

# G.H RAISONI INSTITUTE OF ENGINEERING AND TECHNOLOGY, PUNE

## Computer Science Department

### Market Basket Analysis



# Market Basket Analysis

## Mentored By:

Ayush Chhajer

## Presented By:

Saibaba V More

Pragya Chandra

Prathamesh Dherange

Pratiksha Repaswal

# 1.SRS(Software Requirements Specification)

## Overview

Market Basket Analysis is a technique which identifies the strength of association between pairs of products purchased together and identifies patterns of co-occurrence. A co-occurrence is when two or more things take place together.

Market Basket Analysis creates If-Then scenario rules, for example, if item A is purchased then item B is likely to be purchased. The rules are probabilistic in nature or, in other words, they are derived from the frequencies of co-occurrence in the observations.

## Purpose

It shows what combinations of products most frequently occur together in orders. These relationships can be used to increase profitability through cross-selling, recommendations, promotions, or even the placement of items on a menu or in a store.

## Intended Audience

- Retailers, Business man, customers

## Intended Device

- Both Mobile and Computer

## Proposed Features/ Functionalities

### 1. Store Layout:

Based on the insights from market basket analysis you can organize your store to increase revenues. **Items that go along with each other should be placed near each other to help consumers notice them.** This will guide the way a store should be organized to shoot for best revenues.

### 2. Marketing Messages:

Whether it is email, phone, social media or an offer by a direct salesman, market basket analysis can improve the efficiency of all of them. **By using data from MBA you can suggest the next best product which a customer is likely to buy.** Hence you will help your customers with fruitful suggestions instead of annoying them with marketing blasts.

### 3. Maintain Inventory:

**Based on the inputs from MBA you can also predict future purchases of customers over a period of time.** Using your initial sales data, you can predict which item would probably fall short and maintain stocks in optimal quality. This will **help you improve the allocations of resources to different items of the inventory.**

#### 4. Content Placement:

In case of **e-commerce businesses**, website content placement is very important. If goods are displayed in right order than it can help boost conversions. **MBA can also be used by online publishers and bloggers to display content which consumer is most likely to read next.** This will **reduce bounce rate, improve engagement and result in better performance** in search results.

#### 5. Recommendation Engines:

Recommendation engines are already used by some popular companies like **Netflix, Amazon, Facebook**, etc. If you want to create an effective recommendation system for your company then you will also need market basket analysis to efficiently maintain one. **MBA can be considered as the basis for creating a recommendation engine.**

## Operating Environment / Dependencies

### 1. Operating System Required:

An Intel-compatible platform running Windows 10 /8.1/8 /7 /Vista /XP /2000  
Windows Server 2019 /2016 /2012 /2008 /2003

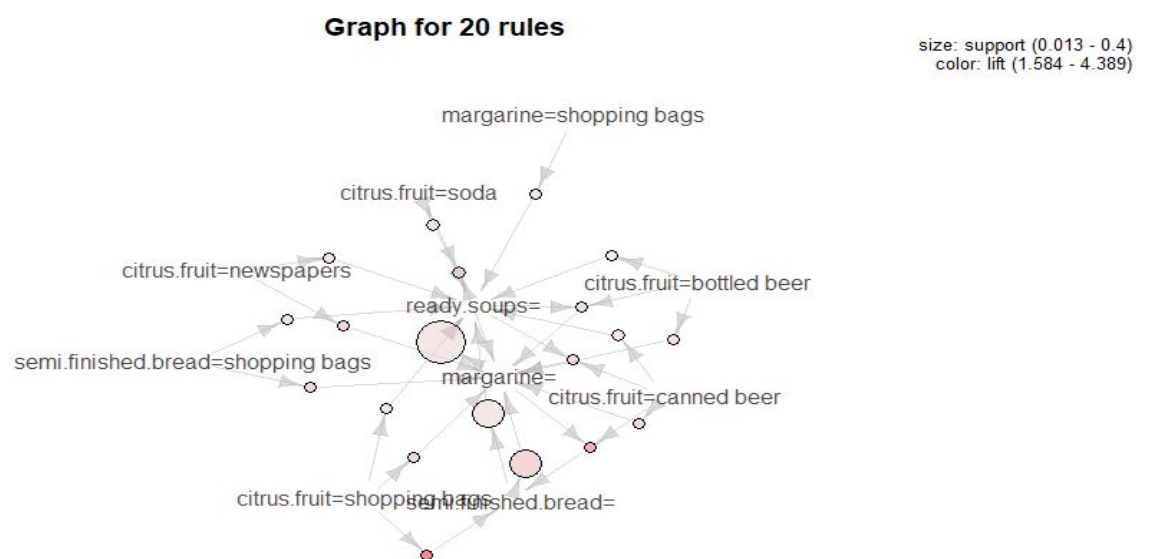
### 2. Tools:

