MARKET BASKET ANALYSIS

By Association Rule Mining

Association rules

Apriori Algorithm

Aaron Antonio Dias Barreto 18JE0003

Ajay Jagannath 18JE0049

IIT(ISM) Dhanbad

Abstract

This project aims to determine the purchasing patterns of customers i.e. which items are most frequently purchased together. In this project, we have used a dataset from Kaggle which consists of 9835 customer transactions.

From this dataset, we determine which items are most frequently bought together. To do this we use the Market Basket Analysis (MBA) technique using apriori algorithm and association rules.

Market Basket Analysis determines if there is any relationship between different products by looking at customer purchases. Association rule mining finds patterns in the data and based on these patterns, rules are formed of the form, if A happens, then B is likely to happen. The Apriori algorithm is used to determine the most frequently occurring items. If an item does not meet the minimum support it is discarded.

First, we use the apriori algorithm to calculate the support of each item in the dataset. We neglect items with low support (below 3%). Next using association rules, we calculate the confidence for two items being purchased together and again we neglect low values of confidence (below 20%).

After applying the above-mentioned algorithm to the dataset, the results we get are as follows, the items with the highest confidence are whipped/sour cream and whole milk followed by root vegetables and whole milk.

From these results, we can conclude that if a customer purchases whipped/sour cream, he/she is most likely to purchase whole milk. Similarly, if a customer purchases root vegetable, he/she is most likely to purchase whole milk.

Introduction

Food and other household products are the inevitable needs of every family throughout the world. Hence, this industry is incredibly saturated with heavy competition. So, Retailers can no longer depend only on traditional means of marketing to pull customers in. The advent of technology has allowed for a new means of understanding the wants of customers that have far exceeded the capacity of customer feedback. The collection and processing of data to further business prospects has allowed for an exact science that is business analytics. We are going to study a branch of business analysis that deals with designing the very fabric of the shop based on maximizing profits based on collected information and it is called Market Basket Analysis.

Market Basket Analysis (MBA) gives an idea of purchasing patterns of customers of different demographics, age, and several other aspects. The name itself refers to the shopping basket that the customers use to shop with. It deals with best-optimizing everything related to the functioning of the shop starting from the location, layout, marketing, pricing, etc.

MBA is a kind of methodology based on the observation that if you buy a certain set of items, more probable you will also buy another set of linked items. A famous case of MBA is the" Beer and Diaper" case, where the store management identified that middle aged male shoppers who purchased diapers also purchased beer with it. So, they came up with the idea that placing these two items will increase the sales exponential.

MBA is done using several algorithms that can be both manual and automated. Out of the several dozen aspects involved in MBA, the most important are Association Rule Mining and Apriori Algorithm.

MBA can be keenly made by the wide use of Association Rules which is in turn a part of Data Mining.

Data mining can be defined as the processing of huge amounts of raw data by observing patterns and hence gaining knowledge which can be further used for future tasks. These patterns must be useful, understandable, valid and reasonable. The process of data mining includes intersection of various fields like statistics, Artificial intelligence, Database technology and Machine learning.

Association rule mining gives the view of patterns and relations among the large set of data collected. This rule shows how frequently an item set occurs in transaction. A typical example is Market Basket Analysis. MBA is the one of the efficient methods to show relations between items. It helps the shop owners to relate several products that people purchase together on regular basis despite them not having any direct similarity.

There are various methods, techniques and algorithms viz. Apriori Algorithm, Éclat Algorithm, F.P. Growth Algorithm etc. Of all these, Apriori algorithm is easier and widely used.

Apriori Algorithm is a series of commands that results in the most occurring item in the given data set. To find the most repeated item, it uses join and prune steps repeatedly.

The subtle concepts lying beneath the names of above discussed topics are discussed and explained in a very broad sense in the following sections.

