Food Sales Prediction System using Apriori Algorithm

Swapnil Dhanawde

University of Mumbai, Fort, Mumbai, India

Abstract- **Newly setup restaurant management face difficulty in knowing that which menu items are most in demand by consumers with the large sales data in hand which usually uses manual methods or feedback forms/verbal feedback to get the most loved menu item by people. This creates an such that sometimes customers cannot order the favourite order because some of the menus cannot be made available due to shortage of some raw materials. Apriori algorithm is one algorithm that searches frequent itemset by using the association rule technique.**

Based on the results of implementing the association rule method using Apriori algorithm to find the linkages between the level of sales menu item ordered at restaurant. With this association rule, it is expected to help the restaurant owners to find out the which menus are most in demand by consumers, so a system can be used to restock the raw material required for most demanding menu item using data mining association rule Apriori algorithm using.

Keywords- Apriori, Association Rules, Data Mining

1. INTRODUCTION:

The focus of this research paper to point out the problem commonly faced by most of the newly opened restaurants. The restaurants usually have a situation of shortage of raw material for menu list option selected by the customer. There is high probability that people may like to order their favorite menu item and restaurant cannot fulfil it because of inventory not been filled according to the demand. This problem can affect the sales which can sometime be a huge problem to newly opened restaurants to sustain the competition around them. As seeing the current market people have become very conscious about their health and eating habits, people now sees that for the money they spent on food should benefit their health as well as it should be quickly available and also affordable.

New restaurants near the office environment especially near business parks is always filled with customers, ranging from office workers to the normal people. During this time the management difficulties in knowing the menu most in demand by consumers and still use manual, and sometimes customers cannot order the desired order for some menu contained in the menu list has been exhausted or some lack of raw material inventory.

So, to solve the problem we need to analysis the sales of items from the menulist. We can implement different algorithms. By performing data analysis company and want to know the number of food and beverage sales, existing sales data to be processed or analyzed to determine the level of consumer trends in each product on an interest factor. From the data processing will be obtained a product consumption pattern of the food sales. The data analysis is done by creating an application that helps the restaurant in acquiring knowledge in the form of patterns of food sales transaction within a certain period of months.

2. PURPOSE:

The purpose concerning this research search out implement patterns of association rules utilizing Apriori algorithm to find linkages level of sales were scheduled in the menu card. With association rules proper to help the administrative order to experience what the menu is the most urgent purchasers. Apriori invention that is a form of data mining algorithms will specify news about the connection between parts in the database that maybe second-hand widely in trade processes exceptionally in the process to foresee the result of the purchase of meal in this paper.

3. SCOPE:

The most popular algorithm known as Apriori to create and test example, that is to say the manufacture of applicant associations of articles that may be contingent on sure rules and therefore tested either an alliance of these parts qualifies minimum support.

The sphere searches out act data reasoning of guest and ask about the number of cooking and beverage demand, existent buying data expected treated or analysed to decide the level of consumer flows in each commodity on an interest determinant. From the data processing will be acquired a merchandise use pattern of the food sales and marketing.

4. FLOW STRUCTURE OF APRIORI ALGORITHM:

Apriori algorithm that uses previously established attribute frequency information to process more data.

The following flowchart illustrates the steps involved in deciding the study's rules of association:
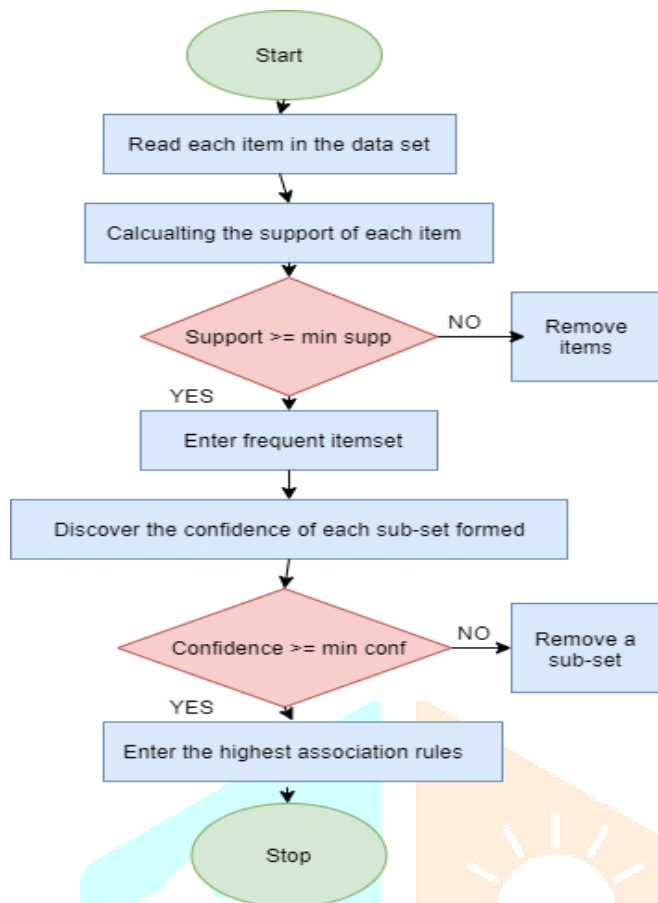
The Apriori algorithm is a method for determining a relation between an object and other items in a high frequency pattern. The iterations of the Apriori algorithm are divided into several stages:

