



Palestine Technical University (Kadoorie)
Faculty of Engineering and Technology
Department of Computer Systems Engineering

USING AI WITH SUPERMARKETS

By:

Mays Khalil Najjar 201811598

Supervised:

Dr. Yazeed Sleet.

Tulkarem, Palestine

Semester: 2nd Academic

Year: 2021-2022

ABSTRACT

Market Basket Analysis is an effective tool in retail industry which will help the market owner to increase the business and improve sales distribution techniques. This is totally done by association rule mining in which it analyses the customer behavior against the purchasing item from market. It analyses the customer purchasing pattern and generate frequent itemsets.

Data mining consists of artificial intelligence, machine learning and database management to extract new patterns for huge data sets and the knowledge associated with these patterns. So, we can use data mining in supermarket application, by using the association rule, which contain the Apriori Algorithm.

Generation of frequent itemset will enhance the market strategy, placement of goods and many more. Which results in increase in sales of goods and any one can do profitable business.

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

TABLE OF CONTENTS:

ABSTRACT.....	2
Introduction	4
1.2 Problem Statement	4
1.3 Objectives	5
Literature Review	5
Methodology.....	6
3.1 Data collection method	6
3.2 Technology and tools used.....	7
Software Diagrams.....	7
Conclusion and Future Work	10
5.1 Use Case Diagram.....	11
5.2 Activity Diagram	12
References.....	14

1. Introduction:

Market Basket Analysis is a data mining technique that is used widely to find the associations among products. It provides the retailer useful information on products are brought together by its customers. This information may be used to influence customers' purchases, to re-organize the current layout of the retail shop, allocating the products on the shelves, etc.

Its results allows companies make easier decisions on how to bundle products and how to reorganize stores in a better way to increase sales and increase customer satisfaction. Whenever a customer enters a retail store, she walks around with the intention of buying the items in mind. During grocery shopping, some unlisted items might grab the customer's attention, and she will most probably get the item, which is most likely not directly related to any of the other items in the basket.

There's a high probability that amongst various customers we will find similar associations between items. When checking out at the counter the items bought look unrelated to each other, but that may not be the case. The items bought may be highly dependent on each other, where if a customer buys item "A" there is a high chance, she would buy item "B".

Market Basket Analysis is done for decision making in choosing the appropriate allocation of products in the retail's layout. This is done by finding how the items are dependent on each other.

1.2 Problem Statement:

Nowadays people buy daily goods from supermarket nearby. And with the changing in time the habits or behavior of the customer also changes.

So, we will help the market owner to increase the business and improve sales distribution techniques by using the artificial intelligence.

The placement of the items is one of the challenges for the retailers. They are unaware of the purchasing habits of the customer so they don't know which items should be placed together in their store. 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.

1.3 Objectives:

The project's purpose is to get better efficiency of market and sales strategy using consumer transactional data collected during the sales transaction, identifying the frequent items according the transaction from the dataset and the basis of min_support and min_confidence, and generating of the association rule from the frequent item sets.

2. Literature Review:

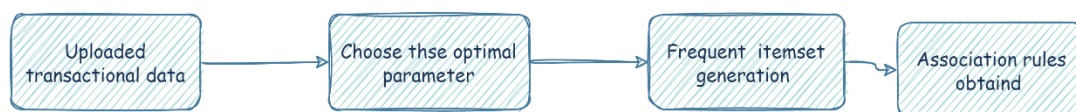
Many research and paper publish in this field:

- A Survey: On Association Rule Mining *Jeetesh Kumar Jain, **Nirupama Tiwari, ***Manoj Ramaiya
- Rakesh Agarwal proposed the Apriori algorithm. Apriori was the first associative algorithm proposed and future development in 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.
- Hilage, and Kulkarni (2011) applied association rule mining technique, rule induction technique and Apriori algorithm to the database of shopping malls. From the calculations, they observed

that Sunflower Oil, Sugar and Rava are highly correlated to each other.

- S. Kamley, S. Jaloree, R.S. Thakur in 2014 have developed an association rule mining model for finding the interesting patterns in stock market dataset. This model is helpful in predicting the price of share which will be helpful for stock brokers and investors to invest in the right direction by understanding market conditions.
- Deshpande, Shirpada, Patole and Bulsara (2010) showed how different methods of data mining can be applied to different fields. To conclude, it is very challenging to design and improve a data mining system, which can work correctly for different domains.
- In 2009 E. Ngai, L.Xiu and D. Chau presented how data mining in customer relationship management is an emerging trend, which helps in identification, attraction, retention and development of a customer. Customer retention and development are important to maintain a long 9 term and pleasant relationship with the customers which is very much useful in maximizing the organization's profit.

