iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.44835

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Smart Profitable Solutions with Recommendation Framework

Saurav Bawiskar[1], Onkar Joshi[2], Raju Choudhary[3], Kamlendra Tripathi[4]
[1, 2, 3, 4]*Student, Dr. Babasaheb Ambedkar Technological University, Loner, India*

Abstract: ***Discovering the frequent patterns in transactional databases is one of the crucial functionalities of apriori algorithm. Apriori algorithm is an algorithm which works on the principle of association rule mining. It is a dynamic and skillful algorithm used for discovering frequent patterns in a database, hence proving out to be efficient and important in data mining. Apriori algorithm finds associations between different sets of data. Every different set of data has a collective number of items and is called a transaction. The accomplishment of apriori is the set of rules that expose us how often any particular item or a set of items is contained in a set of data. In our proposed system, to provide efficiency, our basic aim is to implement apriori algorithm by setting up a threshold value and a varying support count which will act as a filter for our recommendation data. We can adjust the threshold value in order to increase or decrease the accuracy of the system. We have used apriori algorithm keeping in mind, its application in retailing industry and its capability of computing and handling large datasets and especially for the purpose of market basket analysis. The use of apriori algorithm along with analytical tools can provide insights into data and help the user in management and decision making provided that the user feeds the system in a correct way. Our aim is to provide user with recommendations which would ultimately help them in improving their business operations.***
Keywords: Apriori algorithm, Association rule mining, Data mining, Market basket analysis

I. INTRODUCTION

The business model of Retailing is concerned with selling goods and earning profits, the possibility of gaining profits is directly related to the sales number (assuming that the selling price set for the products are above cost price of product). Retail businesses currently use software for their business management. Most commonly used software for purpose of business management carries an ERP interface and certain tools which only help in carrying out transactional entries and analysis of sales and profits. Such software's are widely used today and are useful for retailers in managing their business operations. Most of these software's help the user in inventory management, billing and sales data visualization, but they lack in implementation of certain new technologies like prediction, forecasting, recommendation, etc. If the user is provided with a simple tool to analyze their sales data and provide them with some meaningful information out of that data, it can be very useful for the user. One such way in which we can provide this functionality is by combining the existing management systems along with a data mining system. Data mining is an important research domain nowadays that focuses on knowledge discovery in databases. Data mining uses data from the database to find out meaningful and useful information such that this information can be used in improve any existing system. Its objective is prediction and description.

One of the characteristic of data mining is the association rule mining. It consists of two procedures: first, finding the frequent item set in the database using a minimum support and constructing the association rule from the frequent item set with specified confidence. It relates to the association of particulars where for every event of a, there exists an event of b. This mining is more useful in performing the market basket analysis. Apriori is an algorithm for mining data from databases which shows items that are related to each other it showcases that for every item that an individual bought, what would be the possible items associated with the purchased item. Elaborate techniques, e.g., compressing the data, eliminating the redundant information within or between files that is reduplication, storing only updated parts of data, have been developed to effectively address the original objective of reducing the data size. In data mining apriori is an algorithm for finding and analyzing association between items. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation) association rule mining is a significant technique of data mining. This technique emphasizes more on finding relationships between two or more entities. For understanding these relationships, a technique called market basket analysis has been popularized in data mining. This helps in understanding the business organizations. Our approach was to provide a simple and easy management and recommendation tool to those small scale retailers which will help them in recommendation of products depending upon their sales history and also analyze the sales and profit numbers using graphs and charts.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

II. PROPOSED METHODOLOGY

To generate recommendations based on previous sales history, it is required that the inventory, billing and transaction data are fed properly into the system. To get the accurate data entries from the user we have created an inventory management system and a billing system which will help user keep a track of their stock summary. The billing system will generate the transactions which would be stored across the transaction database. Our recommendation system will use data from this database to generate a list of recommended products which were sold in the highest numbers in the given time span.

