

Market Basket Analysis using Apriori Algorithm

Dr. Yojna Arora¹, Ms. Neha Bhateja², Ms. Vanshita Goswami³, Mr. Rohan Kukreja⁴,
and Ms. Amisha Rajput⁵

^{1,2} Assistant Professor, Computer Science Engineering, ASET, Amity University Haryana, India

^{3,4,5} Student, Computer Science Engineering, ASET, Amity University Haryana, India

Correspondence should be addressed to Dr. Yojna Arora; yojana183@gmail.com

Copyright © 2022 Dr. Yojna Arora. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- A Technique that check for dependency for one Data item to another is *Association Rule* which is an old Data mining approach. Which is used to identify the next product that might interest a customer. The *Apriori* Algorithm is applied in this for mining frequent products sets and relevant Association rule. With this algorithm we can use this for up-sell and also in cross-sell to show the Association rule with the help of the algorithm. These methods are widely used in global companies, so for the good understanding the companies used the methods to remain up to date that what customers demands with which products. The results helps the big retailers to identify a trend for customers buying patterns, which is very helpful information for the retailers to plan their big business operations.

KEYWORDS- Data mining, market basket analysis, Apriori, Association

I. INTRODUCTION

In today's World the Globalization has a huge impact on business environment. And the business place become widely available, which affecting the customers demands in different way. These business marketplace also opens a lot of business opportunities for lot of business owners, shop retailers to do better business. For the betterment the shop and complex owners try to modify the arrangement of the products for the customers betterment, after figure out that what people want on what occasions. And to succeed this is the best challenging strategies to deal with, business means to compete with others and find a strategic way. And this management gives benefit to both customers and the business owners.

Market Basket Analysis is known for association rule which is a data mining approach, means the knowledge of the relation between the items or the products in a data transaction is the best way for the betterment of business. Now a days it is the best technique used by the inventory staff to know the co relation between the frequently purchased items. This technique is used mainly for mostly daily need items like for the grocery or for dairy products. With the advancement in business the market basket analysis has widely used for many business to help them in better decision making.

Market Basket Analysis is good to help businesses in making for better decision, mostly in marketing level. A

survey shows that this technique is widely implemented in most of the big enterprise companies. The purpose of market basket analysis implementation is get a better understanding of customer preferences based on customer daily transactions.

Widespread programs of Market Basket evaluation in retail are:

- Contrasting income, geared toward customers to promote on the way to spend extra money on their onetime shopping baskets call the search engines like google on the website as clients who buy this additionally buy this.
- Product placement or catalog composition, tell placement of content material merchandise on their media websites, or a product in its catalog
- Shop constructing, placed merchandise that show up collectively next to every other, to enhance the patron buying revel in 4. evaluation of the Loss leader, the loss chief is fee approach anyplace the product is sold at complete fee underneath its marketplace fee to inspire exceptional income for added worthwhile income or services. Market Basket Analysis :

Data Mining project may be divided into two classes, first: clustering and association pattern and second : type and regression. If we are saying we've item X and item Y.

An associated rule

X→Y indicate that every customer that purchases X will purchase Y too

There are three types of mining in association rule mining types as shown in the figure 1:

- Frequent Itemset mining used to find a frequent object which appears inside the entire transaction.
- While application Itemset Mining focuses on finding the set above a software threshold set via the consumer.
- The utility threshold can be any parameter like value, time and so on uncommon Itemset Mining is opposite to frequent Itemset Mining, as it attempts to find a uncommon set that exists inside the dataset. introduced collection market basket analysis, in which as opposed to locating the cooccurrence itemset, the author was focusing on analysing the shopping collection from large set of dataset.

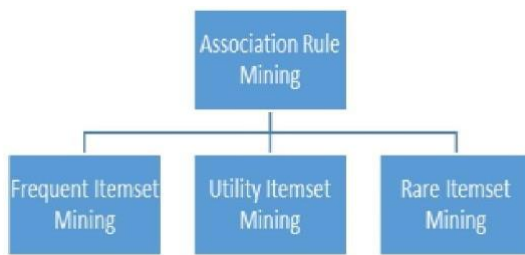


Figure 1: Association Mining

A. Data Mining [4][5]

The process of digging statistics to find hidden connections and expect future developments has an extended history. occasionally referred to as "facts acquisition information," the term "data mining" was not coined until the Nineteen Nineties. however it's far based totally on 3 integrated clinical disciplines: haematics (statistical evaluation of statistics relationships), practical intelligence (human intelligence demonstrated by means of software program and / or gadget) and device getting to know (algorithms can study from statistics to make predictions).

