Research Journal of Recent Sciences ISSN 2277-2502
Vol. 1(2), 19-27, Feb. (2012) Res.J.Recent Sci.

Association Rule – Extracting Knowledge Using Market Basket Analysis

Raorane A.A.[1] Kulkarni R.V.[2] and Jitkar B.D.[3]
[1]Department of computer science, Vivekanand College, Tarabai park Kolhapur, MH, INDIA
[2]Chh. Shahu Institute of business Education and Research Centre Kolhapur, MH, INDIA
[3]Department of the Computer Science, D.Y. Patil College of engineering, Kolhapur, MH, INDIA

Available online at: www.isca.in
(Received 30th December 2011, revised 10th January 2012, accepted 11th January 2012)

Abstract

Decision making and understanding the behavior of the customer has become vital and challenging problem for organizations to sustain their position in the competitive markets. Technological innovations have paved breakthrough in faster processing of queries and sub-second response time. Data mining tools have become surest weapon for analyzing huge amount of data and breakthrough in making correct decisions. The objective of this paper is to analyze the huge amount of data thereby exploiting the consumer behavior and make the correct decision leading to competitive edge over rivals. Experimental analysis has been done employing association rules using Market Basket Analysis to prove its worth over the conventional methodologies.

Keywords: Decision making, data mining, association rule, market basket analysis.

Introduction

Database technology since the mid- 1980 has been characterized by the popular global adaptation of relational model and drastic change of research and development activities on new and powerful database systems. These employ advanced data model.

The exponential growth of computer hardware and system software technology in the past three decades has led to large supplies of powerful and cost effective computers, data collection equipment and storage media. This technology provides a great boost to the database and information industry and makes a huge number of databases and information repositories available for transaction management information retrieval and data analysis.

Data can be now stored in many different types of databases. One database architecture that has recently emerged and is widely commercialize is the data warehouse, a repository of multiple heterogeneous data sources, organized under a unified schema at a single site in order to facilitate management decision making. Data warehouse is the technology in campuses data cleansing, data integration and On-line Analytical processing (OLAP). OLAP incorporates analysis techniques with functionalities such as summarization, consolidation and aggregation as well as the ability to view information from different angles. OLAP tools have been commercially used for in depth analysis such as data classification, clustering and characterization of data changes over time.

