Jordan Tompkins

wgu  D212 Data Mining II Task 3

WGU D212 TAsk 3

A1, Proposal of Question: The question that was being explored using Market Basket Analysis in this project was if there was any products that would be associated with the purchase of the 10ft iPhone charger cable 2 pack?

A2, Defined Goal: The goal of this analysis is to see if there were any products that were purchased along side the 10ft iPhone charger cable 2 pack. This could be helpful in determining what products should be kept near the charger cables, or what items should be promoted along side this product.

B1, Explanation of Market Basket: Market Basket Analysis (MBA) is a useful analysis process that helps retailers with understanding customer purchase patterns. This will allow retailers to thus increase their sales if they are able to get an understanding of customers behaviors. MBA analyses the large data sets, such as purchase history, and from there, you can reveal groupings and patterns of items to help better understand which items to put together. With MBA, association rules are created, and these “rules” show the frequency in which items are bought together, and can be used to predict the likelihood of those events occurring (TechTarget, 2023). To do MBA, you would first start off by creating an “itemsets” from the transactions data. It could be as simple as the {Bread, Peanut butter}, etc. if you were to look at a grocery store for example. From there you would want to measure the frequency of the itemsets as well as the strength of the associations between those items. This is done by looking at things such as confidence, support and lift, defined later in this project. One of the most common ways to determine these association rules is with the Apriori algorithm that was used for this analysis. This algorithm first generates the frequent itemsets mentioned above with a certain number of items that you can set. From there it will decide which itemsets are frequent due to a required minimum level of support. After, the frequent itemsets are partitioned and re-combined repetitively and the support is determined. This is done until there are no more possible itemsets to be created. After getting the frequent itemsets, the association rules are then created from this list. It splits the list apart into the antecedents and the consequents and calculates the confidence. These split itemsets are your association rules. (Dr. Currie Sivek, 2020) In this analysis, we are expecting to see the likelihood of scenarios involving the 10ft iPhone charger cable 2 pack, regardless of if the charger pack was the antecedent (the initial item) or the consequent (the item purchased after the first item).

B2, Transaction Example:

A screenshot of a computer

Description automatically generated

The above screenshot is an example of a transaction in the dataset. It shows that the customer purchased an Apple Lighting to Digital AV Adapter, a TP-Link AC1750 Smart Wifi Router and an Apple Pencil.

B3, Market Basket Assumption: There are many assumptions of market basket analysis. One of the main assumptions is that “…customers who buy a certain item (or group of items) are more likely to buy another specific item (or group of items)” (SmartBridge, 2022). You are basically assuming that buying specific items are cause other items to be purchased. When someone buys ones item, are they more likely to buy another item or not? Market Basket Analysis assumes that yes, there are certain items that when bought, are more likely to be bought in conjunction with the first. Market basket analysis goes through the steps to determine the likelihood of this being the case.

C1, Transforming the Dataset: The dataset used was the telecom\_market\_basket csv file provided by WGU. The dataset contains 7,500+ rows of transactions and 20 columns of items. One thing to note, however, is that in the initial csv file uploaded, every second row is blank. Due to this, the data had to be formatted in a way that got rid of all these blank rows between each transaction. See code provided within jupyter notebook file uploaded. After that, a list of lists had to be created of the transactions in order for us to then instantiate the transaction encoder and fit and transform that list using that encoder to then create a new dataframe that was going to be used for the analysis. See jupyter notebook attached for the code as well as the csv file attached separately for the final dataset used. This final dataset has 7,501 rows of transactions and 119 columns, which represent each individual product that had been bought.

C2, Code: See below for screenshots showing the execution and functionality of the code used to generate the association rules with the Apriori algorithm. A value of 0.02 was used for the minimum support attribute of the Apriori algorithm due to it not being too inclusive or too restrictive.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

C3, Association Rules table: See screenshot below for association rules table

A screenshot of a computer

Description automatically generated

C4, Top Three Rules: See table below for the top three rules:

A screenshot of a computer

Description automatically generated

Within this table, the top three noted rules are “If VIVO Dual LCD Monitor Desk mount then SanDisk Ultra 64GB Card”, “If SanDisk Ultra 64GB Card then VIVO Dual LCD Monitor Desk mount” and “If FEIYOLD Blue Light Blocking Glasses then VIVO Dual LCD Monitor Desk mount”.

D1, Significance of Support, Lift and Confidence Summary: There are three main components of the analysis that was looked at: The support, the lift and the confidence. The support is basically the probability that an event will occur, ie the number of times that rule occurs divided by the total number of transactions. When looking at specifically the 10 ft iPhone Charger cable 2 pack, we get two rules that occur, one where the cable pack is the antecedent and one where it is the consequent. In both rules, the total support was 0.023, or roughly 2.3%. This means that the rule “If 10 ft iPhone Charge Cable 2 Pack then “Dust-off Compressed Gas 2 pack” and vice versa, had a probability of happening 2.3% of the time. The next metric looked at was the confidence, which is the measure of conditional probability, or the rule divided by the proportion of transactions with the antecedent. In our case here, with the iPhone charger pack as the antecedent, the confidence was 0.46. As the consequent it was 0.097. Finally, the last metric explored was the lift. The lift is the probability of all items occurring together divided by the product of the proportion of antecedents times the proportion of the consequent. In the rules including the iPhone charger pack, it would be the proportion of transactions with that rule divided by the proportion of transactions with the iPhone charger pack times the proportion of transactions with the Dust-off gas pack. A lift above one indicates that the buying the antecedent does increase the likelihood of the consequent being bought. With this rule, the lift was 1.9 which gives strong indication that the likelihood of someone buying the Dust-off gas pack was indeed increased due to buying the iPhone charger pack. Dr. Susan Currie Sivek published an article on towardsdatascience.com that explains the different metrics extremely well.

D2, Practical Significances of Findings: Market basket analysis can tell a retailer a lot about the customer’s behaviors in their store. With the analysis performed, the company is able to tell that they only sell the iPhone charger pack roughly 5% of the time and that percentage decreases if the transaction is an iPhone charger pack as the antecedent and dust-off compressed gas as the consequent. However, despite the low percentage of purchases with that specific rule and/or product, the store is able to confidently say that the likelihood of buying the Dust-Off compressed air does in fact increase when the customer buys the iPhone charger pack. So the iPhone charger pack impacts the sales of the dust-off compressed gas.

D3, Course of Action: The recommended course of action for the store would be to include these two items close to each other. They aren’t going to want to have the iPhone charger at the back of the store and the Dust-off pack at the front, or vice versa. They should keep the two items together due to the fact that when someone buys the iPhone charger pack, the likelihood of them buying the compressed gas pack increases. So to sell more of either product, they should put the two items in the same part of the store.

E, Panopto Recording: <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=2f9d1317-e779-4cb9-916c-b0430164f4b6>

F, Web Sources: There were no outside sources used for data or code

G, Sources:

Smartbridge. (2022, July 5). *Market basket analysis 101: Anticipating customer behavior*. Smartbridge. https://smartbridge.com/market-basket-analysis-101/

Susan Currie Sivek, Ph. D. (2020, November 17). *Market basket analysis 101: Key concepts*. Medium. https://towardsdatascience.com/market-basket-analysis-101-key-concepts-1ddc6876cd00

TechTarget. (2023, January 31). *What is Market Basket Analysis? definition from whatis.com*. Customer Experience. https://www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis?Offer=abt\_pubpro\_AI-Insider