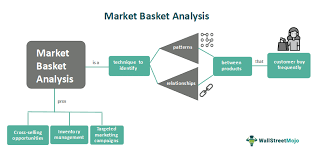
**MARKET BASKET INSIGHTS**

**Name : S.Srigomathi**

**Registernumber : 913321104054**

**Phase -5 : Complete Project**

**Project : Market Basket analysis**

****

**Problem Definition:**

The problem is to perform market basket analysis on a provided dataset to unveil hidden patterns and associations between products. The goal is to understand customer purchasing behaviour and identify potential cross selling opportunities for a retail business. This project involves using association analysis techniques, such as Apriori algorithm, to find frequently co-occuring products and generate insights for business optimization.

**Problem Statements:**

Nowadays people buy daily goods from super market nearby. There are many supermarkets that provide goods to their customer. The problem many retailers face is the placement of the items. They are unaware of the purchasing habits of the customer so they don’t know which items should be placed together in their store. With the help of this application shop managers can determine the strong relationships between the items which ultimately helps them to put products that co-occur together close to one another. Also decisions like which item to stock more, cross selling, up selling, store shelf arrangement are determined.

**Design Thinking**:

1.Data Source

2.Data Preprocessing

3.Association Analysis

4.Insights Generation

5.Visualization

6.Business Recommendations

**Implementing Market Basket Analysis in Python:**

* First, define the minimum support and confidence for the association rule.
* Find out all the subsets in the transactions with higher support(sup) than the minimum support.
* Find all the rules for these subsets with higher confidence than minimum confidence.

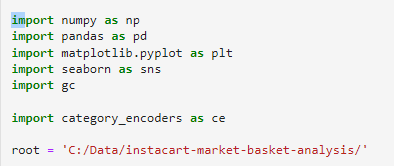
Dataset Link :

(https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis )

**Data Preprocessing:**

Data preprocessing is a crucial step in preparing data for market basket analysis. Properly cleaned and organized data ensures the accuracy and effectiveness of the analysis. Here are the key steps involved in data preprocessing for market basket analysis:

**Feature Extraction**





**Choice of Algorithm:**

Algorithms used in Market Basket Analysis, there are multiple data mining techniques and algorithms used in Market Basket Analysis. One of the important objectives is “to predict the probability of items that are being bought together by customers.”

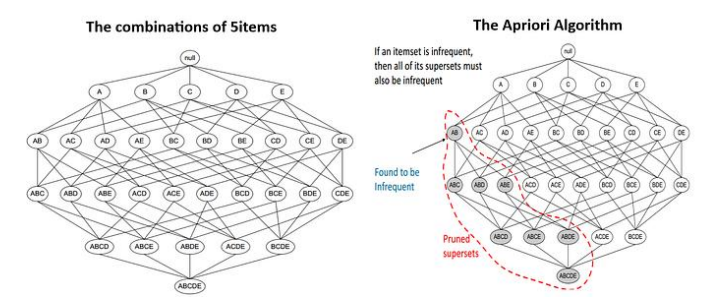
1. Apriori Algorithm

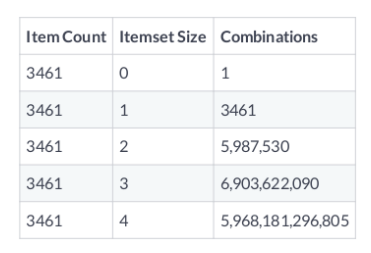
2. AIS Algorithm

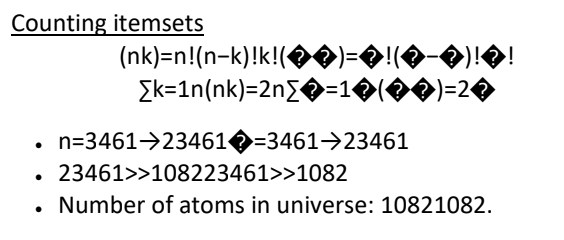
3. SETM Algorithm

4. FP Growth

**Apriori Algorithm:**









● Not possible to consider all itemsets.

▪ Not even possible to enumerate them.

● How do we remove an itemset without even

evaluating it?

▪ Could set maximum k� value.

● Apriori algorithm offers alternative.

▪ Doesn't require enumeration of all itemsets.

▪ Sensible rule for pruning.

**Apriori principle:**

▪ Subsets of frequent sets are frequent.

▪ Retain sets known to be frequent.

▪ Prune sets not known to be frequent.

**Ex:**

● Candles = Infrequent

▪ -> {Candles, Signs} = Infrequent

● {Candles, Signs} = Infrequent

▪ -> {Candles, Signs Boxes} = Infrequent

● {Candles, Signs, Boxes} = Infrequent

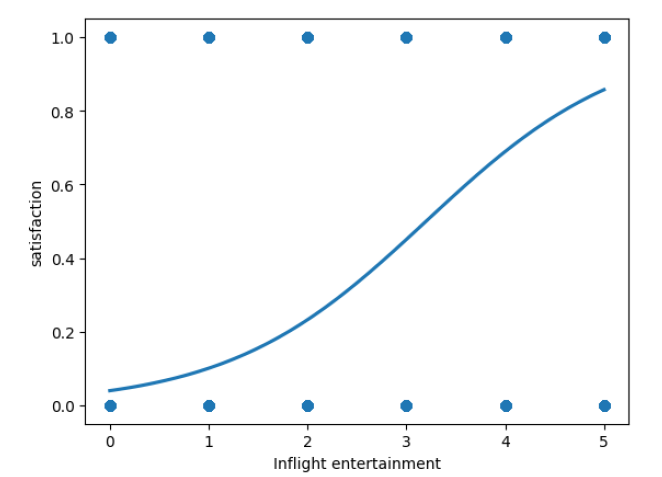
▪ -> {Candles, Signs, Boxes, Bags} = Infrequent

