Market Basket Analysis using Apriori Algorithm for Retail Optimization

Vikram Kotlo Kotlovikram@gmail.com 940-977-4733 11608418	ShivaSai Kumar Tummala shivasaikumart@gmail.com 940-595-3309 11699034	Swetha Vahana Reddy Morreppa reddyswethuvahana@gmail.com 940-843-6774 11715081	Sai Ritwik Reddy Nagireddy nsairitwikreddy@gmail.com 940-344-2006 11704178

Motivation:

Our drive for this project stems from the pressing need of retailers to delve into customer purchasing patterns. Through Market Basket Analysis (MBA) employing the Apriori algorithm, we aim to equip retailers with profound insights into customer behaviors. The motivation lies in the potential to refine marketing strategies, product placements, and promotions in the dynamic retail landscape.

Significance:

The significance of our project lies in its transformative potential for retailers, enabling them to make decisions grounded in data. Market basket analysis, facilitated by the Apriori algorithm, emerges as a potent tool. Its significance is underscored by the prospect of optimizing retail strategies, resulting in elevated sales and heightened customer satisfaction.

Objectives:

Identifying Itemset Associations:

Goal: Unearth associations between items in customer transactions. Goal: Unearth associations between items in customer transactions.

Objective: Implement the Apriori algorithm to reveal frequent itemsets, forming the bedrock for market basket analysis.

Optimizing Retail Strategies:

Goal: Derive actionable insights for retail optimization from customer purchasing behaviors.

Objective: Generate association rules by fine-tuning parameters (minimum support, minimum confidence, minimum lift, minimum length, and maximum length) for extracting meaningful insights.

DataTransformation and Preprocessing:

Goal: Ensure seamless application of the Apriori algorithm through dataset

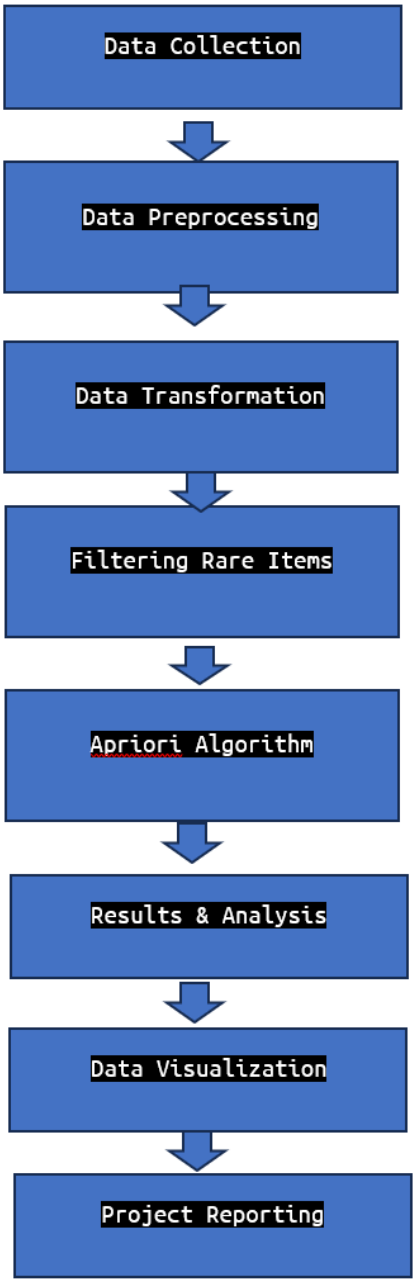
Objective: Implement the Apriori algorithm to reveal frequent itemsets, forming the bedrock for market basket analysis.

Optimizing Retail Strategies:

Goal: Derive actionable insights for retail optimization from customer purchasing behaviors.

Objective: Generate association rules by fine-tuning parameters (minimum support, minimum confidence, minimum lift, minimum length, and maximum length) for extracting meaningful insights.

WORK-FLOW DIAGRAM



Evaluation Metrics:

Goal: Assess the effectiveness of generated association rules.