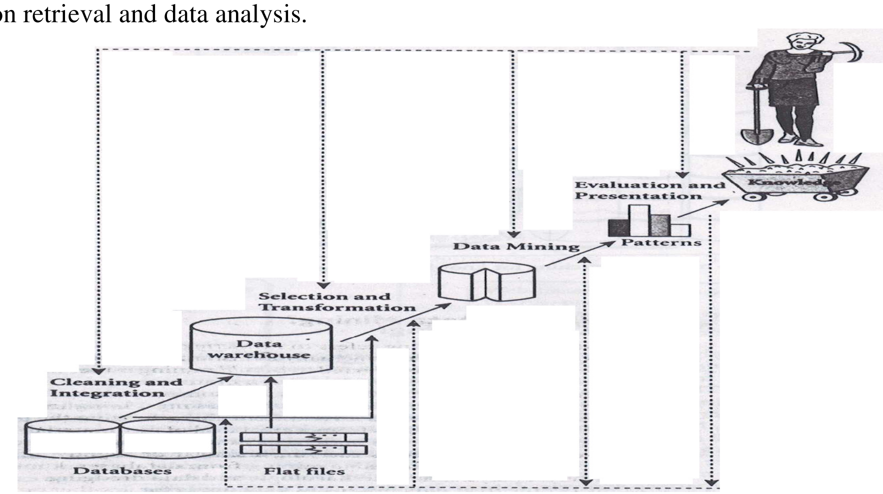


Figure -1
Data mining as a step in the process of knowledge discovery

The vast amount of data coupled with the need for powerful data analysis tools has been described as a data rich but information poor situation. The exponential rise in collection of data as well as storage, has lead to the necessity for powerful data analysis tools. As a result data collected in large databases has become "Data mountains", which are rarely visited. Consequently, the high valued decisions are often made based not on the information rich data stored in databases but rather on a decision makers intuition simply because the decision maker does not have the tools to extract the valuable knowledge inside vast amount of data. In addition consider current expert systems technologies which typically rely on users of domain experts to manually input knowledge into knowledge bases. Unfortunately this procedure is is prone to biases and is extremely time consuming and costly. Data mining tools perform data analysis and may uncover important data patterns, contributing greatly to business strategies and scientific and medical research. The ever widening gap between data and information, calls for a systematic development of data mining tools that will turn data tombs into "golden nuggets" of knowledge.[9]

Data Mining refers to extracting or mining knowledge from huge amounts of data. It means that the mining gold from rocks or sand is referred to as gold mining rather than rock or sand mining. Thus data mining should have been aptly called as "Knowledge mining from data" which unfortunately somewhat long. "Knowledge mining" a shorter term may not describe the importance of mining from large amount of data. Nevertheless mining is characterizing the process that finds a small set of precisions nuggets from a great deal of raw material. There are many other terms conveying similar or slightly different meaning to data mining such as knowledge mining from databases, knowledge extraction, data/pattern analysis, and data archaeology.

Data mining is simply an essential step in the process of knowledge discovery from the databases. Knowledge discovery as a process consists of an iterative sequence of the following steps: i) Data Cleansing (to remove noise and inconsistent data). ii) Data integration (where multiple data sources may be combined). iii) Data Selection (Where data relevant to he analysis task are retrieved from the database). iv) Data transformation (Where data are transformed or consolidated into forms appropriate for mining by performing summery or aggregation operations for instance. v) Data mining (an essential process where intelligent methods are applied in order to extract data patterns). vi) Pattern evolution (to identify the truly interesting patterns representing knowledge based on some interestingness measures). vii) Knowledge presentation (Where visualization and knowledge representation techniques are used to present the mined knowledge to the user)[6]