1. Setting a minimum level of assistance
2. Candidate itemset formation, candidate k-itemset is generated by combining (k-1) -itemsets from previous iterations. Apriori algorithms have a variety of characteristics. Trimming candidate k-itemset subsetnya containing k-1 items are not included in the high frequency pattern with a duration of k-1 is one of the Apriori algorithm's characteristics.
3. The k-itemset help calculation for each nominee. Every candidate k-support itemsets is determined by scanning a database and calculating the number of transactions that include all of the items in the candidate k-itemset. It is also a feature of the Apriori algorithm, in which the necessary calculations to search the entire database as k-itemset longest are performed.
4. Build a pattern with a high frequency. A pattern with a high frequency that includes the k-k-itemset item or a set of candidates k-itemset objects.
5. Repeat the procedure until no k-itemset meets the minimum support requirement [2].

5. RESULT ANALYSIS:

The data used is the sales transaction data newly form restaurant (Fast Food Chain) restaurant in Dec 2021.

Table 01. Sample Data Month Dec 2021

Date of Transaction	Vadap av	Samo sa	Mix Bha ji	Chat Ite ms	Chicken Burge r	Egg Oml et
01-12-2021	5	2	3	0	0	2
02-12-2021	2	4	2	2	1	0
03-12-2021	7	0	4	4	3	0
04-12-2021	0	0	0	0	0	0
05-12-2021	0	0	0	0	0	0
06-12-2021	1	1	5	1	0	1
07-12-2021	1	1	0	0	0	3
08-12-2021	2	0	3	3	2	0
09-12-2021	6	0	3	1	6	0
…	…	…	…	…	…	…
30-12-2021	2	0	3	3	3	2
31-12-2021	2	4	2	2	2	3

Can be seen from the above table shows the title line and the next line-item group is the number of sales transactions. Samples above data used as an example in sing the association rule Apriori algorithm. The next steps to be able to enter data on the form RapidMiner format then converted into a numerical binomial format which contains the digits 1 to the sales transaction and the number 0 for the transaction without the sale, so that it becomes like to see in table 2.

Table 02. Transaction Sales in binomial format

Serial Number	Vadap av	Samo sa	Mix Bha ji	Chat Ite ms	Chicken Burge r	Egg Oml et
1	1	1	1	0	0	1
2	1	1	1	1	1	0
3	1	0	1	1	1	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	1	1	1	1	0	1
7	1	1	0	0	0	1
8	1	0	1	1	1	0
9	1	0	1	1	1	0
…	…	…	…	…	…	…
30	1	0	1	1	1	1
31	1	1	1	1	1	1

Shown in table 02 the results of the numerical format to format conversion binomial, number 1 for the sales transaction numbers from 0 to transactions without sales. At this stage, the search support value on each item. Support here means the number of transactions in the database that contains the item. Here's how the settlement of the case study above using Apriori algorithm:

a) formation itemset

Of format support table 2 can be calculated from the combined category or categories by adding the transaction containing that category divided by the total number of transactions, which can be broken down into the following table:

Table 03. Support of each item

Item	Total	Support
Vadapav	22	71%
Samosa	15	48%
Mix Bhaji	22	71%
Chat Items	20	65%
Chicken Burger	18	58%
Egg Omlet	14	45%