Objective: Utilize support, confidence, and lift as evaluation metrics, providing a quantitative measure of frequency, reliability, and strength of associations.

Parameter Tuning and Comparison:

Goal: Understand the impact of different parameter settings on rule generation.

Objective: Compare association rules under varied parameter settings, such as min_support and min_confidence, shedding light on their influence.

Features:

Apriori Algorithm Implementation:

Feature: Deploy the Apriori algorithm as the foundational method for market basket analysis.

Benefit: Efficiently uncover frequent itemsets and association rules.

Data Preprocessing:

Feature: Transform and preprocess the dataset to ensure compatibility with the Apriori algorithm.

Benefit: Enable a smooth application of the algorithm to the dataset.

Experiments and Analysis:

Feature: Conduct experiments using the Apriori algorithm on the provided dataset.

Benefit: Derive meaningful insights from customer transactions through the analysis of association rules.

Evaluation Metrics:

Feature: Evaluate association rules using metrics such as support, confidence, and lift.

Benefit: Quantify the effectiveness of the generated rules in providing valuable information about customer purchasing behaviors.

Parameter Tuning and Comparison:

Feature: Tune key parameters and compare their impact on rule generation.

Benefit: Provide guidance on parameter settings for optimizing the extraction of association rules.

Related Work:

In the context of our project, the related work provides a foundational understanding of market basket analysis, association rule mining, and the Apriori algorithm. Key references include the pioneering work of Agrawal and Srikant (1994) on the Apriori algorithm, the use of essential libraries such as NumPy, Pandas, and Matplotlib, and insights from prior studies utilizing similar Market Basket Optimization datasets.

DATASET:

Design of Features/Labels:

In Market Basket Analysis, the dataset doesn't have traditional features and labels as in supervised learning tasks. Instead, the focus is on identifying associations between items within transactions. The items themselves act as features, and the association rules that express relationships between itemsets act as labels.

Features:

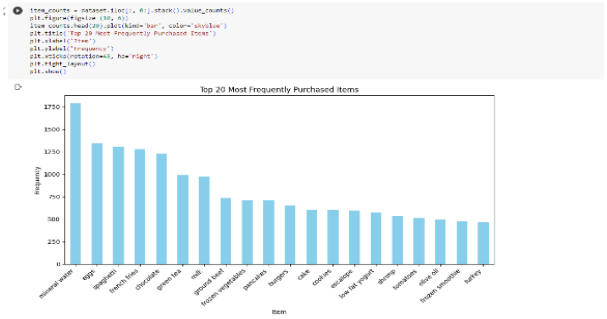
The individual items (products) listed in the dataset form the features. In the dataset, each column represents a different item that can be purchased. For example, "shrimp," "almonds," "avocado," etc., are features.

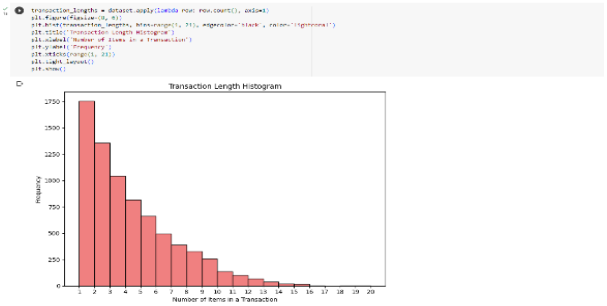
Labels (Association Rules):

The association rules represent relationships between itemsets. An association rule is in the form of "If {itemset A} then {itemset B}". The left-hand side (antecedent) contains the items that are present in a transaction, and the right-hand side (consequent) contains the items that are likely to be purchased together with the antecedent items. The association rules are generated using the Apriori algorithm.
For example, one association rule could be: "If {avocado, green tea} then {almonds}". This rule suggests that customers who buy "avocado" and "green tea" together are likely to buy "almonds" as well.

