**MARKET BASKET INSIGHT**

**INTRODUCTION:**

One of the challenges for companies that have invested a lot In consumer data

collection is how to mine important Information from their vast customer

databases andProduct feature databases, in order to gain economical

Advantage. Several aspects of market basket analysis haveBeen studied in

academic literature, such as using customerInterest profile and interests on

particular products for one To one marketing, purchasing patterns in a

multi-store Environment to improve the sales. Market basket analysis Has been

intensively used in many companies as a means to Discover product associations

and base a retailer’s Promotion strategy on them. A retailer must know the needs

of customers and adapt to the Them. Market basket analysis is one achievable

way to find which items can be place together. Market basket Analyses give

retailer good information about related sales On group of goods basis Customers

who buys bread often Also buy several products related to bread like milk, butter

Or jam. It makes sense that these groups are placed side by Side in a retail center

so that consumers can access them Promptly. Such related groups of goods also

must be located Side-by-side in order to remind customers of related items And to

lead them through the center in a logical manner. Therefore the Market consumer

behaviors need to be Analyzed which can be done during dissimilar data mining

Technique

**ABSTRACT**

Market Basket Analysis is an important part of the analytical system in the retail

Organization to determine the placement of goods, designing sales promotion for

different Segments of customers to improve customer satisfaction and hence the

profit of the Supermarkets.MBA is well known activity of ARM ultimately used for

business Intelligent decisions. Mining frequent item sets and hence deduce rules

to build classifiers With good accuracy is essential for efficient algorithm.

issues for a leading Supermarket are addressed here using frequent item set

mining. The project uses file as database. Here, the itemsets and transactions

items are kept in a Matrix form representing rows as list of items and column as

transactions. The frequent item sets are mined from database using the Apriori

algorithm and then the Association rules are generated. The project is beneficial

for supermarket managers to determine the relationship between The items that

are purchased by their customers.

Keywords: Market Basket Analysis, Association Rule Mining, Apriori Algorithm

**Problem Statement**

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.

**Objectives**

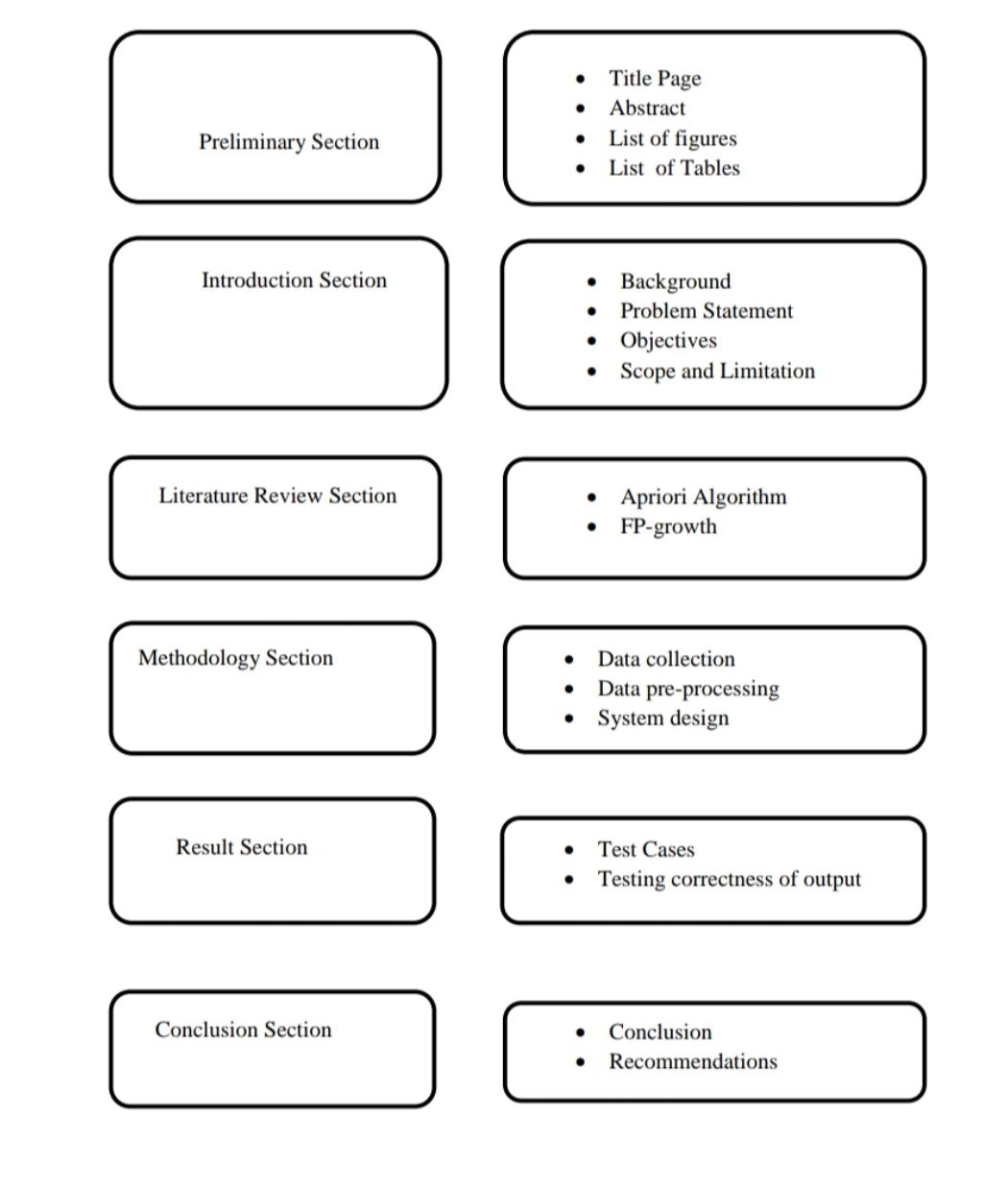
1. To identify the frequent items from the transaction on the basis of support and Confidence
2. To generate the association rule from the frequent item sets.

