**UNIVERSITY OF WESTMINSTER**

**Msc Big Data Technologies**

Module Code: 7BUIS008W.2

Module Title: Data Mining and Machine Learning

Module Leader: Dr. Panagiotis Chountas

Module Tutor: Dr. Natalia Yerashenia

Assignment: Coursework 1

Submission Date: 23 Mar. 2022

Student ID: W1848650

Student Name: Mohammadamin Shams

Table of Contents

[1 Abbreviations 3](#_Toc125309923)

[2 Radio Listeners Analysis for Retail purposes 3](#_Toc125309924)

[2.1 Data Analysis 3](#_Toc125309925)

[2.1.1 Understanding The Data and Pre-processing 3](#_Toc125309926)

[2.2 Market Basket Analysis 5](#_Toc125309927)

[2.2.1 Change Format of the Data 5](#_Toc125309928)

[2.2.2 Apriori Algorithm 5](#_Toc125309929)

[2.2.3 Business insights 6](#_Toc125309930)

[3 References 7](#_Toc125309931)

# Abbreviations

|  |  |
| --- | --- |
| MBA | Market Basket Analysis |

# Radio Listeners Analysis for Retail purposes

In this part we want to talk about the second task of the coursework 1. In this part we have a data set that contains data about radio listeners from all over the world and the artists that they like to listen.

In this part the goal is to find the frequent artists and groups that people listen to and after finding the frequent item set with the minimum support provided create baskets to offer to listeners for marketing purposes this objective can be done with the help of algorithms like Apriori that perform market basket analysis.

Market basket analysis is a modelling technique that is used by retailers to increase sales by better understanding customer purchasing patterns. It is based on this assumption that if a person buys a certain set of items, they are likely to buy another set of items. In order to this analysis, we need the history of the purchased items (SearchCustomerExperience. 2022).

To perform MBA and understand the pattern of artist that people buy their music Apriori algorithm is used.

In the next section first, we understand the data and pre-process and cleanse it before the Apriori and then find out the frequent items and after that we create rules based on frequent itemset.

## Data Analysis

To perform MBA first we need to understand the data in order to do that pandas profiling is used and other codes to do the first part of pre-processing.

### Understanding The Data and Pre-processing

The raw data set contains four variables and 289955 observations. The first value gives us the user id and the next one gives artist that they listen and the other one is the listeners sec and the fourth is the country of the listener. We have 1004 distinct artist and 15000 distinct users. About the 70 percent of the listeners are male and 30 percent is female. The most transaction contain four artist and they are Radiohead and the Beatles and Coldplay and red-hot chili peppers. The large amount of people is belonged to United States and Unitec Kingdom and Germany.

Graphical user interface, application

Description automatically generated

Figure 9– Frequent countries

Graphical user interface, application

Description automatically generated

Figure 10– Frequent artists



Figure 11– First rows of the data

The data type of the 4 variables is listed below in fig 12.

Text

Description automatically generated

Figure 12– Data types

#### Missing Values and Duplicates

In the data as can be seen in figure 10 we don’t have any missing values nut on the other hand we have two duplicates, and we need to remove them before analysis.

Graphical user interface, application

Description automatically generated

Figure 13– Duplicates

## Market Basket Analysis

### Change Format of the Data

To do Apriori method on our data we need to change our dataset from only transaction to one hot-encoded table or create a list for each user because Apriori algorithm can work for these inputs. The one hot-encoded table has all the artist as columns and all the users as rows and if a user listens to a artist we put one for that value and for others that he doesn’t listen we put zero and this is done for all the users.

A screenshot of a computer

Description automatically generated with low confidence

Figure 13–Hot-encoded table

### Apriori Algorithm

The next part is performing the Apriori algorithm on the hot-encoded table. Apriori is the most established algorithm for finding frequent items. The basic principle that Apriori uses is that any subset of a frequent itemset must be frequent and based on these frequent itemset we create association rules. In order to find the frequent itemset we need to run the Apriori with minimum support of 0.08 which the outcome is not good enough for creating rules so as Professor suggested we change it to 0.03 in order to find more frequent items and also get not only single frequent items but also has more than one value, in our case we only get frequent items with length of two.

Text

Description automatically generated

Figure 14–Hot-encoded table

Then we need to create our rules for that we put the minimum threshold for confidence at 40 percent level and order our rules by their lift. In this order having these two together and ordering them gives us the three top rules at the above of our association rule.

A screenshot of a computer

Description automatically generated with medium confidence

Figure 15–Top three rules

### Business insights

With the selected parameters for minimum support and confidence level Apriori created 7 rules and we can see top three in the fig 15. We could interpret this rule in the way if listener heard antecedent then consequent is predicated that listener listens to it by the level of support and confidence and lift in the table.

If we change the minimum support and confidence, we can have more rules but with these rules we can suggest that the radio stations should give advertisement based of rules created for example lets only consider top three rules. First rule implies that if someone listens to led zeppelin its probably want to listen to pink Floyd too and we should suggest this item for those who listen to led zeppelin and for the other two rules is the same of the first rule that we explained.

# References

SearchCustomerExperience. 2022. *What is market basket analysis? Definition from WhatIs.com*. [online] Available at: <https://www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis#:~:text=Market%20basket%20analysis%20is%20a,likely%20to%20be%20purchased%20together.> [Accessed 23 March 2022].