Product Analysis Support For Customers Using Association Rule Mining

Prof. Aruna Varanasi[1], Shaik Khasim Saheb[2], Prof.T V Narayana Rao[3]
Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India
Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India
Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India
[1]*arunavaranasi@sreenidhi.edu.in*
[2]*shaikkhasims@sreenidhi.edu.in*
[3]*tvnrbobby@yahoo.com*

***Abstract*— As the technology is rapidly growing mining the data is always plays a vital role. The Data Mining technology is still going to be expanded in much number of disciplines. Data Mining is commercially used in large number of applications. It occupies a major role in many of business aspects. If we see the applications of data mining evident it was everywhere like marketing, banking, healthcare, insurance, real-time industries, environmental, geographical, remote sensing, scientific, and in academic purposes also it has its own importance. Always it is used to make smart decisions for the business purposes in order to improve their growth in the business. On the other hand it is also helpful for customers. In that aspect we contribute this paper especially for customers to make a good decision. Huge number of products are coming into the market and going out. The customer can't expect the right product always; sometimes the product may be going to be failed within a short period of time. Upon consideration of all those things we came up with a a paper in such we In this paper we proposed architecture based on transactional data chosen from different data sources. We observed two factors whether to purchase a product or not based on association rule mining which is a popular domain of Data Mining. This implementation would suggest the customers to make a smart decision.**

***Keywords*— Product Analysis, Customer Support, Transactional Data, Association Rule Mining, Concurrent Apriori.**

I. INTRODUCTION

Data Mining is relatively new technology which is always focused to make smart decisions in many fields of businesses organizations like retail stores, banks, insurance companies, hospitals etc. All these organizations combine data mining statistics, recognize the patterns and decision making as also one of great aspect they are including. Data Mining can be used to find patterns which were very much helpful to make marketing strategies. There are so many applications of data mining used that are not only in one business aspects but tend to increase profits by creating new businesses and trends. The theme which was applied in one business using Data Mining also can be used in other business it was the one of the Data Mining major challenge. The prediction applied in one area has to be work out in case of other area also it's a challenging aspect in Data Mining i.e., we can name it as an automatic prediction in terms of Data Mining technology [2].

In one of its challenges like automatic recognition of patterns [12] it can be used to extract good amount of information which can be used in the business aspects. In this paper, the last section presents results of the prediction and patterns related to which we have created our own dataset on products and sales in the area of market basket analysis on sales discipline [1]. This paper is organized as follows. Section II describes the work related to data mining. The section III gives a brief overview of data mining and association rule mining. Section IV views the transactional data information which is applied in the current research. Section V presents the architecture which determines how customer can choose a decision to purchase a product. Section VI describes ARM on product data. Section VII examines the factors associated with products and sales. Section VIII presents the experimental results which we have obtained in our research work. Finally in the later part of the paper future work is presented, conclusion then some references which were helped us to do this research.

II. RELATED WORK

This section describes the related work which is appropriated with the association rule mining. In and around computer science and engineering we have so many disciplines where data mining is one of the major disciplines where in which always tremendous changes are to be taken care. In this domain the interested well-known and traditionally used example which we came across is market basket analysis. Along with the association rule mining it is also included with the other applications like marketing, financial, telecommunication [8] etc. The general definition to association rule mining is given in paper [4] where they were explained complete terminology which was related to ARM. The widely used and popularly known association rule mining algorithm is Apriori specified in [4]. To avoid difficulties in apriori we have FPGrowth [5]. Much research is going around data mining technology and to handle this people are proposing many algorithms and developing some new tools to maintain such kind of transactional data from a huge number of data sources. For processing the huge amount of data still research is going on, but if we see the current related work we have two solutions for processing such huge amount of data i.e., complex event processing (CEP). Secondly, the parallel distributed processing i.e. Hadoop framework. The question arises is now can we integrate the customer order data i.e., transactional data with customer feedbacks information in real time. For example, if we choose one transactional data related to configuration of the product, data set is available brand names, and available model data set. These three data sets can be taken concurrently. With this we can say that data is impending from

different data sources and finally to be integrated for making decisions. In this study we proposed architecture based on the above problem statement. The solution depicted here is theoretical as the transactional data which was in real-time data source had restricted access. The literature reveals that publicly available data usually come from a single data source [10]. So we proceeded to introduce an architecture based on different data sources, theoretically we proposed an architecture in the following section IV with sample data.