What was new is also new as statistics mining technology is constantly evolving to fit the unlimited energy of huge information and low cost laptop strength.

II. LITERATURE REVIEW

Data Mining has played an vital a part of marketing books of the last few many years. marketplace basket evaluation is one of the oldest web sites inside statistics mining field and is an tremendous instance of policies of mining association. Diverse algorithms of The affiliation Rule Mining (ARM) has been developed via researchers to assist customers acquire their dreams targets . Ramakrishnan Srikant and Rakesh Agrawal an proposed apriori algorithm that is one of classical algorithms for locating not unusual patterns regulations of Boolean association. The authors explain in element the concept of the guidelines of mass mining for length dating tables. By analysing the percentage client shopping a selected product as nicely a percentage of all income generated with the aid of this product. through making such organizations, you will effortlessly find the fine products and what is their It income aspect. evaluating which products are the best as crucial because the maximum cost clients experience those unique merchandise sorts every day. because the main departments products generate a variety of site visitors in the shop, it is vital that use this facts to set a few unique points merchandise close by. any other crucial broadcast of studies within the discipline of experimental evaluation is that this the technique of producing organizational policies. Berry and Linoff are destined to discover styles by excluding institutions or joint activities from store made data. clients bread buyers commonly buy a few bread-associated merchandise including milk, butter or jam. It it makes experience that those agencies of group s are set apart on the side of the income middle so that customers can access them right now. Such agencies are related to products in addition it should be set aside to remind

customers of associated products and steering within the center in a totally logical way.

III. METHODOLOGY

A. Market Basket Analysis [1][2]

Market basket analysis defines a combination of sales that usually take place in a joint venture in a trade. Suppose, those who buy bread and eggs, in addition they often get butter. Encouraging team it should target customers who buy bread and eggs by being given butter, to encourage them to pay more in their basket. Also called "Affinity Analysis" or "Association Rule Mining".

B. Association Rule [1][3]

Association Rule is associated with the announcement "what why ". This issue can be within the shape of a transactional announcement made through supermarket clients. as an example, at a retail save store, a hundred clients visited final month to to buy merchandise. It became stated that out of one hundred clients, 50 of whom have purchased Product A, 40 of them bought a Product B and 25 of them bought each Product A and Product B. importance of The law of integration may be measured before 3 obstacles, namely support, confidence and lift. Support of a product or set of products is that the fraction of transactions in our information set that contain that product or set of products.

Support A=

No. of Transactions that contain A/ Total Transactions

In example, Support of Product A is 50%, guide for Product B is 40% and help of Product A and B is 25% self belief is a conditional opportunity that customer buy product A will also purchase product B.

Confidence(A => B) = P(A | B)=

The Number of Transaction tat Contains A and B / Total Transaction that Contains A

Of the 50 customers who purchased Product A, 25 were bought with Product B. It means when someone buys product A, 50% chance of buying Product B too. Lift ratio shows how well the law works in acquisition results, compared to the random selection of a transaction. As a general rule, the lifting rate is greater than another proposes a particular app within the law. Great lift there is one indication that the presence of A in him increases the likelihood that product B may occur in this interaction. A smaller lift than one shows that the presence of A reduced the chances of product B may occur in this interaction.

Lift (A => B)= (Confidence(A => B))/ (Support(B))

A lift value of 1.25 implies that chance of purchasing product B would increase by 25%.

C. Apriori Algorithm [4][8]

The Apriori algorithm divides consumers segments / groups / collections are completely different start. Then find the normal object sets again organization rules for each category. The Apriori algorithm makes the search effort Clients behave as teams, to those specific groups of

individuals are usually successfully satisfied. The Apriori algorithm works in two steps:

- Generate all sets of common objects - A regular occurrence set is a set of object support in addition to a little support.
- Produce all the rules of a trusted organization from standard sets - The surest rule of thumb is a law with more confidence than a little self-confidence.

