Market Basket Analysis

In today's competitive world cost of acquiring new customer base has increased. So it is mandatory to maintain the existing customer base. In order to do so companies practice Cross Selling and Up selling.

Cross Selling : Cross-selling involves selling related, supplementary products or services based on the customer's interest in, or purchase of, one of your company's products.

But how do you know what products to promote together?

Market basket analysis gives you the product affinity insights you need to recommend the items that customers are most likely to buy alongside any given product.

Market basket analysis helps retailers **discover the buying patterns** and the **relationships between items.**

Classic example of Market Basket Analysis are **Burger and Fries, Coffee and Cookies**

But not all relationships are so obvious.

By analysing the contents of transactions and **looking for attributes that appear together**, retailers **learn what combinations of items occur frequently**

With a better understanding of those correlations, they can improve their Upsell and **cross-sell** strategies.

Upon check-out, the items in the cart or basket are recorded.

The same holds true for e-commerce organizations, like Amazon. Compared to traditional bookstores, **Amazon has a much better insight in buying patterns**. From purchasing patterns, Amazon is able to detect which titles are frequently ordered by the same buyers. This knowledge enables Amazon to effectively cross-sell items, by recommending titles to buyers under headings like **"customers who bought this item also bought ..."**

Frequent itemset mining leads to the discovery of associations and correlations among items in large transactional or relational data sets. With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining such patterns from their databases. The discovery of interesting correlation relationships among huge amounts of business transaction records can help in many business decision-making processes such as catalogue design, cross-marketing, and customer shopping behaviour analysis.

This information can lead to increased sales by helping retailers do selective marketing and plan their shelf space.

Market Basket analysis allows seller to

1. Plan the shelf space - knowing customer who buys bread also buys milk – these two products can be place little apart so on the way customer may encounter other products and may buy them.
2. Which items to put on sale at reduced price.
3. If customers tend to purchase computers and printers together, then having a sale on printers may encourage the sale of printers as well as computers.

These patterns can be represented in the form of **association rules**. For example, the information that customers who purchase computers also tend to buy antivirus software at the same time is represented in the following association rule:

computer ⇒ antivirus software [support = 2%, confidence = 60%]

SUPPORT and **CONFIDENCE** are two measures of rule interestingness. They respectively reflect the usefulness and certainty of discovered rules.

A support of 2% for above rule means that 2% of all the transactions under analysis show that computer and antivirus software are purchased together.

A confidence of 60% means that 60% of the customers who purchased a computer also bought the software.

Association rules are considered interesting if they satisfy both a minimum support **threshold** and a minimum confidence threshold. These thresholds can be a set by users or domain experts. Additional analysis can be performed to discover interesting statistical correlations between associated items.

Rules that satisfy both a minimum support threshold (min sup) and a minimum confidence threshold (min conf) are called strong. By convention, we write support and

confidence values so as to occur between 0% and 100%, rather than 0 to 1.0.

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

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

A set of items is referred to as an itemset. An itemset that contains k items is a k-itemset. The set {computer, antivirus software} is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply, as the frequency, support count, or count of the itemset.

Note that the itemset support defined above is sometimes referred to as **relative support**, whereas the **occurrence frequency** is called the **absolute support**. If the relative support of an itemset I satisfies a prespecified minimum support threshold (i.e., the absolute support of I satisfies the corresponding minimum support count threshold), then I is a frequent itemset. The set of frequent k-itemsets is commonly denoted by Lk

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

$$= support\ count(A \cup B) / support\ count(A) .$$