In Table 03 support which has a threshold of more than 50%, there are four categories: Vadapav, Mix Bhaji, Chat Item and Chicken Burger.

b) Combination 2 itemset

The process of formation of C2 or called with two itemsets with the minimum amount of support = 50% can be described in table 4. Establishment combination is obtained by mounting one item to another. Create a couple of items starting from the first item and then proceed to the second item

Table 04. Combination of 2 Item sets

Category	Total	Support
Vadapav, Samosa	14	47%
Vadapav, Mix Bhaji	21	70%
Vadapav, Chat Items	19	63%
Vadapav, Chicken Burger	17	57%
Vadapav, Egg Omlet	8	27%
Samosa, Mix Bhaji	14	47%
Samosa, Chat Items	12	40%
Samosa, Chicken Burger	10	33%
Samosa, Egg Omlet	13	43%
Mix Bhaji, Chat Items	20	67%
Mix Bhaji, Chicken Burger	18	60%
Mix Bhaji, Egg Omlet	13	43%
Chat Items, Chicken Burger	18	60%
Chat Items, Egg Omlet	11	37%
Chicken Burger, Egg Omlet	9	30%
Chat Items, Chicken Burger	18	60%
Chat Items, Egg Omlet	11	37%

In Table 04 support which exceeds the threshold of 50% there

are six categories, namely Vadapav, Mix Bhaji, Vadapav, Chat Items, Vadapav, Chicken Burger, Mix Bhaji, Chat Items, Mix Bhaji, Chicken Burger, Chat Items, Chicken Burger.

Formation with 3 itemset or 3 combinations of items can be described in table 5.

Table 05. Support of the combination of three categories

Category	Total	Support
Vadapav, Samosa, Mix Bhaji	13	43%
Vadapav , Samosa , Chat Items	11	37%
Vadapav , Samosa , Chicken Burger	9	30%
Vadapav , Samosa , Egg Omlet	12	40%
Vadapav , Mix Bhaji , Chat Items	19	63%
Vadapav , Mix Bhaji , Chicken Burger	17	57%
Vadapav , Mix Bhaji , Egg Omlet	11	37%
Vadapav , Chat Items , Chicken Burger	17	57%
Vadapav , Chat Items , Egg Omlet	10	33%
Chat Items , Samosa , Chicken Burger	10	33%

From Table 05, the combined data from the three categories that meet the threshold of more than 50%, there are 3 data that Vadapav + Mix Bhaji + Chat Items, Vadapav + Mix Bhaji + Chicken Burger, Vadapav + Chat Items + Chicken Burger

c) Formation of Association

The next step after determining the support from the combination of the category is to calculate the confidence to find the association rules. For the formation of the association refers to the combination of 3 itemset exceeds the threshold of 50%. By using the formula of calculation confidence, then obtained the data as shown in Table 6.

Table 06. Confidence 3 items

Category	Total	Support	Confidence
Vadapav, Mix Bhaji, Chat Items	19/27	63%	70%
Vadapav, Mix Bhaji, Chicken Burger	17/27	57%	63%
Vadapav, Chat Items, Chicken Burger	17/27	57%	63%

From Table 06 it can be seen that the association rules that form that there are three rules that can be used as a reference to make and know the rules associated with the highest confidence value.

Having in mind the confidence value of 3 items formed the next step make the final association rules and determine the highest confidence score of each combination described in Table 07 below. For the formation of the set of item association rules that have been obtained will be formed into a pattern of "If X, then Y".