IV. PROPOSED WORK

The approach used is explained in this newsletter. There are two steps involved first is to apply association rule and second is to use Apriori algorithm. Python libraries are been used in this. Numpy supports large matrices and multi-dimensional records. It consist of inbuilt mathematical characteristic. Pandas are an crucial library for statistics scientists. it is an open-source gadget getting to know library that offers bendy excessive-level records systems and a ramification of analysis equipment. Matplotlib is chargeable for plotting numerical records.

A. Data Preparation. [6], [7]

In this phase, the dataset used in the research is checked , including daily order sheets from the supermart. We have thirteen months data from December 2010 until December 2011 have been collected which consist of 541911 rows and 8 columns.

Then the data was prepared for the process with included 3 main steps-

i) data cleaning, ii) data transformation iii) data reduction
The first is data cleaning , in which data set data values are processed in advance. Missing, missing and incomplete data identified and managed. Then in the data conversion, the data was converted into code numbers to simplify the algorithm. The last step is data reduction. In this step, the data set is reduced to suit the mining process of common materials and requirements.

B. System Design [9]

System Design can be explained in figure 2.

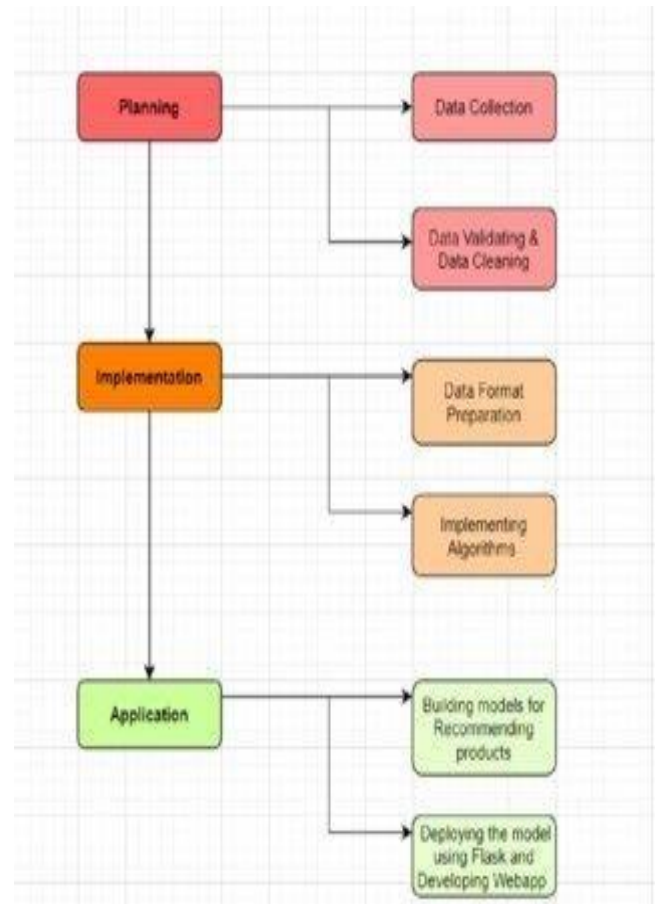


Figure 2: System Design

V. RESULTS AND DISCUSSION

Now after the data preparation and understanding we will apply the algorithms.

In applying the apriori algorithm, we are able to define the frequent data that we wanted by giving the support value.

After making use of the apriori set of rules and finding the often bought object, it's miles now the time for us to apply the association rules. From association regulations, we should extract facts and even discover information about which items that is greater powerful to be sold collectively We have created three baskets for three different countries. So we generate three results for each country. For France, From the association rules results, it could be seen in Figure 3.

