**MARKET BASKET ANALYSIS**

**Abstract**

Market basket analysis is a crucial component of the analytical system used in retail organizations to put products and create sales promotions for various customer categories in order to increase customer happiness and, consequently, the profit of the supermarkets. The well-known ARM activity known as MBA is what business intelligence judgments are ultimately based on. An effective algorithm must mine frequent item sets and derive rules to create classifiers with high accuracy. Here, regular item set mining is used to address the problems facing a major retailer. Files are used as the project's database. The rows of this matrix form represent lists of things, while the columns represent transactions for the item sets and transactions of items. The Apriori algorithm is used to extract the frequent item sets from the database, and after that, the association rules are produced. The study helps store management understand how the products that people buy are related to one another.

**Introduction**

Retail has progressed from regular corner stores in the 1900s to modern e-commerce, which has rattled the retail sector to its core. A new era of limitless opportunities for trade and consumers has emerged as a result of this altering process. Today's consumers can choose from a variety of solutions independently in practically every field. When a customer needed to purchase something in the past, his sole option was to select a product from the store's catalogue. The number of possibilities has, however, grown dramatically with the advent of the information age and globalization. Nowadays, buyers have a vast selection of items and variations to pick from. Geographical, seasonal, and other restrictions are no longer a problem. Once seen as luxury items, these objects are now thought of as everyday items. All of this gave businesses today's seemingly endless opportunities. But because of this boundless opportunity, a vast number of brand-new rivals entered the market. Retail businesses look for marketing tactics to draw in new clients or retain existing ones. Only new marketing approaches, which include effective advertising and sound product planning, could ameliorate the problem. Market basket analyses have demonstrated exceptional success in other nations where they have been used. As a result, international retailers like Tesco and Walmart have begun employing market basket research to increase profits. Yet in order to use market basket analysis to gain insights, we must have knowledge about our customers' purchases, specifically what they buy and when they buy it. As a result, the information on customer purchases that is based on their behavior becomes important.

Over the past two decades, there has been a massive increase in the amount of data available, although not all of it is relevant. To extract valuable information, companies have turned to data mining, also known as the Knowledge Discovery and Data (KDD) process. Data mining involves searching through large volumes of data to discover useful information, and it is widely used in a variety of fields, including manufacturing, marketing, CRM, retail trade, psychology, and education.

There are numerous techniques for data mining, such as Neural Networks, Artificial information, Classification, Association, Prediction, Clustering, Regression, Sequence discovery, and Visualization, all of which can help organizations find solutions to their problems and extract meaningful knowledge.

Market basket analysis is a technique used to analyze the itemsets that customers purchase together on a single visit to a store, known as the market basket. During a visit to a supermarket, customers often purchase products from different categories and put them all together in one basket, which is considered a single transaction. Market basket analysis involves analyzing these baskets collectively to uncover associations and connections between specific objects, discover customer behaviors, and identify relationships between items. In retail, this technique is used to understand if a customer who buys a certain group of items is more or less likely to buy another group of items. For example, it is common for customers who buy beer to also purchase chips. Companies are interested in analyzing these purchase behaviors to create new marketing and sales strategies that can benefit the company and improve customer experiences. Supermarkets and sellers use market basket analysis to identify which items are frequently purchased together, and then use this information to create targeted promotions, optimize store layouts, and improve inventory management.

**Related Work**

|  |  |
| --- | --- |
| S.no | Title |
| 1. | Market basket analysis in a multiple store environment Yen-Liang Chen a, Kwei Tang b, Ren-Jie Shen a, Ya-Han Hu a |
| 2. | Association Rule – Extracting Knowledge Using Market Basket Analysis  Raorane A.A.1 Kulkarni R.V.2 and Jitkar B.D.3 |
| 3. | Market Basket Analysis Robert C. Blattberg, Byung-Do Kim & Scott A. Neslin |

**Methods**

The Apriori algorithm is a popular data mining technique used to identify frequently occurring item sets in a dataset. The algorithm is based on the principle that if an item set is frequent, then all of its subsets must also be frequent. By leveraging this observation, the algorithm is able to narrow down the search space and avoid searching through all possible item sets, which can be computationally expensive.

The Apriori algorithm is comprised of two main steps. The first step involves generating candidate item sets that meet a minimum support threshold. The support of an itemset is defined as the proportion of transactions in the dataset that contains the itemset. To generate candidate item sets, the algorithm first identifies all frequent 1-itemsets (i.e., items that meet the minimum support threshold) and then combines these frequent 1-itemsets to generate candidate 2-itemsets, and so on, until no new candidate item sets can be generated.

The second step involves pruning infrequent item sets from the list of candidate item sets. To do this, the algorithm scans the dataset to count the support of each candidate item set. Any candidate item set that does not meet the minimum support threshold is removed from the list of candidate item sets. The remaining item sets are considered frequent and are outputted by the algorithm as the frequent item sets.