3. Methodology:



3.1 Data collection methods:

We use a United States supermarket dataset for marketing analysis purposes. From Kaggle website. Supermarket XYZ has been operating since 2008 and business flourished until 2016. They have a large database. These datasets are used to analysis a supermarket in United States for the purpose of increasing revenue. This data for one week.

3.2 Technology and tools used:

- Flutter:

Excellent technology to build a mobile app, it has tools that we need to build the app. And it can use python language as backend, by writing the API in python then from flutter we can connect to this API using http package.

- Python:

Is mostly known as the best AI language. It is commonly used due to its ease of syntax as well as the large variety of libraries available for AI.

- Numpy:

Fast and versatile, the NumPy vectorization, indexing, and broadcasting concepts are the de-facto standards of array computing today. offers comprehensive mathematical functions, random number generators, linear algebra routines, Fourier transforms, and more.

- Seaborn:

Is a library for making statistical graphics in Python. It builds on top of matplotlib and integrates closely with pandas data structures.

- Matplotlib:

Is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

- MLxtend:

A Python Library with Interesting Tools for Data Science Tasks. And we use it for Apriori implementation.

How:

In the beginning, the user will import his data which has the market basket data shape. Then:

1. Data pre-processing:

The data set used is related to retail industry. It contains information of each transaction with transaction ids. Each row represents a single transaction, i.e., information of a single customer. There are 7500 total transactions. For example, if a row presents the data like {Potpie, coffee creamers, fluid Milk, Egg, Butter}, it means this customer has taken those mentioned items in a single transaction.

	A	B	C	D	E	F	G	H	I	J	K	L
1	shrimp	almonds	avocado	vegetables mix	green grapes	whole weat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea
2	burgers	meatballs	eggs									
3	chutney											
4	turkey	avocado										
5	mineral water	milk	energy bar	whole wheat rice	green tea							
6	low fat yogurt											
7	whole wheat pas	french fries										
8	soup	light cream	shallot									
9	frozen vegetable	spaghetti	green tea									
10	french fries											
11	eggs	pet food										
12	cookies											
13	turkey	burgers	mineral water	eggs	cooking oil							
14	spaghetti	champagne	cookies									
15	mineral water	salmon										
16	mineral water											
17	shrimp	chocolate	chicken	honey	oil	cooking oil	low fat yogurt					
18	turkey	eggs										
19	turkey	fresh tuna	tomatoes	spaghetti	mineral water	black tea	salmon	eggs	chicken	extra dark chocolate		
20	meatballs	milk	honey	french fries	protein bar							

we have to prepare the data in a sparse matrix format where products are in column and id as index. Initially we group by based on the quantity purchased and later we encode it with 0s and 1s.

2. Applying Apriori Algorithm:

2.1 Choosing optimal parameters:

- The list that we want to extract rules from, and it's the data set that we have.
- The min_support parameter: is used to select the items with support values greater than the value specified by the parameter.

- c. The min_confidence parameter: to filter those rules that have confidence greater than the confidence threshold specified by the parameter.
- d. The min_lift parameter: specifies the minimum lift value for the short-listed rules.
- e. The min_length parameter: specifies the minimum number of items that we want in our rules.

We want rules for only those items that are purchased at least 5 times a day, or $7 \times 5 = 35$ times in one week, since our dataset is for a one-week time period.

The support for those items can be calculated as $35/7500 = 0.0045$.

The minimum confidence for the rules is 20% or 0.2.

Similarly, we specify the value for lift as 3 and finally min_length is 2 since we want at least two products in our rules.

These values are mostly just arbitrarily chosen, so you can change them as we want.

2.2 Building the rules:

We can build the rules using methods with the parameters that we choose.

2.3 Rule Strength Measure:


*Support (A) = frequency of (A) / number of transactions in the dataset.

*Confidence measures “conditional probability” i.e., how often a particular item is dependent on another

Confidence (if A then C) = Support (if A then C) / Support (A)

According to this we can accept or reject the rule. After the specification of the mentioned thresholds by the user, the association rules and the frequent item sets are generated side by side.

2.4 Visualization the rules:



This will represent the entire Association rules graphically, which will help us to understand the entire process in a single snapshot. In this graph, the support values for the Body and Head portions of each association rule are indicated by the sizes and colors of each circle.

