D212 Association rules and lift analysis

Nhi Le

Western Governors University

October 29, 2023

**Part I: Research Question**

**A1. Proposal of question**

Can we use market basket analysis to identify how items should be grouped based on customers’ purchasing habits?

**A2. Defined goal**

By understanding customer’s characteristics and purchasing habits, stakeholders can have marketing strategies to increase profits. The goal of this analysis is to understand the relationships among the items, item groups are most likely to be bought together. Sales and marketing can use this information to create sales strategies, or create product bundles, or offer discounts into appropriate items to influence customer’s buying habits.

**Part II: Market Basket Justification**

**B1. Explanation of market basket**

Market Basket Analysis (MBA) is a data mining approach used to increase sales by understanding customer’s purchasing habits. With customer’s purchase history, MBA can identify groups of products when the customers buy a product in a group, they will likely buy other products in that group. Apriori algorithm is used in MBA to discover relationships between items in a transaction database. Apriori algorithm utilizes a breadth-first search approach discover frequent itemset. It accomplishes this by generating candidate item sets and pruning those that do not meet the minimum support threshold (Software Testing Help, 2023). The expected outcome is to find the groups of items that are likely bought together or should be sold together.

**B2. Transaction example**

A list of computer parts

Description automatically generated

**B3. Market basket assumption**

In market basket analysis, the assumption of Apriori algorithm is that “All subsets of a frequent itemset must be frequent” and “If an itemset is infrequent, all its supersets will be infrequent” (GeeksforGeeks, 2022).

**Part III: Data Preparation and Analysis**

**C1. Transforming the data set**

The cleaned data set is attached with this submission as 'df\_prepared3.csv'.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated



**C2. Code execution**

A screenshot of a computer

Description automatically generated

**C3. Association rules table**

A screenshot of a computer

Description automatically generated

**C4. Top three rules**

A screenshot of a computer

Description automatically generated

Based on support metric, the top rule is that 5.97% of transactions have both VIVO Dual LCD Monitor Desk mount and Dust-Off Compressed Gas 2 pack. 17.41% of transactions VIVO Dual LCD Monitor Desk mount is purchased and 23.84% of transactions Dust-Off Compressed Gas 2 pack is purchased.

Based on lift metric, the top rule shows that 2.291162 indicates the strength of association between VIVO Dual LCD Monitor Desk mount and SanDisk Ultra 64GB card. Since the lift value is positive, it suggests that when VIVO Dual LCD Monitor Desk mount is purchased, the likelihood of customers buying a SanDisk Ultra 64GB card increases significantly. Moreover, when a SanDisk Ultra 64GB card is purchased, the likelihood of customers also buying VIVO Dual LCD Monitor Desk mount also increases.

Based on confidence metric, the top rule shows the likelihood that Dust-Off Compressed Gas 2 pack is also bought if 10ft iPhone Charger Cable 2 Pack, or FEIYOLD Blue light Blocking Glasses, or SanDisk Ultra 64 GB card is bought. They all have confidence values over 0.41.

**Part IV: Data Summary and Implications**

**D1. Significance of support, lift, and confidence summary**

Support measures how frequently an item appears in all transactions and helps to evaluate if a rule is worth considering for further analysis (Garg, 2019). In the analysis, there are a few item sets with support values over 5%. That means out of 7501 transactions, some item sets appear together over 375 times.

Lift shows the frequency of consequent by calculating the conditional probability of occurrence of consequents given antecedents (Garg, 2019). This metric can suggest how items are placed in the store. The range of lift values is 1.071 to 2.291. For those item sets with lift values close to 1, the rule happens likely due to commonality rather than association. For the item sets with lift values more than 1.5, it suggests that when an item is bought, the likelihood another item will be bought increases.

Confidence measures the likelihood of items being purchased alongside the antecedents. Since confidence doesn’t really indicate a relationship between the items rather than by chance, lift should be considered to help with the limit of confidence (Garg, 2019). In this analysis, the range of confidence runs from 8% to 46%.

**D2. Practical significance of findings**

I think the results from the analysis are practically significant. By understanding the customer’s purchasing habits, the company can possibly increase the profit. The stakeholders can consider top rules from the analysis to reorganize the items in the stores to make it more visible for customers. They can also create appropriate product bundles with small discounts to influence customer’s buying behaviors. The results from the market basket analysis will be helpful for stakeholders to see which item sets to focus on.

**D3. Course of action**

Through the analysis, I would recommend to the stakeholders the importance of grouping items together to increase a customer’s buying behavior. The stakeholders can look into the item sets with significant lift, support, and confidence values. With the efforts of reorganizing the items in the stores and creating appropriate product bundles, the company can have significant outcomes. It will also increase profits by making the products visible for customers to buy more than one product. Moreover, the company can create marketing strategies for the item sets with high lift and confidence values. For example, when customers buy an ‘antecedent’ item, they can receive a small percentage of discount on the ‘consequent’ item if they buy both.

**Part V: Attachments**

**E. Panopto video**

<https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=0fe1c0e7-fb4f-4ffd-9e73-b0aa017580ab>

**F. Sources for third-party code**

Kimnaruk, Y. (2022, October 15). *Market basket analysis: Step-by-step coding*. Medium. https://medium.com/mlearning-ai/market-basket-analysis-step-by-step-coding-cd13ce1f8de9

**G. Sources**

GeeksforGeeks. (2022, January 13). *Apriori algorithm*. GeeksforGeeks. https://www.geeksforgeeks.org/apriori-algorithm/

Software Testing Help. (2023, June 27). *Apriori algorithm in Data Mining: Implementation with examples*. Software Testing Help. https://www.softwaretestinghelp.com/apriori-algorithm/

Garg, A. (2019, February 7). *Complete Guide to Association rules (1/2)*. Medium. https://towardsdatascience.com/association-rules-2-aa9a77241654