- R Language
- R Studio

## Technology

**Language:** R Programming Language

## Plot



## Algorithms

### Apriori Algorithm:

The apriori principle can reduce the number of itemsets we need to examine. Put simply, the apriori principle states that

*| if an itemset is infrequent, then all its **supersets** must also be infrequent*

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

The probability that item I is not frequent is if:

- $P(I) < \text{minimum support threshold}$ , then I is not frequent.
- $P(I+A) < \text{minimum support threshold}$ , 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.

## Rules

### Association Rule:

Association rules analysis is a technique to uncover how items are associated to each other. There are three common ways to measure association.

### Basic Concept of Association Rule:

- >Itemset: Collection of one or more items. K-item-set means a set of k items.
- >Support Count: Frequency of occurrence of an item-set
- >Support (s): Fraction of transactions that contain the item-set 'X'

**Measure 1: Support.** This says how popular an itemset is, as measured by the proportion of transactions in which an itemset appears.

**Measure 2: Confidence.** This says how likely item Y is purchased when item X is purchased, expressed as  $\{X \rightarrow Y\}$ . This is measured by the proportion of transactions with item X, in which item Y also appears.

**Measure 3: Lift.** This says how likely item Y is purchased when item X is purchased, while controlling for how popular item Y is.

$$\begin{array}{l}
 \text{Rule: } X \Rightarrow Y \\
 \begin{array}{l}
 \nearrow \text{Support} = \frac{\text{freq}(X, Y)}{N} \\
 \rightarrow \text{Confidence} = \frac{\text{freq}(X, Y)}{\text{freq}(X)} \\
 \searrow \text{Lift} = \frac{\text{Support}}{\text{Supp}(X) \times \text{Supp}(Y)}
 \end{array}
 \end{array}$$