**Scope**

The scope of the application is limited to desktop application right now. The

application Is targeted towards a supermarket of Nepal.

**Report Organization**



**REQUIREMENT ANALYSIS AND FEASIBILITY**

**Literature Review**

Data Mining provides a lot of opportunities in the market sector. Decision making

And Understanding the behavior of the customer has become vital and challenging

problem for The organization in order to sustain in this competitive environment.

The challenges that The organization faces is to extract the information from their

vast customer databases, in Order to gain competitive advantage. Yanthy et al [1] in

this paper author states about the important goal in data mining is to Reveal hidden

knowledge from data and various algorithms have been proposed for, but The

problem is that typically not all rules are interesting –only small fraction of the

Generated rules would be of interest to any given users. Hence numerous methods

such as Confidence, support, and lift have been proposed to determine the best or

most interesting Rules. However some algorithms are good at generating rules high

in one measure but bad In other.

Rakesh Agarwal [2] proposed the Apriori algorithm. Apriori was the first associative

Algorithm proposed and future development in assocuaton ,classification

,associative .

Classification algorithms have used apriori as part of the technique. Apriori

algorithm is a Level-wise, breadth-first algorithm which counts transactions Apriori

algorithm uses prior Knowledge of frequent item set properties.

To improve The efficiency of the level-wise generation of frequent item sets Apriori

property is used Here. Apriori property insists that all non-empty subsets of a

frequent item set must also Be frequent. This is made possible because of the anti-

monotone property of support Measure – the support for an item set never

exceeds the support for its subsets. A two-step Process consists of join and prune

actions are done iteratively It is one of the Data Mining Algorithm which is used to

find the frequent items/item set From a given data repository.

The algorithm involves 2 steps

1. Pruning
2. Joining

The Apriori property is the important factor to be consider before proceeding with

The Algorithm Apriori property states that If an item X is joined with item Y,

Support (XUY) =min (Support(X), Support(Y))

Basically when we are determining the strength of an association rule i.e. how string

The Relationship is between the transaction of the items we measure through the

use of the Support and confidence.

The support of an item is the number of transaction containing the item. Those

items that Do not meet the minimum support are excluded from the further

processing. Support

P(RHS/LHS)=P(RHS∩LHS)/P(LHS)=support(RHS∩LHS)/support(LHS)

While determining the rules we must measure these two components as it is

Very important to us. A rule that has very low support may occur simply by chance.

Confidence on the other hand, measures the reliability of the inference made by

the rule.Han [4, 5] presented a new association rule mining approach that does not

use candidate Rule generation called FP-growth that generates a highly condensed

frequent pattern tree (fptree) representation of the transactional database. Each

database transaction is Represented in the tree by at most one path. FP-tree is

smaller in size than the original Database the construction of it requires two

database scans, where in the first scan, Frequent item sets along with their support

in each transaction are produced and in the Second scan, FP-tree is constructed

The mining process is performed by concatenating the patterns with the ones

produced From the conditional FP-tree.