The Apriori algorithm can be extended to generate association rules, which are rules of the form "if A then B", where A and B are item sets. To generate association rules, the algorithm iterates over the frequent item sets and generates all possible rules for each item set. The quality of the rules can be evaluated using metrics such as support, confidence, and lift.

Overall, the Apriori algorithm is a straightforward and powerful technique for discovering frequent item sets and generating association rules in a dataset.

The ECLAT (Equivalence Class Clustering and Bottom-Up Lattice Traversal) algorithm is another widely used algorithm in data mining for discovering frequent item sets in a dataset. Similarly, to the Apriori algorithm, it relies on the idea that if an item set is frequent, then all of its subsets must also be frequent. However, the ECLAT algorithm has a different approach for generating frequent item sets.

The ECLAT algorithm consists of two primary steps. The first step involves constructing an itemset tree to represent the item sets in the dataset. The tree is built by going through each transaction in the dataset and adding it to the tree. Each node in the tree corresponds to an item, and edges between nodes indicate the presence of an itemset containing those items. For instance, if transaction 1 contains items A, B, and C, and transaction 2 contains items A and B, then the tree would have nodes for A, B, and C, and edges between A and B, A and C, and B and C.

The second step involves traversing the itemset tree to identify frequent item sets. The algorithm starts from the root of the tree and examines each node in turn. For each node, it recursively finds all the nodes that are linked to it in the tree and creates a new itemset that includes the current node and all of its connected nodes. The support of the itemset is then calculated by counting the number of transactions that contain the itemset. If the item set meets the minimum support threshold, it is considered frequent and added to the list of frequent item sets. The algorithm then continues the traversal by recursively calling itself on each child node of the current node.

The ECLAT algorithm can also be used to generate association rules in a similar way to the Apriori algorithm. Overall, the ECLAT algorithm can be more efficient than the Apriori algorithm in certain scenarios, particularly when the item sets in the dataset have low cardinality (i.e., few unique items), since the itemset tree can be pruned more aggressively than the candidate itemset generation in Apriori.

The FP-Growth algorithm is a well-known approach for finding frequent item sets in large datasets, which is similar to the Apriori and ECLAT algorithms. FP stands for "Frequent Pattern", which is a set of items that commonly appear together in a dataset.

The FP-Growth algorithm consists of two primary steps they are Constructing the frequent pattern tree: In this step, the algorithm builds a tree structure called the frequent pattern tree (FP-tree) by scanning the dataset twice. During the first scan, it identifies all the frequent single items that satisfy the minimum support threshold and creates a header table that stores each item along with a pointer to the first occurrence of that item in the dataset. In the second scan, the algorithm reads the dataset again and constructs the FP tree by adding each transaction to the tree, where each transaction is ordered by the frequency of its items in descending order. Mining frequent item sets: In this step, the algorithm recursively generates frequent item sets from the FP-tree. It starts at the bottom of the tree and iterates over each item in the header table in descending order of frequency. For each item, it follows its link to the corresponding subtree in the FP-tree and generates a conditional pattern base by removing the item from each transaction in the subtree and keeping track of the support count of each remaining item. It then recursively constructs a new FP-tree from the conditional pattern base and mines frequent item sets from this tree. This process continues until all frequent item sets have been generated.

The FP-Growth algorithm is known for its efficiency in certain scenarios, particularly when the item sets in the dataset have high cardinality (i.e., many unique items) since it avoids the costly candidate itemset generation phase of the Apriori algorithm.

**Results**

Distribution of the features:

Icon

Description automatically generated with medium confidence

The results of apriori algorthm

support itemsets

0 0.045803 (6 RIBBONS RUSTIC CHARM)

1 0.031124 (60 CAKE CASES VINTAGE CHRISTMAS)

2 0.040339 (60 TEATIME FAIRY CAKE CASES)

3 0.046928 (ALARM CLOCK BAKELIKE GREEN)

4 0.035142 (ALARM CLOCK BAKELIKE PINK)

... ... ...

126 0.030535 (JUMBO BAG RED RETROSPOT, JUMBO BAG BAROQUE B...

127 0.042053 (JUMBO BAG RED RETROSPOT, JUMBO BAG PINK POLKA...

128 0.035196 (JUMBO SHOPPER VINTAGE RED PAISLEY, JUMBO BAG ...

129 0.037392 (JUMBO BAG RED RETROSPOT, JUMBO STORAGE BAG SUKI)

130 0.032517 (LUNCH BAG RED RETROSPOT, LUNCH BAG BLACK SKU...

