# INTRODUCTION TO ASSOCIATION RULES

Association rules are if-then statements that show the probability of relationships between data items within large data sets in various types of databases.

What is association mining?

- Association rule mining

  Finding frequent patterns, associations, correlations or casual structures among item set in transaction databases, relational databases and other information repositories.

- Applications

  Market basket analysis, cross marketing, catalog design, shelf space layout design, etc.

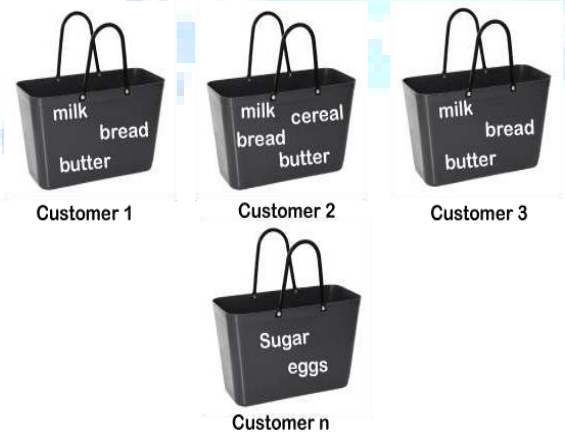
- Examples

  Rule form: Body→ Head[ Support, Confidence]

  buys ( x,"Computer") →buys (x,"Software") [2%,60%]

- Association rule mining is a popular and well researched method for discovering interesting relations between variables in large databases.
- It is intended to identify strong rules discovered in databases using different measures of interestingness.
- Based on the concept of strong rules, Rakesh Agrawal et al. introduced association rule

For example, if a customer buys bread, he most likely can also buy butter, eggs, or milk, so these products are stored within a shelf or mostly nearby. Consider the below diagram:



Association rule mining can be divided into three types of algorithms

Apriori
Eclat
F-P Growth Algorithm

Association rule learning works on the concept of If and Else Statement, such as if A then B.


Here the If element is called antecedent, and then statement is called as Consequent. These types of relationships where we can find out some association or relation between two items is known *as single cardinality*. It is all about creating rules, and if the number of items increases, then cardinality also increases accordingly. So, to measure the associations between thousands of data items, there are several metrics. These metrics are given below:

- **Support**
- **Confidence**
- **Lift**

**Support**

Support is the frequency of A or how frequently an item appears in the dataset. It is defined as the fraction of the transaction T that contains the itemset X. If there are X datasets, then for transactions T, it can be written as:

$$\text{Supp(X)} = \frac{Freq(X)}{T}$$

**Confidence**

**Confidence** indicates how often the rule has been found to be true. Or how often the items X and Y occur together in the dataset when the occurrence of X is already given. It is the ratio of the transaction that contains X and Y to the number of records that contain X.

$$\text{Confidence} = \frac{Freq(X,Y)}{Freq(X)}$$

**Lift**

It is the strength of any rule, which can be defined as below formula:

$$\text{Lift} = \frac{Supp(X,Y)}{Supp(X) \times Supp(Y)}$$

It is the ratio of the observed support measure and expected support if X and Y are independent of each other. It has three possible values:

- If Lift= 1: The probability of occurrence of antecedent and consequent is independent of each other.
- Lift>1: It determines the degree to which the two itemsets are dependent to each other.
- Lift<1: It tells us that one item is a substitute for other items, which means one item has a negative effect on another.