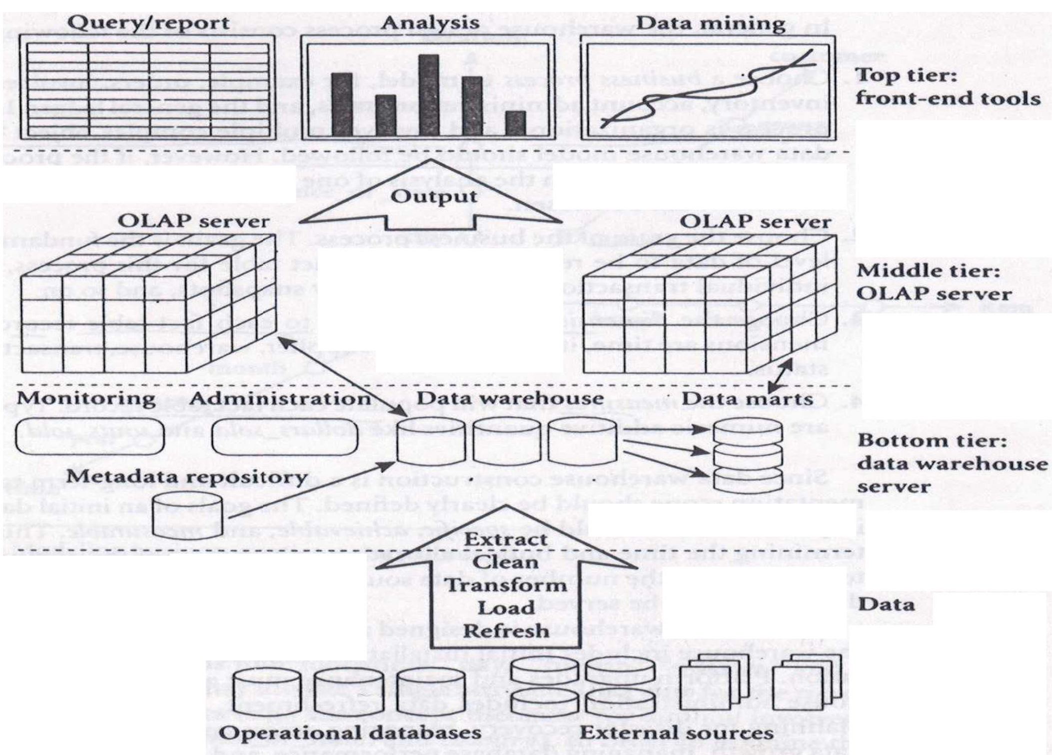
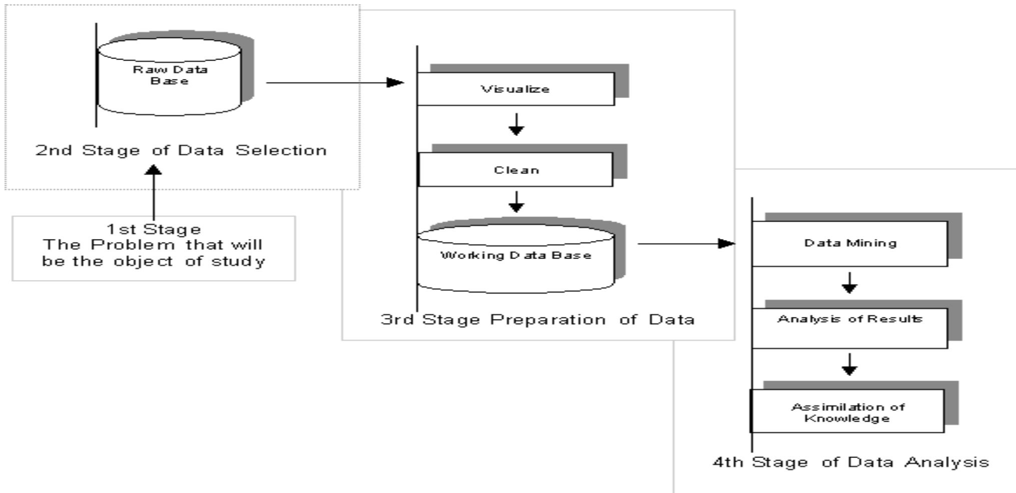


Figure – 2
A Three – tier data warehousing architecture



Figure–3
Sequence of steps in Data mining

There are various Data base terminology that can be used to represent knowledge from the data. Some of them are described in the following way,

Knowledge base - This is the domain knowledge that is used to guide the search or evaluate the interestingness of result pattern. Such knowledge can include concept hierarchies, used to organize attribute or attribute values into different levels of abstraction. Knowledge such as user benefits, which can be used to assess a pattern's interestingness based on its unexpectedness, may also be included. Other examples of domain knowledge are additional interestingness constraints or thresholds, and metadata

Data mining engine- This is essential to the data mining system and ideally consists of a set of functional modules for tasks such as characterization, association, classification, cluster analysis, and evolution and deviation analysis.

Pattern evaluation module - This component typically employs interestingness measures and interacts with the data mining modules so as to focus the search towards interesting patterns. It may use interestingness thresholds to filter out discovered patterns. Alternatively, the pattern evaluation module may be integrated with the mining module, depending on the implementation of the data mining methods used. For efficient data mining, it is highly recommended to push the evaluation of pattern interestingness as deep as possible into the mining process so as to confine the search to only the interesting patterns.

Graphical user interface- This module communicates between users and the data mining system, allowing user to interact with the system by specifying data mining query or task, providing information to help focus the search and performing exploratory data mining based on the intermediate data mining results. In addition this component allows the user to browse database and data warehouse schemes or data structures, evaluate mined patterns and visualize the patterns in different forms.