Table 07. Association Rules

Category (X => Y)	Support XUY	Support X	conf
Vadapav, Mix Bhaji => Chat Items	63%	70%	90%
Vadapav => Mix Bhaji, Chat Items	63%	71%	89%
Mix Bhaji, Chat Items => Vadapav	63%	67%	94%
Mix Bhaji => Chat Items, Vadapav	63%	71%	89%
Chat Items, Vadapav => Mix Bhaji	63%	63%	100%
Chat Items => Mix Bhaji, Vadapav	63%	65%	97%
Vadapav => Chicken Burger, Mix Bhaji	57%	71%	80%
Vadapav, Mix Bhaji => Chicken Burger	57%	70%	81%
Chicken Burger => Vadapav, Mix Bhaji	57%	58%	98%
Chicken Burger, Vadapav => Mix Bhaji	57%	57%	100%
Mix Bhaji => Chicken Burger, Vadapav	57%	71%	80%
Mix Bhaji, Chicken Burger => Vadapav	57%	60%	95%
Chat Items => Mix Bhaji, Chicken Burger	57%	65%	88%
Chat Items, Mix Bhaji => Chicken Burger	57%	67%	85%
Mix Bhaji => Chat Items, Chicken Burger	57%	71%	80%
Mix Bhaji, Chicken Burger => Chat Items	57%	60%	95%
Chicken Burger => Chat Items, Mix Bhaji	57%	58%	98%
Chicken Burger, Chat Items => Mix Bhaji	57%	60%	95%

According to the table 07, the final result of the formation of the association rules, namely with the highest confidence score of 100% Chat Items, Vadapav => Mix Bhaji and Chicken Burger, Vadapav => Mix Bhaji.

Advantages/ Disadvantages:

Apriori algorithm is the most natural algorithm to implement and answer the real-world questions.
It uses large itemset accordingly this algorithm maybe used accompanying abundant databases.
Usually, restaurants and food joints have large databases accompanying various differences like type.
Once the frequent itemset are known stocks or raw materials used for that itemset are always kept in buffer.
The authors used the data mining tools or requests that once survive. While the database is used to store data namely utilizing a Microsoft Excel database.
It is simplest but it uses huge memory for database. The items in the menulist can changes additional.

Inference:

Uses association rules to find the cuisine that is great demand apiece buyers.
Combinations of items from the cuisines are create than it is appointed count of transactions in the support.
Itemset bearing support beneath minimum support are rejected and therefore forwarded for the next round.
We can too use the algorithm to find the frequent itemset migratory reasonable. Such as that cuisine is chosen in different seasons.

Usefulness:

The method used in the process of collecting data should to ask straightforwardly about some problems that are frequently confronted by restaurants particular about what the menu is the most demanding consumers and documentation is done to find data on things or variable in the form of a file or document.
The authors used a most demanding consumers and documentation is done to
database is used to store data that is using a Microsoft Excel database

Limitations of study: -

It is simplest but it uses huge memory for database.
The items in the menulist may changes overtime.

6. CONCLUSION

- In Based on the results of the implementation of Apriori algorithm method with association rules on the sales transaction data in a newly form restaurant (Fast Food Chain) generate some rules that:
  - If a customer ordered Vadapav Chat Items and then would order Mix Bhaji, also if customer ordered Chicken Burger and Vadapav then would order Mix Bhaji with the highest confidence score of 100%,
  - If a customer ordered the Chicken Burger, then be ordered Vadapav and Mix Bhaji or Chat Item and Mix Bhaji with confidence value of 98%.
  - If a customer ordered a Chat Items, then will order the Mix Bhaji and Vadapav with a confidence value of 97%.

And has been developed in the form of web-based applications. This data was just made small to easily manageable for predictions.

References:

1. Priyanka Joshi, Gori Shanker St. Margaret Engineering College Neemrana, India "Railway Accident Factors Analysis using Modified Checkpoint Apriori Algorithm " "IJSRD - International Journal for Scientific Research & Development| Vol. 5, Issue 10, 2017 | ISSN (online): 2321-0613"

2. Shubhangi D. Patill, Dr. Ratnadeep R. Deshmukh2 and D.K. Kirange3 1 Department of Information Technology, Government Polytechnic, Jalgaon, India 2 Department of Computer Science and IT, Dr. 8abasaheb Ambedkar Marathwada University, Aurangabaad, India "Adaptive Apriori Algorithm for Frequent Itemset Mining "Conference on System

Modeling & Advancement in Research Trends, 25th_27'h November, 2016

3. 1Research Scholar, Department of Information Technology,Hindustan University,Chennai, ,Tamilnadu "CLASSIC APRIORI: IMPLEMENTATION OF MODIFIED APRIORI ALGORITHM IN MATLAB FOR PADDY CROPS DISEASE PREDICTION USING DATA MINING TECHNIQUES" Associate Professor, Department of Information Technology, Hindustan University,Chennai, Tamilnadu " "Journal of Theoretical and Applied Information Technology 10th July 2015. Vol.77. No.1 2015 "