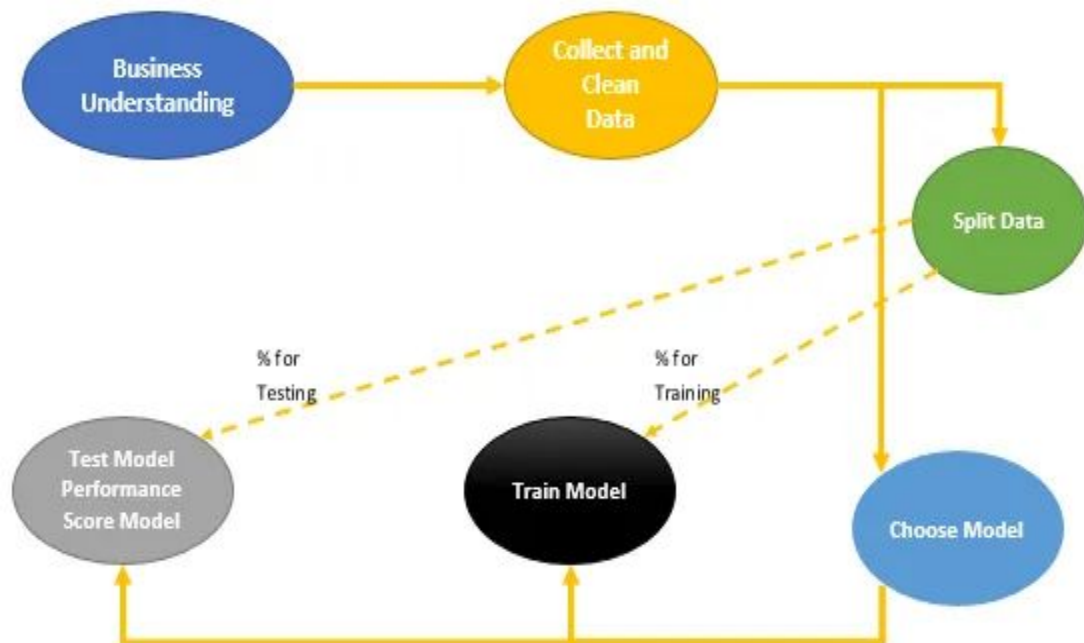
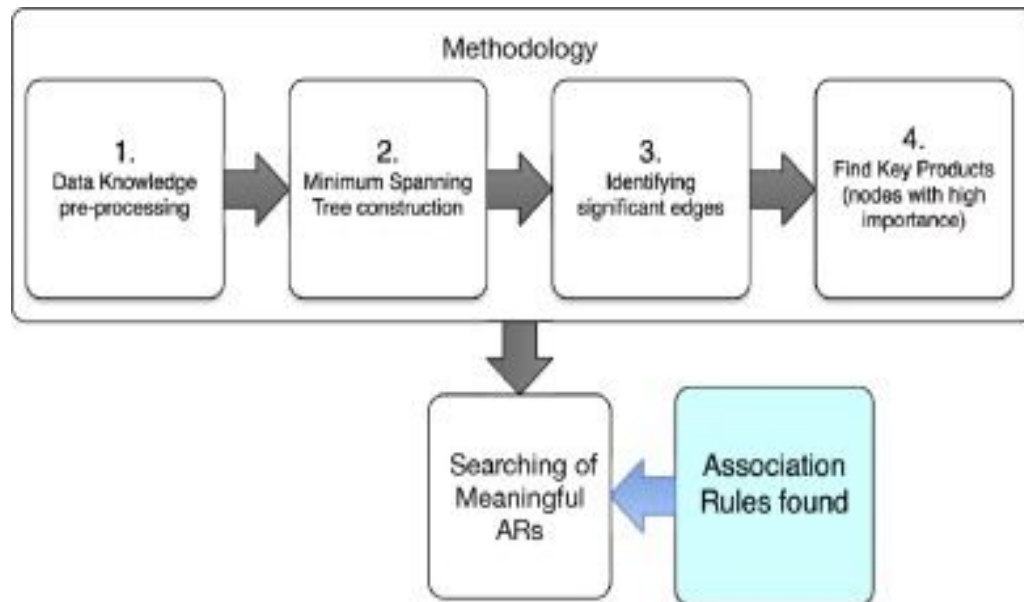
## Details

### 1. Pseudocode:

- Start
- initialise the association library
- initialise the association library visualization
  - Allocate the dataset from the local address(groceries dataset)
- Download apriori packages
- apply apriori algorithm and allote the value to rules variable
- apply data mining using apriori algorithm
  - initialise support as low for better result
  - initialise confidence as low for better result
- show the rules using inspect element
- Get the summary of the rules variable
- Manipulate the dataset:
  - Sort the data by support in descending order
  - Remove Redundancy
- Visualize the dataset according to applied condition
  - plot the graph for top 20 rules
  - plot graph with interactive as true