One constraint of FP-growth method is that memory may Not fit FP-tree especially

in dimensionally large database. Liu [6] proposed CBA the first Associative

Classification (AC) algorithm.CBA Implements the famous Apriori algorithm[3] in

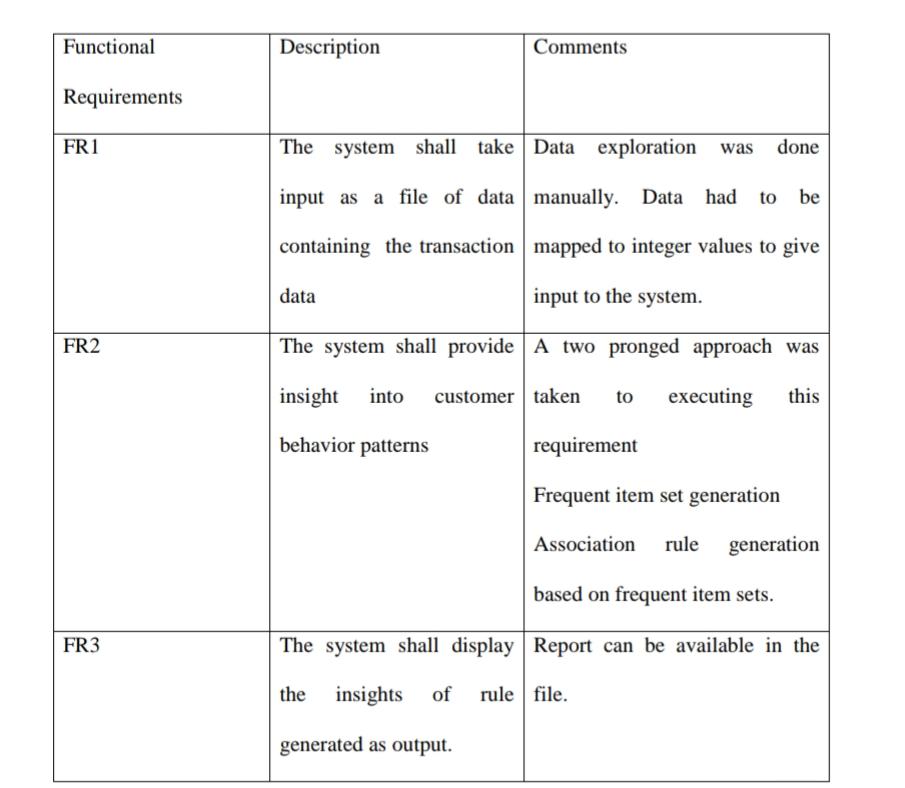
order to discover frequent rule items. The Apriori algorithm consists of three main

steps.

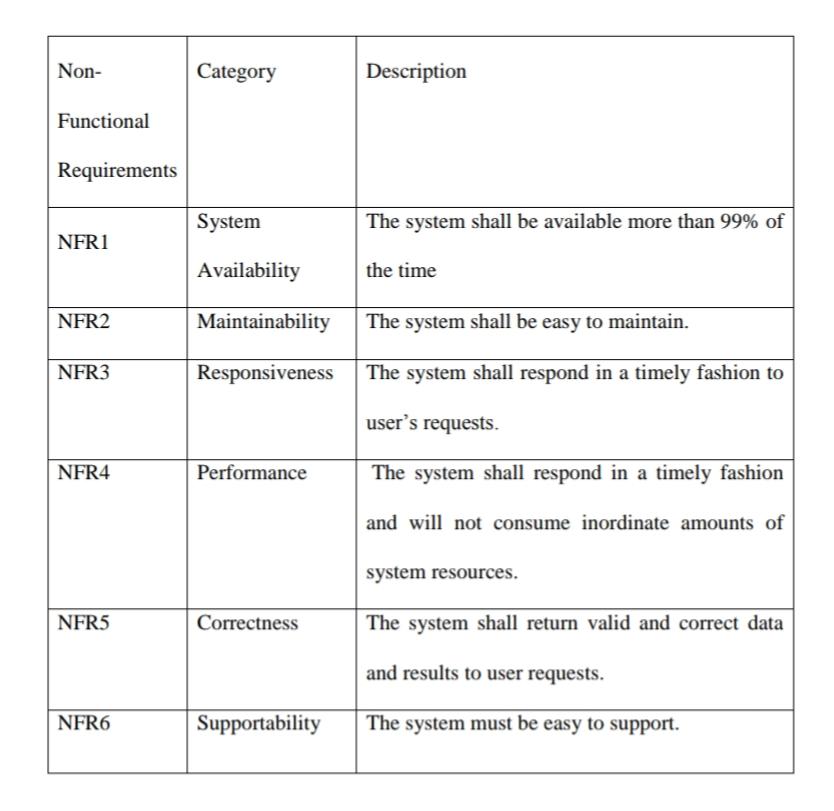
1. Continuous attribute in the training data set gets discredited.
2. Frequent rule items discovery
3. Rule generation

