**Assignment 5**

**Aim:**

To implement the Apriori algorithm on retail grocery data for Market Basket Analysis.

**Objective:**

To understand and apply market basket analysis using the Apriori algorithm by identifying frequent itemsets and deriving association rules based on customer purchasing behavior from transaction data.

**Theory:**

The **Apriori algorithm** is a foundational data mining technique used to discover frequent itemsets and generate association rules. It is widely used in **Market Basket Analysis**, where the goal is to identify items that are often bought together.

**Working of Apriori:**

* It generates candidate itemsets and incrementally expands them.
* It eliminates itemsets that do not meet the specified **minimum support threshold**.
* From the frequent itemsets, it generates association rules based on **confidence** and **lift** metrics.

**Importance of Apriori in Data Mining:**

* Identifies frequently purchased item combinations.
* Helps in decision-making such as:  
    
  + *"Customers who buy milk often also buy bread."*
* Supports:  
    
  + Store layout optimization
  + Product bundling
  + Inventory and promotional planning

**Dataset:**

* **Filename**: Groceries\_dataset.csv
* **Columns**:  
    
  + Member\_number: Unique ID for each customer
  + Date: Date of purchase
  + itemDescription: Item purchased

**Description:**

Each row represents a single item bought by a customer on a specific date. Transactions with the same Member\_number and Date represent multiple items bought in a single purchase.

**Example Transactions:**

* Bread, Milk
* Bread, Diaper, Beer, Eggs
* Milk, Diaper, Beer, Coke
* Bread, Milk, Diaper, Beer
* Bread, Milk, Diaper, Coke

**Steps for Implementation:**

**1. Importing Libraries**

* pandas: For data handling
* mlxtend: For implementing Apriori and generating association rules
* matplotlib: For visualizing results

**2. Loading the Dataset**

* Load Groceries\_dataset.csv using pandas.
* Create unique transactions by combining Member\_number and Date.
* Group items based on this unique transaction ID.

**3. Preprocessing**

* Convert each transaction into a list of items.
* Use TransactionEncoder (from mlxtend) to one-hot encode the data.
* Convert the encoded data into a Boolean DataFrame for Apriori.

**4. Applying the Apriori Algorithm**

* Use the apriori() function to identify frequent itemsets.
* Set a **minimum support threshold** (e.g., min\_support = 0.01).
* The result is a DataFrame showing itemsets with their support values.

**5. Generating Association Rules**

* Use association\_rules() to generate rules from frequent itemsets.
* Filter rules based on metrics like **confidence** and **lift**.
* Output includes:  
    
  + antecedents, consequents
  + support, confidence, and lift

**6. Visualization**

* Plot a **bar chart** for the top frequent itemsets by support.
* Use **scatter plots** to show:  
    
  + Support vs Confidence
  + Support vs Lift

These visualizations help assess the quality and relevance of the rules.

**Results & Insights**

**Frequent Itemsets (Examples):**

* {whole milk}, {tropical fruit}, {other vegetables}

**Association Rules (Examples):**

* {yogurt} → {whole milk} – confidence: 0.45, lift: 1.22
* {root vegetables} → {other vegetables} – confidence: 0.41, lift: 1.34

**Business Insights:**

* Dairy and vegetables are often bought together.
* Results can guide:  
    
  + Product placement
  + Combo deals
  + Targeted marketing

**Conclusion:**

The **Apriori algorithm** effectively identified frequent item combinations and meaningful associations in the grocery dataset. These insights can help retailers make strategic decisions regarding store layout, promotions, and personalized offers.

**References:**

* [GeeksforGeeks: Apriori Algorithm](https://www.geeksforgeeks.org/apriori-algorithm/)
* [MLXtend Documentation](https://mlxtend.github.io/)