The thickness of each line indicates the confidence value (conditional probability of Head given Body) for the respective association rule

4. CONCLUSION AND FUTURE WORK:

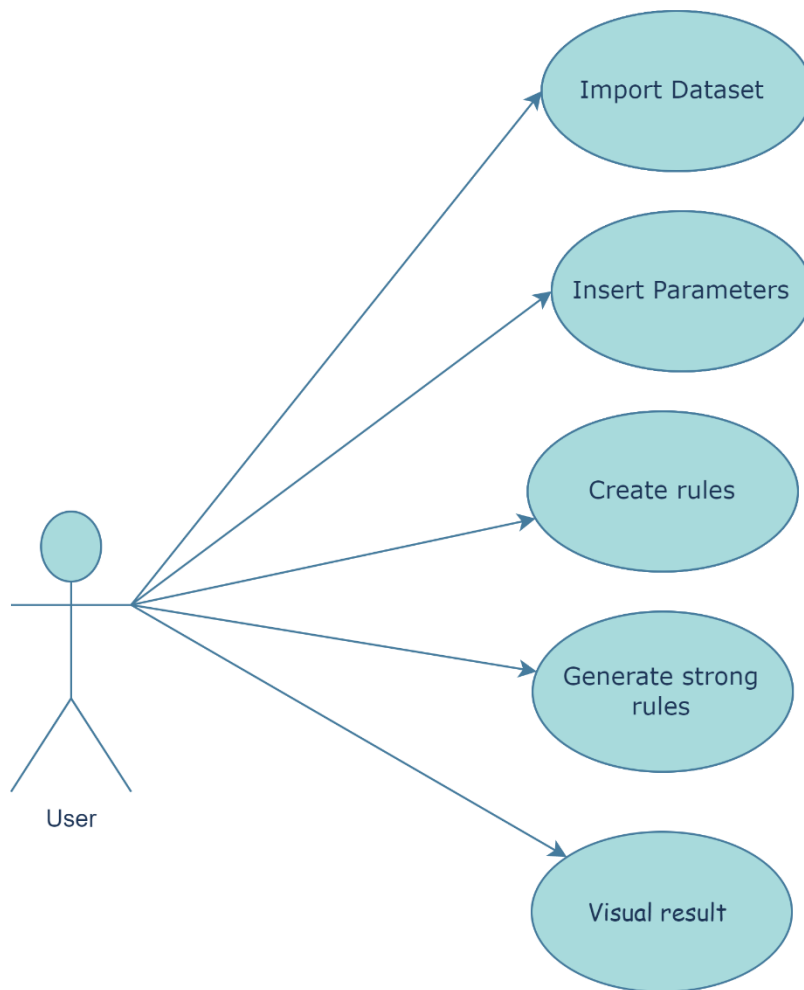
Mining Association Rules is one of the most used functions in data mining. Association rules are of interest to both database researchers and data mining users.

All the above information can be utilized for better marketing strategies. Retailer can arrange those frequently bought items very close to each other in the shelf so that customer can get all these items easily.

Some new products (related to previous items) can also be placed nearby which will attract to the customers. In future references the project can be further analyzed where the support threshold can be decreased and more rules can be obtained. Additionally, more transactions can be inputted in order to increase the range of customers.

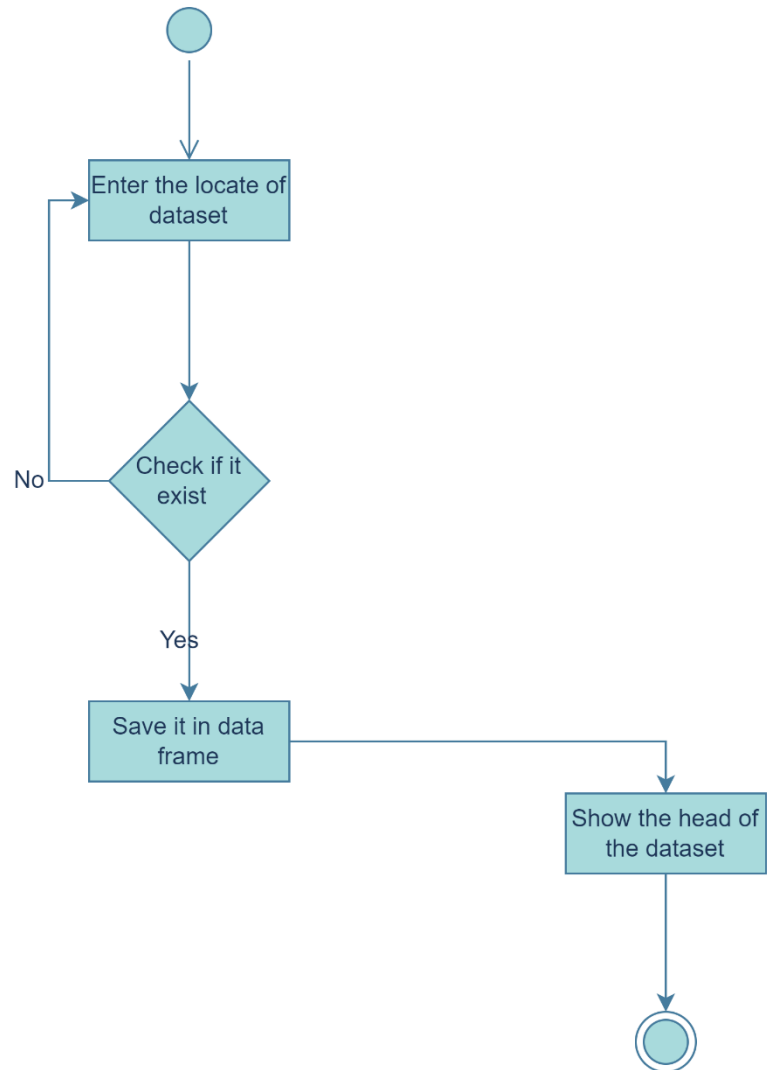
5. Software Diagrams:

5.1 Use Case Diagram:

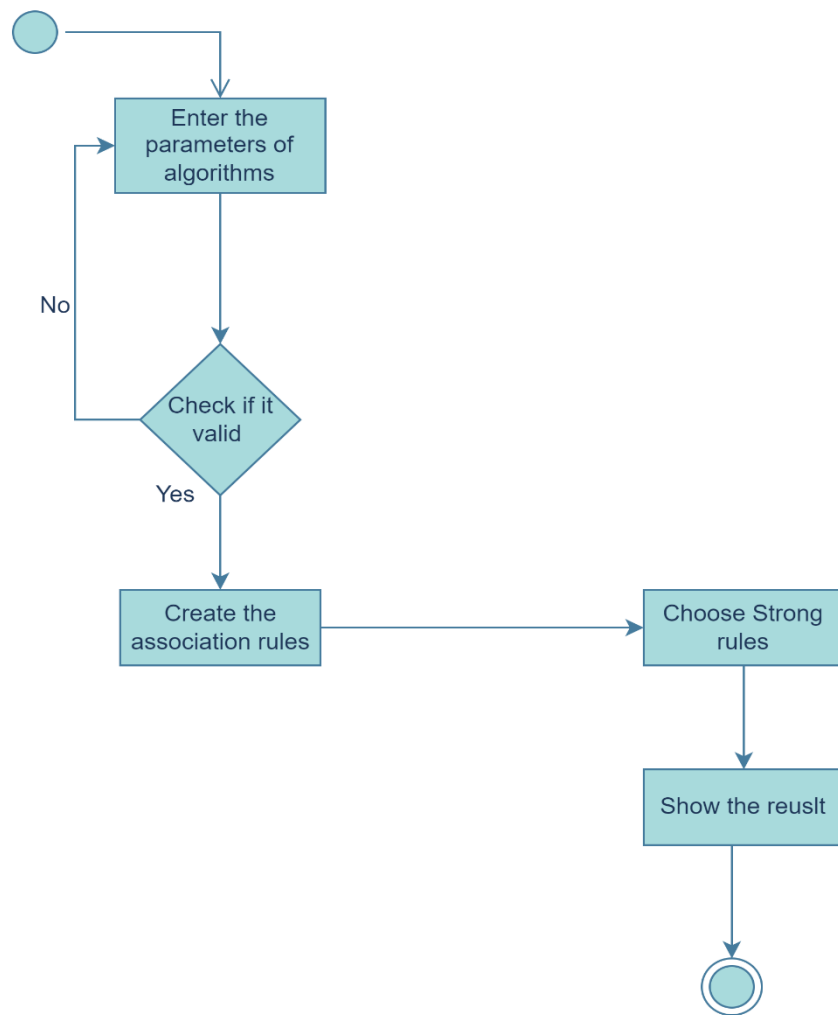


5.2 Activity Diagram:

5.2.1 Import Dataset.




5.2.2 Applying the Aprori Algorithm.



6. References:

1. Zaki, M. J. (2000). Scalable algorithms for association mining. Transactions on Knowledge and Data Engineering, 12(3), 372–390.
2. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2 PART 2), 2592–2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
3. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications - A decade review from 2000 to 2011. Expert Systems with Applications. <https://doi.org/10.1016/j.eswa.2012.02.063>
4. [Book] Data mining techniques: for marketing, sales and CRM. (Third Edition) by Gordon S. Linoff, Michael J.A. Berry
5. Mushtaq, A. (2015). Data Mining For Marketing. Journal of Marketing and Consumer Research, 9, 17–29.
6. Yen, S.-J., Wang, C.-K., & Lee, Y.-S. (2016). Mining Consumption Association Rules.
7. Gangurde, R., Kumar, B., & Gore, D. (2017). Building Prediction Model using Market Basket Analysis. International Journal of Innovative Research in Computer and Communication Engineering, 5(2). <https://doi.org/10.15680/IJIRCCE.2017.0502136>
8. J. Han, J. Pei, Y. Yin, and R. Mao, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach"



9. Phai Prasad J, Murlidher Mourya, A Study On Market basket Analysis Using Data

10. Yen, S.-J., Wang, C.-K., & Lee, Y.-S. (2016). Mining Consumption Association Rules.