- A. *Set/20 Red Retrosport Paper Napkins and Pack of 12 London Tissues, Set/6 Red Spotty Paper Plates are the items that has the highest association each other since these two items has the highest “lift” value.*

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED ...)	(PACK OF 20 SKULL PAPER NAPKINS, SET/6 RED SPO...)	0.033163	0.035714	0.030612	0.923077	25.846154	0.029428	12.535714
1	(PACK OF 20 SKULL PAPER NAPKINS, SET/6 RED SPO...)	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED ...)	0.035714	0.033163	0.030612	0.857143	25.846154	0.029428	6.767857
2	(PACK OF 20 SKULL PAPER NAPKINS, SET/6 RED SPO...)	(SET/6 RED SPOTTY PAPER CUPS, PACK OF 6 SKULL ...)	0.035714	0.040816	0.035714	1.000000	24.500000	0.034257	inf
3	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED ...)	(SET/6 RED SPOTTY PAPER CUPS, PACK OF 6 SKULL ...)	0.030612	0.040816	0.030612	1.000000	24.500000	0.029363	inf
4	(SET/20 RED RETROSPOT PAPER NAPKINS, SET/6 RED ...)	(SET/6 RED SPOTTY PAPER PLATES, PACK OF 6 SKUL...)	0.030612	0.040816	0.030612	1.000000	24.500000	0.029363	inf
...
433	(PLASTERS IN TIN WOODLAND ANIMALS, ROUND SNACK...)	(PLASTERS IN TIN SPACEBOY, PLASTERS IN TIN CIR...)	0.048469	0.089286	0.030612	0.631579	7.073684	0.026285	2.471939
435	(SET/6 RED SPOTTY PAPER CUPS, MINI PAINT SET V...)	(SET/20 RED RETROSPOT PAPER NAPKINS)	0.038265	0.132653	0.035714	0.933333	7.035897	0.030638	13.010204
436	(SPACEBOY LUNCH BOX, LUNCH BAG SPACEBOY DESIGN)	(LUNCH BAG DOLLY GIRL DESIGN)	0.056122	0.084184	0.033163	0.590909	7.019284	0.028439	2.238662
437	(LUNCH BAG DOLLY GIRL DESIGN)	(SPACEBOY LUNCH BOX, LUNCH BAG SPACEBOY DESIGN)	0.084184	0.056122	0.033163	0.393939	7.019284	0.028439	1.557398
438	(PACK OF 20 SKULL PAPER NAPKINS, SET/20 RED RE...)	(SET/6 RED SPOTTY PAPER PLATES, SET/6 RED SPOT...)	0.035714	0.122449	0.030612	0.857143	7.000000	0.026239	6.142857

395 rows x 9 columns

Figure 3: Results for France

For Germany, Here in Figure 4 we can see that,

B. Woodland charlotte bag and red Retrosport Charlotte Bag have highest lift value.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(WOODLAND CHARLOTTE BAG)	(RED RETROSPOT CHARLOTTE BAG)	0.126915	0.070022	0.059081	0.465517	6.648168	0.050194	1.739959
1	(RED RETROSPOT CHARLOTTE BAG)	(WOODLAND CHARLOTTE BAG)	0.070022	0.126915	0.059081	0.843750	6.648168	0.050194	5.587746
2	(PLASTERS IN TIN CIRCUS PARADE)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.115974	0.137856	0.067834	0.584906	4.242887	0.051846	2.076984
3	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN CIRCUS PARADE)	0.137856	0.115974	0.067834	0.492063	4.242887	0.051846	1.740427
4	(PLASTERS IN TIN SPACEBOY)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.107221	0.137856	0.061269	0.571429	4.145125	0.046488	2.011670
5	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN SPACEBOY)	0.137856	0.107221	0.061269	0.444444	4.145125	0.046488	1.607002
6	(ROUND SNACK BOXES SET OF 4 FRUITS)	(ROUND SNACK BOXES SET OF 4 WOODLAND)	0.157549	0.245077	0.131291	0.833333	3.400298	0.092679	4.529540
7	(ROUND SNACK BOXES SET OF 4 WOODLAND)	(ROUND SNACK BOXES SET OF 4 FRUITS)	0.245077	0.157549	0.131291	0.535714	3.400298	0.092679	1.814509

Figure 4: Results for Germany

And lastly for United Kingdom, the results are depicted in figure 5. Results for United the Kingdom.