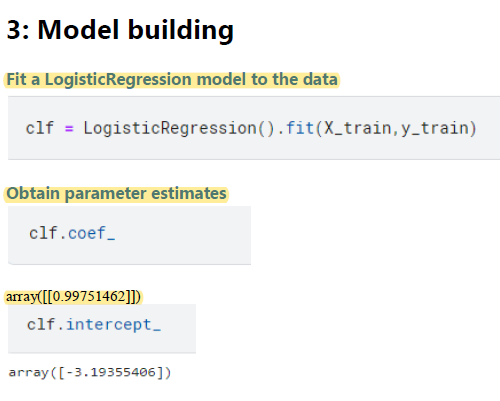
Model Creation:

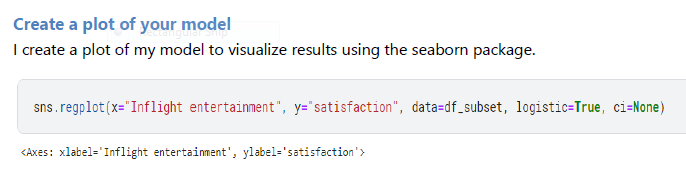
Once created the features, the machine learning model is trained. Models are feed using the data provided. Algorithm can be supervised or unsupervised and depending the objective, the project will be a classification or regression task. In order to capt the changing behaviour of the data, those machine learning models have to be retrained periodically. This time has to be defined with the client and according to the needs of the problem.

Model Evaluation:

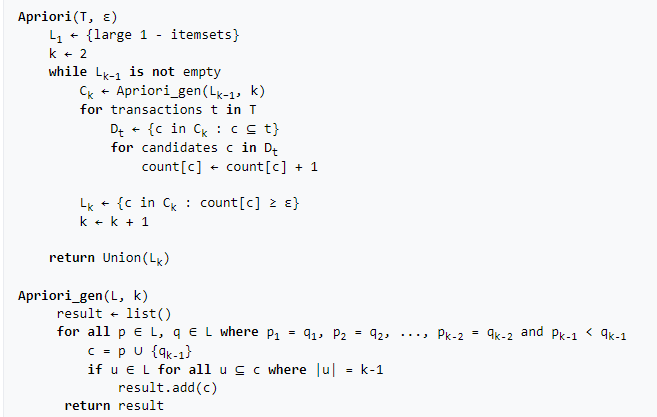
The last step in the standard procedure is the model evaluation. The performance of it is the result of all the work done along the process. Depending the type of the problem there are some metrics to evaluate the performance of a model.

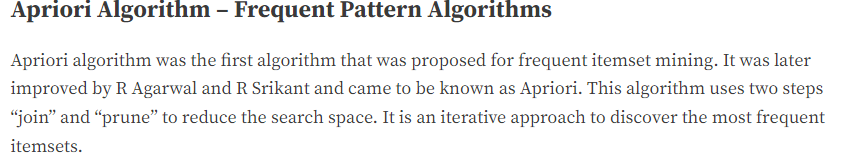












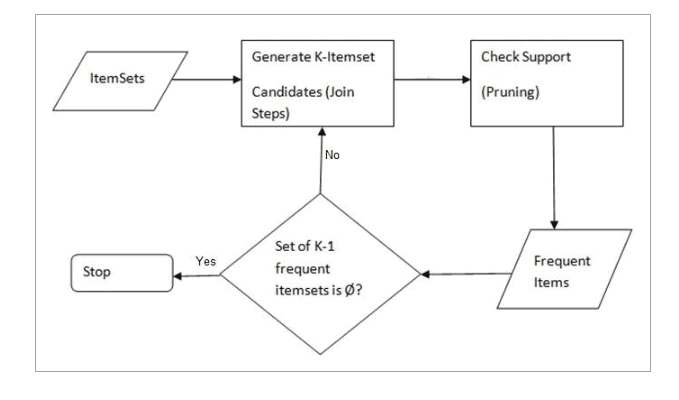
**Apriori says:**

The probability that item I is not frequent is if:

* P(I) < minimum support threshold, then I is not frequent.
* P (I+A) < minimum support threshold, then I+A is not frequent, where A also belongs to itemset.
* If an itemset set has value less than minimum support then all of its supersets will also fall below min support, and thus can be ignored. This property is called the Antimonotone property.

**The steps followed in the Apriori Algorithm of data mining are:**

1. **Join Step**: This step generates (K+1) itemset from K-itemsets by joining each item with itself.
2. **Prune Step**: This step scans the count of each item in the database. If the candidate item does not meet minimum support, then it is regarded as infrequent and thus it is removed. This step is performed to reduce the size of the candidate itemsets.



**Conclusion:**

The strength or dependability of association rule mining is critical to consider. As association rule mining uncovers interesting associations and relationships among large sets of data items, it reveals how frequently a particular item set appears in a transaction. Association rules are useful in data mining for analyzing and forecasting customer behavior.