**THE QUEST TO ANALYZE CUSTOMERS PURCHASING TRENDS THROUGH MARKET BASKET ANALYSIS?**

**OFM3 — OFM3 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS**

**DATA MINING II — D212**

**PRFA — OFM3**

**STUDENT NAME: IBRAHIM SULEIMAN**

**DATA ANALYTICS**

**STUDENT NUMBER: 001429984**

**COLLEGE OF INFORMATION TECHNOLOGY, WESTERN GOVERNORS UNIVERSITY**

**NAME OF INSTRUCTOR: KEIONA MIDDLETON**

**NOVEMBER 20th, 2022**

**Part 1: Research Question**

Section A: Description of the Report

Section A1: The quest to analyze customers purchasing trends through market basket analysis?

This can be resolved via market basket analysis and is pertinent to a situation involving a firm in the real world.

By predicting their shopping habits, "market basket analysis" helps businesses better understand and ultimately serve their customers. Market basket analysis, which is used in recommendation systems, is the most common application of artificial intelligence that customers encounter.

We'll go over the processes needed to implement a market basket analysis system as well as the mechanics of market basket analysis.

Section A2: Goal of the data analysis

Market basket analysis seeks to comprehend consumer behavior by establishing links between the things that consumers purchase. The goal is to perform a market basket analysis to look at customer information and discover crucial connections between consumer purchases, allowing for more efficient tactical and corporate judgment (Jyotsna Vadakkanmarv. 2021).

Part II: Method Justification

Using the market basket data mining technique, we can examine what customers purchase, how and why they make their purchases, and what they purchase collectively.

Section B1: How Market Basket Analyzes Teleco Dataset

Even after three decades, market analysis is still a useful tool for learning about the telecoms and eCommerce industries.

Increasing market share: Once a company hits its peak growth, it might be challenging to find new strategies to do so. Market basket analysis, which combines demographic and gentrification data, can be used to pinpoint the locations of new businesses or the geographic focus of geo-targeted marketing. For instance, market basket analysis can probably explain why there is a BestBuy electronic superstore there if you've ever wondered why it exists.

Understanding consumer behavior patterns is one of the fundamental tenets of marketing. We will pivot the products to the columns and use one-hot encoding to first encode the data in a way that will facilitate analysis.

The Apriori algorithm will then be used to trim the results and look at the top associations. In order to restrict sets, Apriori will first consider the frequency of individual items, use that frequency to further limit larger sets depending on the frequency of smaller sets. This enables us to develop a useful set of associations to examine (GeeksforGeeks, 2020).

Section B2: Transaction in the dataset

The following transactions are included in the dataset: TP-Link AC1750 smart WiFi router, Apple Pencil, and Apple Lightning digital AV adapter. Based on the initial purchase, we might infer that they own an Apple product. As a result, seeing the Apple pencil at the conclusion is consistent with the notion that they own an Apple device and would likely acquire several Apple products at once.

Section B3: Assumption

Market basket analysis' basic tenet is that when two or more products are present in the majority of baskets, it means that they are complementary in terms of consumption (if not purchase), and that buying one will lead to buying the others. Market basket analysis makes decisions, as we'll see later.

Market basket analysis bases its findings on the fundamental premise that there are connections or linkages between the products. To gauge these correlations and identify the most likely sets of products that are frequently bought together, we employ a variety of indicators (Kamakura, 2012)

An essential concept in the Apriori algorithm is the antimonotonicity of the support measure. It assumes that every subset of a frequently occurring item set must also be often occurring. Similar to this, any infrequent item set must likewise have all of its supersets be infrequent.

As a first step, make a frequency table of each item that appears in each transaction.

Step 2: Since only those factors are important, we can infer that the support is more than or equal to the threshold support.

Step 3: The next step is to make all possible pairs of the important elements, bearing in mind that the order doesn't matter because AB and BA are the same.

Step 4: At this point, we'll tally the instances of each pair over all transactions.

Step 5: Once more, only those item sets that surpass the support criterion are noteworthy.