Literature review

- An important part of Market Basket Analysis is The Association rule mining. Rakesh Agrawal and Usama Fayyad have developed a large number of algorithms for data mining. One of the first papers published on association rule mining was by Rakesh Agarwal, Tomasz Imielinski and Arun Swami in 1993, where they put forward an effective algorithm that generated all the association rules between items. The paper also showed the results when the algorithm was applied to the sales data of several of the large retailing company.
- In 1994 Rakesh Agarwal and Ramakrishnan Srikant introduced two new algorithms for discovering association rules between items in large datasets that were not only fundamentally different from the previous algorithms but were also much faster than the previous algorithms. They then combined the best features of both the above-mentioned models into one model called the "Apriori Hybrid" algorithm.
- In 1997 Gary J.Russell, Wagner A. Kamakura used consumer data to segregate consumers based on their brand preferences. This research paper gave insights into the competing brands for each product. This paper also helped stores realize which brands were being preferred by which type of consumers.
- In May 2000 Jiawei Han, Jian Pei, and Yiwen Yin proposed a new algorithm for association rule mining called the frequent pattern (FP) growth algorithm. This algorithm was different from the previously used apriori algorithm.
- In the 2000s many researchers tried to improve on the existing market basket analysis algorithms, some notable improvements are mentioned below:

- In 2003 Reinhold Decker and Katharina Monien developed an algorithm for market basket analysis using neural networks.
- In 2006 Nanda Kumar and Ram Rao developed an algorithm that used the supermarket sales data to help the supermarket price items intelligently, in order to maximize sales and hence maximize overall profit.
- In the same year, Yasemin Boztug and Thomas Reutterer proposed a model that combined the two types of models of market basket analysis i.e the exploratory model and the explanatory model. These models usually existed independently. This was the first time the two types of models were combined.

Market Basket Analysis

Basics of Market Basket Analysis:

Market Basket Analysis determines what items are frequently bought together or placed in the shopping cart together by the customers. It establishes any possible relationship between different products by looking for combinations of products that frequently occur together. Market Basket Analysis is a correlation study, it is not a cause and effect study. Like every other field of data analytics, MBA involves classification, association, prediction, clustering and outlier analysis.

A classic example of market basket analysis is the beer and diaper example: It was found that in many stores across the U.S. on weekends people purchased beers and diapers together. Not many stores would have realized this if not for market basket analysis. The reason for this purchase though is a very logical one, it's just that most of the purchases were made by middle-aged males who wanted to do the shopping for themselves and their children for the entire weekend. So, an obvious change the retailer can make from this analysis is to keep diapers and beer near one another for the ease of shoppers and to increase sales.

Market Basket Analysis has spread its influence to the online sphere as well, where retailers make purchase suggestions to users based on information from both data from the demographic as well as this particular customer's previous purchase history.

There are two main branches of MBA

- **Predictive MBA** used to classify items and predict models based on the obtained information
- **Differential MBAs** used to remove insignificant data and results. It also compares information between different stores, demographics, season, days and other factors.

The insights gained from Market Basket analysis can be implemented in several ways to multiply customer satisfaction and maximize profits.

- **Product Placement**: The store should be designed in such a way that the products that are frequently bought together are kept close to each other. So that the customer finds it easier to pick the products and therefore the sales of the store increase. Thus increasing the total profit. Ex: Beer and Diapers.
- **Online recommendations**: The retailer can suggest products to a customer who has already bought a certain product by adding "customers who purchased this product also purchased these products" section. Retailers can also send discount coupons to regular customers to increase the amount of purchases and hence increase the total profit of the retailer.
- **Cross-Selling**: Suggesting products that complement the products purchased by the customer to increase the sales of both of them. Ex. Offering Fries when a Burger is purchased.
- Loss Leader Analysis-Selling a product at a lower price that tempts the customer to buy another product which is sold at a higher than normal price such the profits are normalized. Ex: selling toothbrushes at a lower price and toothpaste at a higher price.

