Market Basket Analysis

By

Sree Vikas Sabbani

Abstract:

Nowadays, Machine Learning helps the Retail Industry in many ways. You can imagine that from forecasting the performance of sales to identifying the buyers, there are many applications of AI and ML in the retail industry. Market Basket Analysis (MBA) is a data mining technique which retailers use to increase their sales by better understanding customer buying patterns. It involves analyzing large data sets, such as purchase history, to reveal product groupings and products likely to be bought together.

In this article, we will cover the topic of Market Basket Analysis and different elements related to it.

The History of Market Basket Analysis:

The start of Market Basket Analysis dates to the early 1990s, when major retailers realized how important it was to understand consumer buying behavior. Retailers evaluated purchase combinations by hand before the era of big data, using manual methods that laid the foundation for today's complex algorithms.

Data Mining:

A data mining technique used to uncover purchase patterns in any retail environment is called Market Basket Analysis. In simple terms, Market Basket Analysis is a data mining technique to analyze the combination of products bought together.

This is a technique that gives careful study of purchases made by a customer in a supermarket. This concept finds the pattern of frequent purchases of items by customers. This analysis can help to promote deals, offers, and sales by the companies, and data mining techniques help to achieve this analysis task.

Example:

- Data mining concepts are in use for sales and marketing to supply better customer services, to improve cross-selling opportunities, and to increase direct mail response rates.
- Customer Retention in the form of pattern identification and prediction of likely defections is possible by Data mining.
- The Risk Assessment and Fraud area also uses the data-mining concept for finding inappropriate or unusual behavior etc.

What is Market Basket Analysis?

Market Basket Analysis is a data mining technique which is used by retailers to increase sales by better understanding customer buying patterns. It involves analyzing large data sets, such as buy history, to reveal product groupings, as well as products that are likely to be bought 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.

Example 1:

When you shop on Amazon and look at a product, you often see suggestions like "Customers also bought." This is because Amazon uses Market Basket Analysis to see what items people usually buy together. They then show these suggestions to encourage you to buy more. This is a common tactic to sell more products.

Example 2:

Banks, like Citibank, look at what you buy with your credit card to offer exclusive deals. They sometimes set up booths in big shopping centers to give instant discounts. They also team up with food delivery apps like Swiggy and Zomato to give you exclusive offers when you pay with their credit card.

Association Rule:

Association rule is a relationship between 2 item-sets, and it tells how strong the relationship is.

For example, A and B are 2 sets of different items where A is antecedent, and B is consequent then the relationship between them is called **Association Rule,** and it is denoted by A -> B.

Here, Itemset like A and B can have any number of items i.e. A = {'Banana', 'Milk'} and B = {Vanilla Ice-cream}. Similarly, A could also be {'Milk'} and B = {'Bread'}. To be noted, for association rule A->B, A means items which are bought together or something which is already happened, and B means items which could be recommended or something that is likely to happen next.

There are different metrics that quantify the strength of the relationship of item set in an association rule. These are **Support, Confidence, Lift, Leverage, and Conviction**. Let us discuss them in detail.

Support:

Support of an item-set is the probability of the occurrence of the item-set out of the total number of transactions. It quantifies the share of transactions for an item-set out of all the transactions.

If A is an item-set, then support(A) = n(A)/n(transactions)

Confidence:

If the association is A->B, then the probability of A and B occurring together out of all the transactions where A has already occurred is called confidence.

confidence = n (A ∩ B)/n(A) = P(B|A

Lift:

It talks about the strength of the relationship A->B. Lift values above 1 signify that item

A and B tend to be purchased together, while values close to 1 indicate independence, and values below 1 imply that item A's presence discourages the purchase of item B.

lift = support (A ∩ B)/(support(A) x support(B)) = confidence(A->B)/support(B)

Example:

Assume there are 100 customers.

10 of them bought milk, 8 bought butter and 6 bought both. Support = P (Milk & Butter) = 6/100 = 0.06.

Confidence = support/P(Butter) = 0.06/0.08 = 0.75. Lift = confidence/P(Milk) = 0.75/0.10 = 7.5.

In the example provided, the lift value is 7.5, which exceeds 1. This shows the presence of item A positively influences the likelihood of item B being in the same transaction, suggesting that customers tend to buy these items together.

Features of Market Basket Analysis:

- **Product placement:** Finding products that may often be bought together and arranging the placement of those items (such as in a catalog or on a website) close by to encourage the purchaser to buy both items.
- **Physical shelf arrangement:** An alternate use for physical product placement in a store is to separate items that are often bought at the same time to encourage individuals to wander through the store to find what they are looking for to potentially increase the probability of additional impulse purchases.
- **Up-sell, cross-sell, and bundling opportunities:** Companies may use the affinity grouping of multiple products as a sign that customers may be predisposed to buying the grouped products at the same time. This enables the presentation of items for cross-selling or may suggest that customers may be willing to buy more items when certain products are bundled together.
- **Customer retention:** When customers contact a business to sever a relationship, a company representative may use Market Basket Analysis to figure out the right incentives to offer to keep the customer's business.

Requirements of Market Basket Analysis:

- **Data Granularity**: Market Basket Analysis thrives on detailed transactional data, highlighting individual product purchases.

- **Computational Infrastructure**: Powerful computational tools are essential to processing vast datasets.

- **Domain Expertise**: Knowledge of the retail landscape enhances the quality of insights derived.

