 **GRT INSTITUTE OF ENGINEERING AND**

**TECHNOLOGY, TIRUTTANI – 631209**

**Approved by AICTE, New Delhi Affiliated to Anna University, Chennai**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**PHASE 4**

**PROJECT TITLE**

***Market Basket Insights***

**COLLEGE CODE : 1103**

**JABASTION GNANARAJ G**

3rd yr, 5th sem

Reg no. : 110321121010

**gjabastion@gmail.com**

**4.1 PERFORMING ASSOCIATION ANALYSIS :**

1. **DATA PREPARATION:**

Download the dataset from the Kaggle link you provided.

The dataset likely contains information about customer transactions, where each row represents a transaction, and the items purchased in that transaction are listed in columns.

**2.DATA EXPLORATION:**

Begin by exploring the dataset to understand its structure. Look at the column names and their meanings.

Check for any missing data or anomalies in the dataset.

**3.DATA PREPROCESSING:**

Data preprocessing involves cleaning and formatting the data to make it suitable for association analysis. This typically includes:

Handling missing values.

Encoding the data into a suitable format (e.g., a binary matrix where each column represents an item, and each row represents a transaction).

**4.ASSOCIATION RULE MINING:**

Use association rule mining algorithms, such as Apriori or FP-Growth, to identify associations or patterns in the data.

These algorithms look for sets of items that frequently appear together in transactions.

**5.SETTING SUPPORT AND CONFIDENCE THRESHOLDS:**

Define minimum support and confidence thresholds. These thresholds help filter the results to focus on the most significant associations.

Support is the percentage of transactions that contain a particular itemset, and confidence is the probability that an item B is purchased when item A is purchased.

**6.GENERATING ASSOCIATION RULES:**

Apply the chosen algorithm to the dataset to find frequent itemsets and association rules. Association rules typically take the form "if A, then B," indicating that if one item or set of items is present in a transaction, another item or set of items is likely to be present as well.

**7.INTERPRETING THE RESULTS:**

Examine the generated association rules. Look for interesting and meaningful patterns. For example, you might find rules like "Customers who buy item A are likely to buy item B as well."

Assess the support, confidence, and lift (a measure of the strength of the association) for each rule.

**8.BUSINESS INSIGHTS:**

Translate the discovered patterns into actionable business insights. For example, you might use these insights to:

Optimize product placement in stores or on e-commerce websites.

Create targeted marketing campaigns.

Suggest product bundling opportunities.

Enhance inventory management.

**9.VISUALIZATION (OPTIONAL):**

Create visualizations, such as scatter plots or network diagrams, to represent the associations and make them more understandable to non-technical stakeholders.

**10.ITERATE AND REFINE:**

Association analysis can be an iterative process. You may need to adjust the support and confidence thresholds or explore different data preprocessing techniques to refine your results.

**4.2 PERFORMING ASSOCIATION ANALYSIS WITH PROGRAMMING:**

**DATA LOADING:** The program loads a dataset (specified by 'your\_dataset.csv') that contains transaction data.

**DATA PREPROCESSING**: Data preprocessing is a crucial step in association analysis. It involves converting the dataset into a suitable format for analysis, where columns represent items, and rows represent transactions.

**ASSOCIATION ANALYSIS:** The program uses the Apriori algorithm to find frequent itemsets in the dataset. Frequent itemsets are combinations of items that appear together often in transactions.

**GENERATING ASSOCIATION RULES**: After identifying frequent itemsets, the program generates association rules. These rules describe relationships between items and are often in the format "if A, then B." They indicate the likelihood of one item or set of items being purchased when another item or set of items is purchased.

**DISPLAYING THE RESULTS:** The generated association rules, along with their support, confidence, and lift values, are displayed in the output.

**PROGRAMMING:**

import pandas as pd

from mlxtend.frequent\_patterns import apriori

from mlxtend.frequent\_patterns import association\_rules

# Load your dataset

data = pd.read\_csv('your\_dataset.csv')

# Data Preprocessing

# You may need to preprocess your dataset to create a binary matrix

# where columns represent items, and rows represent transactions.

# Perform Association Analysis

# Use Apriori to find frequent itemsets

frequent\_itemsets = apriori(data, min\_support=0.1, use\_colnames=True)

# Generate Association Rules

association\_rules = association\_rules(frequent\_itemsets, metric="lift", min\_threshold=1.0)

# Display the association rules

print(association\_rules)

**OUTPUT:**



**CONCLUSION:**

Association analysis, often referred to as Market Basket Analysis, is a powerful data mining technique used to uncover patterns and relationships within transactional data. This method helps organizations identify associations between items or products that are frequently purchased together. By analyzing these associations, businesses can gain valuable insights that can be used to:

**Optimize Sales**: Understanding which products are frequently bought together allows businesses to optimize product placement and promotions, potentially increasing sales.

**Customer Insights**: Association analysis can reveal customer preferences and buying behaviors, helping businesses tailor their marketing and product offerings.

**Inventory Management:** It aids in better inventory management, ensuring that popular products are well-stocked, while avoiding overstocking of less popular items.

**Cross-Selling and Upselling:** Businesses can use these associations to implement effective cross-selling and upselling strategies, suggesting complementary products to customers.

**Targeted Marketing:** By knowing what items tend to be purchased together, businesses can create targeted marketing campaigns and recommendations for customers.