

MARKET BASKET INSIGHTS USING MACHINE LEARNING

510521205039:S.SARAN KIUMAR

INTRODUCTION:

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.

The adoption of market basket analysis was aided by the advent of electronic point-of-sale (POS) systems. Compared to handwritten records kept by store owners, the digital records generated by POS systems made it easier for applications to process and analyze large volumes of purchase data.

Implementation of market basket analysis requires a background in statistics and data science, as well as some algorithmic computer programming skills. For those without the needed technical skills, commercial, off-the-shelf tools exist.

ABSTRACTION OF MARKET BASKET ANALYSIS:

In the retail industry, market basket analysis explores the relationship between products by considering the co-occurrence of purchases in previous transactions. Association analysis is a generalization of applications, like market basket analysis, and is now commonly applied in clickstream analysis, cross-selling recommendation engines, and information security. Association analysis is an unsupervised data science technique where there is no target variable to predict. Instead, the algorithm reviews each transaction containing a number of items (products) and extracts useful relationship patterns among the items in the form of rules. The challenge in association analysis is to differentiate a significant observation against unscrupulous rules. The Apriori and Frequent Pattern Growth algorithms offer efficient approaches to extract these rules from large datasets in the transaction logs.

MODULE FOR THE MARKET BASKET INSIGHTS:

A module for market basket insights is a software component that can be used to analyze customer purchase data to identify patterns and trends. This information can then be used to develop marketing strategies, improve product placement, and increase sales.

Market basket analysis modules typically work by using a variety of algorithms to identify frequently purchased items, item sets, and association rules. For example, an algorithm might be used to identify all pairs of items that are purchased together more than a certain percentage of the time. This information could then be used to create a promotion on the two items, or to place them next to each other in the store.

Market basket analysis modules can also be used to identify customer segments. For example, a module might be used to identify a group of customers who frequently purchase diapers and baby wipes. This information could then be used to target these customers with special promotions or to develop new products that appeal to their need

Market basket analysis modules are used by a wide variety of businesses, including retailers, e-commerce companies, and financial institutions. They can be used to improve the efficiency of operations, increase sales, and improve customer satisfaction.

Here are some examples of how a market basket insights module can be used:

A grocery store could use the module to identify frequently purchased items and place them together in the store. For example, if the module identifies that milk and bread are frequently purchased together, the store could place them next to each other in the dairy aisle.

An e-commerce company could use the module to recommend products to customers based on their past purchases. For example, if a customer has purchased a pair of shoes, the module could recommend other shoes and accessories that are frequently purchased together.

A financial institution could use the module to identify customer segments who are at risk of churning. For example, the module might identify a group of customers who have recently stopped using their credit cards. The financial institution could then contact these customers and offer them special incentives to keep using their cards.

Market basket insights modules are a powerful tool that can be used by businesses of all sizes to improve their operations and increase their profits.

Dataset Link: <https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

Design Thinking:

DATA SOURCE FOR MARKET BASKET INSIGHTS:

Obtaining data for market basket insights typically involves collecting transactional data from retail or e-commerce businesses. This data is often sourced from point-of-sale (POS) systems, e-commerce platforms, or customer relationship management (CRM) systems. Here are some common data sources for market basket analysis:

1. **Surveys and Feedback:** Customer surveys and feedback can provide qualitative data that complements quantitative transactional data. This can help in understanding customer motivations and preferences.

2. **Third-party Data Providers:** Some companies purchase data from third-party providers that aggregate transactional data from various retailers. These datasets can be used for broader market basket analysis across multiple businesses.
3. **Online Marketplaces:** If you operate an online marketplace, you can collect data on transactions and user behavior on your platform.
4. **Social Media and Web Analytics:** While not transactional data, social media and web analytics data can provide insights into customer sentiment, trends, and preferences that may be relevant to market basket analysis.
5. **Open Data Sources:** In some cases, publicly available data, such as government statistics or industry reports, can provide insights into market trends and consumer behavior.

DATA PREPROCESSING FOR MARKET BASKET INSIGHTS:

1. Data Collection and Integration:

- Collect data from various sources, such as POS systems, e-commerce platforms, or CRM systems.
- Integrate data from different sources if necessary, ensuring that it's in a consistent format.

2. Data Cleaning:

- Remove duplicates: Check for and remove duplicate transactions or records to ensure data accuracy.
- Handle missing values: Determine how to handle missing data, either by imputing values or removing rows with missing information.
- Outlier detection: Identify and address any outliers that could distort the analysis.

3. Data Transformation:

- Transaction-level data: Convert the data into a transaction-level format where each row represents a unique transaction with items purchased.
- Binning or discretization: Group continuous numerical data (e.g., item prices) into discrete bins if needed.
- Encoding categorical data: Convert categorical variables (e.g., product names) into numerical representations (e.g., one-hot encoding or label encoding).