III. OVERVIEW TO DATAMINING AND ASSOCIATION RULE MINING

Generally, the term Data Mining can be defined as a process of distinguishing data from different perspective and encapsulating it into some profitable information. The information may be used to improve revenue or to increase costs or both. Technically we can say Data Mining is a procedure to find correlations and patterns among many fields over the large Relational Data Base. In the domain of Data Mining we use three key terms Data, Information, and Knowledge [11]. The term Data may be a fact or a number or text that can be handled by a computer machine. Today most of the organizations are processing huge amount of data through different formats and in various Data Bases resulting in transactional data, non-operational data, and metadata. The term information provides the patterns, associations and relationships which are adapted with the data. And the information which we obtained can be converted into knowledge to produce historical patterns and future trends. In the domain of Data Mining the association rule mining is a well-known research approach to find out relationships among the variables in a large scheme of databases [5]. It was dedicated to find strong association rules [3] by applying different measures over the dataset which we considered for doing research analysis. Many algorithms are implemented so far for achieving association rules in Data Mining. In our research we have considered most commonly used algorithm i.e. Apriori to extract the rules for the transactional data which is considered for analysis purpose.

IV TRANSACTIONAL DATA INFORMATION

Because of having some restricted access [8] over the publicly available transactional data, in our research we have framed our own transactional data and proceeded to do analysis on that data. The product chosen for the analysis purpose was laptop device. So the transactional data have following list of items:

1. RAM having three associated values like 2GB, 4GB, and 8GB.
2. Hard Disk Drive attribute associated with 320GB, 500GB, and 800GB.
3. Processor varies in i3, i5, and i7.
4. Graphic Card having a value 1 GB.

The study follows as below:

Phase1: In this phase we apply the basic Apriori algorithm with some rules. The rules will be discussed in results section.

Phase2: Based on the results with the above transactional data resulted in important some information with the specified rules

Phase 3: Here we considered one more dataset and extracted some other rules which will actually help in decision making whether to purchase a product or not.

This reason to name this approach as concurrent Apriori process based mining approach for customer to make a decision to buy a product or not.

V. ARCHITECTURE

In this study we proposed an architecture which will support to the customer to make a decision. The architecture which we proposed based on our little research. Below is the architectural diagram which represents concurrent apriori process based approach

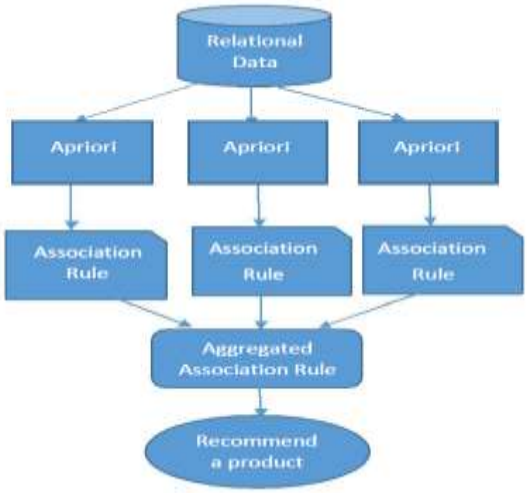

Figure 1 : Concurrent Apriori Process

In the above architecture because of having some real-time issues when we try to access the transactional data it doesn't allow to have access. Due to that we framed our own data to further proceed with the research. The data which we framed was totally related to product and sales only. The data which we considered is included in the section IV called Transactional Data Information. In that section we have provided on which product we made analysis such that the customer can choose a better decision. The data which we considered on top of that we have applied apriori algorithm individually then we came up with some association rules. Finally we combined all those individual association rules and we have got an aggregated rule. Based on that aggregated rule which we got we recommended the product to the customer so that the customer can choose the product i.e., customer can make a decision smartly.

