DATA MINING TECHNIQUES TO PERFORM MARKET ANALYSIS

[1]G.Sivabharathi, [2]Dr.K.Chitra
[1]Research Scholar, [2]Assistant Professor
[1]Department of Computer Science, [2]Department of Computer Science,
[1]Madurai Kamaraj University, [2]Goverment Arts College, Melur ,Madurai, India.

Abstract: Market analysis is an important part of analytic device in retail organizations to set up the income for special segments of customers, to enhance customer satisfaction and to growth profit of the organization which has different channels and different areas. These problems for a popular shopping center are addressed the use of repeated object set mining and decision tree method. The repeated item sets are mined from the market basket database the use of the green. Apriori set of rules and hence the association rule are generated.

Index Terms – Association Rules, Repeated Item sets, Apriori, Market Basket Analysis.

I. INTRODUCTION

One of the challenges for organizations that have invested a set of consumer records series is the way to extract essential information from their huge purchaser databases and product feature databases, so that it will gain aggressive gain. Several factors of marketplace basket analysis were studied in instructional literature, such as the usage of consumer hobby profile and pastimes on particular merchandise for one-to-one advertising and marketing, buying styles in a multi-shop environment to improve the sales . Market basket analysis has been intensively used in many companies as a method to discover product associations and base a retailer's promoting strategy on them. Knowledgeable choice can be made without difficulty approximately product placement, pricing, merchandising, profitability and additionally reveals out, if there are any successful merchandise that don't have any large related elements. Similar merchandise may be observed so those can be placed close to every different or it can be move-bought. A store has to recognize the wishes of clients and adapt to them. Market basket analysis is one possible way to find out which objects may be prepare. Market basket analysis gives store proper statistics about related income on institution of goods foundation and additionally it is essential that the retailer should understand wherein channel and wherein area the products can be offered more and which session (i.e) morning or evening.

Market basket analysis is one of the facts mining techniques focusing on discovering shopping patterns through extracting associations or co-occurrences from a shop's transactional records. Market basket analysis determines the products which might be bought together and to reorganize the supermarket format and additionally to layout promotional campaigns such that products' buy can be progressed. Association rules are derived from the common object sets the use of aid and self assurance as threshold tiers. The sets of items that have minimal guide are known as frequent item set. The guide rely of an item set is defined as the percentage of transactions within the facts set which contain the object set. Self belief is defined as the measure of actuality or trustworthiness related to every found pattern. Association regulations derived depends on self assurance.

II. RELATED WORK

A number of approaches have been proposed to implement facts mining techniques to carry out market analysis. Loraine et al. in their work proposed a marketplace basket analysis using common item set mining. They as compared Apriori with okay-Apriori algorithm to find the common gadgets]. Vishal et al. carried out information mining in on line shopping gadget the use of Tanagra tool. They made decision approximately the location of product, pricing and promoting.

Sudha and Chris et al. proposed the impact of customer's perception and crm on Indian retailing in the changing business scenario using data mining techniques. Comparing to the works discussed above, our work is different by using apriori and decision tree to perform market basket analysis.

III. RELATED WORK

A. Customer data set

The Wholesale customer data provided by the UCI Machine Learning Repository is used for analysis of this work . The dataset has 8 continuous and 1 numeric input attributes namely channel, region, fresh, milk, grocery, frozen, detergents, delicatessen and session. It also has the predicted attribute i.e) the class label. Here the channel1 represents horeca (hotel/restaurant/café), channel2 represents retail shops. Region1 represents Lisbon, region2 represents Oporto, region3 represents the others.

The description of the dataset is tabulated in Table 1.