From a data warehouse perspective, data mining can be viewed as an advanced stage of on-line analytical processing (OLAP). However data mining goes far beyond the narrow scope of summarization-style analytical processing of data warehouse systems by incorporating more advanced techniques for data understanding.[5,7]

Related work: S. vijaylaxmi, V, Mohan, S. Suresh Raju in their paper "Mining of users' access behavior for frequent sequential pattern from web logs" explains, In sequential pattern Mining comes in association rule mining. For a given transaction database T, an association rule is an expression of form $X \rightarrow Y$ holds with confidence τ % of transaction set T if σ % of transaction set T support X U Y. Association rule Mining can be divided in to two steps. Firstly frequent pattern with respect to support threshold minimum support are mined. Secondly association rules are generated with respect to confidence threshold minimum confidence.[12]

Parvinder S. Sandhu, Dalvinder S. Dhaliwal and S. N. Panda in paper "Mining utility-oriented association rules" explains, An efficient approach based on profit and quantity" Association rule mining has been an area of active research in the field of knowledge discovery and numerous algorithms have been developed to this end. In this paper, they propose an efficient approach based on weight factor and utility for effectual mining of significant association rules. Initially, the proposed approach makes use of the traditional Apriori algorithm to generate a set of association rules from a database. The proposed approach exploits the anti-monotone property of the Apriori algorithm, which states that for a k-itemset to be frequent all (k-1) subsets of this itemset also have to be frequent. Subsequently, the set of association rules mined are subjected to weightage (W-gain) and utility (U-gain) constraints, and for every association rule mined, a combined utility weighted score (UW-Score) is computed. Ultimately, they determine a subset of valuable association rules based on the UW-Score computed. The experimental results demonstrate the effectiveness of the proposed approach in generating high utility association rules that can be lucratively applied for business development.[8]

B. Yıldız and B. Ergenç (Turkey) in "Comparison of Two Association Rule Mining Algorithms without Candidate Generation" Association rule mining techniques play an important role in data mining research where the aim is to find interesting correlations among sets of items in databases. In this study, they compare Matrix Apriori and FP-Growth algorithms. Two case studies analyzing the algorithms are carried out phase by phase using two synthetic datasets generated in order i) to see their performance with datasets having different characteristics, ii) to understand the causes of performance differences in different phases. Their findings are i) performances of algorithms are related to the characteristics of the given dataset and threshold value, ii) Matrix Apriori outperforms FP-Growth in total performance for threshold values below 10%, iii) although building matrix data structure has higher cost, finding itemsets is faster.[10]

Extraction of Interesting Association Rules using Genetic Algorithms, Peter P. Wakabi-Waiswa* Venansius Baryamureeba, In this paper they describes the process of discovering interesting and unexpected rules from large data sets is known as association rule mining. The typical approach is to make strong simplifying assumptions about the form of the rules, and limit the measure of rule quality to simple properties such as support or confidence. Support and confidence limit the level of interestingness of the generated rules.[4]

Jyothi Pillai in "User centric approach to itemset utility mining in Market Basket Analysis" describes Business intelligence is information about a company's past performance that is used to help predict the company's future performance. It can reveal emerging trends from which the company might profit . Data mining allows users to sift through the enormous amount of information available in data warehouses; it is from this sifting process that business intelligence gems may be found.[1]

"Efficient Association Rule Mining for Market Basket Analysis" Shrivastava A., Sahu R. writes in that Data mining is an attitude that business actions should be based on learning, that informed decisions are better than uninformed decisions, and that measuring results is beneficial to the business. Data mining is also a process and a methodology for applying the tools and techniques. So the focus of this paper is to enhance these algorithms in a way that it provides frequent profitable patterns which help market analyst to make the best informed decisions for improving their business.[2]

Material and Methods

Data Collection Method: From the supermarket called Shetkari Bazar in kolhapur city in Maharashtra, India, the day today transactional data is gathered. The sale of products are the actual transactions made by the customer, which consist of various products. This various products transactions made by the customer are stored in the secondary storage medium i.e. hard disk or CD's. It is stored in different categories i.e. department wise sale, counter wise sale, country wise sale, day wise sale, month wise sale. Depending on the necessary statistical analysis, required data is filtered through these databases.