4. Data Aggregation:

- Group transactions: Group transactions by unique identifiers (e.g., transaction ID or customer ID) to analyze market baskets at various levels (e.g., basket analysis per customer or per store).
- Calculate item frequencies: Count the frequency of each item's appearance in transactions to determine its popularity.

5. **Market Basket Analysis Data Structure:**

- Create a binary item-item matrix: Create a matrix where rows represent transactions or customers and columns represent items. Each cell indicates whether an item was present (1) or absent (0) in a transaction.
- Calculate item support: Calculate the support for each item (the percentage of transactions containing that item) to identify frequent items.
-

ASSOCIATION ANALYSIS FOR MARKET BASKET INSIGHTS:

Association analysis, often used for market basket insights, is a data mining technique that discovers patterns, relationships, and associations within transactional data. It helps businesses understand which items are frequently purchased together, enabling them to make informed decisions regarding product placement, cross-selling, and promotional strategies. The Apriori algorithm and FP-Growth algorithm are commonly used for association analysis. Here's how you can perform association analysis for market basket insights:

1. **Data Preprocessing:**

- As mentioned earlier, clean and preprocess your transactional data to get it into a suitable format for analysis.
- Convert your data into a transaction-item format, where each row represents a transaction, and columns represent items, with binary values indicating whether an item was purchased in that transaction.

2. **Setting Parameters:**

- Determine the minimum support and minimum confidence thresholds based on your specific business goals and dataset characteristics.
 - **Minimum Support:** The minimum percentage of transactions that an itemset must appear in to be considered frequent. This helps filter out infrequent itemsets.

3. **Apriori Algorithm:**

- Apriori is one of the most widely used algorithms for association analysis.
- It employs a breadth-first search strategy to generate frequent itemsets and discover association rules.
- The algorithm starts with individual items as candidates and progressively builds larger itemsets by joining frequent itemsets from the previous iteration.
- Prune infrequent itemsets to reduce computational complexity.

4. **FP-Growth Algorithm** (Optional):

- FP-Growth is an alternative algorithm that is more efficient than Apriori, especially for large datasets.

- It constructs a compact data structure called an FP-tree and uses it to mine frequent itemsets.
- FP-Growth is often faster than Apriori but may require more memory.

5. **Mining Frequent Itemsets:**

- Run the chosen algorithm to discover frequent itemsets based on the specified support threshold.
- Frequent itemsets are sets of items that occur together frequently in transactions.

INSIGHTS GENERATION FOR MARKET BASKET INSIGHTS:

Generating actionable insights from market basket analysis is a crucial step in making the most of the patterns and associations discovered in transactional data. Here's a guide on how to generate insights from market basket analysis:

1. **Understand Your Business Objectives:**

- Clearly define the specific business objectives or questions you want to address through market basket analysis. What are you trying to achieve? For example, are you looking to increase cross-selling, optimize product placement, or enhance customer experience?

2. **Segmentation:**

- Segment your data: Divide your data into meaningful segments, such as customer segments or store locations. Analyzing different segments separately can reveal specific insights tailored to each group.

3. **Identify Key Performance Indicators (KPIs):**

- Determine the KPIs that are relevant to your objectives. Common KPIs for market basket insights include:
 - **Basket Size:** Average number of items in a transaction.
 - **Support:** Percentage of transactions containing a specific item or itemset.
 - **Confidence:** The likelihood that a rule is true.
 - **Lift:** A measure of how much more likely an item is to be purchased when another item is purchased, compared to its standalone probability.

4. **Visualizations:**

- Use data visualization techniques to make insights more accessible:
 - **Bar Charts:** Display the most frequent items or itemsets.
 - **Heatmaps:** Show item associations using colors and cells.
 - **Network Graphs:** Visualize item relationships and connections.
 - **Scatter Plots:** Display relationships between different KPIs, such as support vs. confidence.

5. Interpretation:

- Dive into the generated association rules and patterns:
 - Identify strong rules: Focus on rules with high confidence and lift values.
 - Look for unexpected relationships: Sometimes, the most valuable insights come from unexpected associations.
 - Consider context: Understand the context behind the patterns. For example, is there a seasonal or promotional factor influencing certain associations?
 - Check for temporal trends: Analyze how patterns change over time, which can lead to time-specific marketing strategies.

VISUALIZATION FOR MARKET BASKET INSIGHTS:

Visualizing market basket insights can help make complex patterns and associations more understandable and actionable. Here are several visualization techniques you can use to present market basket analysis results effectively:

1. Bar Charts:

- **Frequent Items:** Create a bar chart showing the most frequently purchased items. This provides a quick overview of popular products.
- **Item Support:** Display a bar chart showing the support (percentage of transactions containing an item) for various items. This can help identify the most common items.

2. Heatmaps:

- **Item Co-occurrence:** Create a heatmap that shows the co-occurrence or association between items. Cells with darker colors represent stronger associations. This is useful for visualizing item relationships.