A. Architecture

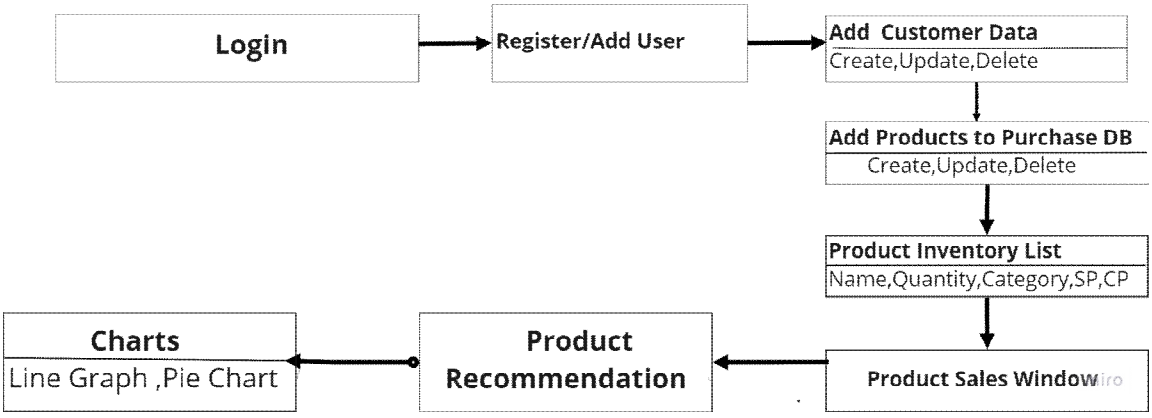


Fig. 1 Architecture

B. Algorithms

Apriori algorithm is the most established association rule mining algorithm. It is based on the apriori principle that all the nonempty (at least one) subsets of a frequent itemset must be frequent. It is a two-step process.

1) Step 1: The prune step

It scans the entire database to perceive the count of each candidate in Ck where Ck represents candidate k- itemset. The count of each itemset in Ck is match up with a predefined minimum support count to find whether that itemset can be arranged in frequent k-item set Lk.

2) Step 2: The join step

Lk is natural joined with itself to generate the next candidate k+1-itemset Ck+1. The main step here is the prune step which requires scanning the whole 1database for finding the count of each itemset in whole candidate k-itemset. If the database is enormous then it requires more time to find all the frequent item sets in the DB.

Input: D, Database of transactions; min sup, minimum support threshold

$$\text{Support (A)} = \frac{\text{Number of transaction in which A appears}}{\text{Total number of transactions}}$$

$$\text{Confidence (A}\longrightarrow\text{B)} = \frac{\text{Support(AUB)}}{\text{Support(A)}}$$

Fig. 2 Support and confidence

Fig.2 shows how support and confidence values are measured, in the above figure we can see that there are two products A and B. Support is the frequency in which product A appears in all transactions and Confidence is the combined occurrence of pair of A and B with respect to occurrence of A in transaction.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Output: L, frequent item sets in D.
Method:
(1) L1=find_frequent_1-itemsets(D);
(2) for(k=2; Lk-1≠Φ; k++){
(3) Ck = apriori_gen(Lk-1, min_sup);
(4) for each transaction t∈D{
(5) Ct=subset (Ck,t);
(6) for each candidate c∈Ct
(7) c.count++ ; (8) }
(8) Lk={c∈Ck |c.count≥min_sup }
(9) }
(10) return L=UkLk ; Procedure apriori_gen(Lk-1:frequent(k-1)-itemsets)
(11) for each itemset l1∈ Lk-1{
(12) for each itemset l2∈ Lk-1{
(13) if(l1 [1]= l2 [1])∧ (l1 [2]= l2 [2]) ∧…∧(l1 [k-2]= l2 [k-2]) ∧(l1 [k-1]< l2 [k-1]) then {
(14) c=l1∞l2;
(15) ifhas_infrequent_subset(c, Lk-1) then
(16) delete c;
(17) else add c to Ck ;
(18) }}}
(19) return Ck

