© May 2024 | IJIRT | Volume 10 Issue 12 | ISSN: 2349-6002

Data Analysis of Business

ACHAL BUTALE[1], RASHMI GARGAM[2], ABHINAV GANDHEWAR[3], GANESH BUKKAWAR[4], DR. DEVASHRI KODGIRE[5]

[1, 2, 3, 4, 5] Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering Research and Technology, Chandrapur, Maharashtra, India

Abstract— The secret of doing successful business lies in the accuracy of the decisions taken for the inventory management, production plans, being customer centric and being agile for the market developments. The business data processing for any business is huge one and may contain many hidden things, which must be revealed out intelligently and with optimization with respect to the time and other source constraints. Many times, it is beyond the scope of the human mind to figure out and relate the interdependencies of the multiple factors embedded in the business data and hence the machines could help in this context to make the task easy. When it comes to find the Association rules between different products of any shop or store, the Apriori algorithm tops the choice. The current review work depicts the attempts to use the Apriori algorithm in an optimized way and implementing the same according to the prevailing conditions. Because of the fierce competition in the market, everyone is busy with getting the maximum attention of people. For that producer must have products which satisfies the needs of customers. Huge scale research is going in this field. In such situations, customer requirements are very important. The value of a production plan can be modelled as a function that reflects the communication of the company with different agents, for example, customers and competitors. The issue concentrated in this system is to recognize the production plan with the maximum utility for a company, where expected number of the customers for the chosen products assesses the utility of a production plan in the plan. The solution is achieved using Apriori Algorithm in Data Mining.

I. INTRODUCTION

Because of the fierce competition in the market, everyone is busy with getting the maximum attention of people. For that producer must have products which satisfies the needs of customers. Huge scale research is going in this field. In such situations, customer requirements are very important. The value of a production plan can be modeled as a function that reflects the communication of the company with different agents, for example, customers and competitors. The issue concentrated in this system is to recognize the production plan with the maximum utility for a company, where the utility of a production plan is assessed by expected number of the customers for the chosen products in the plan.

Consider a supermarket with a large collection of items. Typical business decisions that the management of the supermarket has to make includes, what to put on sale, how to design coupons, how to place merchandise on shelves in order to maximize the profit, etc. Analysis of past transaction data is a commonly used approach in order to improve the quality of such decisions. Until recently, however, only global data about the cumulative sales during some time period (a day, a week, a month, etc.) was available on the computer. Progress in bar-code technology has made it possible to store the so called basket data that stores items purchased on a per-transaction basis. Basket data type transactions do not necessarily consist of items bought together at the same point of time. It may consist of items bought by a customer over a period of time. Examples include monthly purchases by members of a book club or a music club.

Several organizations have collected massive amounts of such data. These data sets are usually stored on tertiary storage and are very slowly migrating to database systems. One of the main reasons for the limited success of database systems in this area is that current database systems do not provide necessary functionality for a user interested in taking advantage of this information.

This project introduces the problem of mining" a large collection of basket data type transactions for association rules between sets of items with some minimum specified confidence, and presents an efficient algorithm for this purpose.

An example of such an association rule is the statement that 90% of transactions that purchase bread and butter also purchase milk. The antecedent of this rule consists of bread and butter and the consequent consists of milk alone. The number 90% is the confidence factor of the rule. We Propose Apriori algorithm for finding the k-least products which is also important for production plan.

II. LITERATURE REVIEW

A) Frequent Itemset Mining

High utility itemset mining finds all high utility itemsets with utility qualities higher than the base utility edge in a transaction database [14]. The utility of an itemset alludes to its related esteem, for example, benefit, amount or some other related measure. Some standard techniques for mining affiliation rules [1, 7] that is finding frequent itemsets depend on the bolster certainty demonstrate. They locate all frequent itemsets from given database. The issue of frequent itemset mining [1, 2] is finding the entire arrangement of itemsets that show up with high event in transactional databases. However the utility of the itemsets is not considered in standard frequent itemset mining calculations. Frequent itemset mining just considers whether a thing has happened frequently in database, however overlooks both the amount and the utility connected with the thing. In any case, the event of an itemset may not be a sufficient pointer of intriguing quality, since it just demonstrates the quantity of transactions in the database that contains the itemset. It doesn't uncover the genuine utility of an itemset, which can be measured regarding cost, amount, benefit, or different articulations of client inclination [17]. In any case, utility of an itemset like benefit, amount and weight are imperative for tending to certifiable choice issues that require expanding the utility in an association. In numerous regions of professional retail, stock, advertising research and so on basic leadership is essential. So it can help in examination of offers, advertising procedures, and outlining diverse sorts of index.