Methodology: In a supermarket, suppose as a manager, he may like to learn more about the buying habits of the customers. "Which groups or sets of items customers are likely to purchase on a given trip to the store? To answer this question, Market Basket Analysis from Association Rule mining may be performed on the retail data of customer transaction at store. The result may be used to plan marketing or advertising strategies as well as catalog design different store layouts. In one of the strategy, items that are frequently purchased together can be placed in close proximity in order to further encourage the sale of such items together. Market Basket Analysis can help retailers to plan which items to put on sale at reduced prices.

If we think of the universe as the set of items available at the store, then each item has a Boolean variable representing by a Boolean vector of values assigned to these variables. The Boolean vector can be analyzed for buying patterns that reflect items that are frequently associated or purchased together. These patterns can be represented in the form of association rules. For example, the information that customers who purchase computers also tend to buy printer at the same time is represented in Association Rule below.
Computer = Printer
Support = 20%, Confidence = 80%

Rule support and confidence are two measures of rule interestingness they reflect the usefulness and certainty of discovered rules. A support of 20% means that 20% of all the transactions under analysis show that computer and printer are purchased together. A confidence of 60% means that 60% of the customers who purchased a computer also bought the printer. Typically association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold can be set users or domain experts.[17,3]

Basic Concepts: Let J = { i1, i2, …..i_m) be set of items. Let D, the task- relevant data, be a set of database transactions where each transaction T is set of items such that T ⊆□J. Each transaction is associated with an identifier, called TID. Let be a set of items. A transaction ∩∪⊆→│T is set to contain A if and only if A ⊆ T. An association rule is implication of the form A ⇒ B, where A ⊂ J, B ⊂ J, and A ∩ B = ϕ . The rule A ⇒B holds in the transaction set D with support where is the percentage if transaction D. That contain A ∪ B(i.e. both A and B) this is taken to be the probability, P(A ∪ B) the rule A ⇒ B has confidence C in the transaction set D if C is the percentage of transactions in D containing A that also contain B. that is taken to be the conditional probability, P(B│A) that is
Support = (A ⇒ B) = P (A ∪ B)
Confidence = (A ⇒ B) = P (B│A)

Rules that satisfy both a minimum support threshold(min_sup) and minimum confidence threshold (min_conf) are called strong. By convention we write support and confidence values so as to occur between 0% and 100% rather than 0 to 1.0.[11]

A set of items is referred to as an itemset. An itemset that contain K items is a K-itemset. The set (computer, printer) is a 2-itemset. The occurrence frequency of an itemset is the number of transaction that contain the itemset, This is also known as simply as the frequency support count or count of the itemset. An itemset satisfies minimum support if the occurrence frequency of an itemset is greater than or equal to the product to min_sup and the total number of transaction in D. The total number of transaction required for the itemset to satisfy minimum support is therefore refered to as the minimum support count. If an itemset satisfies minimum support then it is a frequent itemset. The set of frequent itemset is commonly denoted by L_k.
Association rule mining is a two step process: i) Find all frequent itemset. By definition each itemset will occur at least as frequently as a pre-determined minimum support count. ii) Generate strong association rules from the frequent item set: these rules must satisfy minimum support and minimum confidence.[13,15]

Actual Work: Market Basket Analysis is a tool of knowledge discovery about co-occurrence of nominal or categorical items. Market Basket Transaction of Market Basket Analysis is a data mining technique to derive association between data sets. We have categorical data of transaction records as input to the analysis and the output of the analysis are association rules as a new knowledge directly from stored data.

Let us start with an example. Suppose we have transaction data from an organization and the number of transaction in one day are limited as the data shown below

Table–1
Transaction Dataset

Transaction ID	Items from customers who brought more than 1 item
1	Sugar, Wheat, Pulses. Rice
2	Sugar, Pulses
3	Wheat, Pulses
4	Pulses, Wheat, Rice
5	Wheat, Pulses
6	Sugar, Wheat
7	Sugar, Rice, Pulses

Based on the above data, we can derive the following output of association rules using Market Basket Analysis: i) Step by step computation of Market Basket Analysis is as follow. ii) Generate all possible association rules. iii) Compute the support and confidence of all possible association rules. iv) Apply two threshold criteria: minimum support and minimum confidence to obtain the association rule.

