

# A Study of Sales Prediction Analysis in a Business Organization using Data Mining Technique

<sup>1</sup>Dr. Lakhmi Prasad Saikia, <sup>2</sup>Jaya Choudhary

<sup>1</sup>Professor, Computer Sc& Engg, Assam downtown University, India

<sup>2</sup>PhD Research Scholar, Computer Sc & Engg, Assam downtown University, India

<sup>1</sup> lp\_saikia@yahoo.co.in : <sup>2</sup>choudhary.jaya70@gmail.com

**Abstract:-** Various studies have been presented on sales prediction using datamining technique. The data mining technique has advantages & disadvantages however datamining techniques are more effective tool for analyzing sales prediction. The main objective of this paper is to give insights about customer's experience of buying pattern, mining the database and association using sales data.

**Keywords:-** Sales prediction, data mining, Association rule, frequent itemset generation, FP(frequent pattern) algorithm, MFP(Most frequent pattern), clusters.

\*\*\*\*\*

## I. INTRODUCTION

Data is a subfield of computer science in the computational process for finding some pattern from the large set of data. These large data set is used for further use by extracting some important information, which we call it as datamining process. A business organization every day collect large amount of data and data mining techniques are implemented on it, rather than it becomes in the consuming process to retrieve the data from the large dataset without datamining technique. A business organization is supported by three technologies

- A huge data collection
- Data mining algorithm
- A powerful computing process to implement.

There are various types of dataset that can be encountered in datamining. During marketing of sales of data we generally observe the purchase decision made over many time period of thousands of individuals who select among several products under a variety of price and advertising condition. It becomes very interesting concept to study the process who wants to buy certain products, his psychological mindset during purchase which is being converted into statistical format. This study will see any technical format is there to analyze customer's buying behaviour. The proposed research will make use of FP- Growth algorithm" to generate a set of association rules from a database. This algorithm will first analyze the data provided thus looking for specific types of patterns or trends.

## II. CONCEPT OF DATA MINING TECHNOLOGY

Data mining is the extraction of information from large databases from which future trends and behaviours of any

business can be made. The value of existing information can be enhanced and can be integrated with new product online using data mining techniques.

Data mining techniques have evolved as a result of lot of research and development. It has evolved with the evolution of computers, improvements in data access techniques etc.

The most commonly used techniques in datamining are :-

- Artificial neural networks:- They are non-linear predictive models that learn through learning.
- Decision trees:- A set of decisions i.e. represented by tree shaped structure.
- Genetic algorithm :- It can be use for datamining. These are optimized techniques that use processes such as genetic combination, mutation and natural selection in a design.
- Rule induction:- The extraction of useful information that rules from data based on statistical significance.

## III. AREA OF THE STUDY

Small business organization is a privately owned operated business that is limited in size and have small carrying loan cost of employees. Its revenue depends on the organization. These types of organization always have a problem of carrying loan cost. This drains the capital that puts pressure on the other that effects the predicted revenue. Due to this, there is always a gap between estimated & predicted results. This hampers the quality and growth of the business.

The research work will be for the small business organization using FP growth algorithm. A decision tree will be made on the basis of which a company can

make a good decision on the items sale for the profitable business.

#### IV. OBJECTIVE OF THE RESEARCH WORK

The following objectives are classified as follows:-

- 1) To give insights about customer's experience of buying pattern. This will make a prior prediction of database.
- 2) To classify the database through frequent itemset generation which helps to prepare a future prediction.

This study on customer will help the business organization improve their marketing strategies by understanding issues like

- a) Behaviour of customer.
- b) Customer motivation and decision strategy that differ between products and that differ in the level of importance.
- c) How business management can adjust and improve their sales and marketing ideas to more effectively reach customer.

#### V. REVIEW OF THE STUDY

Previous studies on customer behavior have been presented and used in real problem. Data mining technique are expected to be more effective tool for analyzing customer behaviour.

Junzo watada and Kozo Yanashi in their paper entitled "A data mining approach to customer behavior" tried to improve data mining analysis by applying several methods including fuzzy clustering principal component analysis. Many defects included in the conventional methods are improved in this paper.

In "Market basket analysis in multi store environment", the author Yen-ling chen, Kwei Tang, Reawolet jie shen, Yahan HU find out that there are two main problem in using the existing methods which are used in a multi-store environment. The first is caused by the temperol nature of purchasing pattern. An apparent example is seasonal products. The second problem is associated with finding common association pattern in subset of store. To overcome this problem, the authors developed an apriori like algorithm for automatically extracting association rules in multistore environment. Business domains have got a serious relationship with datamining and association rule mining for better profitability. Developments of business depends on

values, number of items sold, quality, preferences and likability of customers.

Sandhu et al(2011) propose an algorithm to evaluate the association between items or transaction based on weightage and utility factors. The product of these two metrics produces an easier and user friendly approach to derive association between customers, products and transactions on a stated period.

