**Market Basket Analysis of Churn Dataset**

Kelseyann Wright

Student ID: 00565814

College of Information Technology, Western Governors University

Data Mining II - D212

Kesselly Kamara

January 15, 2024

**Market Basket Analysis of Churn Dataset**

**A1. PROPOSAL OF QUESTION**

The proposed question is, “

**A2. DEFINED GOAL**

The goal of this analysis is to

**B1. EXPLANATION OF MARKET BASKET**

*The submission logically explains how market basket analyzes the selected data set and includes expected outcomes.*

Market basket analysis (MBA) is a popular machine learning method. MBA identifies useful association rules to discover relationships between different transactions based on the co-occurrence of data values such as items, products, or services (Lim, 2022). For example, MBA could be used to identify grocery items that are frequently purchased together, such as cereal and milk, and use that information to optimize shelf placement or sale promotions.

The measures of support, confidence, and lift are used to discover association rules (Lim, 2022). Support is the probability of an event, such as the purchase of a specific item or multiple items together and is measured as the proportion of transactions in which that event occurs (Lim, 2022). Confidence is the probability of a consequent given an antecedent, or the probability of event B given event A (Lim, 2022). For example, it could be the probability that a customer purchases milk given that they purchase cereal. It is calculated as *P(A and B) / P(A),* where *P* is probability calculated as the proportion of transactions meeting the specified condition. Lift is similar to confidence, except lift controls for how popular both items are (Lim, 2022). Lift is calculated as *P(A and B) /[P(A) \* P(B)].* A lift value greater than 1 indicates that there is an association between the items A and B, with higher values indicating stronger associations (Lim, 2022). These three measures are used in the Apriori Algorithm during MBA to efficiently identify associations in large transactional datasets. The Apriori Algorithm starts by identifying individual items in a dataset and then iteratively combines items into larger item sets, only if the item sets are frequent, to identify item associations (Lim, 2022).

**B2. TRANSACTION EXAMPLE**

Include one accurate example of transaction in the dataset

A screen shot of a computer

Description automatically generated

**B3. MARKET BASKET ASSUMPTION**

Summarize one assumption

**C1. TRANSFORMING THE DATA SET**

The submission transforms the data set to make it suitable for market basket analysis and includes a copy of the cleaned data set.

SHOULD HAVE 7501 rows and 119 columns

**C2. CODE EXECUTION**

The submission executes the code used to generate association rules with the Apriori algorithm and provides screenshots that demonstrate the error-free execution of the code.

A screen shot of a computer

Description automatically generated

**C3. ASSOCIATION RULES TABLE**

The complete association rules table is attached (ruletable\_T3.csv). A preview of the first few rows of the table is provided below.

A screenshot of a computer

Description automatically generated

**C4. TOP THREE RULES**

A black screen with white text

Description automatically generatedThe submission includes a screenshot of the top 3 relevant rules and accurately explains the top 3 relevant rules generated by the Apriori algorithm

**D1. SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY**

The submission accurately summarizes the significance of support, lift, and confidence from the results of the analysis.

**D2. PRACTICAL SIGNIFICANCE OF FINDINGS**

The submission accurately discusses the practical significance of the findings from the analysis.

**D3. COURSE OF ACTION**

The submission accurately recommends a course of action for the real-world organizational situation from part A1 based on the results from part D1.

**E. PANOPTO VIDEO OF CODE AND E1. PANOPTO VIDEO OF PROGRAMS**

A Panopto video providing full documentation of the code and description of programs used is attached. The link is also provided here: LINK

**F. SOURCES OF THIRD-PARTY CODE**

No sources of third-party code were used.

**G. SOURCES**

Lim, Y. (2022). Data Mining: Market Basket Analysis with Apriori algorithm. Towards Data Science. https://towardsdatascience.com/data-mining-market-basket-analysis-with-apriori-algorithm-970ff256a92c

Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825–2830.