MBA aims to find relationships/patterns across purchases to deduce a model in the form of a condition algorithm:

IF {A}: then {B} or {Ai} -> {Ci}

This is done based on the concept of Association.

Association Rule Mining (ARM):

While concepts of association rules were prevalent from earlier times, ARM was accurately explained at the end of 20th century, when three computer science Researchers Rakesh Agrawal, Tomasz Imieliński and Arun Swami together developed an automated way to find the correlation among various products using point-of-sale (POS) systems. Using these algorithms to markets, the data-scientists were discovering the connection between different products bought, which are called association rules, and hence were able to use that information to anticipate the frequency of various items being bought together.

This ARM finds is helpful in varied fields in day-to-day life. Some of the main examples are as listed below:

Market Basket Analysis (MBA):

This is the most widely used example of ARM. Data of purchased items is collected from the retailers. This dataset is known as the "market basket" dataset and collection of several records. Each record describes all the products purchased by a customer in a sale. Analysing the data by observing "which groups are inclined towards which set of

items", where groups can classified in a wide variety of ways, providing the shops a strategy to alter the store layout and to design a structure by which the customers are able to access related products easily.

Medical Diagnosis:

ARM in medical field is used for recommending doctors in treating patients. Diagnosis is no simple process and high probability of error may result in unconfirmed outcomes which in most cases means disaster for everyone involved. Using ARM, we can identify the likeliness of the occurrence of any illness in relation to various symptoms and associated factors. However, Association rules are just correlative and not cause and effect. Hence, they cannot be the endgame for any diagnosis.

Census Data:

It is the responsibility of every government to collect huge quantities of census data every once in awhile. The collected information can be used to plan efficient public services and help major corporations by understanding the demographic of the population thoroughly. This application of ARM and data mining has great scope in supporting better public policies and is hence an important tool to maintain efficient functioning of a democratic society.

Association rule mining (ARM):

The basic idea of ARM is to first analyse the regular patterns that takes place in the given dataset and based on the regular patterns that has been analysed, obtain rules of association which are of the form

[if **A occurs** it implies, then **B** is likely to occur.]

So, it is a conditional dependence relation that says that if A occurs that makes B more probable to occur as well. Such patterns are observed in sequences of time series data, like financial data or looking at fault analysis, where one thing causes the fault to occur, or you can look at in the transactional data column context which is where it was originally proposed and that is what we will look at in more detail.

So, the transaction are said to be collection of products that were bought simultaneously. So, the goal here is to find first frequent item sets, then you would say that an item set **A** implies item set **B**.

There are two elements of these rules:

Antecedent (if): This is an item/group of items that are typically found/occur in the datasets.

Consequent (then): items likely to occur if the antecedents occur.

There are 2 essential measures that help determine association:

Support- it gives the percentage of the data which contains both items A and B. Basically, support tells us about the frequency with which both items occur together.

So, the support of an association rule is the % of item-sets that has **A** union **B**.

$$S=P(A \cap B)$$

Confidence - It tells us the probability of B to occur provided that A has already occurred. So, the confidence of a rule is a % of item-sets having A, which also contained A union B. So, essentially this tells you how confident we are in about the association.

$$C=P(A \cap B)/P(A)$$

So, typically we look for rules with both high support and confidence that tell us that 2 items have a definite correlation.

There are 2 more important terms that we need to understand in Association rules:

Lift- It indicates the strength of a association rule over the random occurrence of A and B. It basically gives us the strength of any rule.

Item-Set: A sequence of items in a collection is called an item set. It is a type of dataset. If any item set has k-items it is then called a k-item set. An item set must consist of more than one item. An item set that occurs most repeatedly is called a frequent item-set.

ARM consists of 2 steps:

1. Finding frequent item sets.
2. Generate association rules for the frequent item sets.

Apriori Algorithm:

Apriori algorithm is said to be the foremost algo which was introduced for frequent ARM. It was further developed later by R Srikanth and R Agarwal and then widely known as The Apriori Algorithm. The algorithm has 2 steps (1): Join (2): Prune for reduction of the space getting searched. It follows a repetitive approach for finding the frequent item sets.