Li Xiaohui(2012) proposes a new kind of association rule mining algorithm and points out the limits of apriori algorithm on the basis of researching apriori algorithm. The improved algorithm deletes useless itemsets after generating candidate itemsets every time, reduces the number of itemsets generated in the next step, thereby reduces the times of database scanning, saving storage space required during algorithm and reduces the computational time. Verification results also show that the improved apriori algorithm can make the scanning fewer and reduces to about half. The time of scanning and comparison is even shorter when the database scale is quite large.

"Efficient association rule mining for market basket analysis" Shrivastava A, Sahu R, defines 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. Datamining is also a process and a methodology for applying the tools and techniques. Association rule mining is also one among the most commonly used techniques in datamining. A typical and the most running example of association rule mining is market basket analysis. This process of analysis customer buying habits by finding association between the different items that customer place in their "Shopping baskets". The discovery of such associations can help retailers develop marketing strategies by gaining insights into which items are frequently purchased together by customer and which items bring them better profits when placed in close proximity. For single dimensional association rule mining, FP-tree algorithm are in greater use today. Since candidate set generation in Apriori is still time consuming and costly.

#### VI. PROPOSED SYSTEM

With reference to literature work, the proposed solution introduces a better way for FP-growth algorithm by doing enhancement. The algorithm is designed in such a way, it executes effectively and efficiently. It should take less time and minimum number of scans to generate frequent item sets and strong association between them.

The Algorithm works as follows:-

Step 1:- Find minimum support for each item.

Step 2:- Order frequent itemset in descending order (consider only items with high or equal to minimum support.

Step 3: - Draw FP- Tree.

Step 4:- Minimum Frequent Pattern.

For this purpose a property matrix containing counted values of corresponding properties of each product has been used as shown below.

Let we have set X of N items in a dataset having set Y of attributes. The algorithm counts maximum of each attribute values for each item in the dataset.

INPUT : Dataset(DS)

OUTPUT: Matrix

FREQUENT PROPERTY

FPP(DS)

BEGIN

For each item  $X_i$  in DS

a) For each item  $X_i$  in DS

I) Count occurrences

For  $X_i$

$C = \text{count}(X_i)_i$

ii) Find attribute name of C

Next[End of inner loop]

b) Find Most Frequent pattern

$MFP = (\text{combine } M_i)$

Next[End of outerloop]

END

The above algorithm has been used to generate a property matrix containing counted values of corresponding properties of each item. This procedure receives data sets from clusters. Clusters are formed from data on basis of the quantity sold. The first loop scans all the records of the dataset. The innerloop counts occurrences of the attribute for a given item and placed in the MFP matrix. Finally

maximum occurrences of attributes value within a row give a single pattern.

### VII. EXPERIMENTAL ANALYSIS ON BUSINESS DATA

A business data consists of multiple attributes of item data related with sales process. This may include text, numerical and spatial data. Certainly the final report guide the seller about the status of selling the item. Implementaion on frequent itemset works as follows. Let us consider the frequent itemset with transaction id[vasiljevic Vladica ppt]:-

T_ID	ITEMSETS
1	f,a,c,d,g,m,p
2	a,b,c,f,l,m,o
3	b,f,h,o
4	b,k,c,p
5	a,f,c,l,p,m,n

min support:-3

[ a, b,c,d,f,g,k,l,m,n,o,p]

Now we calculate no. of transactions occurs for each item(i.e. support value)

item	support
a	3
b	3
c	4
d	1
f	4
g	1
k	1
l	2
m	3
n	1
o	2
p	3

Now we have to check which item has less than the minimum support 3 and they are d, g,l,m,n,o which has to be resolved and write the item having greater than or equal to three in the new table

item	support
f	4
c	4
a	3
b	3

m	3
p	3

Now we have to count how many times items are occurring in the pattern like

By using the rule of FP growth algorithm from the pattern.

