ASSIGNMENT NO. 5

AIM: Assignment on Apriori Algorithm using Market Basket Dataset.

OBJECTIVES:

- To understand the concept of the Apriori algorithm and its working principles.
- To implement the Apriori algorithm using Python.
- To discover frequent itemsets and generate association rules from transaction data.
- To analyze the impact of support, confidence, and lift thresholds on the resulting rules.

THEORY: Apriori is a classic algorithm used in market basket analysis to identify frequent itemsets and generate association rules. It operates on the principle that if an itemset is frequent, all of its subsets are also frequent. This "apriori" property helps reduce the search space when finding frequent itemsets.

The algorithm is primarily used in transactional databases to uncover relationships between items in customer transactions. It is widely applied in retail, recommendation systems, and business intelligence.

Algorithm Steps: Given a dataset of transactions, the algorithm proceeds as follows:

1. **Set Minimum Thresholds:** Define minimum support, confidence, and lift values.
2. **Generate Candidate Itemsets:** Start with single items and iteratively build larger itemsets.
3. **Prune:** Eliminate itemsets that do not meet the minimum support threshold.
4. **Generate Rules:** From the frequent itemsets, generate association rules that satisfy the confidence and lift thresholds.
5. **Interpret Results:** Analyze the rules to identify useful insights and patterns.

The support of an itemset is the proportion of transactions containing that itemset.

Confidence measures how often item in Y appear in transactions that contain X.

Lift compares the confidence of a rule with the expected confidence if X and Y were independent.

Dataset:

The dataset used for this assignment is **transactions2.csv**, which contains a list of transactions. Each transaction includes a set of items purchased together. The goal is to uncover frequent combinations of items and generate association rules.

Key Terms in Association Rule Mining:

- **Frequent Itemsets:** Sets of items that appear frequently together in the dataset.
- **Support:** Frequency of an itemset in the dataset.
- **Confidence:** Likelihood that an item Y is purchased when item X is purchased.
- **Lift:** Ratio that measures how much more often X and Y occur together than expected if they were statistically independent.
- **Implementation Summary:** The algorithm was implemented using the mlxtend library in Python.

Methodology:

Step 1: Import Required Libraries

Step 2: Load the Dataset

Step 3: Convert Items to List of Lists

Step 4: One-Hot Encode the Transactions

Step 5: Apply the Apriori Algorithm

Step 6: Generate Association Rules

Step 7: Display the Results

Parameter Tuning:

1. **Support Threshold:** Helps filter out rare combinations. Lower values may find more rules but increase noise.
2. **Confidence Threshold:** Determines rule strength. High values yield more reliable rules.
3. **Lift Threshold:** Values >1 suggest a positive association between items.

Conclusion:

Apriori is a powerful tool for market basket analysis and can reveal valuable insights into customer purchasing habits. By tuning support, confidence, and lift, one can balance rule discovery and relevance. Despite its computational complexity for large datasets, it remains a foundational algorithm in association rule mining.

	Antecedents	Consequents	Support	Confidence	Lift
0	Bread	Butter	0.375	0.5	1.333333
1	Butter	Bread	0.375	1.0	1.333333
2	Milk	Bread	0.375	0.6	0.800000
3	Bread	Milk	0.375	0.5	0.800000

People who buy bread are somewhat likely to buy butter too. The lift > 1 shows a positive association.

Butter buyers always buy bread. This is a strong association and the rule is confident and positively correlated.

Though there's some confidence, the lift < 1 implies a negative or neutral relationship.

Not a strong association. People who buy bread don’t necessarily buy milk together more often than chance.

REFERENCE:

- https://www.geeksforgeeks.org/apriori-algorithm/
- GitHub Repository: https://github.com/Utkarsh-Rane43/ML-LAB---122B1F110