The steps taken in the Apriori Algo for mining of data are listed below:

Join Step: This step is used to generate (K+1) item set from a K-item set by joining each item with itself.

Prune Step: This step is used to scan the total count of each of the items in the database/item set. If the item under evaluation does not meet the minimum support threshold required, then it is regarded as infrequent and thus is removed. This step is performed in order to reduce the size of the candidate item sets.

1) Take every item as a set and count how many times they occur.

2) Compare the value to a predefine value of support and eliminate sets that have value below this predefined value.

3) Combine the sets 2 at a time and count how many time they occur.

4) Keep repeating above steps till the most frequent set is obtained.

The Pseudo code for Apriori Algorithm(can be implemented in any language):

Ck: Size k: candidate Item Set

Lk : Size k: frequent item set

L1 = {frequent items}; (final answer)

k=1;

while(Lk !=Æ)

{

Ck+1 = candidates generated from Lk;

k++; }

count (Ck)++;

Lk+1= candidates with value>min_support

Stop; → marks *the end of the iterative process of join and prune.*

Return Uk Lk; → most frequent item set returned

Apriori Algorithm uses a bottom up type approach, where subsets (that have value above threshold) are increased 1 at a time and this is called candidate generation.

Methodology

- Let us assume, database has 5 transactions
- The Minimum support(defined)=50 %
- The Minimum confidence(defined)=70%

Support= (50/ (50+70)) *5=2.08

TRANSACTION ID	ITEM SETS
A	R, A, K, E, S, H
B	R, A, J, E, S, H
C	L, O, K, E, S, H
D	J, A, D, E
E	R, O, M, E

- CANDIDATE **1**

ITEM SET	SUPPORT
R	3
A	3
K	2
E	5
S	3
H	3
J	2
L	1
O	2
D	1
M	1

So, We Eliminate All the Cases Which Have Support Less Than 2.08

- **L-1**

ITEMSET	SUPPORT

R	3
A	3
E	5
S	3
H	3

Taking Two Item sets At a Time From **L-1** We Get

- Candidate **2**

ITEM SET	**SUPPORT**
R A	2
R E	3
R S	2
R H	2
A E	3
A S	2
A H	2
E S	3
E H	3

S H	3

Again We Eliminate the Cases Having Support Less Than 2.08

- L-2

ITEM SET	SUPPORT
R E	3
A E	3
E S	3
E H	3
S H	3

- CANDIDATE 3

ITEM SET	SUPPORT
R A E	2
R E S	2
R E H	2
A E S	2
A E H	2
E S H	3

- L-3

ITEMSET	SUPPORT
E S H	3

ASSOCIATION RULES

ASSOCIATION RULE	SUPPORT	CONFIDENCE	CONFIDENCE %
E->S^H	3	3/5=0.6	60
S->E^H	3	3/3=1	100
H->E^S	3	3/3=1	100

implementation

Now, we shall perform market basket analysis for a give data set with the help of Apriori theorem and **association rules.**

We consider a sample data set from Kaggle for our analysis

From: Kaggle- Groceries market basket dataset

Dataset= **groceries – groceries.csv**

For our analysis, we consider a small support of 3% and a small confidence of 20% because of the very large dataset we have at our hands (9835 records with 158 unique items).