f,c,a,b,m,p

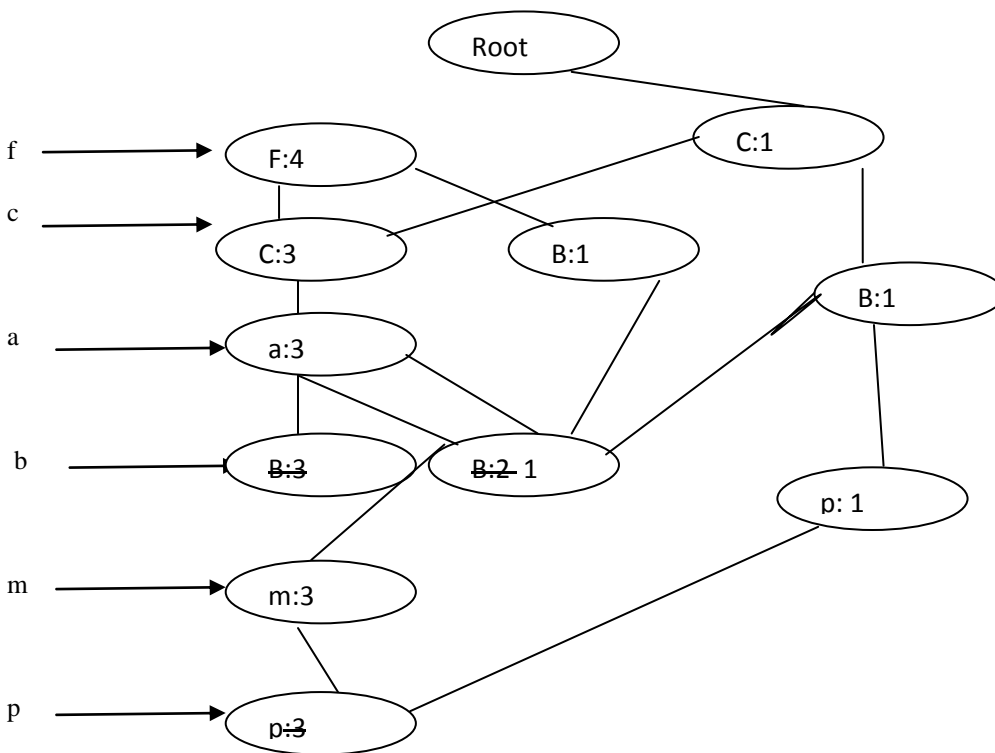
we create another table by using the pattern

tid	itemsets	ordered items
1	f,a,c,d,g,m,p	f,e,a,m,p
2	a,b,c,f,l,m,p	f,c,a,b,m
3	b,f,h,o	f,a
4	b,h,e,p	c,b,p
5	a,f,c,l,p,m,n	f,c,a,b,m,p

f : 4  
 c : 4  
 a : 3  
 b : 3  
 m : 3  
 p : 3

This part is used for creating FP tree.

**FP TREE**



**VIII. CONCLUSION**

In this system, clustering find associated patterns of sale. From the experimental results, It is clear that the approach is very efficient for mining patterns and predicting factors affecting the sales of items. We formulate most frequent pattern of items. We identify the trends of selling items through their known attributes. Our technique is very simple by using matrix and counting of attribute value. But this study has left with some work to implement decision making through online for customer. But our future work

will be extended. Here it takes less time and minimum number of scans to generate frequent itemsets.

**REFERENCES**

[1] Junzo Watada and Kozo Yanashi, “ A data mining approach to consumer behavior”, IEEE, 2006.  
 [2] Yen-ling chen, Kwei Tang, Reawalt jie shen, Ya-han HU, “ Market basket analysis in multistore environment”, Decision support system 40(2005) pg 339-354

- 
- [3] Sandhu et.al, “ Mining utility-oriented association rules: An efficient approach based on profit and quantity”. International journal of the physical sciences VOL 6(2), pp 301-307, 18 January 2011
- [4] Li xiaohui, “ Improvement Apriori Algorithm for association rules”, World automation congress(WAC), IEEE, 2012
- [5] A Shrivastava, R Sahu, “Efficient Association Rule mining for market basket analysis”, Global Journal of e-business & knowledge management 2007, VOL 3, NO-1, PP 21-25.
- [6] Wei Zhang, Hongzhi Liao, Na Zhao, “ Research on the FP-growth algorithm about association rule mining”, Business and information management , 2008, ISBIM '08, IEEE, 26 june 2009
- [7] Z Zheng, R Kohavi, L Mason, “ Real world performance of association rule algorithms”. 2001
- [8] B.Sangameshwari, P. Uma , A Survey on Data Mining Techniques In Business Intelligence, International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 3 Issue 10 October, 2014 Page No. 8575-8582
- [9] “Association rules mining” <https://www.vskills.in/certification/tutorial/data-mining-and.../association-rules-mining>.
- [10] Jiauei Han, Michele Kamber” Data mining concepts and technique “, Simon Fraser University ISBN 1-55860- 489-8-2001.
- [11] Aditya Joshi, Nidhi Pandey, Rashmi Chawla, Pratik Patil- “ Use of data mining techniques to improve the effectiveness of sales & marketing “ International journal of computer science and mobile computing VOL 4 ISSUE 4, April 2015, Pg 81-87
- [12] Aurangzeb Khan, Khairullah Khan, Behram B. Baharuddin “ Mining frequent patterns mining of stock data using hybrid clustering Association algorithm” International conference 2009.
- [13] J. R. Quinlan, “ Induction of decision trees, “ Mach. Learn ; VOL 1, PP. 81-106, 1986.
- [14] Prashant Palvia, Dwight B. Means, Jr; and Wade M. Jackson. “ Determinants of computing in very small business “ information & management VOL 27, 1994, PP. 161-174.
- [15] Robert Jay Dilger. “ Small Business size standards: A Historical analysis of contemporary issue “ Senior Specialist in American national government January 21, 2016.
- [16] Vasiljevic Vladica . “FP- Growth algorithm powerpoint presentation”.