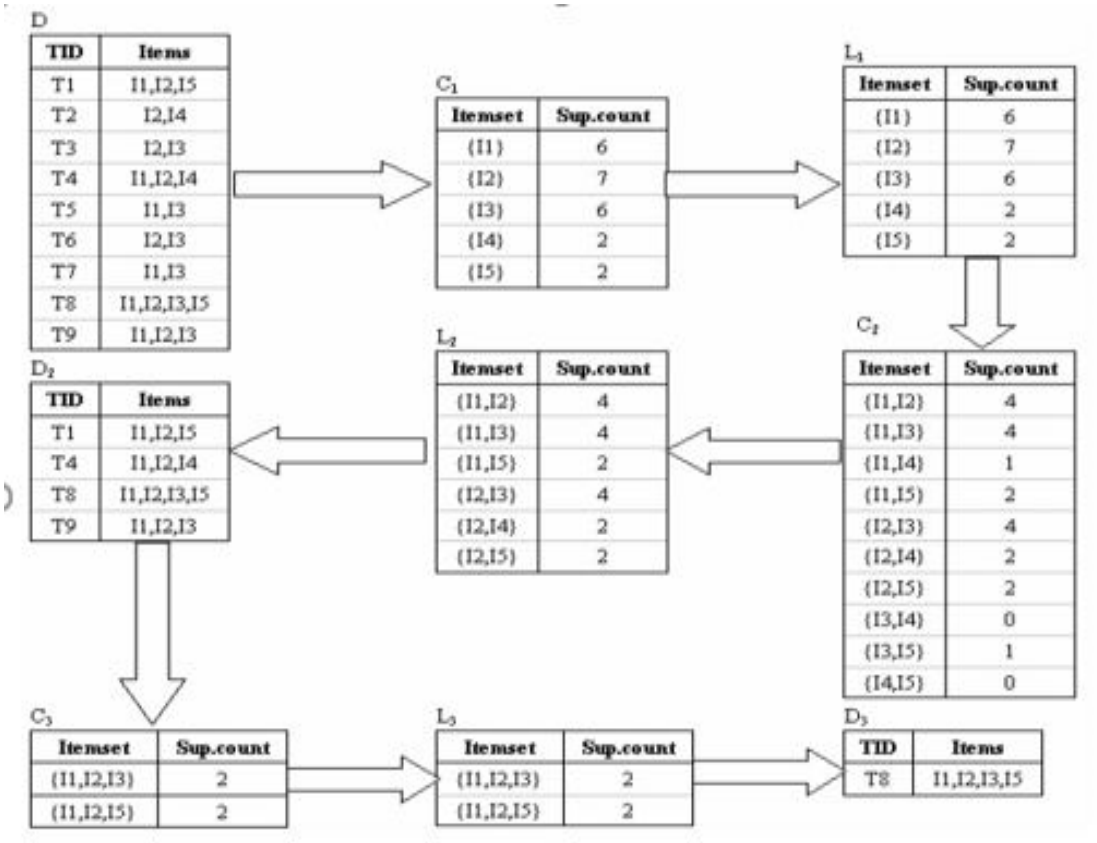


Fig. 3 Basic Diagram of how apriori mining algorithm associates items.

As observed in figure 3, there are 9 transactions (T1, T2, T3, T4, T5, T6, T7, T8, T9) and 5 item sets (I1, I2, I3, I4, I5). We have taken 2 as minimum support. As executing apriori algorithm, first pass consists (I1, I2, I3, I4, I5) as candidate generation items i.e. c1. By comparing with a minimum support frequent items (L1) as (I1, I2, I3, I4, I5). Further steps executed as per same methodology. Goal of the Apriori Algorithm is to find associations between different sets of data. Every distinct set of data has a number of items and is called a transaction. The accomplishment of Apriori is sets of rules that disclose us how often items are contained in sets of data. In order to find more valuable rules, our basic aim is to implement apriori algorithm using multithreading approach which can utilization our system hardware power to improved algorithm is reasonable and effective, can extract more value information. Serial mining consume time and reduce performance for mining. In proposed system, apriori algorithm is implemented in serial and parallel manner and comparisons of both on the basis of varying support count and time using parallel programming technique.