The results of ECLAT

Frequent itemsets of size 1:

('LUNCH BAG SUKI DESIGN',)

('RED RETROSPOT CHARLOTTE BAG',)

('GARDENERS KNEELING PAD KEEP CALM',)

('PAPER CHAIN KIT VINTAGE CHRISTMAS',)

('LUNCH BAG CARS BLUE',)

('LUNCH BAG PINK POLKADOT',)

('LUNCH BAG APPLE DESIGN',)

('PACK OF 72 RETROSPOT CAKE CASES',)

('REGENCY CAKESTAND 3 TIER',)

('POPCORN HOLDER',)

('JUMBO BAG RED RETROSPOT',)

('LUNCH BAG RED RETROSPOT',)

('6 RIBBONS RUSTIC CHARM',)

('SET OF 3 CAKE TINS PANTRY DESIGN',)

('SET/5 RED RETROSPOT LID GLASS BOWLS',)

('VINTAGE SNAP CARDS',)

('60 TEATIME FAIRY CAKE CASES',)

('PACK OF 60 PINK PAISLEY CAKE CASES',)

('NATURAL SLATE HEART CHALKBOARD',)

('JAM MAKING SET PRINTED',)

('HEART OF WICKER SMALL',)

('BAKING SET 9 PIECE RETROSPOT',)

('PLEASE ONE PERSON METAL SIGN',)

('SET OF 6 SPICE TINS PANTRY DESIGN',)

('SPOTTY BUNTING',)

('CHARLOTTE BAG SUKI DESIGN',)

('CHOCOLATE HOT WATER BOTTLE',)

('PARTY BUNTING',)

('LUNCH BAG BLACK SKULL.',)

('ASSORTED COLOUR BIRD ORNAMENT',)

('JUMBO BAG VINTAGE LEAF',)

('WOODEN PICTURE FRAME WHITE FINISH',)

('RETROSPOT TEA SET CERAMIC 11 PC',)

('JUMBO BAG ALPHABET',)

('JAM MAKING SET WITH JARS',)

('JUMBO BAG APPLES',)

('HEART OF WICKER LARGE',)

('RABBIT NIGHT LIGHT',)

('JUMBO SHOPPER VINTAGE RED PAISLEY',)

('JUMBO BAG PINK VINTAGE PAISLEY',)

('GREEN REGENCY TEACUP AND SAUCER',)

('RECIPE BOX PANTRY YELLOW DESIGN',)

('HOT WATER BOTTLE KEEP CALM',)

('REX CASH+CARRY JUMBO SHOPPER',)

('JUMBO BAG PINK POLKADOT',)

('GARDENERS KNEELING PAD CUP OF TEA',)

('ALARM CLOCK BAKELIKE RED',)

('SET/20 RED RETROSPOT PAPER NAPKINS',)

('JUMBO STORAGE BAG SUKI',)

("PAPER CHAIN KIT 50'S CHRISTMAS",)

('WOODEN FRAME ANTIQUE WHITE',)

('LUNCH BAG WOODLAND',)

('HOME BUILDING BLOCK WORD',)

('ROSES REGENCY TEACUP AND SAUCER',)

('JUMBO BAG BAROQUE BLACK WHITE',)

('VICTORIAN GLASS HANGING T-LIGHT',)

('JUMBO BAG STRAWBERRY',)

('LUNCH BAG SPACEBOY DESIGN',)

('ALARM CLOCK BAKELIKE GREEN',)

('WHITE HANGING HEART T-LIGHT HOLDER',)

('SET OF 4 PANTRY JELLY MOULDS',)

('LUNCH BAG SUKI DESIGN', 'RED RETROSPOT CHARLOTTE BAG')

('GARDENERS KNEELING PAD KEEP CALM', 'LUNCH BAG SUKI DESIGN')

('JUMBO BAG PINK VINTAGE PAISLEY', 'PACK OF 72 RETROSPOT CAKE CASES')

('GREEN REGENCY TEACUP AND SAUCER', 'PACK OF 72 RETROSPOT CAKE CASES')

('PACK OF 72 RETROSPOT CAKE CASES', 'RECIPE BOX PANTRY YELLOW DESIGN')

The results of FP Growth

* It resulted in no frequency items

**Conclusion**

* The apriori algorithm and ECLAT algorithms are the best algorithms whereas in the ECLAT we can define the support values same as the apriori algorithm and for the fpgrowth algorithm the result is a null set which means there are no frequency items, so the Apriori and the ECLAT are the best algorithms to perform the operations.
* Later these algorithms can be deployed and can be seen the changes in the revenue of the organization.

**Reference**

* Market Basket Analysis Robert C. Blattberg, Byung-Do Kim & Scott A. Neslin
* Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining panel Manpreet Kaur, Shivani Kang
* Market basket analysis with association rules Yüksel Akay Ünvan