MARKET BASKET ANALYSIS FOR SAMPLE DATA SET WITH MINIMUM SUPPORT OF 3% AND MINIMUM CONFIDENCE OF 20%

[73]:
```
import numpy as np # linear algebra
import pandas as pd
# for reading the dataset
from mlxtend.frequent_patterns import apriori, association_rules
#inbuilt fuctions corresponding to MBA
import matplotlib.pyplot as plt
# for graphs
from scipy.special import comb
```

[74]:
```
df = pd.read_csv('/kaggle/input/groceries/groceries - groceries.csv')
#storing the dataset in the form of a table/dataframe
```

[75]:
```
df.shape
```

Out[75]
```
(9835, 33)
```

[76]:
```
products = (df['Item 1'].unique())
#storing the unique items as list
```

[77]:
```
products.shape
#158 unique products
```

Out[77]
```
(158,)
```

[78]:
```
print(products)
```

```
['citrus fruit' 'tropical fruit' 'whole milk' 'pip fruit'
 'other vegetables' 'rolls/buns' 'potted plants' 'beef' 'frankfurter'
 'chicken' 'butter' 'fruit/vegetable juice' 'packaged fruit/vegetables'
 'chocolate' 'specialty bar' 'butter milk' 'bottled water' 'yogurt'
 'sausage' 'brown bread' 'hamburger meat' 'root vegetables' 'pork'
 'pastry' 'canned beer' 'berries' 'coffee' 'misc. beverages' 'ham'
 'turkey' 'curd cheese' 'red/blush wine' 'frozen potato products' 'flour'
 'sugar' 'frozen meals' 'herbs' 'soda' 'detergent' 'grapes'
 'processed cheese' 'fish' 'sparkling wine' 'newspapers' 'curd' 'pasta'
 'popcorn' 'finished products' 'beverages' 'bottled beer' 'dessert'
 'dog food' 'specialty chocolate' 'condensed milk' 'cleaner' 'white wine'
 'meat' 'ice cream' 'hard cheese' 'cream cheese' 'liquor'
 'pickled vegetables' 'liquor (appetizer)' 'UHT-milk' 'candy' 'onions'
 'hair spray' 'photo/film' 'domestic eggs' 'margarine' 'shopping bags'
```

```
'dish cleaner' 'baking powder' 'specialty cheese' 'salty snack'
'Instant food products' 'pet care' 'white bread'
'female sanitary products' 'cling film/bags' 'soap' 'frozen chicken'
'house keeping products' 'spread cheese' 'decalcifier' 'frozen dessert'
'vinegar' 'nuts/prunes' 'potato products' 'frozen fish'
'hygiene articles' 'artif. sweetener' 'light bulbs' 'canned vegetables'
'chewing gum' 'canned fish' 'cookware' 'semi-finished bread' 'cat food'
'bathroom cleaner' 'prosecco' 'liver loaf' 'zwieback' 'canned fruit'
'frozen fruits' 'brandy' 'baby cosmetics' 'spices' 'napkins' 'waffles'
'sauces' 'rum' 'chocolate marshmallow' 'long life bakery product' 'bags'
'sweet spreads' 'soups' 'mustard' 'specialty fat' 'instant coffee'
'snack products' 'organic sausage' 'soft cheese' 'mayonnaise'
'dental care' 'roll products' 'kitchen towels' 'flower soil/fertilizer'
'cereals' 'meat spreads' 'dishes' 'male cosmetics' 'candles' 'whisky'
'tidbits' 'cooking chocolate' 'seasonal products' 'liqueur'
'abrasive cleaner' 'syrup' 'ketchup' 'cream' 'skin care'
'rubbing alcohol' 'nut snack' 'cocoa drinks' 'softener'
'organic products' 'cake bar' 'honey' 'jam' 'kitchen utensil'
'flower (seeds)']
```

[79]:

```
values1 = []
for index, row in df.iterrows():
    labels = {}
    uncommons = list(set(products) - set(row))
    commons = list(set(products).intersection(row))
    for uc in uncommons:
        labels[uc] = 0
    for com in commons:
        labels[com] = 1
    values1.append(labels)
values1[0]
new_df = pd.DataFrame(values1)
#iterating through each and every element of the dataset and finding intersections in purchases
```

[80]:

```
freq_items = apriori(new_df, min_support=0.03, use_colnames=True)
#setting minimum support to value of 3% ,narrowing down the products by neglecting products whose sales are comparitively low
freq_items
#items with support > 3%
```