C. Woodland charlotte bag and red retrospot Charlotte bag

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(WOODLAND CHARLOTTE BAG)	(RED RETROSPOT CHARLOTTE BAG)	0.126915	0.070022	0.059081	0.465517	6.648168	0.050194	1.739959
1	(RED RETROSPOT CHARLOTTE BAG)	(WOODLAND CHARLOTTE BAG)	0.070022	0.126915	0.059081	0.843750	6.648168	0.050194	5.587746
2	(PLASTERS IN TIN CIRCUS PARADE)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.115974	0.137856	0.067834	0.584906	4.242887	0.051846	2.076984
3	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN CIRCUS PARADE)	0.137856	0.115974	0.067834	0.492063	4.242887	0.051846	1.740427
4	(PLASTERS IN TIN SPACEBOY)	(PLASTERS IN TIN WOODLAND ANIMALS)	0.107221	0.137856	0.061269	0.571429	4.145125	0.046488	2.011670
5	(PLASTERS IN TIN WOODLAND ANIMALS)	(PLASTERS IN TIN SPACEBOY)	0.137856	0.107221	0.061269	0.444444	4.145125	0.046488	1.607002
6	(ROUND SNACK BOXES SET OF 4 FRUITS)	(ROUND SNACK BOXES SET OF 4 WOODLAND)	0.157549	0.245077	0.131291	0.833333	3.400298	0.092679	4.529540
7	(ROUND SNACK BOXES SET OF 4 WOODLAND)	(ROUND SNACK BOXES SET OF 4 FRUITS)	0.245077	0.157549	0.131291	0.535714	3.400298	0.092679	1.814509

Figure 5: Results for UK

The higher the lift value, the higher the association between the items will. If the lift value is more than 1, it is enough for us to say that those two items are associated each other.

VI. CONCLUSION

This studies proposes an implementation of MBA technique .The objective of this studies changed into to apprehend a patron shopping for pattern from any super mart patron’s daily order. 3 unique experiments with datasets had been conducted. As a end result, we can see the associated menu ordered by means of purchaser. for example, 70% of customers who order 12 PENCILS TALL TUBE SPOTTY will also order 12 PENCILS TALL forest . this is primarily based on 25% of the complete information. This statistics may be used for destiny planning in phrases of inventory and advertising and marketing approach. in this research, best 13 months information were used. For future research, more statistics have to be used. this will growth the accuracy of the results.

REFERENCES

- [1] Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia computer science*, 85, 78-85.
- [2] Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). Market basket analysis in a multiple store environment. *Decision support systems*, 40(2), 339-354.
- [3] Raorane, A. A., Kulkarni, R. V., & Jitkar, B. D. (2012). Association rule–extracting knowledge using market basket analysis. *Research Journal of Recent Sciences ISSN*, 2277, 2502.
- [4] Hilage, T. A., & Kulkarni, R. V. (2011). Application of data mining techniques to a selected business organization with special reference to buying behaviour. *International Journal of Database Management Systems*, 3(4), 169.
- [5] Arora, Y., Singhal, A., & Bansal, A. (2014). A study of applications of RBF network. *International Journal of Computer Applications*, 94(2).
- [6] Arora, Y., Singhal, A., & Bansal, A. (2014). PREDICTION & WARNING: a method to improve student's performance. *ACM SIGSOFT Software Engineering Notes*, 39(1), 1-5.
- [7] Sharma, N., Arora, Y., Makkar, P., Sharma, V., & Gupta, H. (2021). Real Estate Price’s Forecasting through Predictive Modelling. In *Machine Learning for Predictive Analysis* (pp. 589-597). Springer, Singapore.
- [8] Arora, Y., & Kohli, R. (2022). A Study for Making Universities Intelligent Using Data Analytics and Artificial Intelligence. In *Intelligent System Algorithms and Applications in Science and Technology* (pp. 147-155). CRC Press.
- [9] Musungwini, S., Zhou, T. G., Gumbo, R., & Mzikamwi, T. (2014). The relationship between (4ps) & market basket analysis: a case study of grocery retail shops in Gweru Zimbabwe.