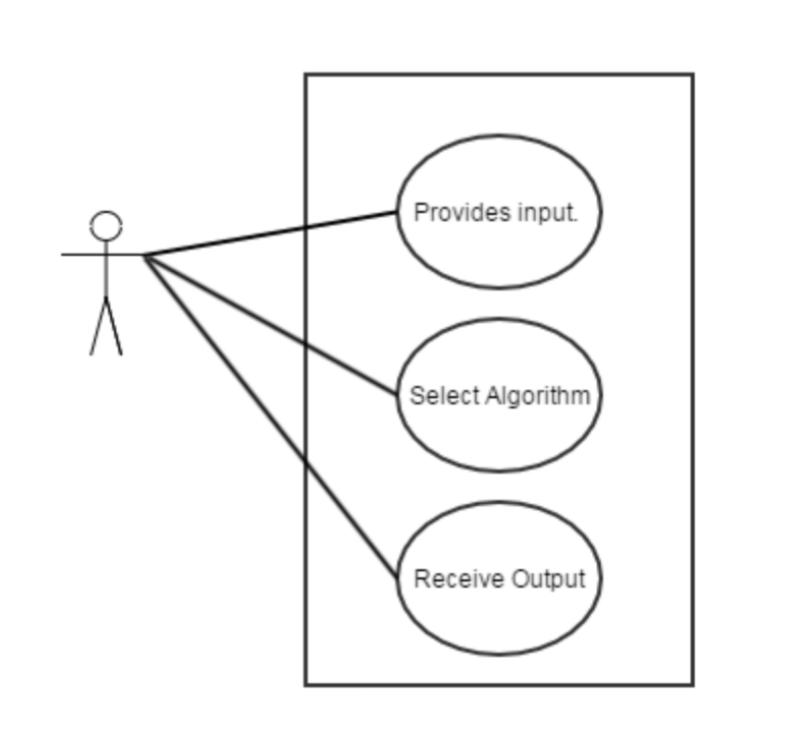
**Requirement Analysis**

**Functionality Requirements :**

****

**Non functionality requirements**



The user will provide input to the application .The input is a text file where items

are Mapped into integer’s value. The row represents the items that were purchased

in one Single transaction and column represents the transaction.

The user then can select the algorithm to run and provide the necessary parameters

i.e. Confidence and support.

After processing the user will receive output in to the desired path that the user

wants. The output will be a text file containing association rules.

**SYSTEM DESIGN**

**Methodology**

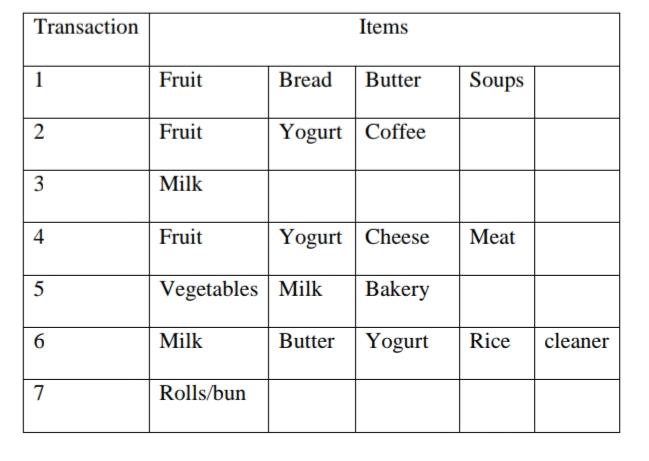
**1.Data collection**

The data collection process involves continuously recording transactions over a

period of time. The more data you collect, the more accurate your market basket

analysis will be.