Exploratory data analysis:

Data Visualization





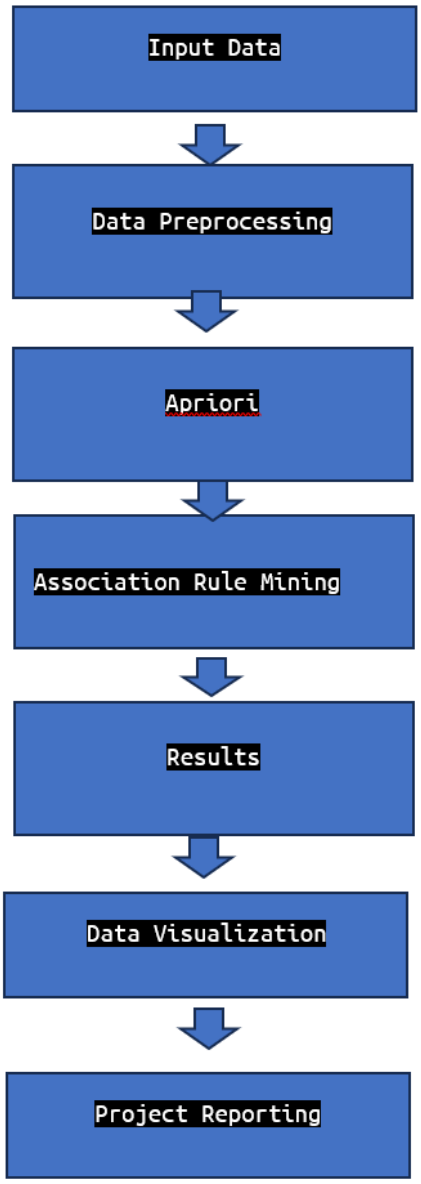
Detailed Design of Features:

Our project features an implementation of the Apriori algorithm, leveraging essential libraries like NumPy, Pandas, and Matplotlib. The algorithm is designed to efficiently discover frequent itemsets and association rules. Data preprocessing is a crucial aspect, ensuring the dataset aligns with the Apriori algorithm's input requirements. Additionally, the project incorporates parameter tuning, allowing customization of key parameters such as minimum support, minimum confidence, and others.

Analysis:

The analysis phase involves scanning the dataset multiple times using the Apriori algorithm, iteratively generating candidate itemsets, evaluating their support, and deriving frequent itemsets. The results include a set of association rules, each accompanied by support, confidence, and lift values. These rules offer insights into item associations, aiding in understanding customer preferences and guiding marketing strategies.

ARCHITECTURE DIAGRAM



Implementation:

The implementation of the Apriori algorithm is a central component of our project. It involves the efficient mining of frequent itemsets through an iterative approach of candidate generation and pruning. The algorithm scans the dataset, generating association rules that provide valuable information about customer purchasing behaviors.

Preliminary Results:

As of the current status, the implementation has yielded preliminary results, showcasing association rules with corresponding support, confidence, and lift values. These results offer a glimpse into the potential of the Apriori algorithm in uncovering meaningful insights from customer transactions.

Project Management - Implementation Status Report:

Work Completed:

- Vikram – 60%Data Collection and algorithm choosing and design
- Shiva Sai – 40% Data Collection and algorithm design
- Sai Ritwik – 80% Data Visualization and Pre-processing
- Swetha Vahana – 20% Data Visualization and took responsibility for output.

References/Bibliography:

Agrawal, R., & Srikant, R. (1994). "Fast algorithms for mining association rules." In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94), 487-499.

NumPy: Harris, C. R., et al. (2020). "Array programming with NumPy." Nature, 585(7825), 357-362.

Pandas: McKinney, W. (2010). "Data structures for statistical computing in Python." In Proceedings of the 9th Python in Science Conference, 51-56.

Matplotlib: Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment." Computing in Science & Engineering, 9(3), 90-95.

Apyori: Rasckhaert, F. (2020). "Apyori: A simple implementation of Apriori algorithm in Python." Retrieved from https://github.com/ymoch/apyori

Github link:
https://gist.github.com/Vikramkotlo/bb7f2c90845e61baa552b790fe498082