

Lab Report on
Market Basket Analysis



Submitted to
Department of Computer Science and Engineering
Nepal Engineering College

in Partial Fulfillment of the
Requirements for the Degree of B.E. in Computer

Submitted By
Dipson Thapa (020-313)

Date: 28/07/2024

Market Basket Analysis

Introduction:

The Apriori algorithm is a cornerstone in data mining, specifically designed for uncovering frequent itemsets and association rules within extensive transaction databases. By identifying groups of items that frequently appear together in transactions, referred to as frequent itemsets, the algorithm uses a minimum support threshold to ensure the significance of these groupings. From these itemsets, association rules are derived, such as "If {itemset A}, then {itemset B}," which elucidate relationships between different itemsets. These rules are assessed based on metrics like confidence, indicating the likelihood of the rule being valid, and lift, which evaluates the strength of the association. The algorithm iteratively expands itemsets and prunes those not meeting the minimum support threshold, making it highly efficient for large datasets. Apriori is extensively employed in market basket analysis to help retailers understand customer purchasing patterns, enhancing strategies like product placement and cross-selling.

Dataset:

The dataset utilized for this market basket analysis is the Grocery Store Data Set, which details sales of various items. You can access it

Link: <https://www.kaggle.com/datasets/mittalvasu95/the-bread-basket>

Notebook name: Market Basket Analysis_020313.ipynb

Libraries Used:

1. Pandas:

- Utilized for data manipulation and analysis. It offers data structures like DataFrames that are ideal for handling tabular data.

2. Pathlib:

- A standard library for handling filesystem paths in a more readable and efficient way.

3. mlxtend.preprocessing.TransactionEncoder:

- This utility from the mlxtend library is used to convert a list of lists (where each inner list represents a transaction) into a one-hot encoded format, which is necessary for applying association rule mining algorithms.

4. mlxtend.frequent_patterns.apriori:

- This function is part of the mlxtend library and is used for identifying frequent itemsets in a dataset.

5. **mlxtend.frequent_patterns.association_rules:**

- This function generates association rules from a DataFrame of frequent itemsets. It calculates metrics like confidence, lift, and others to evaluate the strength of the rules.