Project:MARKET BASKET INSIGHTS

Phase3:Development part 1

**Introduction:**

Basket analysis, also known as market basket analysis or association rule learning, is a popular technique used in data mining to discover relationships between items in large-scale transaction datasets. This paper provides an in-depth analysis of the theoretical features of basket analysis, with a focus on the most critical steps in the process. In addition, R markdown code examples are provided to demonstrate how these steps can be implemented in practice. Basket analysis is a widely-used method for identifying patterns and associations between items purchased together in retail transactions. This technique has proven to be invaluable for marketing, inventory management, and product placement strategies. The Apriori algorithm and Eclat algorithm are two prominent approaches for generating association rules in basket analysis. This paper will outline the crucial steps in the process and provide R markdown code examples to facilitate understanding and implementation.

**Market basket analysis :**

Market basket analysis is a data mining technique used by retailers to increase sales by better understanding customer purchasing patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings, as well as products that are likely to be purchased together.

**Examples of Market Basket:**

Analysis Another example might be an online store examining customer purchase data to see which goods are often purchased together. The study may indicate that customers who buy laptops also buy mouse pads, extra hard drives, and extended warranties.

**Application of Market Basket Analysis** Market basket analysis has several applications in different industries. Some of the applications are: Retail Industry: Market basket analysis is widely used in the retail industry to identify the relationship between different products and how they are purchased together.

**Market basket transaction:**

In simple terms Basically, Market basket analysis in data mining is to analyze the combination of products which been bought together. This is a technique that gives the careful study of purchases done by a customer in a supermarket. This concept identifies the pattern of frequent purchas.

**Data preprocessing and transaction encoding:**

The first step in basket analysis is data preprocessing and transaction encoding. This step involves converting raw transaction data into a suitable format for analysis. This usually involves transforming the data into a binary matrix, where each row represents a transaction, and each column represents an item. The value in each cell is 1 if the item is present in the transaction and 0 if not.

*# Load required libraries*

**library**(arules)

## Loading required package: Matrix

##

## Attaching package: 'arules'

## The following objects are masked from 'package:base':

##

## abbreviate, write

**library**(datasets)

*# Load sample data*

data(Groceries)

*# Convert to transactions object*

groceries\_trans <- as(Groceries, "transactions")

**Generating frequent itemsets:**

The second step is to generate frequent itemsets, which are sets of items that occur together frequently in the dataset. This is typically done using a user-defined support threshold. The support of an itemset is the proportion of transactions that contain the itemset. In the case of the Apriori algorithm, this step involves iteratively finding frequent k-itemsets and pruning infrequent itemsets.

*# Define support threshold*

support\_threshold <- 0.01

*# Generate frequent itemsets*

frequent\_itemsets <- apriori(groceries\_trans, parameter = list(support = support\_threshold, target = "frequent itemsets"))

## Apriori

##

## Parameter specification:

## confidence minval smax arem aval originalSupport maxtime support minlen

## NA 0.1 1 none FALSE TRUE 5 0.01 1

## maxlen target ext

## 10 frequent itemsets TRUE

##

## Algorithmic control:

## filter tree heap memopt load sort verbose

## 0.1 TRUE TRUE FALSE TRUE 2 TRUE

##

## Absolute minimum support count: 98

##

## set item appearances ...[0 item(s)] done [0.00s].

## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].

## sorting and recoding items ... [88 item(s)] done [0.00s].

## creating transaction tree ... done [0.00s].

## checking subsets of size 1 2 3 4 done [0.00s].

## sorting transactions ... done [0.00s].

## writing ... [333 set(s)] done [0.00s].

## creating S4 object ... done [0.00s].

**Generating association rules:**

Once frequent itemsets have been identified, the next step is to generate association rules. An association rule is an expression of the form X => Y, where X and Y are disjoint itemsets. The strength of an association rule is measured using metrics such as confidence, lift, and leverage. Confidence is the conditional probability of finding Y in a transaction given that X is present, while lift measures how much more likely Y is to be present in a transaction with X compared to a random transaction.

*# Define confidence threshold*

confidence\_threshold <- 0.5

*# Generate association rules*

association\_rules <- apriori(groceries\_trans, parameter = list(support = support\_threshold, confidence = confidence\_threshold, target = "rules"))

## Apriori

##

## Parameter specification:

## confidence minval smax arem aval originalSupport maxtime support minlen

## 0.5 0.1 1 none FALSE TRUE 5 0.01 1

## maxlen target ext

## 10 rules TRUE

##

## Algorithmic control:

## filter tree heap memopt load sort verbose

## 0.1 TRUE TRUE FALSE TRUE 2 TRUE

##

## Absolute minimum support count: 98

##

## set item appearances ...[0 item(s)] done [0.00s].

## set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].

## sorting and recoding items ... [88 item(s)] done [0.00s].

## creating transaction tree ... done [0.00s].

## checking subsets of size 1 2 3 4 done [0.00s].

## writing ... [15 rule(s)] done [0.00s].

## creating S4 object ... done [0.00s].

**Evaluating and selecting rules:**

The final step in basket analysis is evaluating and selecting association rules based on their quality metrics. Rules with high lift and confidence values are usually of greater interest, as they indicate strong relationships between itemsets. The choice of rules to retain depends on the specific problem context and the desired outcomes.

*# Sort rules by lift*

sorted\_rules <- sort(association\_rules, by = "lift", decreasing = TRUE)

*# Inspect the top 5 rules*

inspect(sorted\_rules[1:5])

## lhs rhs support

## [1] {citrus fruit, root vegetables} => {other vegetables} 0.01037112

## [2] {tropical fruit, root vegetables} => {other vegetables} 0.01230300

## [3] {root vegetables, rolls/buns} => {other vegetables} 0.01220132

## [4] {root vegetables, yogurt} => {other vegetables} 0.01291307

## [5] {curd, yogurt} => {whole milk} 0.01006609

## confidence coverage lift count

## [1] 0.5862069 0.01769192 3.029608 102

## [2] 0.5845411 0.02104728 3.020999 121

## [3] 0.5020921 0.02430097 2.594890 120

## [4] 0.5000000 0.02582613 2.584078 127

## [5] 0.5823529 0.01728521 2.279125 99

**Conclusion:**

Basket analysis with association rules is a powerful technique for discovering relationships between items in large-scale transaction data. The key steps in this process include data preprocessing and transaction encoding, generating frequent itemsets, generating association rules, and evaluating and selecting rules.

.