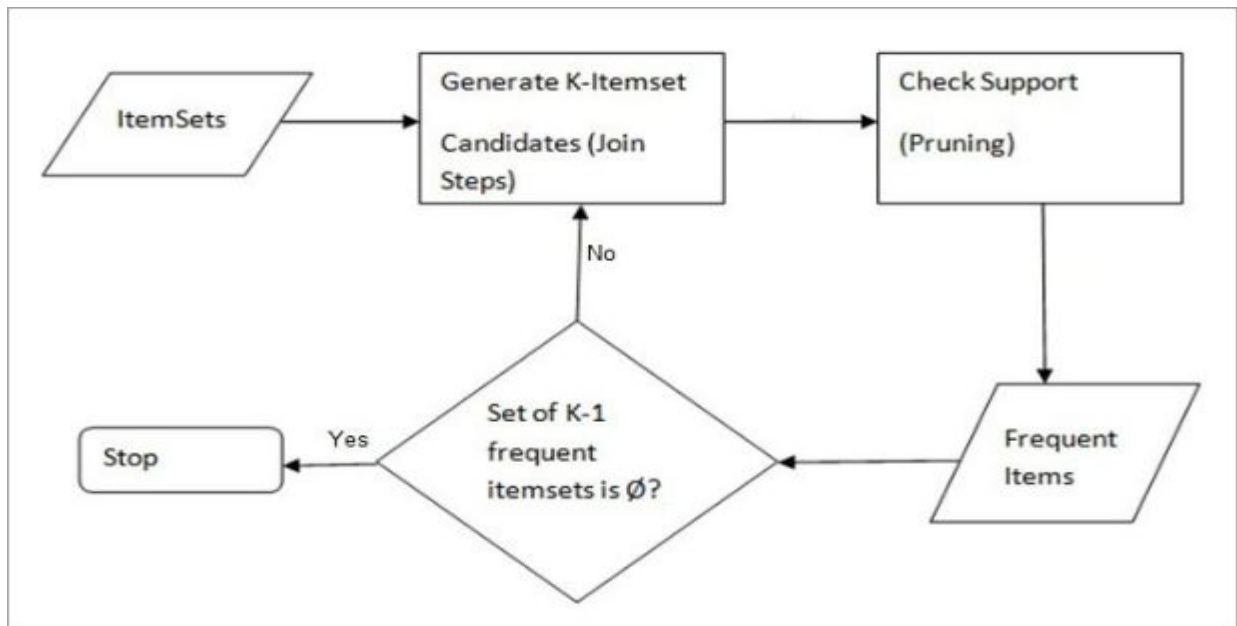


## 2. Mind Map:

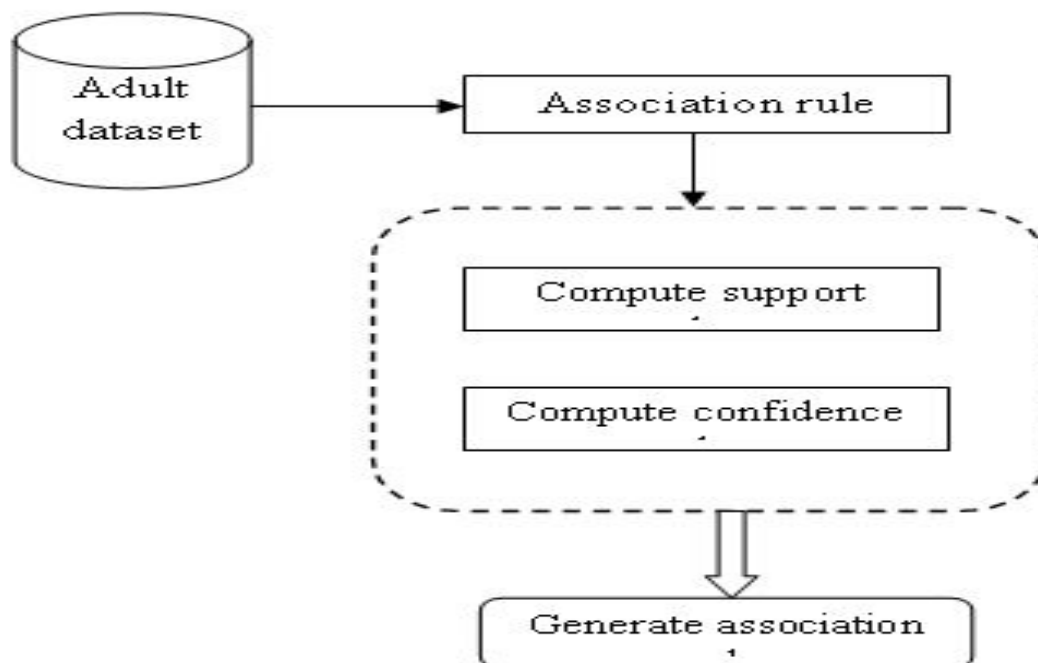


### 3. Workflow Diagram:

#### Apriori Algorithm



#### Association Rule





## 2) Developed Code

[https://drive.google.com/file/d/1Chsb42aTgte3QAcB\\_6XLbZWDpkNKIutu/view?usp=sharing](https://drive.google.com/file/d/1Chsb42aTgte3QAcB_6XLbZWDpkNKIutu/view?usp=sharing)

## 3) Github uploaded Link

[https://github.com/saimore027/Market\\_basket\\_analysis.git](https://github.com/saimore027/Market_basket_analysis.git)

## 4) 5 min Explanatory video on YouTube link

<https://youtu.be/XVfcci4VjUk>