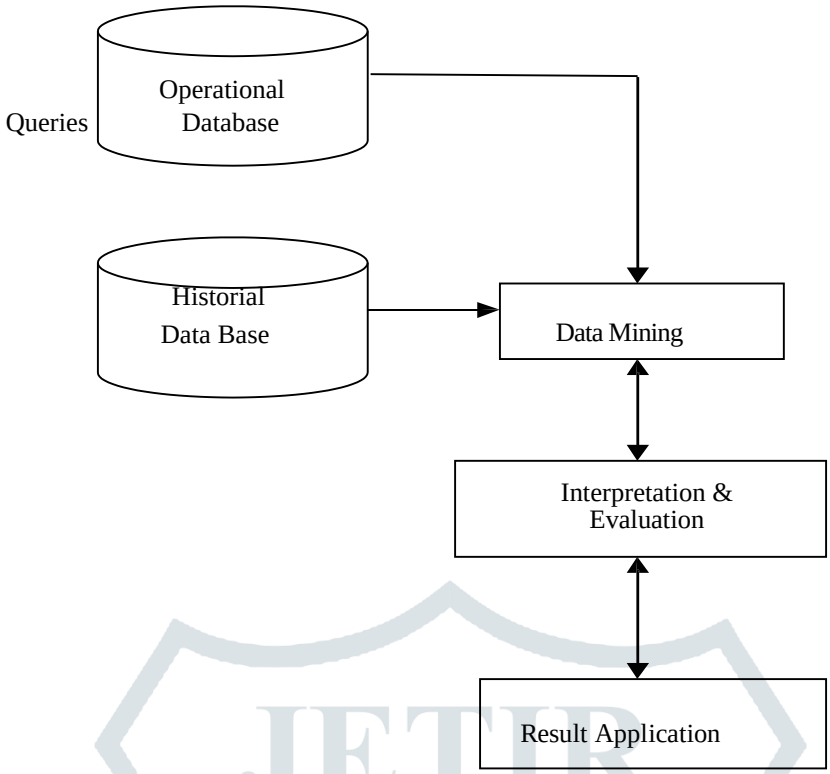

Fig.1 Block Diagram of Proposed system.

Table.1 Summary of Market Analysis Dataset

Attribute	Description
Channel	1. Horeca 2. Retail
Region	1. Lisbon 2. Oporto 3. Others
Fresh	Annual spending on fresh products
Milk	Annual spending on milk products
Grocery	Annual spending on grocery products
Frozen	Annual spending on frozen products
Detergents	Annual spending on detergents products
Delicatessen	Annual spending on delicatessen products

B. Association Rules:

Association rules are of the form if X then Y. Frequent patterns is patterns (such as item sets, subsequences, or substructures) that appear in a data set frequently . Frequent pattern mining searches for recurring relationships in a given data set. Association rules are not always useful, even if they have high support, confidence and lift > 1. Association rules can also be improved by combining purchase items. Items often fall into natural hierarchies. In this Section, repeated item set can be generated using apriori algorithm and associate outliers also be generated according to the given support count and confidence level.

IV. SIMULATION RESULTS

The whole dataset was given to the data mining tool like Tanagra. Then repeated item set is found using apriori algorithm in the association technique . This paper is mainly focused to find out whether the products can be sold more at morning session or evening session.

Table2. Statistical analysis of wholesale customer data

Attribute	Min	Max	Mean	Standard deviation
Fresh	3	112151	12000.30	12647.329
Milk	55	73498	5796.27	7380.377
Grocery	3	92780	7951.28	9503.163
Frozen	25	60869	3071.93	4854.673
Detergents	3	40827	2881.49	4767.854
Delicatessen	3	47943	1524.87	2820.106

The statistical analysis of the whole dataset is given in Table 3. It gives the mean and accuracy of the product sold in two sessions.

Table 3. Statistical analysis of session

Description of Session							
Session = Morning				**Session =Evening**			
Examples [47.5%]				Examples [52.5%] 231			
Att – Desc	Test value	Group	Overall	Att – Desc	Test value	Group	Overall
Continuous attributes : Mean (StdDev)				Continuous attributes : Mean (StdDev)			
Channel	1.54	1.36(0.48)	1.32(0.47)	Fresh	0.80	12462.13 (14302.89)	12000.30 (12647.33)
Milk	0.39	5941.77 (7921.05)	5796.27 (7380.38)	Grocery	0.54	8186.32 (9870.68)	7951.28 (9503.16)
Delicatessen	0.26	1562.14 (1931.80)	1524.87 (2820.11)	Region	0.31	2.55 (0.78)	2.54 (0.77)
Frozen	0.13	3102.73 (5626.43)	3071.93 (4854.67)	Detergents	0.06	2894.07 (4868.62)	2881.49 (4767.85)
Detergents	-0.06	2867.59 (4665.58)	2881.49 (4767.85)	Frozen	-0.13	3044.07 (4043.96)	3071.93 (4854.67)
Region	-0.31	2.53 (0.77)	2.54 (0.77)	Delicatessen	-0.26	1491.15 (3435.49)	1524.87 (2820.11)
Grocery	-0.54	7691.50 (9096.26)	7951.28 (9503.16)	Milk	-0.39	5664.62 (6869.39)	5796.27 (7380.38)

Fresh	-0.80	11489.85 (10530.36)	12000.30 (12647.33)	Channel	-1.54	1.29 (0.45)	1.32 (0.47)
Discrete attributes : [Recall] Accuracy				Discrete attributes : [Recall] Accuracy			

A receiver operating characteristic (ROC) curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings.