Let us call the items currently seen by the customer as X(independent variable) and other items associated to those current items as Y(dependent variable). If you have only two items, namely A and B, we have only two possible association rules:
[A] →[B] and [B] →[A]

If you have three items name A, B and C, we have 12 possible association rules:

Table– 2
Combination of purchased items

	X	**Y**
1	[A] →	[B]
2	[A] →	[C]
3	[A] →	[B,C]
4	[B] →	[A]
5	[B] →	[C]
6	[B] →	[AC]
7	[C] →	[A]
8	[C] →	[B]
9	[C] →	[AB]
10	[A,B] →	[C]
11	[A,C] →	[B]
12	[B,C] →	[A]

Notice that the independent variable(X) is combination of items up to d-1, where d is number of items. Dependent variable (Y) is the combination of the set difference between all items and items listed on the dependent variable.[16,17]

For our demonstration example, we have 4 items that generate 50 possible association rules as shown below.

Table–3
Association Rules

No	X	Y	No	X	Y	No	X	Y	No	X	Y	No	X	Y
1	A	B	11	B	A C	21	C	A B D	31	A B	C D	41	BD	A
2	A	C	12	B	C D	22	D	A	32	A C	B	42	BD	C
3	A	D	13	B	A D	23	D	B	33	A C	D	43	BD	A C
4	A	B C	14	C	A C D	24	D	C	34	A C	B D	44	CD	A
5	A	C D	15	C	A	25	D	A B	35	A D	B	45	CD	B
6	A	B D	16	C	B	26	D	B C	36	A D	C	46	CD	A B
7	A	B C D	17	C	D	27	D	A C	37	A D	B C	47	A B	D
8	B	A	18	C	A B	28	D	A B C	38	B C	A	48	A B	C
9	B	C	19	C	B D	29	A B	C	39	B C	D	49	A C	B
10	B	D	20	C	A D	30	A B	D	40	B C	A D	50	B C	A

Total number of items, d	1	2	3	4	5	10	100	500
Total possible association rules, R	0	2	12	50	180	57002	5.15378E+47	3.636E+238

In general, the total number of possible association rules, R is exponential to the number of items, d, which is according to the following formula: R= 3d- 3d+1+1. Clearly if we have thousand number of items, it is impossible to compute the frequency of all possible association rules and rules and most of them will not have enough support any way. To compute support and confidence, we first set our transaction data into binary data.

Table–4
Conversion of transaction data into Binary data

Tran ID	Items from the customers Who bought more than 1 item	Tran. ID	A	B	C	D
1	Sugar, Wheat, Pulses, Rice	1	1	1	1	1
2	Sugar, Pulses	2	1	0	0	1
3	Wheat, Pulses	3	0	1	0	1
4	Wheat, Rice, Pulses	4	0	1	1	1
5	Wheat, Pulses	5	0	1	0	1
6	Sugar, Wheat	6	1	1	0	0
7	Sugar, Rice, Pulses	7	1	0	1	1
		Sum	4	5	3	6

For simplicity we call the items by its first letter (A for Sugar, B for Wheat, C for Pulses, D for Rice). Let us give name Bin Record to the binarized transaction record table. Each column in Bin Record are named as A_, B_, C_ and D_. We also set an array of 1 on the left of Bin Record to help computing th support count and name it Bin One.

The support count denoted by $n(X \cup Y)$ can be compute. Number of transaction, N is always 7 in our denomination example because we have only 7 data. Support (in percent for each association rule is simply a ratio between support count and the number of transaction.

$$\text{Support } (X \rightarrow Y) = \frac{n(X \cup Y)}{N}$$

Support for the independent variable, n(x), is easier to compute than the support count for the union because it is only dependent on the sum of the binarised transaction record. Confidence is computed easily by taking ratio of support counts of the union of the dependent variable to the support count of dependent variable.

$$\text{Confidence } (X \rightarrow Y) = = \frac{n(X \cup Y)}{n(X)}$$

The third step to compute association rules of Market Basket Analysis is to apply two threshold criteria : minimum support and minimum confidence. Thus we set these two threshold into cells for example we set 40% minimum support and 80% minimum confidence.[17,18] Minimum Support = 40%, Minimum Confidence = 80%

Different threshold value will produce broader or stricter rules. For example, if we set the minimum support to 25% and minimum confidence to 60%. We can obtain 8 association rules as shown in the figure below.

