Name:A.Vinoth

Artificial intelligence

Market basket analysis

Phase 5

Problem Statement for Market Basket Analysis

Given Problem Statement

"The problem is to perform market basket analysis on a provided dataset to unveil hidden patterns and associations between products. The goal is to understand customer purchasing behavior and identify potential cross-selling opportunities for a retail business. This project involves using
association analysis techniques, such as Apriori algorithm, to find frequently co-occurring products and generate insights for business optimization."

Problem Analysis and Inference

From analyzing the given problem, the task is to perform market basket analysis with the given dataset to find hidden patterns and relations between products if they exist. This is to understand the purchasing behavior of customers and check for potential cross-selling opportunities for retailers. This problem requires the usage of association analysis techniques, to generate insights for business improvements.

Dataset Details

This is the link to the dataset:
https:/www.kaggle.com/datasets/aslanahmedov/market-basket-analysis

The dataset contains transactional data from a retail store over a specified time period. Each row represents a unique transaction, listing the items purchased by a customer. The dataset includes information such as transaction ID, customer ID, and a list of
purchased products.

Product details include product names or IDs, categories, and prices.

It also contains additional metadata such as country of purchase, date, time and other information.

Key Objectives of the Problem

Discover frequent itemsets: Apply the Apriori algorithm to identify which products are frequently purchased together in customer transactions.

Calculate association rules: Establish association rules, including support, confidence, and lift, to quantify the relationships between products.

Uncover cross-selling opportunities: Identify product pairs or sets that exhibit strong associations, enabling the retail business to strategically promote and bundle related products.

Understand customer purchasing behavior: Gain insights into customer preferences and behaviors based on the discovered patterns.

Optimize business strategies: Utilize the analysis findings to enhance product placement, marketing campaigns, and overall business operations.

Expected Results

A comprehensive report summarizing the results of the market basket analysis, highlighting significant patterns and associations among products.

Visualizations, including charts and graphs, to illustrate the frequent itemsets, association rules, and other relevant insights.

Recommendations for the retail business to improve its product offerings, enhance customer experience, and boost revenue based on the analysis results.

Preprocessing:

Preprocessing steps in market basket analysis typically include:

Data Collection:

Gather transaction data that includes details of items purchased by customers.

Data Cleaning:

Remove any duplicate records, missing values, or inconsistencies in the data.

Data Transformation:

Convert the transaction data into a suitable format, such as a binary matrix where rows represent transactions, and columns represent unique items with binary values (1 for presence, 0 for absence).

Support Threshold Setting:

Determine a minimum support threshold, which represents the minimum frequency an itemset must occur in the dataset to be considered for analysis.

Itemset Generation:

Identify frequent itemsets, which are combinations of items that meet the support threshold.

Rule Generation:

Generate association rules from frequent itemsets, which include metrics like confidence and lift to assess the strength of relationships between items.

Rule Pruning:

Remove rules that do not meet desired confidence or lift criteria to focus on more meaningful associations.

Post-Processing:

Interpret and validate the generated rules to extract actionable insights for marketing strategies.

These preprocessing steps help identify patterns and associations among items purchased by customers, which can be valuable for optimizing product placement, promotions, and marketing strategies.

Feature Extraction Techniques:

In market basket analysis, feature extraction techniques are used to transform transaction data into meaningful features that can be used for pattern recognition and rule generation. Some common feature extraction techniques include:

Binary Encoding:

This technique represents each item in a transaction as a binary value (1 for presence, 0 for absence). The binary ma**trix of transactions can serve as a feature representation.**

Count Encoding:

It counts the frequency of each item in a transaction. This can be used to assess the popularity of individual items.

TF-IDF (Term Frequency-Inverse Document Frequency):

Originally used in text analysis, TF-IDF can be adapted to market basket analysis by treating items as "terms" in a "document." It gives higher weight to items that are less frequent across all transactions but common in a specific transaction.

One-Hot Encoding:

This technique creates binary columns for each unique item and marks the presence or absence of each item in a transaction. It's effective when you want to maintain item specificity in the feature representation.

Association Rule Metrics:

Metrics derived from association rules can also be used as features. These include support, confidence, lift, and other rule evaluation measures. They can be used to quantify the strength and importance of item associations.

Sequential Patterns:

In scenarios where the order of item occurrence matters, you can extract sequential patterns as features to capture temporal relationships in transactions.

Market Basket Metrics:

Metrics like "average basket size" or "number of items per transaction" can be used as features to describe transaction characteristics.

Customer Segmentation Features:

Features like customer demographics, purchase history, and loyalty program data can be incorporated to segment customers and identify patterns within each group.

The choice of feature extraction technique depends on the specific goals of the market basket analysis and the nature of the data. It's common to use a combination of these techniques to gain a comprehensive understanding of customer purchase behavior and item associations.

What is Association Rule for Market basket Analysis?

Let I = {I1, I2,..., Im} be an itemset. Let D, the data, be a set of database transactions where each transaction T is a nonempty itemset such that T 谨 I. Each transaction is associated with an identifier, called a TID(orTid). Let A be a set of items(itemset). T is the Transaction which is said to contain A if A 谨 T. An Association Rule is an implication of the form A 院 B, where A 军 I, B 军 I, and A nB = φ.The rule A 院 B holds in the data set(transactions) D with supports, where 's' is the percentage oftransactions in D that contain A U B (that is the union of set A and set B, or, both A and B). This is taken as the probability, P(A U B). Rule A 院 B has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contains B. This is taken to be the conditional probability, like

P(B|A). That is,

support(A 院 B) =P(A U B)

confidence(A 院 B) =P(B|A)

Rules that satisfy both a minimum support threshold (called min sup) and a minimum confidence threshold (called min conf) are called "Strong".

support count(A U B) / support count(A)

Generally, Association Rule Mining can be viewed in a two-step process:-

1. Find all Frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a pre-established minimum support count, min sup.

2. Generate Association Rules from the Frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence.

Association Rule Mining is primarily used when you want to identify an association between different items in a set, then find frequent patterns in a transactional database, relational databases(RDBMS).

Algorithms used in Market Basket Analysis

There are Multiple Techniques and Algorithms are used in Market Basket Analysis. One of the important objectives is “to predict the probability of items that are being bought together by customers”.

Algorithms used in Market Basket Analysis

There are Multiple Techniques and Algorithms are used in Market Basket Analysis. One of the important objectives is “to predict the probability of items that are being bought together by customers”.

=> AIS

=> SETM Algorithm

=> Apriori Algorithm

=> FP Growth