Illustration:

Consider the little case of transaction database, a client purchases numerous things of various amounts in a deal transaction. All in all, everything has a specific level of benefit. For example, expect that in an electronic superstore, the benefit (in INR) of 'Printer Ink' is 5, and that of 'Laser Printer' is 30. Assume 'Printer Ink' happens in 6 transactions, and 'Laser Printer' happens in 2 transactions in a transactional database. In frequent itemset mining, the event recurrence of 'Printer Ink' is 6, and that of 'Laser Printer' is 2. 'Printer Ink' has a higher recurrence. In any case, the aggregate benefit of 'Laser printer' is 60, and that of 'Printer ink' is 30; in this manner, 'Laser Printer' contributes more to the benefit than 'Printer Ink'. Frequent itemsets are essentially itemsets with high frequencies without considering utility. Be that as it may, some infrequent itemsets may likewise contribute more to the aggregate benefit in the database than the frequent itemsets. This case demonstrates the way that frequent itemset mining methodology may not generally fulfill the retail business objective. In all actuality a most profitable clients who may purchase full valued things or high edge things which may not present from substantial number of transactions are vital for retail business since they don't purchase these things frequently.

B) High Utility Itemset Mining

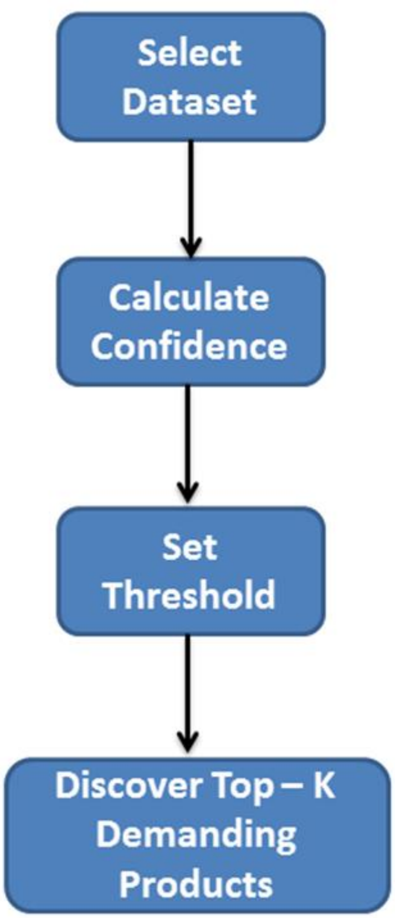
The constraint of frequent itemset mining lead scientists towards utility based mining approach, which permits a client to helpfully express his or her points of view concerning the value of itemsets as utility and after that find itemsets with high utility qualities higher than given limit [3]. Amid mining process we ought not recognize either frequent or uncommon itemsets but rather distinguish itemsets which are more valuable to us. Our point ought to be in distinguishing itemsets which have higher utilities in the database, regardless of whether these itemsets are frequent itemsets or not. This prompts to another approach in information mining which depends on the idea of utility called as utility mining. High utility itemset mining alludes to the disclosure of high utility itemsets. The principle target of high utility itemset mining is to distinguish the itemsets that have utility values above given utility edge [14]. The term utility alludes to its related benefit or some other related measure.

III. METHODOLOGY

The Entire Project is divided into 3 Modules:

1) User Interface Design
2) Dataset Selection & Preprocessing
3) Calculation of Confidence Value using Apriori Algorithm

After calculating the confidence value we will set a threshold call as *minconf*. Depending upon the *minconf we* will discover the Top-K demanding Products.



A DFD shows what kind of information will be input to and output from the system, where the data will come from and go to, and where the data will be stored. It does not show information about the timing of processes, or information about whether processes will operate in sequence or in parallel.

DFD Level-0

Fig 2 DFD Level-0

DFD Level-1

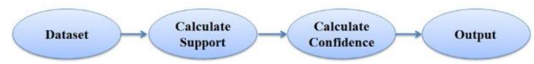
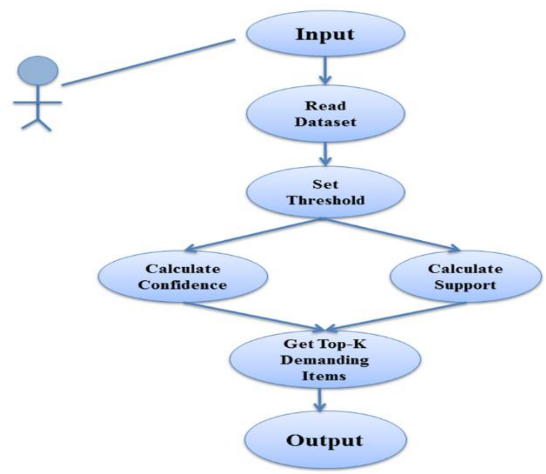