Association rule mining is interested in finding frequent rules that describe association between unrelated frequent items in databases, and it has two main measurements: support and confidence values. The frequent item sets is defined as the item set that have support value greater than or equal to a minimum threshold support value, and frequent rules as the rules that have confidence value greater than or equal to minimum threshold confidence value. These threshold values are generally assumed to be feasible for mining frequent item sets [1]. Association Rule Mining is all about finding all rules whose support and confidence outstrip the threshold, minimum support and minimum confidence values.

Association rule mining advance on two main steps. The first step is to find all item sets with adequate supports and the second step is to generate association rules by combining these frequent or large item sets.

Any given association rule has a support level and a confidence level. Support is the percentage of the population which fascinates the rule or in the other words the support for a rule R is the ratio of the number of occurrence of R, given all occurrences of all rules. The support of an association pattern is the percentage of task relevant data transactions for which the pattern is true.

C. Dataset

To get the results we have considered a list of 21 products and assigned them with a unique id. Out of those selected 21 products we have created a pseudo transaction dataset in csv format which has more than 5500 transactions.

D. Applications

This kind of recommendation systems can be used for all retail based operations. It can mainly be used for small scale retail operations for instance, a grocery store, pharmacy store, takeaway food outlet.

E. Results

Fig.4 given below show the csv transaction file which consists of set of transactions which have occurred in a period of time. The numbers which have been displayed in csv file are the assigned product id for every product.

	A	B	C	D	E	F	G	H	I
1	01-01-2020	21	5	6	2				
2	01-01-2020	11	20	19					
3	01-01-2020	21	19	21	1	14	5		
4	01-01-2020	19	20	14	18	7	15	2	
5	01-01-2020	21	19	4	1	15	12	12	
6	01-01-2020	6	13	11					
7	01-01-2020	1	6	17	3	5	8		
8	01-01-2020	4	13	16	17				
9	01-01-2020	13	15	6	4	18	15		
10	01-01-2020	6	9	8	15	6	9	20	
11	01-01-2020	21	14	12	10				
12	01-01-2020	21	4	20	4				
13	01-01-2020	5	19	8	7	4	14	6	
14	01-01-2020	10	12	10	13	19			
15	01-01-2020	1	9	18	14	3	7	5	21
16	01-01-2020	15	18	3	3	21	4	12	13
17	01-01-2020	11	18	7	4				
18	01-01-2020	14	5	3	14	9	12		
19	01-01-2020	13	7	9	20	10	3	16	
20	02-01-2020	4	12	11	7	4	19		
21	02-01-2020	10	21	14	6	9	16	5	

Fig. 4 transaction file as input (in csv format)

