

Project 5: Market Basket Insights: Unveiling Customer Behaviour Through Association Analysis

PHASE 2

The Market Basket Analysis workflow using the IBM Sample Dataset begins with data collection and integration of the dataset. Data preprocessing follows to clean and structure the data. Exploratory Data Analysis (EDA) helps uncover transaction patterns. Association rule mining (Apriori or FP-Growth) identifies item associations, and clustering methods (K-Means or DBSCAN) segment customers based on their purchasing behaviour. Recommendation systems, predictive modelling, and data visualization provide insights. These insights are interpreted to align with business goals, and strategies are developed. Testing, continuous improvement, and documentation are integral, ensuring insights are effectively communicated and compliant with data privacy regulations.

1. Data Collection and Integration:

Start by acquiring the IBM Sample Dataset for Market Basket Analysis. You can download it from the source.

Integrate the dataset into your data management system.

2. Data Preprocessing:

Clean the dataset, handling any missing values and outliers.

Format and structure the data for analysis.

3. Exploratory Data Analysis (EDA):

Conduct exploratory data analysis to understand the dataset. This may include:

Summary statistics.

Visualizations to explore transaction patterns.

Identifying top-selling products.

4. Association Rule Mining:

Apply association rule mining to identify item associations and co-occurrences in the transactions.

Algorithms:

Apriori or FP-Growth.

Set appropriate support and confidence thresholds to filter meaningful rules.

5. Clustering and Segmentation:

Group similar customers or transactions based on purchasing behavior.

Algorithms:

K-Means or DBSCAN.

Interpret clusters to identify customer segments.

6. Recommendation Systems:

Develop recommendation systems to suggest items to customers based on their preferences.

Algorithms:

Collaborative Filtering.

Content-Based Filtering.

Hybrid Approaches.

Generate personalized recommendations.

7. Predictive Modelling:

Build predictive models to forecast future purchases or identify trends.

Algorithms:

Time Series Analysis.

Machine Learning Models (e.g., decision trees, random forests).

Train models using historical data and evaluate their performance.

8. Data Visualization and Reporting:

Create interactive dashboards and visualizations to present insights and patterns.

Tools like Tableau, Power BI, or Python libraries (e.g., Matplotlib, Seaborn) can be used.

9. Interpretation and Strategy Alignment:

Interpret the generated insights and align them with business objectives.

Formulate actionable strategies, such as optimizing inventory or designing targeted marketing campaigns.

10. Testing and Validation:

- Implement strategies derived from insights on a smaller scale or through A/B testing.

- Monitor outcomes and measure the impact.

11. Continuous Improvement and Adaptation:

- Continuously monitor customer behaviour and update insights as patterns evolve.

- Collect feedback from stakeholders for refinement.

12. Documentation and Knowledge Sharing:

- Document the entire process, including data sources, preprocessing, and algorithm choices.

- Share findings with relevant stakeholders.

13. Collaboration and Communication:

- Promote collaboration among data scientists, domain experts, and business teams.
- Ensure alignment between insights and business goals.

14. Compliance and Data Privacy:

- Ensure that the project adheres to data protection regulations and maintains customer privacy.

TEAM MEMBERS:

1. J. ARUN KUMAR-813821205002
2. C. MURUGANANTHAN-813821205033
3. T. SRIHARIHARAN-813821205049
4. S. VENGADASHAN-813821205055
5. S. VISHWA-813821205059