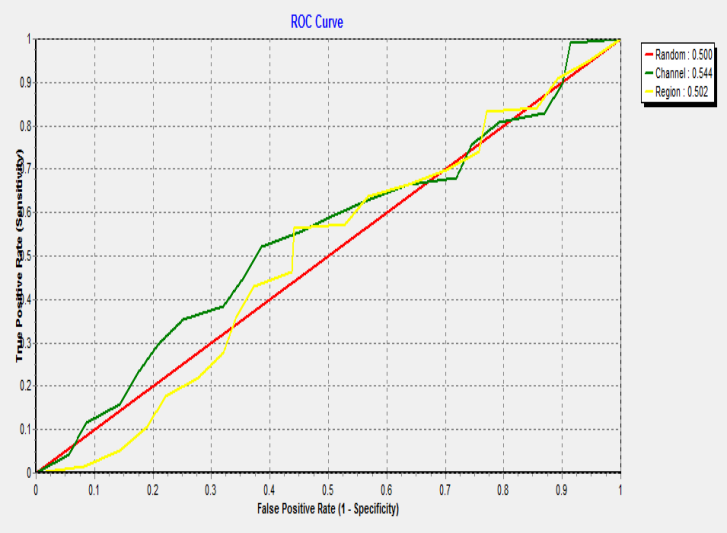

Fig.2 ROC curve

The ROC curve of our work is shown in Fig. 2. Here the positive value should be taken as morning and the result becomes nearly true positive is little bit higher than the false positive . This diagram illustrates at what channel and region our products sends more in the morning and whether it gets true positive or not.

V.CONCLUSION

In this paper, a framework for decision tree and repeated item set is evolved for the analysis of wholesale information. The wholesale customer dataset is taken and analyzed to understand the consultation at which the product may be offered more the usage of decision tree set of . The statistics within the dataset is preprocessed to make it suitable for classification. The preprocessed facts is used for classification and we received excessive class accuracy.

REFERENCES

[1] Loraine Charlet Annie M.C.1 and Ashok Kumar D, "Market Basket Analysis for a Supermarket based on Frequent Itemset Mining", IJCSI International Journal of Computer Science Issues, Vol. 9, No. 3, pp.257-264, 2012.

[2] Vishal jain, Gagandeep singh narula & Mayank singh, "Implementation of data mining in online shopping system using Tanagra tool", International journal of computer science And engineering Vol. 2, No. 1, 2013.

[3] Sudha vemaraju, "Changing waves in indian retailing: Impact of customers perception and crm on indian retailing in the changing business scenario", International Journal of Multidisciplinary Research , Vol.1, No.8, 2011.

[4] Chris Rygielski, Jyun-Cheng Wang b, David C. Yen, "Data mining techniques for customer relationship management", Technology in Society,2002.

[5] P Salman Raju, Dr V Rama Bai, G Krishna Chaitanya, "Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, No.1, 2014.

[6] Bharati M Ramager, "Data Mining techniques and Applications", International Journal of Computer Science and Engineering, Vol. 8, No.12,2009.

[7] P. Nancy, and Dr. R. Geetha Ramani, "A Comparison on Data Mining Algorithms in Classification of Social Network

Data",International Journal of Computer Applications, Vol.32, No.8, 2011.

[8] Sheikh, L Tanveer B. and Hamdani, "Interesting Measures for Mining Association Rules", IEEE Conference-INMIC , 2004.

[9] Sonali Agarwal, Neera Singh, Dr. G.N. Pandey, "Implementation of Data Mining and Data Warehouse in E-Governance", International Journal of Computer Applications, Vol.9, No.4, 2010.

[10] Chen, Y.-L., Tang, K., Shen, R.-J., Hu, Y.-H.:"Market basket analysis in a multiple store environment", Journal of Decision Support Systems, 2004.

[11] Berry, M.J.A., Linoff, G.S.: "Data Mining Techniques: for Marketing, Sales and Customer Relationship Management" (second edition), Hungry Minds Inc., 2004.

[12] C. Rygielski, J. C. Wang, and D. C. Yeh, "Data mining techniques for customer relationship management," Technology in Society, vol. 24,2002.