**Association Rule**

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an item set occurs in a transaction. A typical example is Market Based Analysis.

Market Based Analysis is one of the key techniques used by large relations to show associations between items. It allows retailers to identify relationships between the items that people buy together frequently.

The Association rule is very useful in analyzing datasets. The data is collected using bar-code scanners in supermarkets. Such databases consists of a large number of transaction records, which list all items bought by a customer on a single purchase. So the manager could know if certain groups of items are consistently purchased together and use this data for adjusting store layouts, cross-selling, promotions based on statistics.

| TID | Items |
| --- | --- |
| 1 | Bread, Milk |
| 2 | Bread, Diaper, Beer, Eggs |
| 3 | Milk, Diaper, Beer, Coke |
| 4 | Bread, Milk, Diaper, Beer |
| 5 | Bread, Milk, Diaper, Coke |

Association rules are created by thoroughly analyzing data and looking for frequent if/then patterns. Then, depending on the following two parameters, the important relationships are observed:

1. **Support**: Support indicates how frequently the if/then relationship appears in the database.
2. **Confidence**: Confidence tells about the number of times these relationships have been found to be true.

The story goes like this: young American men who go to the stores on Fridays to buy diapers have a predisposition to grab a bottle of beer too. However unrelated and vague that may sound to us laymen, association rule mining shows us how and why!  
Let’s do a little analytics ourselves, shall we?  
**Suppose an X store’s retail transactions database includes the following data:**

* Total number of transactions: 600,000
* Transactions containing diapers: 7,500 (1.25 percent)
* Transactions containing beer: 60,000 (10 percent)
* Transactions containing both beer and diapers: 6,000 (1.0 percent)

However, as surprising as it may seem, the figures tell us that **80% (=6000/7500) of the people who buy diapers also buy beer**.  
This is a significant jump of 8 over what was the expected probability. This factor of increase is known as Lift – which is the ratio of the observed frequency of co-occurrence of our items and the expected frequency.

So, for our example, one plausible association rule can state that the people who buy diapers will also purchase beer with a Lift factor of 8. If we talk mathematically, the lift can be calculated as the ratio of the joint probability of two items x and y, divided by the product of their probabilities.  
***Lift = P(x,y)/[P(x)P(y)]***