ASSOCIATION RULE MINING

Association Rule Mining for Market Basket Analysis

What is association rule mining?

Association rule mining is a rule-generating machine learning method where rules tell us about the strength of the relationship between variables in a large dataset. We mainly find usage of association rules in market basket analysis where a strong positive relation between two products makes the seller sell them together and earn more profit. Even the name of this machine learning method explains what we are trying to do. We are finding association rules between variables from a large dataset.

Market Basket Analysis

"Market Basket Analysis" is one of the best applications of machine learning in the retail industry. By analyzing the past buying behavior of customers, we can find out which are the products that are bought frequently together by the customers.


taroonreddy.com

Market Basket Analysis

Frequent item set mining leads to the discovery of associations and correlations between items in huge transactional or relational datasets. With vast amounts of data continuously being collected and stored, many industries are becoming interested in mining such kinds of patterns from their databases.
The disclosure of "Correlation Relationships" among huge amounts of transaction records can help in many decision-making processes such as the design of catalogs, cross-marketing, and behavior customer shopping Analysis.

A popular example of frequent item set mining is Market Basket Analysis.

This process identifies customer buying habits by finding associations between the different items that customers place in their "shopping baskets". The discovery of this kind of association will be helpful for retailers or marketers to develop marketing strategies by gaining insight into which items are frequently bought together by customers.

Market Basket Analysis
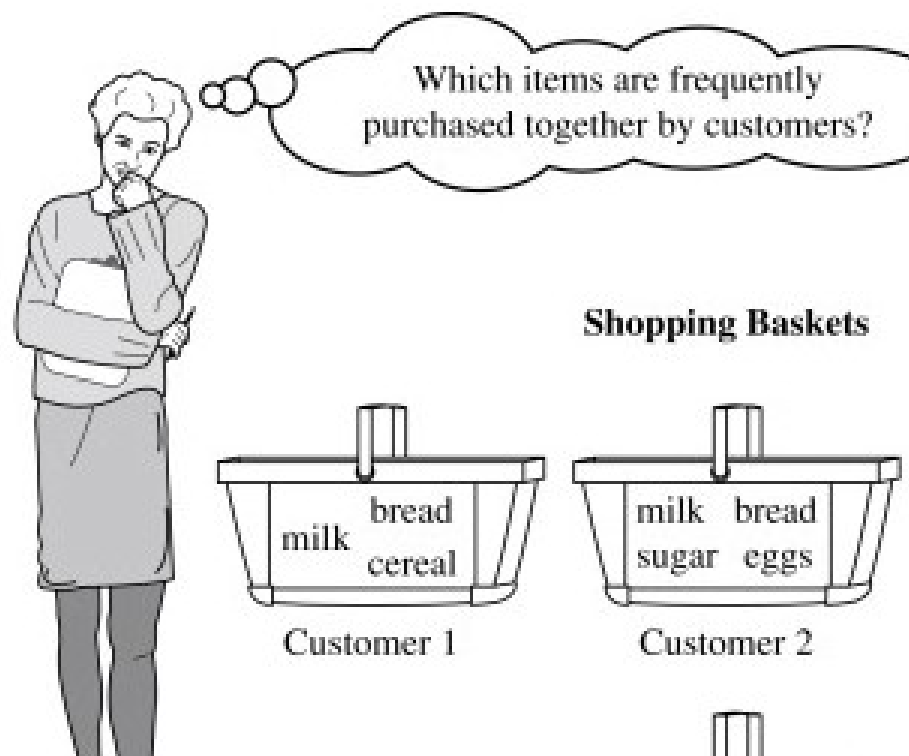
Association Rule Mining for Market Basket Analysis

Let I = {I1, I2,…, Im} be an item set. Let D, the data, be a set of database transactions where each transaction T is a nonempty item set such that **T ⊆ I**.

Each transaction is associated with an identifier, called a TID(or Tid).
Let A be a set of items(item set). T is the Transaction which is said to contain A if **A ⊆ T**.
An **Association Rule** is an implication of the form **A ⇒ B**, where **A ⊂ I, B ⊂ I**, and **A ∩B = ϕ**.

The rule **A ⇒ B** holds in the data set(transactions) D with supports, where 's' is the percentage of transactions in D that contain **A ∪ B** (that is the union of set A and set B, or, both A and B). This is taken as the probability, **P(A ∪ B)**. Rule **A ⇒ 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,

$$support(A \Rightarrow B) = P(A \cup B)$$
$$confidence(A \Rightarrow B) = P(B|A)$$

Association Rule Mining for Market Basket Analysis

$$support(A \Rightarrow B) = P(A \cup B)$$
$$confidence(A \Rightarrow B) = P(B|A)$$

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

$$Confidence(A \Rightarrow B) = P(B|A) =$$
$$support(A \cup B) / support(A) =$$
$$support\ count(A \cup B) / support\ count(A)$$

Association Rule Mining for Market Basket Analysis

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

1. Find all Frequent item sets: *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.*

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".

Apriori Algorithm:

Apriori Algorithm is a widely-used and well-known Association Rule algorithm and is a popular algorithm used in market basket analysis.

It helps to find frequent itemsets in transactions and identifies association rules between these items. The limitation of the Apriori Algorithm is frequent itemset generation.

It needs to scan the database many times which leads to increased time and reduce performance as

It is a computationally costly step because of a huge database.
It uses the concept of Confidence, Support.

FP Growth Algorithm:

FP Growth is known as Frequent Pattern Growth Algorithm.
FP growth algorithm is a concept of representing the data in the form of an FP tree or Frequent Pattern.

Hence FP Growth is a method of Mining Frequent Itemsets.

There is no need for candidate generation to generate the frequent pattern.
This frequent pattern tree structure maintains the association between the itemsets.

A Frequent Pattern Tree is a tree structure that is made with the earlier itemsets of the data. The main purpose of the FP tree is to mine the most frequent patterns.

Every node of the FP tree represents an item of that itemset.

The root node represents the null value whereas the lower nodes represent the itemsets of the data. The association of these nodes with the lower nodes that is between itemsets is maintained while creating the tree.

taroonreddy.com

Association Rule Mining
Case Studies

Association Rule Mining - Concepts

Constraints on below measures are used to select useful and best rules of all the rules given by R
After analyzing these values for all the rules, best rules for WB have been obtained.

Support	Confidence	Lift
• The support Supp(X)=proportion of transactions in the data set which contain the interest.	• The confidence of a rule: Conf(x=>y)= Supp(X U Y)/Supp(X)	• The lift of a rule: Lift(X=>Y)= Supp(X U Y) -------------------------- (Supp(X) X Supp(Y))

E.g.:- Consider rule: {Jack the Ripper (1988)} => {Strawberry Blonde}
Let Jack the Ripper =X and Strawberry Blonde =Y, Then

Support(X U Y)= No of transactions involving both Jack the Ripper and Strawberry Blonde/ Total no of transactions

Confidence= No of transactions where Strawberry Blonde was also bought when Jack the Ripper was bought/ No of transactions where Jack the Ripper was bought

Lift = Ratio of observed support to the expected support

Association Rule Mining - Concepts

Association rule generation is usually split up into two separate steps:

Step #1:
Minimum support is applied to find all frequent itemsets in a database.

Step #2:
These frequent itemsets and the minimum confidence constraint are used to form rules.