Out[80]:

	support	itemsets
0	0.030402	(specialty chocolate)
1	0.033249	(berries)
2	0.077682	(canned beer)
3	0.255516	(whole milk)
4	0.063447	(domestic eggs)
...	...	...
58	0.038332	(rolls/buns, soda)
59	0.030605	(sausage, rolls/buns)
60	0.047382	(root vegetables, other vegetables)
61	0.035892	(other vegetables, tropical fruit)
62	0.042603	(other vegetables, rolls/buns)

63 rows × 2 columns

[81]:

```
final = association_rules(freq_items, metric="confidence", min_threshold=0.2)
#finding association based on the metric of confidence between the 2 products being > 20%)
final.head(5)
#sample of output
```

Out[81]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(yogurt)	(whole milk)	0.139502	0.255516	0.056024	0.401603	1.571735	0.020379	1.244132
1	(whole milk)	(yogurt)	0.255516	0.139502	0.056024	0.219260	1.571735	0.020379	1.102157
2	(bottled water)	(whole milk)	0.110524	0.255516	0.034367	0.310948	1.216940	0.006126	1.080446
3	(soda)	(whole milk)	0.174377	0.255516	0.040061	0.229738	0.899112	-0.004495	0.966533
4	(pip fruit)	(whole milk)	0.075648	0.255516	0.030097	0.397849	1.557043	0.010767	1.236375

[82]:
```
final=final.sort_values(by=['confidence'], ascending=False)
#the final data set is obtained where if a customer chooses an antecedent,we can say they are likely to also choose the consequent with some confidence.
```

[83]:
```
final.count()
```

Out[83]:
```
antecedents           25
consequents           25
antecedent support    25
consequent support    25
support               25
confidence            25
lift                  25
leverage              25
conviction            25
dtype: int64
```

[84]:
```
final
```

Out[84]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
9	(whipped/sour cream)	(whole milk)	0.071683	0.255516	0.032232	0.449645	1.759754	0.013916	1.352735
5	(root vegetables)	(whole milk)	0.108998	0.255516	0.048907	0.448694	1.756031	0.021056	1.350401
20	(root vegetables)	(other vegetables)	0.108998	0.193493	0.047382	0.434701	2.246605	0.026291	1.426693
10	(tropical fruit)	(whole milk)	0.104931	0.255516	0.042298	0.403101	1.577595	0.015486	1.247252
0	(yogurt)	(whole milk)	0.139502	0.255516	0.056024	0.401603	1.571735	0.020379	1.244132
4	(pip fruit)	(whole milk)	0.075648	0.255516	0.030097	0.397849	1.557043	0.010767	1.236375
6	(other vegetables)	(whole milk)	0.193493	0.255516	0.074835	0.386758	1.513634	0.025394	1.214013
8	(pastry)	(whole milk)	0.088968	0.255516	0.033249	0.373714	1.462587	0.010516	1.188729
13	(citrus fruit)	(whole milk)	0.082766	0.255516	0.030503	0.368550	1.442377	0.009355	1.179008
22	(tropical fruit)	(other vegetables)	0.104931	0.193493	0.035892	0.342054	1.767790	0.015589	1.225796
19	(sausage)	(rolls/buns)	0.093950	0.183935	0.030605	0.325758	1.771048	0.013324	1.210344
14	(yogurt)	(other vegetables)	0.139502	0.193493	0.043416	0.311224	1.608457	0.016424	1.170929
2	(bottled water)	(whole milk)	0.110524	0.255516	0.034367	0.310948	1.216940	0.006126	1.080446
12	(rolls/buns)	(whole milk)	0.183935	0.255516	0.056634	0.307905	1.205032	0.009636	1.075696
7	(whole milk)	(other vegetables)	0.255516	0.193493	0.074835	0.292877	1.513634	0.025394	1.140548
16	(yogurt)	(rolls/buns)	0.139502	0.183935	0.034367	0.246356	1.339363	0.008708	1.082825
21	(other vegetables)	(root vegetables)	0.193493	0.108998	0.047382	0.244877	2.246605	0.026291	1.179941
24	(rolls/buns)	(other vegetables)	0.183935	0.193493	0.042603	0.231620	1.197047	0.007013	1.049620
3	(soda)	(whole milk)	0.174377	0.255516	0.040061	0.229738	0.899112	-0.004495	0.966533
15	(other vegetables)	(yogurt)	0.193493	0.139502	0.043416	0.224383	1.608457	0.016424	1.109436
11	(whole milk)	(rolls/buns)	0.255516	0.183935	0.056634	0.221647	1.205032	0.009636	1.048452
23	(other vegetables)	(rolls/buns)	0.193493	0.183935	0.042603	0.220179	1.197047	0.007013	1.046477
18	(soda)	(rolls/buns)	0.174377	0.183935	0.038332	0.219825	1.195124	0.006258	1.046003
1	(whole milk)	(yogurt)	0.255516	0.139502	0.056024	0.219260	1.571735	0.020379	1.102157
17	(rolls/buns)	(soda)	0.183935	0.174377	0.038332	0.208402	1.195124	0.006258	1.042983