Fig 3 DFD Level-1

DFD Level-2

Fig 4 DFD Level-2

Use Case Diagram



OUTPUT

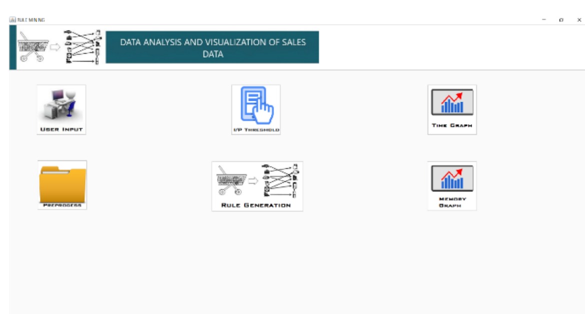


Fig 1 Home Screen

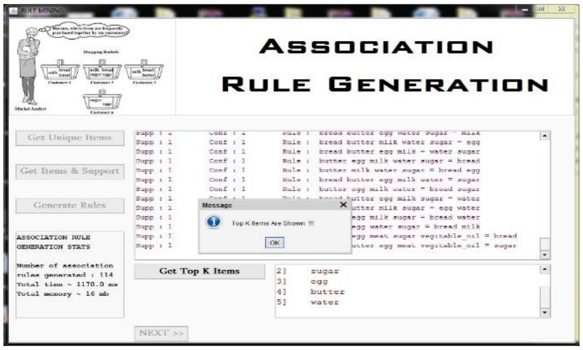


Fig 9 Generate Top-K Items

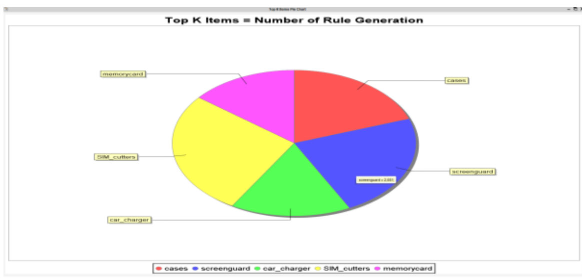


Fig 10 Top K-items pie chart

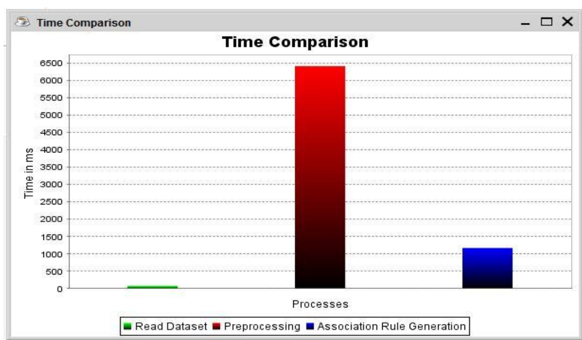


Fig 11 Time Comparison

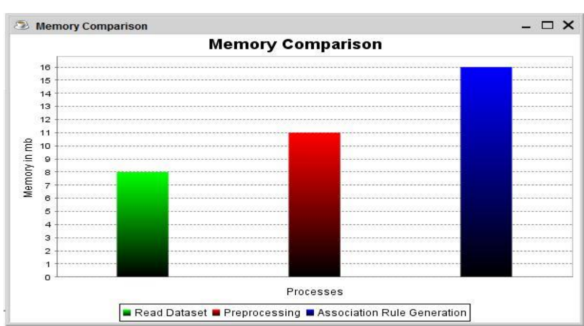


Fig 12 Memory Comparison

CONCLUSION

We implemented Apriori algorithm for association rule mining so as to get the top k demanding items from the transactional dataset. The Apriori principle can reduce the number of itemsets we need to examine. Put simply, the Apriori principle states that if an itemset is infrequent, then all its subsets must also be infrequent. This means that if {beer} was found to be infrequent, we can expect {beer, pizza} to be equally or even more infrequent. So in consolidating the list of popular itemsets, we need not consider {beer, pizza}, nor any other itemset configuration that contains beer. Results show the implementation work and the results generated in terms of Time & Memory Consumption.

REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A., “Mining Association Rules between Sets of Items in Large Database”, In: ACM SIGMOD International Conference on Management of Data (1993).

[2] Yao, H., Hamilton, H.J., Buzz, C. J., “A Foundational Approach to Mining Itemset Utilities from Databases”, In: 4th SIAM International Conference on Data Mining, Florida USA (2004).

[3] Yao, H., Hamilton, H.J., “Mining itemset utilities from transaction databases”, Data & Knowledge Engineering 59(3), 603–626 (2006).

[4] Liu, Y., Liao, W.K., Choudhary, A., “A Fast High Utility Itemsets Mining Algorithm”, In: 1st Workshop on Utility-Based Data Mining, Chicago Illinois (2005).

[5] Erwin, A., Gopalan, R.P., N.R. Achuthan, “CTUMine: An Efficient High Utility Itemset Mining Algorithm Using the Pattern Growth Approach”, In: IEEE CIT 2007.Aizu Wakamatsu, Japan (2007).

[6] Han, J., Wang, J., Yin, Y., “Mining frequent patterns without candidate generation”, In: ACM SIGMOD International Conference on Management of Data (2000).

[7] Erwin, A., Gopalan, R.P., Achuthan, N.R, “A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets”, In: International Workshop on Integrating AI and Data Mining. Gold Coast, Australia (2007).

[8] CUCIS. Center for Ultra-scale Computing and Information Security, North-western University.

[9] Yao, H., Hamilton, H.J., Geng, L., “A Unified Framework for Utility Based Measures for Mining Itemsets”, In: ACM SIGKDD 2nd Workshop on Utility-Based Data Mining (2006).

[10] Pei, J., “Pattern Growth Methods for Frequent Pattern Mining”, Simon Fraser University (2002).S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.