Part III: Data Preparation

Data Preparation Steps

The data must be accessible before the analysis can be done. Making ensuring that none of the columns have any blank columns is the first step. The next step is to confirm that none of the data in the columns are duplicates. We'll also check to see if there are any duplicate columns or rows because we want to be sure there aren't any (false).

The dataset contains several factors that were determined to be useless for the logistic analysis, such as customer demographics that are linked to the location and interaction of the consumer and cannot be modified; as a result, such columns should be removed.

Working with the data is now easier as a result. Any (yes or no) or other alternative category choices cannot be utilized until the categorical variables have been converted to numerical values. The survey columns also need to be renamed to provide a clearer understanding and determination of applicable factors (Peter Grant, 2019).

Section C1: Transformed dataset

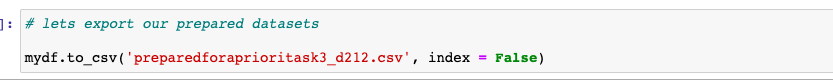
We have already removed the empty records. Next we will use one hot encoding to move products to the columns. This will prepare us for Apriori and Association Rules in the following steps.



Copy of the Cleaned Dataset

The cleaned dataset was provided as csv file in the submission named; We exported our prepared dataset as





Section C2: code used to generate association rules with the Apriori algorithm

Graphical user interface, text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Section C3: Values of the associations rules table

Graphical user interface, text, application, email

Description automatically generated

Section C4: Rules

To extract frequent item sets, association rule learning use the well-known Apriori technique. A database that contains transactions, such as purchases made by customers of a store, is what the apriori approach is designed to deal with. An item is considered "frequent" and added to a frequent itemized list if it reaches a user-specified support level.

Top three rules are as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Support | {Screen Mom Screen Cleaner kit, Dust-Off Compr... | {Screen Mom Screen Cleaner kit} | {Dust-Off Compressed Gas 2 pack} | 0.023997 | 0.370370 | 3.107548 |
| Confidence | {FEIYOLD Blue light Blocking Glasses, Dust-Off... | {FEIYOLD Blue light Blocking Glasses, Nylon Br... | {Dust-Off Compressed Gas 2 pack} | 0.003266 | 0.576471 | 4.836807 |
| Lift | {0, Anker 2-in-1 USB Card Reader, FEIYOLD Blue... | {Anker 2-in-1 USB Card Reader} | {0, FEIYOLD Blue light Blocking Glasses} | 0.003999 | 0.271493 | 8.261544 |

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, email

Description automatically generated

Part IV: Data Summary and Implications

Section D1: Summary of the significance of support, lift and confidence

SUPPORT - Event 1 union event 2 probability, likelihood of a buyer purchasing a desktop computer from an online store if they purchased the Screen Mom Screen Cleaner kit Our highest support looks to be 0.023997, or 2.4% of all transactions.

CONFIDENCE - Probability of one event given the likelihood of another. likelihood that someone will purchase a FEIYOLD Blue light Blocking Glasses and Dust-Off Should they buy a desktop computer. Our highest level of confidence is 0.576471, which indicates that buying FEIYOLD Blue Light Blocking Glasses with Dust-Off appears to occur 58% of the time.

LIFT : Tells us whether or not our presumption that the items have "no relationship"—that is, that they are independent—is accurate. The possibility that the consequent will also appear in a transaction is therefore increased if lift is greater than 1. If lift is less than 1, the possibility of the consequent is decreased by the antecedent. If lift is 1, the likelihood of purchasing the consequent is unaffected by the antecedent. Our top lift is 8.261544, which indicates that there is a high likelihood that the item will be purchased in combination with another item.

Section D2: Significance of the finding

By using consumer data collecting, we were able to discover trends in this study that will improve the effectiveness of marketing and sales initiatives.

Because market basket research enables us to find goods that customers want to buy, the results assist sales and marketing teams to develop more effective product placement, pricing, cross-sell, and up-sell tactics.

The practical value of this report demonstrates that market basket analysis may be used to identify associations between item purchases and that a business might use this sort of study to identify associations between purchases. The information might then be used to inform business decisions.

