Title:** | CA1 |
| **Lecturer Name:** | **Dr. Muhammad Iqbal** |
| **Student Full Name:** | Rachel Mae Lademora |
| **Student Number:** | 2023396@student.cct.ie |
| **Assessment Due Date:** | April 23, 2023 |
| **Date of Submission:** | April 23,2023 |

**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

Business Understanding**1**

Data Understanding2

Data Preparation3

Modelling**4**

Evaluation5

References6

**Introduction**

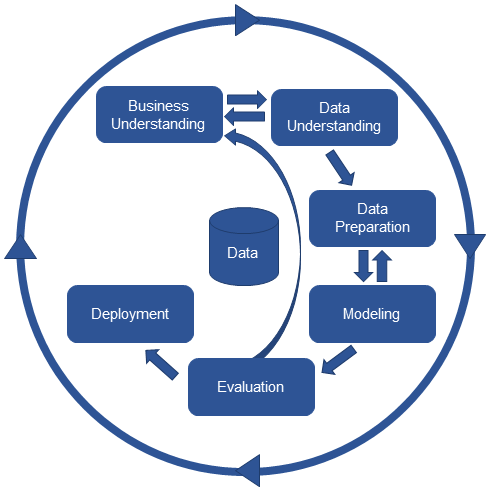
Nowadays, education is the leading way in order for a person can have a better future. Subjects scores examination in education will give the idea if the student performs well in school especially in class. In every test score of each subject, it will determine how student is efficient to handle the progress of every test that they have in the school, this will help the teacher to properly assigned the student if they get a lower or higher score after the test. (Rasul and Bukhsh, 2011).

One of the crucial topics in stocks is determine the influence of the stock prices on hotel segment. This can be related to the inflation of the stock price close that will get a huge impact for the stability of the stocks.

During the end of the month of 2019, The Covid 19 pandemic was faced a wide problem in the whole world. And it causes all of the economic stability of each country in the world get affected. The worldwide pandemic affects in the hotel industry to make the stocks goes down.

The report aims to have a detail visualization of the clustering algorithms. Using these three subjects I can see which machine learning clustering algorithm is best for mitigation of measure the similarity and dissimilarity between different clusters, it will show and compare which clustering segmentation is better to use for the student performance dataset. Furthermore, this report will include the analysis of stock market inflation rate impact during the pandemic of covid 19.

Lastly, in order to properly explain and better understand the report, it will follow the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to explain in detail the stages of Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment.



**Business Understanding**

This phase is vital to understand the business objectives and the appropriate qualification requirements for this project.

The principal objective of this project is to assess the implementation of clustering algorithms and time series autoregressive integrated moving average (ARIMA) model to forecast and identify the seasonality, trend, and residuals in the stocks market data analysis that can make a forecast for the future value using the yahoo finance hotel stock dataset. The hypothesis of this project is to determine which clustering algorithms is best for segmentation, also evaluate the accurate parameters (p,d,q) for (ARIMA) model.

**The requirement of this projects are:**

* Execute and determine appropriate clustering algorithm to be consider for segmentation in the dataset.
* Implement and identify the difference of silhouette score and Davies-Bouldin index using the factors of clustering.
* Determine which clustering algorithm can measure the similarity and dissimilarity of different cluster.

General Goal:

This project aims to equate the results of the of two clustering algorithms from the student performance dataset. Clearly show the inflation rate when the hotel stock market is having a huge impact of inflation rate.

The list of important tools and technologies that will be used in the project are:

* List of the Python Libraries:

1. Pandas
2. Numapy
3. Matplotlib
4. Seaborn and etc.

**Data Understanding**

This second phase of the CRISP-DM methodology, it is very important to look every detail of the data. This will lessen to face possible errors in proceeding to data preparation.

When I use the df.head() function in loading the dataset it shows me the first 5 rows and 8 columns this is clearly visualize in figure 1.

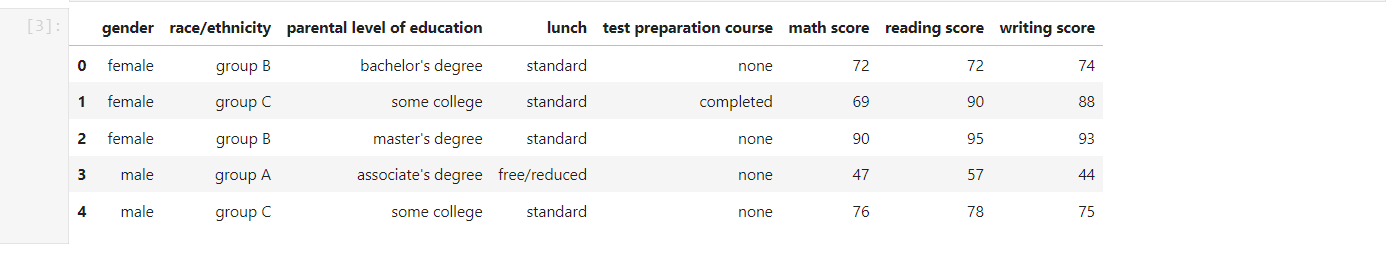


Figure 1 Head of the student performance dataset.

References

Rasul, S. and Bukhsh, Q. (2011). A study of factors affecting students’ performance in examination at university level. *Procedia - Social and Behavioral Sciences*, 15, pp.2042–2047. doi:https://doi.org/10.1016/j.sbspro.2011.04.050.