Table–5
Calculated Support and confidence

No.	X	Y	N(XUY)	N	%Support	n(X)	Confidence	Is in Rule?
1	A	B	2	7	29%	4	50%	0
2	A	C	2	7	29%	4	50%	0
3	A	D	3	7	43%	4	75%	1
4	A	B C	1	7	14%	4	25%	0
5	A	C D	2	7	29%	4	50%	0
6	A	B D	1	7	14%	4	25%	0
7	A	B C D	1	7	14%	5	25%	0
8	B	A	2	7	29%	5	40%	0
9	B	C	2	7	29%	5	40%	0
10	B	D	4	7	57%	5	80%	1
11	B	A C	1	7	14%	5	20%	0
12	B	C D	2	7	29%	5	40%	0
13	B	A D	1	7	14%	5	20%	0
14	B	A C D	1	7	14%	3	20%	0
15	C	A	2	7	29%	3	67%	1
16	C	B	2	7	29%	3	67%	1
17	C	D	3	7	43%	3	100%	1
18	C	A B	1	7	14%	3	33%	0
19	C	B D	2	7	29%	3	67%	1
20	C	A D	2	7	29%	3	67%	1
21	C	A B D	1	7	14%	3	33%	0
22	D	B	3	7	43%	6	50%	0
23	D	B	4	7	57%	6	67%	1
24	D	C	3	7	43%	6	50%	0
25	D	A B	1	7	14%	6	17%	0
26	D	B C	2	7	29%	6	33%	0
27	D	A C	2	7	29%	6	33%	0
28	D	A B C	1	7	14%	6	17%	0
29	A B	C	1	7	14%	9	11%	0
30	A B	D	1	7	14%	9	11%	0
31	A B	C D	1	7	14%	9	11%	0
32	A C	B	1	7	14%	7	14%	0
33	A C	D	2	7	29%	7	29%	0
34	A C	B D	1	7	14%	7	14%	0
35	A D	B	1	7	14%	10	10%	0
36	A D	C	2	7	29%	10	20%	0
37	A D	B C	1	7	14%	10	10%	0
38	B C	A	1	7	14%	8	13%	0
39	B C	D	2	7	29%	8	25%	0
40	B C	A D	1	7	14%	8	13%	0
41	B D	A	1	7	14%	11	9%	0
42	B D	C	2	7	29%	11	18%	0
43	B D	A C	1	7	14%	11	9%	0
44	C D	A	2	7	29%	9	22%	0
45	C D	B	2	7	29%	9	22%	0
46	C D	A B	1	7	14%	9	11%	0
47	A B	C D	1	7	14%	12	8%	0
48	A B	D C	1	7	14%	15	7%	0
49	A C	D B	1	7	14%	13	8%	0
50	B C	D A	1	7	14%	14	7%	0

Output: Association Rule

Table–6
Obtained transaction in association rule

No	X	->	Y	%support	Confidence	Is in Rules?
3	A	->	D	43%	75 %	√
10	B	->	D	57%	80 %	√
15	C	->	A	29%	67 %	√
16	C	->	B	29%	67 %	√
17	C	->	D	43%	100 %	√
19	C	->	B D	29%	67 %	√
20	C	->	A D	29%	67 %	√
23	D	->	B	57%	67 %	√

From the various transaction in the table we establish the market basket analysis by following way:
Support – The frequency of the transaction where product X and Y buy together, Confidence – The frequency of the transaction in which the customer who purchase the product X also purchase product Y i.e. conditional probability.