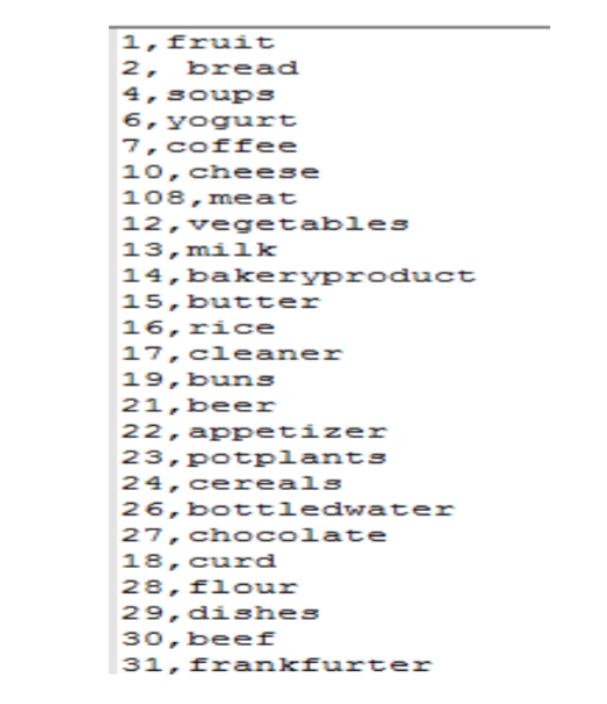
Sample data



**2**.**Data processing**

The data collected was mapped manually as integer values as shown in Figure 4.

For Example the “Fruit” was labeled as 1, “Bread” as 2 “Soups” as 4 and so on.



**Mapped to integers**

The mapped integer’s values were then saved in a text file and given as the input to

Agthe System. It shows the input file that is given to the system.

**3.Apriori algorithm**

Association rule mining finds interesting associations and/or correlation

relationships among large set of data items. Association rules shows attribute value

conditions that occur frequently together in a given dataset. A typical and widely

used example of association rule mining is Market Basket Analysis. For example,

data are collected from the supermarkets. Such market basket databases consist of

a large number of transaction records. Each record lists all items bought by a

customer on a single purchase transaction. Association rules provide information of

this type in the form of “IF-THEN” statements. The rules are computed from the

data, an association rule has two numbers that express the degree of uncertainty

about the rule

1. Support
2. Confidence

**Support**

The support of an item is the number of transaction containing the item. Those

items that do not meet the minimum support are excluded from the further

processing. Support determines how often a rule is applicable to a given data set.

Support (XUY) =min (Support(X), Support(Y))

**Confidence**

Confidence is defined as the conditional probability that a transaction containing

the LHS will also contain the RHS.

Confidence (LHS->RHS->

P(RHS/LHS)=P(RHS LHS)/P(LHS)-support(RHS LHS)/support(LHS).

Confidence determines how frequently item in RHS appears in the transaction that

Contain LHS. While determining the rules we must measure these two components

as it is very important to us. A rule that has very low support may occur simply by

chance.

**Pseudocode**

//Find all frequent itemset

Apriori(database D of transaction, min\_support){

Fl=(frequent 1-itemset)

K-2

While Fk-1/ Empty Set

Ck=AprioriGeneration (Fk-1)//Generate candidate item sets.

For each transaction in the database D

{

Cl=subset (Ck, t)

18

For each candidate c in Ct

Count c++

}

Fk (c in Ck such that countc&gt;min\_support)

K++

}

F-U K&gt;Fk

}

//prune the candidate item sets

Apriori generation (Fk-1) {

//Insert into Ck all combination of elements in Fk-1 obtained by self-joining item sets in Fk-1

//Delete all item sets e in Ck such that some (K-1) subset of c is not in Lk-I

}

//find all subsets of candidate contained in t

Subset (Ck, 1)

}

**INNOVATIONS**

**1.MACHINE LEARNING ALGORITHMS:**

Machine learning algorithms have made market basket

analysis more accessible than ever, allowing retailers to extract valuable insights

from large datasets in real time.

Auto association, specifically in the context of market

basket analysis, refers to a technique used to identify patterns in customer purchase

behavior by analyzing the co-occurrence of items in transactions.

Implementing more advanced machine learning algorithms

to identify hidden patterns and correlations in consumer purchasing data, allowing

businesses to make more precise product recommendations.