Apriori Algorithm:

In this section, we are going to explore the Apriori algorithm, which is a data mining technique used for large datasets. It is used for extracting the relationship within a dataset like a detective trying to find out which things people often buy together when shopping. Rakesh Agrawal and Ramakrishnan Srikant proposed this algorithm in 1994. The Apriori algorithm will help retailers find the relationship between items and help them make strategies that will impact their sales and marketing in a positive way. The Apriori algorithm says that a child item is frequent if its item is frequent.
Apriori algorithm is usually seen as a twostep process:

The first one is frequent itemset mining, which finds groups of items that are bought often. We call these groups "frequent item-sets." If people buy a Phone and Phone case at least 100 times out of 1000 transactions, the algorithm will consider the Phone and Phone case as part of a frequent item set. To do this it involves only two steps, which are Pruning and Joining.

The other one is Rule mining involves finding co-occurrence or correlation within frequent item sets to generate association rules.
Let us understand Apriori algorithm with an example:

Step 1:
Assume that you have a small transaction data of an electronic store:
T1 - {phone, adaptor, phone case}
T2 – {phone case, adaptor}
T3 – {phone, adaptor}
T4 – {phone, phone case}
T5 – {adaptor, phone case}

Step 2:
Now let us calculate support for each item in the above itemset. The minimum support is 2.

- Support for {phone} is 3
- Support for {adaptor} is 4
- Support for {phone case} is 4

As we have the support for each item above 2, we are not eliminating any item to the next iteration.

Step 3:
In this step, let us take ordered pairs of the items.

- Support for {phone, adaptor} is 2
- Support for {phone, phone case} is 2
- Support for {adaptor, phone case} is 3

Since all the order pairs satisfy the support condition, we are not eliminating any item sets here for the next iteration.

Step 4:
We will be creating triplets of the items and check if they meet the minimum support value.

- Support of {phone, phone case, adaptor} is 1
As the support value of above itemset does not meet the minimum support value. Let us consider the ordered pairs for generating association rules.

Step 5:
We will use the confidence for finding out the association rules. We can control the strictness of the rules generated by assuming the minimum confidence value. If the strictness of the rules generated is high, we get more reliable rules and if it is low then we get more rules but those are not reliable. Let us assume the minimum confidence value is 60%.

- Confidence({Phone} -> {Adaptor})
= Support ({Phone, Adaptor}) / Support({Phone})
= 2/3 = 0.67
- Confidence({Adaptor} -> {Phone})
= Support ({Phone, Adaptor}) / Support({Adaptor})
= 2/4 = 0.5
- Confidence({Phone} -> {Phone Case})
= Support ({Phone, Phone Case}) / Support({Phone})
= 2/3 = 0.67
- Confidence ({Phone Case} -> {Phone})
= Support ({Phone, Phone Case}) / Support ({Phone Case})
= 2/4 = 0.5
- Confidence({Adaptor} -> {Phone Case})
= Support ({Adaptor, Phone Case}) / Support({Adaptor})
= 3/4 = 0.75
- Confidence ({Phone Case} -> {Adaptor})
= Support ({Adaptor, Phone Case}) / Support ({Phone Case})
= 3/4 = 0.75

Here we found out the frequent item sets ({Phone} -> {Adaptor}), ({Phone} -> {Phone Case}), ({Adaptor} -> {Phone Case}) and ({Adaptor} -> {Phone Case}) which meets the minimum confidence value.

From the above example problem, we got to know how Apriori algorithm is used to find potential association rules between items in a transaction history of an electronic store.

Other Algorithms Which Are Used in Market Basket Analysis:

As previously discussed, while the Apriori algorithm is a well-known and widely used method in Market Basket Analysis, it is important to note that it is not the only algorithm used in Market Basket Analysis. Market Basket Analysis has a variety of algorithms and techniques, with Apriori being a core one among them. Let us discuss few other algorithms which are used in Market Basket Analysis:

The FP-Growth (Frequent Pattern Growth) algorithm, which is an alternative for Apriori Algorithm, was proposed by Han et al. in 2000. It creates compact data structures which stores the frequency of item sets in a compressed form, which will improve the processing speed. This algorithm is used for large datasets.

Another one is SETM algorithm, it counts the individual item support where it counts the support of individual items and then generate the candidate item sets where the generate the candidate uses relational merge-join operations in generating the candidate item sets.

Lastly, we have Association Rule Mining, which is a very general rule which is used to find the trends between items in an itemset.

There are many more algorithms which serve the purpose of extracting valuable insights from transaction data. The algorithm which should be used totally depends on the requirement of the analysis.

Summary:

In this paper, we have explored the journey from the historical evolution of Market Basket Analysis, progressing to the concept of Association Rules, and finally delving into the realm of associated algorithms.

In this research, we have eliminated less frequent items from the sample dataset by setting minimum values. The outcomes of our study offer valuable insights for organizations seeking to formulate strategies based on customer purchasing behavior.

Limitations:

- **Association rules:** MBA generates association rules based on patterns in the data, but these rules do not guarantee causality.
- **Correlations:** MBA can produce uncertain correlations.
- **Algorithm selection:** Businesses can choose an algorithm that is too slow or complex for their data size, or one that does not capture the nuances of their data relationships.
- **Number of rules:** MBA can produce many.

References:

https://www.linkedin.com/pulse/market-basket-analysis-association-rule-mining-apriori- sharma/
https://medium.com/analytics-vidhya/market-basket-analysis-using-association-rules- 2b0f3e2a897d
https://www.techtarget.com/searchcustomerexperience/definition/market-basket-analysis

https://www.simplilearn.com/what-is-market-basket-analysis-article#:~:text=The%20most%20well%2Dknown%20case,Basket%20Analysis%20cross% 2Dselling%20tactics
https://www.geeksforgeeks.org/market-basket-analysis-in-data-mining/