A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Alina Marrulyn Lozano Flores |
| *Student Number* | 2024068 |
| *Module Title* | Data Visualization Techniques, Machine Learning for Business |
| *Assessment Title* | CA2 |
| *Assessment Due Date* | 24th November 2024 |
| *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.

Abstract

[NOTE: This section is designated for the abstract. Abstracts are not assigned page numbers and should precede the table of contents. If an abstract is unnecessary for your work, please delete this page.]

Attention: All notes must be removed from the document before submission!!

Contents

[Introduction 1](#_Toc158384946)

[Chapter 1 1](#_Toc158384947)

[Chapter 1.1 1](#_Toc158384948)

[Chapter 1.1.1. 1](#_Toc158384949)

[References 2](#_Toc158384950)

[NOTE: The table of contents above has been included for your convenience. To refresh the table, simply click on it, then select 'Update Table' using the mouse. You can choose to update either the page numbers exclusively or the entire table as needed.]

# Introduction

1. Discuss and explain the purpose of a recommendation system for online retail business in machine learning. Briefly compare Content and Collaborative filtering using any dataset of your choice (Datasets used in the class tutorials or exercises are not allowed to use in this CA2). Train and test machine learning models for the user-user and item-item collaborative filtering. Justify your recommendations for the considered scenario by providing a conceptual insight.

In the point 1 I have to use the dataset with clothes and reviews

(60 marks)

2) Perform Market Basket Analysis on the chosen dataset by using Apriori and FP growth algorithms. Can you express major similarities/ divergence between these models? Compare and contrast the machine learning results obtained based on both algorithms.

Market basket Analysis only uses information of the products,there is not dataset yet

(40 marks)

3) Create an interactive Dashboard aimed at younger adults (18 - 35 years) with specific features to summarise the most important aspects of the data and identify through your visualisation why this dataset is suitable for Machine Learning models in an online retail business. Explain how your dashboard is designed with this demographic in mind.

(70 marks)

4) Discuss in detail your rationale and justification for all stages of data preparation for your visualizations.

(30 marks)

2) Perform Market Basket Analysis on the chosen dataset by using Apriori and FP growth algorithms. Can you express major similarities/ divergence between these models? Compare and contrast the machine learning results obtained based on both algorithms.

Association analysis know also as market basket analysis, is an model algorithm usually used to find association between products, commonly used in supermarkets.

This is an unsupervised model, analyses the relation between independent variables to have a new output. The independent variables are categorical/ discrete instead numerical/continuous,

Association analysis starts finding combinations called frequent *itemsets*, this combination could be two or more items. After this you can find the support of each itemset, this number indicates how often these products appear toguether. Support is usually referred as ‘SUPP’ and itemset as ‘X’, and besides we have ‘T’ as the total of transactions or data points. For example if our itemset only appears two times and we have a total of 10 transactions, our formula is:

SUPP = X/T

SUPP = 2/10

SUPP= 0.2

Not all the transactions are useful for the analysis, thus we can discard the transactions with low support that are not relevant, to do this we need to set up a minimal support or minsup, for example to choose that only the items that appear in at least 40% of our transactions are relevant.

The next step is Rule Generation, with statements as *if* and *then,* and we can calculate the level of Confidence that measures how often the rule generated is true. `1q

# References

Refer to my BIB website CA2