**2.INTEGRATION WITH IOT:**

IoT (Internet of Things) plays a significant role in enhancing

Market Basket Insight by providing valuable data and improving the accuracy of

analysis. Here’s how IoT is used in Market Basket Insight.

**3.MOBILE APPS:**

Mobile apps can significantly enhance Market Basket Insight by

providing a direct and convenient channel to collect data on consumer behavior

and preferences. Here’s how mobile apps can be used to improve Market Basket

Insight:

**4.Real time analysis :**

Real-time analysis in Market Basket Insight refers to the capability of

analyzing customer purchasing patterns and product associations as transactions

occur in real time. This approach allows businesses to make immediate decisions

and provide personalized recommendations or promotions to customers while they

are shopping. Here’s how real-time analysis can benefit Market Basket Insight:

**5.Enhanced visualisation:**

Enhanced visualization in Market Basket Insight involves presenting the analysis of

consumer purchasing patterns and product associations in a visually compelling and

easily understandable manner. Effective visualization can help businesses, analysts,

and decision-makers extract actionable insights from the data. Here are some key

aspects of enhanced visualization in Market Basket Insight:

**6.personalisation:**

Personalization in Market Basket Insight involves tailoring product

recommendations, marketing strategies, and shopping experiences to individual

customer preferences and behaviors. It can significantly enhance customer

engagement, increase sales, and improve overall satisfaction. Here are key aspects

of personalization in Market Basket Insight:

Implement recommendation algorithms that analyze a customer’s

purchase history and suggest products they are likely to be interested in based on

their past behavior and the behavior of similar customers.

Adjust product prices in real time based on a customer’s purchase history,

preferences, and behavior. Offer personalized discounts or promotions to

encourage specific purchases.

**7.AI-driven Chatbots:**

Implementing AI-powered chatbots that can engage with customers,

answer their product-related questions, and make real-time product

recommendations.

AI-driven chatbots play a valuable role in Market Basket Insight by engaging with

customers in real-time, answering their questions, making product

recommendations, and providing a personalized and interactive shopping

experience. Here’s how AI-driven chatbots can enhance Market Basket Insight:

**Machine learning algorithms :**

Algorithms Used in Market Basket Analysis

1. Apriori Algorithm
2. AIS Algorithm
3. SETM Algorithm
4. FP Growth
5. **Apriori Algorithm**

Apriori Algorithm is a widely-used and well-known Association Rule algorithm and

Is a popular algorithm used in market basket analysis. It is also considered accurate

And overtop AIS and SETM algorithms. It helps to find frequent itemsets in

Transactions and identifies association rules between these items. The limitation of

The Apriori Algorithm is frequent itemset generation. It needs to scan the database

Many times, leading to increased time and reduced performance as a

Computationally costly step because of a large dataset. It uses the concepts of

Confidence and Support.

1. **AIS Algorithm**

The AIS algorithm creates multiple passes on the entire database or transactional

Data. During every pass, it scans all transactions. As you can see, in the first pass, it

Counts the support of separate items and determines then which of them are

Frequent in the database. Huge itemsets of every pass are enlarged to generate

Candidate itemsets. After each scanning of a transaction, the common itemsets

Between the itemsets of the previous pass and the items of this transaction are

Determined. This algorithm was the first published algorithm which is developed to

Generate all large itemsets in a transactional database.

1. **SETM Algorithm**

This Algorithm is quite similar to the AIS algorithm. The SETM algorithm creates

Collective passes over the database. As you can see, in the first pass, it counts the

Support of single items and then determines which of the are frequent in the

Database. Then, it also generates the candidate itemsets by enlarging large itemsets

Of the previous pass. In addition to this, the SETM algorithm recalls the

TIDs(transaction ids) of the generating transactions with the candidate itemsets.

1. **FP Growth**

FP Growth is known as Frequent Pattern Growth Algorithm. FP growth algorithm is

A concept of representing the data in the form of an FP tree or Frequent Pattern.

Hence FP Growth is a method of Mining Frequent Itemsets. This algorithm is an

Advancement to the Apriori Algorithm. There is no need for candidate generation

To generate a frequent pattern. This frequent pattern tree structure maintains the

Association between the itemsets.

A Frequent Pattern Tree is a tree structure that is made with the earlier itemsets of

The data. The main purpose of the FP tree is to mine the most frequent patterns.

Every node of the FP tree represents an item of that itemset.