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

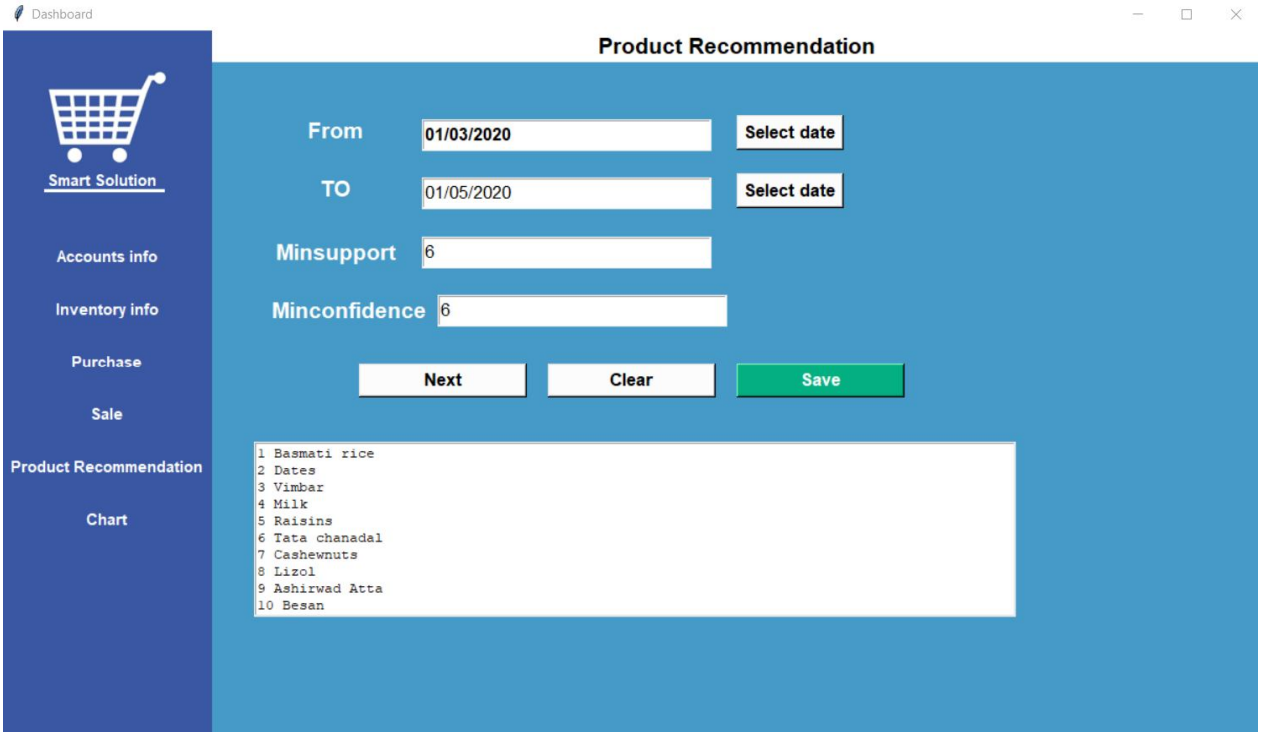
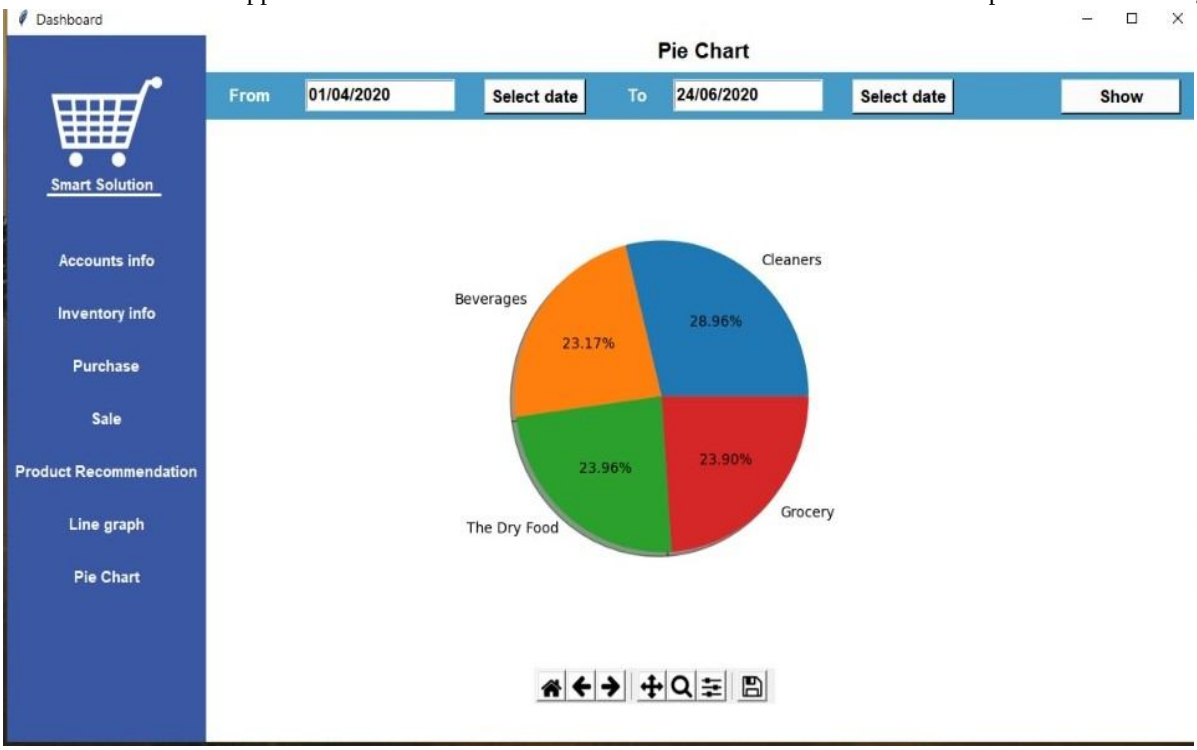