Result:

With the help of python programming language, we were able to perform market basket analysis of a sample dataset. Considering minimum support to be 3% and minimum confidence to be 20%, we were able to come to the following conclusions using market basket analysis

By choosing a threshold for support, we were able to eliminate products whose sales were too negligible to be taken into consideration, hereby saving time and cost.

There was a total of 25 associations that we were able to deduce with reasonable confidence.

Our findings are in the form of antecedents and consequents with various association term's values such as support, confidence, lift, etc.

These association values refer to chance of the consequent item being chosen if the antecedent item is already chosen.

- A customer is likely to take whole milk if they have already taken whipped/sour cream, root vegetables, tropical fruit, yogurt, pip fruit, other vegetables, pastry, citrus fruit, bottled water, rolls/buns or soda.
- A customer is likely to take other vegetables if they have already taken root vegetables, tropical fruit, yogurt, whole milk or rolls/buns.
- A customer is likely to take rolls/buns if they have already taken sausage, yogurt, whole milk, other vegetables or soda.
- A customer is likely to take root vegetables if they have already taken other vegetables.

- A customer is likely to take yogurt if they have already taken other vegetables or whole milk.
- A customer is likely to take soda if they have already taken rolls/buns.

If the confidence and support thresholds were to be to be changed, we would obtain different associations. It is up to the retailer to wisely choose these values.

Thus, based on these association between various products known, the retailer has valuable information about the needs and desires of his customers and so is enabled to provide the best environment for his customer, in order to increase their satisfaction and hence his profits.

Conclusion:-

This project is intended to tell about the importance of analysing the data of itemsets and obtaining the frequent observable patterns from it. There are many data mining techniques from which we can do it, but the most effective way to do this is obtaining strong association rules through applying the apriori algorithm between the items those are bought together. This analysis of data of supermarkets and any other organizations helps them in improving their sales by letting them know the trend of sales as most likely which products will be bought with a particular group of items.

Future scope:

Apriori algorithm, in the present day situation is extensively used in marketing sector. This can also be helpful various factors like medical sector etc., so that this makes a valuable predictions for the well being of society like food habit versus life expectancy. By developing this similar algorithms like this will be more helpful for the commercial and social sector of people in day today life.

Recommendations:

Apriori algorithm only deals with the given attributes. It only visualizes the pattern of attributes and give the result. But in real life the patterns itself does not make a huge sense. Though this algorithm may give the consequences by observing the incidents, this can only be concluded like an empirical form. In the real life, the intuition behind the incident plays a very major role for the incident to be occurred. So, the researchers must be proceed in this way of finding the reasons behind the actions so that the future predictions will be more solid than that of the present day rough analysis using apriori algorithm.