Section D3: Recommendation

In order to analyze the trend of customer transactions and better understand our consumers' buying habits, it is advised to locate all of the association rules from the provided data. This is because the idea that purchasing one set of products increases your likelihood of purchasing another group of goods.

The action a business could take as a result of this study is to identify goods that are frequently purchased together, such a screen cleaner kit and a Dust-Off compressed gas pack, and to provide a sale or discount when a consumer buys both. We already know from the study that it is likely that a consumer will buy both things, but providing a discount for doing so will probably stimulate even more purchases.

Part V: Attachments

Section E: Panopto video presentation

You can view the session using the following link:  
<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=d8e4b9c7-662b-4fef-9767-af57016f21cd>

Sources

Raschka, S. (n.d.). Apriori - mlxtend. Mlxtend. Retrieved November 10, 2021, from <http://rasbt.github.io/mlxtend/user_guide/frequent_patterns/apriori/>

GeeksforGeeks. (2020, April 4). Apriori Algorithm. Retrieved November 10, 2021, from <https://www.geeksforgeeks.org/apriori-algorithm/>  
  
Kamakura, W. A. (2012). Sequential market basket analysis. Marketing Letters, 23(3), 505–516. <https://doi.org/10.1007/s11002-012-9181-6>  
  
Sivek, S. C., PhD. (2020, November 17). Market Basket Analysis 101: Key Concepts - Towards Data Science. Medium. Retrieved November 10, 2021, from

<https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00>

Sneha (2022). Predicting best Association Rule using Apriori Algorithm.

<https://www.youtube.com/watch?v=GSpSKDWR4ko>

Aman A. (2021). Market Basket Analysis using Apriori

<https://www.kaggle.com/code/yekahaaagayeham/market-basket-analysis-using-apriori>

Jyotsna V. (2021). Market Basket Analysis – A Complete Overview,

<https://www.jigsawacademy.com/market-basket-analysis/>

Market Basket Analysis 101: Anticipating Customer Behavior, [Smartbridge](https://smartbridge.com/author/smartbridge/), (2022)

<https://smartbridge.com/market-basket-analysis-101/>

How to reduce dimentionality using PCA in Python?

<https://www.projectpro.io/recipes/reduce-dimentionality-using-pca-in-python>

Jason, B. (2020). Principal Component Analysis for Dimensionality Reduction in Python.

<https://machinelearningmastery.com/principal-components-analysis-for-dimensionality-reduction-in-python/>

Grant, P. (2019). Understanding Multiple Regression; The fundamental basis behind this commonly used algorithm.

Medium.<https://towardsdatascience.com/understanding-multiple-regression-249b16bde83e>

Evgeniy R.(2020). 5 Stages of Data Preprocessing for K-means clustering

Medium. <https://medium.com/@evgen.ryzhkov/5-stages-of-data-preprocessing-for-k-means-clustering-b755426f9932>

Michael G. (2018). Understanding Boxplots.

Medium. <https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51>

Mr D. (2018). K-Mean clustering for Wine Quality Data

<https://www.kaggle.com/code/digvijaysingh16/k-mean-clustering-for-wine-quality-data>

Neelam T. (2020). What is K-means Clustering in Machine Learning?

<https://www.analyticssteps.com/blogs/what-k-means-clustering-machine-learning>

Andrea T.(2016). Introduction to K-means Clustering

<https://blogs.oracle.com/ai-and-datascience/post/introduction-to-k-means-clustering>

How to Swap Between Colorblind and Colorful Dashboards. (2016, July 5). YouTube.

<https://www.youtube.com/watch?v=3iNl7KMK8pM>

Knaflic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley .(Chapter 1, Chapters 3-5, Chapters 7 - 8)

<https://ebookcentral.proquest.com/lib/westerngovernorsebooks/reader.action?docID=4187267&ppg=1>

Knaflic, C. N. (2015). storytelling with data: A data visualization guide for business professionals.

Wiley .

(Chapter 1, Chapters 3-5, Chapters 7 - 8)

<https://ebookcentral.proquest.com/lib/westerngovernors-ebooks/reader.action?docID=4187267&ppg=1>