**Market Basket Analysis Using Apriori and FP-Growth**

**1. Project File**

* **Project Title:** Market Basket Analysis Using Apriori and FP-Growth
* **Submitted By:** Vignesh S
* **Mentor:** Charulatha
* **Submission Date:** 21-03-2025

**2. Abstract**

Market Basket Analysis (MBA) is a data mining method that identifies concealed patterns in customer buying behavior. Analyzing buying patterns allows companies to increase cross-selling, product placement, and recommendations.

This project utilizes MBA through Apriori and FP-Growth algorithms to identify significant product associations, which enable companies to enhance their sales strategy.

**3. Introduction**

Market Basket Analysis assists companies in determining the relationships between items that are often purchased together. This project seeks to examine transactional data to recommend store layout and promotion improvements.We apply Apriori and FP-Growth algorithms to derive meaningful insights.**4. Problem Statement**

It is common for retailers to be unaware of which products are commonly bought together. Without data-driven insights, it is difficult to maximize product placement and promotions.This project seeks to mine product associations fromtransaction data to yield actionable recommendations.

**5. Dataset Description**

* **Dataset:** Groceries Dataset (from Kaggle)
* **Fields:**
  1. Member Number: Customer ID
  2. Date: Date of purchase
  3. Item Description: Product bought
* Preprocessing of data is done by transforming transaction records into One-Hot Encoding format to facilitate association rule mining.

**6. Algorithms Used**

**Apriori Algorithm**

* Generates frequent itemsets based on **support, confidence, and lift**.
* Iteratively expands frequent itemsets using a threshold.
* **Limitation:** Computationally expensive for large datasets.

**FP-Growth Algorithm**

* Uses a tree-based structure to mine frequent itemsets efficiently.
* Avoids candidate generation, making it faster than Apriori.
* **Advantage:** Works better with large datasets.

**7. Implementation Steps**

1. Import Required Libraries
2. Load and Preprocess Dataset
3. Apply Apriori Algorithm & Extract Rules
4. Apply FP-Growth Algorithm & Extract Rules
5. Extract Business Insights & Recommendations
6. Visualize Results Using Bar Charts & Heatmaps

**Code Implementation**

This part contains the Python code for Market Basket Analysis using Apriori & FP-Growth algorithms.

**Step 1: Load and Preprocess the Dataset**

**Importing Necessary Libraries**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from mlxtend.frequent\_patterns import apriori, association\_rules, fpgrowth

from mlxtend.preprocessing import TransactionEncoder

**Loading & Preprocessing Dataset**

df = pd.read\_csv('Groceries\_dataset.csv')

transactions = df.groupby('Member\_number')['itemDescription'].apply(list).tolist()

te = TransactionEncoder()

te\_array = te.fit(transactions).transform(transactions)

basket = pd.DataFrame(te\_array, columns=te.columns\_)

**Step 2: Apply Apriori Algorithm**

frequent\_itemsets\_ap = apriori(basket, min\_support=0.01, use\_colnames=True)

rules\_ap = association\_rules(frequent\_itemsets\_ap, metric='lift', min\_threshold=1)

**Step 3: Apply FP-Growth Algorithm**

frequent\_itemsets\_fp = fpgrowth(basket, min\_support=0.01, use\_colnames=True)

rules\_fp = association\_rules(frequent\_itemsets\_fp, metric='lift', min\_threshold=1)

**8. Results & Insights**

**Key Findings:**

* Frequent itemsets like **Bread & Butter** and **Milk & Cereal** were discovered.
* Apriori was **slower** for large datasets, while FP-Growth was **more efficient**.
* Businesses can use these insights to improve marketing and product placement strategies.

**Visualizations:**

* **Bar Charts:** Top 10 association rules by lift.
* **Heatmaps:** Frequent itemsets appearing together.

**9. Conclusion & Future Scope**

* Market Basket Analysis helps businesses **increase sales and customer engagement**.
* Apriori and FP-Growth effectively generate association rules, but FP-Growth is better for **large datasets**.
* **Future Enhancements:** Implementing AI-based recommendation systems for personalized shopping experiences.

**10. References**

* **Kaggle - Groceries Dataset:** https://www.kaggle.com/datasets/heeraldedhia/groceries-dataset
* **Research Paper:** Market Basket Analysis with Apriori Algorithm and FP-Growth
* **Link:** https://ijersc.org/index.php/go/article/download/45/40/301
* **Python Libraries:** pandas, mlxtend, seaborn, matplotlib.