3. Network Graphs:

- **Association Rules:** Construct a network graph where nodes represent items, and edges represent association rules. The thickness of edges can represent the strength of the association (e.g., lift or confidence).
- **Item Relationships:** Visualize item relationships as a network, making it easy to identify clusters of related products.

4. Scatter Plots:

- **Support vs. Confidence:** Create a scatter plot with support on one axis and confidence on the other. Each point represents an association rule, and you can use size or color to represent lift values.

5. Sankey Diagrams:

- Use a Sankey diagram to illustrate the flow of items from one category to another. This is helpful for understanding how customers move between product categories.

BUSINESS RECOMMENDATION FOR MARKET BASKET INSIGHTS:

Market basket insights can provide valuable guidance for businesses looking to optimize their operations, increase revenue, and enhance customer satisfaction. Here are some business recommendations based on market basket analysis:

1. Cross-Selling and Upselling Strategies:

- **Product Bundling:** Bundle complementary products together and offer them at a slightly discounted price. For example, if customers frequently purchase laptops, bundle them with laptop accessories like bags or mouse pads.
- **Recommendation Engines:** Implement recommendation systems on your website or app to suggest related or frequently co-purchased items to customers during their shopping journey.

2. Store Layout and Product Placement:

- **Optimize Store Layout:** Rearrange in-store displays to place frequently associated items near each other. This encourages customers to pick up additional items as they shop.
- **End-Cap Displays:** Feature frequently purchased products on end-cap displays or at the front of the store to capture shoppers' attention.

3. Inventory Management:

- **Stocking Decisions:** Adjust inventory levels based on the popularity of items and their associations. Ensure that you have sufficient stock of frequently co-purchased items to avoid stockouts.
- **Seasonal Trends:** Recognize seasonal purchasing patterns and adjust inventory accordingly.

4. Promotions and Discounts:

- **Targeted Promotions:** Offer discounts or promotions on items that are frequently associated with one another. For instance, if customers often buy wine and cheese together, create a promotion for both items.
- **Coupon Recommendations:** Send personalized coupons to customers based on their historical purchases and association patterns.

5. Customer Segmentation:

- **Segment Customers:** Group customers based on their purchasing behavior and preferences. Tailor marketing strategies, offers, and recommendations to each segment.

- **VIP Programs:** Create loyalty or VIP programs that reward customers for purchasing associated items or reaching specific spending thresholds.

Implementing Market Basket Analysis in Python:

Here are the steps involved in using the algorithm to implement MBA:

- 1,First, define the minimum support and confidence for the association rule.
- 2.Find out all the subsets in the transactions with higher support(sup) than the minimum support.
- 3.Find all the rules for these subsets with higher confidence than minimum confidence.
- 4.Sort these association rules in decreasing order.
- 5,Analyze the rules along with their confidence and support.

TYPES OF MARKET BASKET INSIGHTS:

- ❖ **Predictive market basket analysis**
- ❖ **Differential market basket analysis**

PREDICTIVE MARKET BASKET ANALYSIS:

This type considers items purchased in sequence to determine cross-sell.

DIFFERENTIAL MARKET BASKET ANALYSIS:

This type considers data across different stores, as well as purchases from different customer groups during different times of the day, month or year. If a rule holds in one dimension, such as store, time period or customer group, but does not hold in the others, analysts can determine the factors responsible for the exception. These insights can lead to new product offers that [drive higher sales](#).

Algorithms for market basket analysis:

In market basket analysis, association rules are used to predict the likelihood of products being purchased together. Association rules count the frequency of items that occur together, seeking to find associations that occur far more often than expected.

Algorithms that use association rules include AIS, SETM and Apriori. The Apriori algorithm is commonly cited by data scientists in research articles about market basket analysis and is used to identify frequent items in the database, then evaluate their frequency as the datasets are expanded to larger sizes.

The arules package for R is an open source toolkit for association mining using the R programming language. This package supports the Apriori algorithm, along with the following other mining algorithms:

- ❖ Opusminer

- ❖ arulesNBMiner
- ❖ RKEEL
- ❖ RSarules

Advantages of Market Basket Analysis:

It helps retailers in the following ways:

- ✓ Increases customer engagement
- ✓ Boosts sales and increases RoI
- ✓ Improves customer experience
- ✓ Optimizes marketing strategies and campaigns
- ✓ Helps in demographic data analysis
- ✓ Identifies customer behavior and pattern

DISADVANTAGE OF MARKET BASKET ANALYSIS:

- ✓ Limited Insight into Causal Relationships
- ✓ Dependence on Data Quality
- ✓ Inability to Account for External Factors
- ✓ Time and Resource Intensive
- ✓ Ethical Concerns

Conclusion:

In this tutorial, we discussed Market Basket Analysis and learned the steps to implement it from scratch using Python. We then implemented Market Basket Analysis using Apriori Algorithm. We also looked into the various uses and advantages of this algorithm and learned that we could also use FP Growth and AIS algorithms to implement Market Basket Analysis.