From the table we can give threshold value to the support and confidence for getting the association rule. From the table we see that when given the minimum value to support i.e. 20% means the frequency of the product X and Y buy together and minimum value of confidence 60% means the frequency of the transaction when customer buy product X also buy product Y.

From the table it is seen that only transaction number 3,10,15,17,19,20 and 23 among 50 transaction have higher than the threshold values of support and confidence i.e. 20% and 60% means the product in the transaction which are there in the rule are certainly purchased by the customer and put these product in the basket.

Conclusion

It is observed from the analysis that the data mining tools can be effectively used for optimizing the patterns associated with dynamic behaviors of the transactions which were made by the customers in purchasing some specific products. I have used the Market basket analysis algorithm, a widely and more pre dominantly used algorithm from association rule in Data Mining. Using this algorithm the frequent transactions made by the customers have been analyzed using the support and confidence of the customers in buying associated items. By using this methodology it is seen that there exists certain association between the products at the time of purchasing the products by the customers.

Further it is observed that this analysis can be best be used in managing the product placement on the shelves in the supermarket. This method can prove to fetch more profit to the seller. Thus the Data Mining tool can be used to improve the strategy in placement of the product on the shelf by using the Data mining tools.

Recommendation

The knowledge generated from Data mining technique is useful for the organizations engaged in Retailing business for their decision making process.

References

1. Pillai Jyothi, User centric approach to itemset utility mining in Market Basket Analysis, *International Journal on Computer Science and Engineering (IJCSE),* **3,** (**2011**)
2. Parvinder S. Sandhu Dalvinder, Dhaliwal S. and Panda S.N., Mining utility-oriented association rules: An efficient approach based on profit and quantity, *International Journal of the Physical Sciences,* **6(2)**, 301-307 (**2011**)
3. Vijaylakshmi S., Mohan V., Suresh Raja S., Mining of users access behavior for frequent sequential pattern from web logs, *International Journal of Database Management System (IJDM),* **2**, (**2010**)
4. Yıldız B. and Ergenç B., (Turkey) in Comparison of Two Association Rule Mining Algorithms without Candidate Generation, *International Journal of Computing and ICT Research,* **674(131)**, 450-457 (**2010**)
5. Nan-chan Hsich, Kuo-Chang cha Enhancing consumer behavior analysis by data mining techniques **(2009)**
6. Peter P. Wakabi-Waiswa Venansius Baryamureeba, Extraction Of Interesting Association Rules Using Genetic Algorithms *International Journal of Computing and ICT Research*, **2(1),** (**2008)**
7. Shrivastava A. and Sahu R., Efficient Association Rule Mining for Market Basket Analysis, *Global Journal of e-Business and Knowledge Management*, **3(1), (2007)**
8. Junzo Watada and Kozo Yamashiro, A Data Mining Approach to consumer behavior- *Procedings of the*

first International Conference on Innovative computing Information (**2006**)

9. Giudici Paulo. Applied Data mining :Statistical Methods for business and industry, -ISBN 9812-53-178-5 (**2003)**

10. Jiauei Han and Michele Kamber, Data mining Concepts and Techniques, Simon Fraser University, ISBN 1-55860-489-8-**(2001)**

11. Aggarwal C.C. and Yu P.S. Mining association with the collective strength approach, *knowledge and data engg., IEEE,* **13(6)** 863-873 **(2001)**

12. Mike Chapple, Data mining an introduction Classification **(2001)**

13. Lars parner -Consumer behavior – The psychology of marketing **(2000)**

14. Sergey Brin, Rajeev Motwani and Craig silvertrin Beyound Market Baskets: Generalizing association rules for correlation, SIGMOD Record, *(ACM Special Interest Group on Management of Data),* **26(2)**, 265, **(1997)**

15. S. Kotsiantis Association Rules mining **(1994)**

16. Agrawal R. and Srikat, Fast Algorithms for Mining Association Rules- l. Sept **(1994)**

17. Agrawal R., Imilienski T. and Swami A., Mining Associations Rules between Sets of Items in large databases. Proc. of the ALM SIGNOD. *Int'l conf. on management of Data*, 207-216 **(1993)**