VI. ARM ON PRODUCT DATA

Although the current existing studies which we have today are working on the transactional data as a classification problem, in our study we have seen this as a knowledge extraction issue further proceeded to the use of association rule mining (ARM). The various experiments have been performed on the transactional data. The purposes of experiments are one is to generate the best configuration of

a product among all available configurations which we have considered. And another experiment is to extract the best brand of the product among all available products in the market. Finally the last experiment is to come out with the best model among all the available models so that we can give suggestions to the customer to make a decision further to choose the best product. Subsequently the transactional data related to product and sales which we have chosen to do analysis was not in a position to obtain from a real world data sources. All the data which we used is our own assumption. The experimental results which we got they may show accurate even if we do on real world data set. In the final experiment the data was used based on number of complaints on appropriated product and the number of units sold in a year within a specified region.

VII. FACTORS ASSOCIATED WITH PRODUCTS AND SALES

In the discipline of retail that is market basket analysis [6] we commonly see the process like how the product is going to be delivered to the customer from a retailer store or from online product delivery sites. We can consider the factors like whether the product is really good or bad. The customer willing to take the product or not. Whether the cost of a product is meaningful or not. And the performance of the product how well it is. Likewise we can observe so many aspects from customer's point of view whether he or she ready to take product with satisfaction. To avoid all those confusions when the customer wants to take a product, in this study we considered the transactional data and applied some of the data mining techniques like association rule mining and prediction etc [14].

VII. RESULTS

The results which we got were totally based on our intentionally considered data set only. The results may vary if we consider other data sets. The data set details which we have chosen is as shown in below table.

Table 1: Configuration Data set

Tid	Item set
1	{ RAM=4GB, HD=320GB, Processor=i3 }
2	{ RAM=2GB, HD=500GB, Processor=i3 }
3	{ RAM=8GB, HD=800GB, Processor=i5, Graphics=1GB }
4	{ RAM=4GB, HD=500GB, Processor=i5 }
5	{ RAM=4GB, HD=320GB, Processor=i7 }
6	{ RAM=2GB, HD=320GB, Processor=i3, Graphics=1GB }
7	{ RAM=2GB, HD=320GB, Processor=i3 }

After applying the basic apriori algorithm on the very first data set the final item set which we obtained is as follows.

Table 2: Best Configuration

Item set	Support
{ RAM = 2GB, Hard Disk = 320 GB, processor = i3}	2

So this is the best configuration which we got when we apply association rule mining among the entire available configuration which we considered in our data set. Now we have best configuration to suggest to the customer. However, only with the help of configuration we cannot justify the product. For that purpose we concurrently choose data set and it produces the best model among all available models in the market. Now at this point of time we can suggest only the configuration and brand whereas the model issues is still to be generated. For that purpose we chose another data set which deals with the available models in the market. Finally we integrated all the rules which we are obtained and offer support to the customer so that he or she can proceed to buy a product. This entire procedure can occur concurrently based on our proposed an architecture called concurrent apriori. While choosing the second and third datasets we found that the transactional data comprise of factors like complaints, number of complaints registered, and number of units sold of a particular model of a particular brand. For the configuration mined we considered limited data set which has only six models from six different companies. We have applied the ARM on complaints data set. Based on study in this paper the following are the results depicted in table III.