Fig. 5 Result window showing recommended products.

The Fig.5 shows the results for the given input transaction, first we have selected a csv file for storing our transaction data. We select the time duration to select the transaction records. The system will analyse records only in this set time frame. The user is then required to enter the minimum support and minimum confidence values after which the recommended product list will be generated.



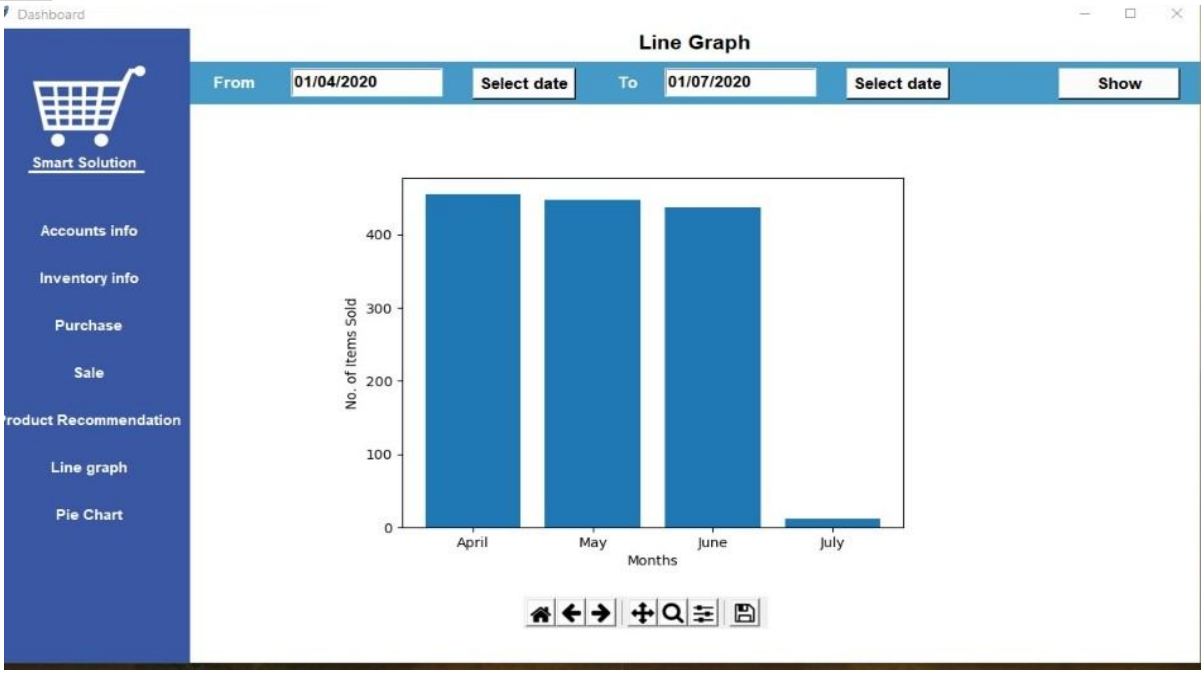


Fig. 6 Pie chart and line graph showing the categorical sales and sales numbers.

The Fig. 6. shows the results as per sales history, the pie chart is used for showcasing which product category comprises contributes what percentage in the overall sales numbers. The line graph shows the total number of items/products sold in any particular month.

III. CONCLUSIONS

The paper represents how the use of data mining along with data management and data visualization systems can help in improving business operations. The use of Apriori association rule mining algorithm can help in creating a frequently sold product item. Considering a retail scenario of use we can use this item set to help the user in their inventory management, and for setting up the pricing, offers and promotions for their products. The user can decide their sales strategy by the data returned by the system. As sales are directly proportional to the profit earned by the user, only if the user sets the selling price of products in a way where selling price is greater than cost price of the product. To get the results we have considered a list of 21 products and assigned them with a unique id. Out of those selected 21 products we have created a pseudo transaction dataset in csv format which has more than 5500 transactions. After selection of this file as input and entering the time duration, minimum support and confidence values from the user, the system returns the frequently sold product set to the user. The end result which is a list of products can also be used to study the seasonal sales pattern of some particular products and help the user in formulating their sales strategy. The paper gives work flow of how the product recommendation system works, the models and algorithm required for performing recommendations is included. The main focus was to demonstrate how data mining can be used along with data management and visualization tools in improving business operations considering the retail sector.

