STUDENT’S NAME: EBENEZER ESSILFIE WILSON

COURSE INSTRUCTOR: DR. KEIONA MIDDLETON

COURSE TITLE: DATA MINING II (TASK 3)

JANUARY 10, 2025

A1.

Do we have items in the dataset that are frequently bought together? If so, how can we maximize sales?

Which item sets do customers normally come for together in the dataset? Businesses use the Market Basket Analysis to help find relationship between items often bought together. This allows them to make profit (Agrawal & Srikant, 1994).

A2.

The goal is to identify strong associations or frequent item sets in the transactional data to inform marketing and sales strategies, such as designing product bundles, personalized promotions, or store layouts that increase overall revenue.

B1.

When a customer buys an item, say, item A and proceeds to buy another item, say, item B because of the purchase of item A, there could be a relationship between A and B. Market Basket Analysis can analyze how often customers buy items A and B together using Apriori algorithm.

Expected Outcomes:

Association rules that reveal meaningful relationships between items. Insights into customer purchasing habits, which can inform marketing strategies, product bundling, or store layouts.

B2.

A screenshot of a computer

Description automatically generated

B3.

When items are often bought together in a transaction, they can impact each other’s probability of being sought after.

C1.

This is how I prepared the data. I used the df.head() to view the summary of the dataset. Upon doing this, I discovered that some rows were empty. All empty rows and columns were dropped. Using the TransactionEncoder(), I saved it in a DataFrame.

A screenshot of a computer

Description automatically generated

Removing empty rows

A screenshot of a computer

Description automatically generated

Running TransactionEncoder and saving it in a new DataFrame

A screenshot of a computer

Description automatically generated

A table with black text

Description automatically generated with medium confidence

Checking for empty column

A screenshot of a computer

Description automatically generated

There is ‘nan’ (empty column) the third from the bottom of the items.

Now getting rid of the empty column leaves us with this.

A screenshot of a computer

Description automatically generated

Saving to prepared csv.

A close-up of a sign

Description automatically generated

C2.

The Apriori algorithm, a fundamental technique in MBA, efficiently discovers frequent item sets in transactional databases by applying support and confidence thresholds (Han, Pei, & Yin, 2000).

A screenshot of a computer

Description automatically generated

C3.

A screenshot of a computer

Description automatically generated

This is the association rules table generated without any errors

C4.

A computer code with text

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

High-lift association rules indicate strong product relationships, which can enhance personalized recommendations and targeted promotions (Tan, Steinbach, & Kumar, 2005). My dataset is sorted by lift first, then confidence as a secondary factor. If two rules have the same lift, then confidence determines ranking. Sorting this way prioritizes the strongest associations for business insights (e.g., bundling products).

D1.

Support helps identify popular rules that appear frequently. The rule (Screen Mom Screen Cleaner kit → Dust-Off Compressed Gas 2 pack) has a support of 0.023997, meaning it appears in 2.4% of all transactions.

Confidence indicates how strongly related two items are. The probability of customers adding compressed gas pack in their SanDisk card purchase is 0.42

Lift measures true association strength beyond just random co-occurrence. The rule (Dust-Off Compressed Gas 2 pack → SanDisk Ultra 64GB card) has a lift of 3.495, meaning customers are 3.5 times more likely to buy the SanDisk card when purchasing the gas pack.

D2.

The results of the Apriori algorithm provide valuable insights into customer purchasing behaviors, which can be leveraged for business decision-making, marketing strategies, and inventory management. Retailers use MBA for inventory management and store layout optimization, ensuring complementary products are placed together to increase sales (Brin, Motwani, Ullman, & Tsur, 1997). Below are the key practical applications of the findings:

Customers usually bought Dust-Off Compressed Gas 2 pack and SanDisk card together. We can bundle these items together in promotional offers. We can also offer discounts on one item if the customer buys the other.

The high confidence (41.7%) for the rule SanDisk Ultra 64GB card → Dust-Off Compressed Gas 2 pack means that almost half of the customers who bought a SanDisk card also purchased the gas pack. We can suggest related products in online stores based on past purchases. We can send personalized emails suggesting these items as “Frequently Bought Together.”

The rule (Screen Mom Screen Cleaner Kit → Dust-Off Compressed Gas 2 Pack) has a lift of 3.107, indicating that customers who purchase the screen cleaner are 3.1 times more likely to also buy the compressed gas pack. We can offer discounts or "Buy One, Get One" deals on these items.

D3.

Provide special promotions or discounts on items that are often bought together. Promote seasonal discounts on these high-lift product pairs. This increases sales volume by encouraging bulk purchases and cross-selling.

E.

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487-499.

Han, J., Pei, J., & Yin, Y. (2000). Mining Frequent Patterns Without Candidate Generation. ACM SIGMOD Record, 29(2), 1-12.

F.

Tan, P.-N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Addison-Wesley.

Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic Itemset Counting and Implication Rules for Market Basket Data. ACM SIGMOD Record, 26(2), 255-264.

G. <https://wgu.hosted.panopto.com/Panopto/Pages/Viewer.aspx?id=142c7c53-4364-470d-80dd-b2750127f786>