Table III: Limited Sales Dataset with six models

Company Name	Model Number
HP	Pavilion G6-2313AX
Acer	NE56R
Samsung	NP300E4V-A01IN
Dell	Inspiron 15321
Lenovo	Essential B490
Sony	VAIO E15131

But we have checked sales data set to the purpose of number of units sold. The ratio of number of complaints to number of units sold is high for the following models as shown in table IV.

Table IV: Products having more number of complaints

Company Name	Model Number
Acer	NE56R
Samsung	NP300E4V-A01IN
Sony	VAIO E15131
HP	Pavilion G6-2313AX

So the other two models which have reminded after doing this concurrent process they qualified for the recommendations. Now we can suggest to the customer to

purchase the product without having any confusions. The two recommended models are as shown in table V.

Table V: Recommended products

Company Name	Model Number
Dell	Inspiron 15321
Lenovo	Essential B490

VII. CONCLUSION AND FUTUREWORK

From the customer's point of view when the customer wants to purchase a product usually undergo ambiguity to buy a product or not. So many factors are included when a product is to be delivered to the end user. It is a general issue in data mining discipline and this study is focused primarily on market basket analysis. So by considering all the factors, in this paper we have considered the data sets and applied some basic algorithms on data sets and extracted some rules associated with the data sets. Finally we integrated all the corresponding rules then we recommended the product to the customer. In the process we made an architecture which was related to a recent issue in data mining. We hope our architecture will be helpful for the researchers who are aspiring for research with real world data emerging from huge number of different data sources. Even though the experimental results has ample outcomes, this work can be appraised by considering meaning full (real world data) related to products sales and customers provided access is given with the data sources.

REFERENCES

[1] Prasadh.K and Sutherr.T: Logic based pattern detection based on multi- level propositionsal process: International journal of computer science & engineering technology(IJCSET), Vol.3 No.12 Dec 2012.

[2]K.Srinivas, G.RaghavenddraRao and A.Govardhan: Mining Association Rules from large datasets towards diseases prediction:International Conference on information and computer Networks(ICICN 2012)IPCSIT vol27,2012..

[3]Jabber, Dekshatlu, Prithvi Chandra," Information detection using associative classification for heart disease prediction" Springer Verlag 2012.

[4]Agarwal R and Srikant R,"Fast algorithm for mining association rules in large data bases" In proceedings of the VLDB .

[5]Han j and yin,"mining frequent patterns without candidate generation", Proceedings of ACM,SIGMOD (2000).

[6]Brin, Motwani and Silverstein, "Beyond Market Baskets: Generalizing Association Rules to Correlations," ACM *SIGMOD international conference on Management of data*, 1997.

[7]Zhong, Li and Wu, "Effective Pattern Discovery for Text Mining," *IEEE Transactions on Knowledge and Data Engineering*, Vol.24, No. 1, pp. 30-44, 2012.

[8]TEJ database, Historical database of finance, Taiwan Economic Journal Co. Ltd., http://www.tej.com./twsite/, 2011.

[9]Charu, Agrawal and Philip, "A New Methodology to Online Group of Association Rules'' IEEE, 2001.

[10] Jesmin Nahar, Tasadduq Imam, Tickle, Yi Phoebe, "Association rule mining to detect factors which contribute to heart disease in males and females" journal of Elsevier.

[I1]. H.Witten, and M.Hall, Data Mining: Real Machine Knowledge Tools and Methods, New Delhi: Morgan. 2011.

[12] Han, Data Mining: Concepts and Procedures New Delhi: Morgan Kaufmann, 2011.

[13]CH.Sandeep Kumar, K.Srinivas, PeddiKishor, T.Bhaskar:An alternative approach to mine Association rules,:2011 IEEE

[14] LUO Junwei and Zhang Bo:Research on mining positive and negative association rules:International conference on computer and communication technologies in agriculture engineering 2010.

[15]S.L. Hershberger, D.G. Fisher, Measures of Association (Encyclopedia of Statistics in Behavioral Science), JohnWiley & Sons, 2005.