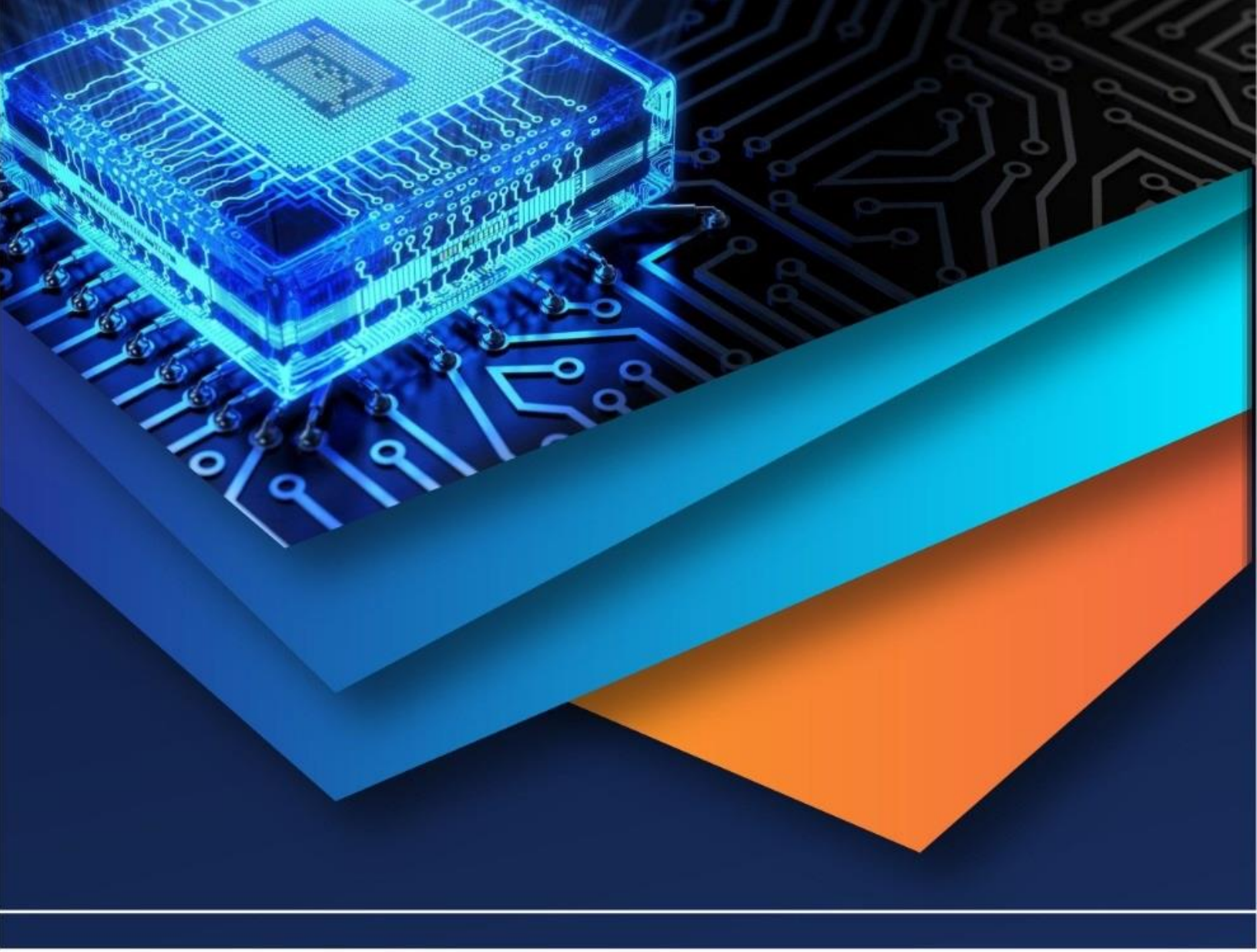
IV. ACKNOWLEDGMENT

We are Thank you to all respected person who helped in discussion for this topic and given their useful suggestion to make better performance. And we also thanks to all persons, references for sharing their knowledge with us.

REFERENCES

[1] Agrawal and R. Srikant. Fast algorithms for mining association rules. IBM Research Report RJ9839, IBM Almaden Research Center, San Jose, California, June 1994.
[2] JugendraDongre, S. V. Tokekar, and GendLalPrajapati, “The Role of Apriori Algorithm for Finding the Association Rules in Data Mining” International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),IEEE Catalogue Number: CFP1463W-DVD ISBN: 978-1-4799-2899-6, 2014.
[3] Sheila A. Abaya, “Association Rule Mining based on Apriori Algorithm in Minimizing Candidate Generation”,In:International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012.

[4] Mamta Dhanda," An Approach To Extract Efficient Frequent Patterns From Transactional Database", In: International Journal of Engineering Science and Technology (IJEST), Vol.3 No.7 July 2011, ISSN:0975-546.
[5] Annie, L. C., & Kumar, A. (2012). Market Basket Analysis for a Supermarket based on Frequent Itemset Mining. International Journal of Computer Science Issues, 9 (5), 257-264.
[6] Conceptual Model of Business Value of Business Intelligence Systems, Ales Popovic, Tomaz Turk, Jurij Jacklic, Journal of Management, Vol. 15, 2010.
[7] Nizar Mabroukeh and C. Ezeiefe, "A Taxonomy of Sequential Pattern Mining Algorithms", ACM Computing Surveys, Vol. 43, No. 1, Article 3, Nov. 2010.
[8] M. Tiwari, "Data Mining: A Competitive tool in Retail Industries", Global Journal of Enterprise Information System, vol. 2, Issue 2, December 2010.
Hongwei Liu, Bin Su and Bixi Zhang, "The Application of Association Rules in Retail Marketing Mix", Proceedings of the IEEE International Conference on Automation and Logistics, Jinan, China, 2007



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)