Application of Retail Analytics Using Association Rule Mining In Data Mining Techniques With Respect To Retail Supermarket

***A.Pappu Rajan**
****S.Suresh**

*Assistant Professor – Systems , St. Joseph's Institute of Management , St.Joseph's College (Autonomous), Tiruchirappalli ,Tamil Nadu, India.
**Assistant Professor – Systems , St. Joseph's Institute of Management , St.Joseph's College (Autonomous), Tiruchirappalli ,Tamil Nadu, India.

Abstract

This is a study on application of retail analytics regarding to the retail outlet in Thiruchirappalli, Tamil Nadu, India. The scope of this research is to identify the products or products of particular category which has association that is products which are likely to buy with other products as a bundle of item sets. This is defined by association rules with the help of support and confidence metrics, thereby finding the frequent market basket bundle item sets. This study was done using the point of sales data that is collected in a retail shop and is more of daily transaction data. Using the association rules and market basket analysis, profitable associated product item sets are found. This identify customer's pattern of buying behavior and it is then used to devise cross selling strategies to improve the sales of the products. Cluster analysis is used to define the relationship between products and identify the factors that influence the buying behavior of the customers. And the results derived is used to draw a visual merchandising technique that will satisfy the customers in terms of service and induce the impulse purchase of products, increasing the profit of the retail outlet and retaining its customers.

Keywords: Data mining, Retail Analytics, Market Basket Analysis, Association Rule Mining.

Introduction

Today information systems are need of the hour in any retail outlets to scale up their performance in terms of operations and customer satisfaction. The various processes which help the customers to procure the desired merchandise from the retail stores for their end use refer to retail management. Retail management includes all the steps required to bring the customers into the store and fulfill their buying needs.

Retail management makes shopping a pleasurable experience and ensures the customers leave the store with a smile. In simpler words, retail management helps customers shop without any difficulty. And the concept of segregating similar products into separate groups is called as category management. The complete range of merchandise available at the retail store is divided into separate product categories consisting of related products. Categories in a retail store refer to the various groups which consist of products belonging to a similar family. The retailer smartly displays all the related products together as distinct categories for his as well as the end-user s convenience. Retail Industry is one of the fast growing industries, meeting the requirements of people at a large scale in terms of products. Today information systems are need of the hour in any retail outlets without which

they fail to perform and satisfy the customers. Business Intelligence is a component of the information systems tracking the behavior and demand of the customer in a retail shop so that consumers are more satisfied by give what they want in an easier manner, increasing the profitability of a retail outlet.

Data mining a concept of business intelligence has the power to harness the hidden knowledge present in the huge data that is got at point of sales. This data gives valuable information, required to understand the customers buying patterns, key performance indicators which helps retailers in making decision such as catalog design, cross-marketing and customer shopping behavior analysis.

Review of Literature

IBM Software Business Analytics (2010) , Paper Titled The Future of Market Research stated that segmentation can greatly improve your clients cross-sell and up-sell activities by making every interaction a golden opportunity. Market research derives customer segments through surveys and demographic research, enabling customer-focused organizations to understand their customer segments better. In this paper they used data mining clustering techniques to find naturally occurring groups within the customer database. While each approach individually provided insight into basic customer groups, combining these approaches yields still deeper insights.

Mourad Ykhlef (2011), in his study paper named Association mining of dependency between time series using Genetic Algorithm and discretisation states that association rule mining is one of the most popular data-mining techniques used to find associations existing between a set of objects or data. A time series is a sequence of observations stamped over the time; Time-series analysis has been used in a variety of applications like: business and health.

Hongyan Liu et al. (2011), in their study titled Methods for mining frequent items in data streams: an overview analyzed that in many real-world applications, information such as web click data, stock ticker data, sensor network data, phone call records, and traffic monitoring data appear in the form of data streams. Estimating the frequency of the items on these streams is an important aggregation and summary technique for both stream mining and data management systems with a broad range of applications.

Marghny H. Mohamed et al. (2011), in their study titled Advanced Matrix Algorithm (AMA): Reducing Number of Scans for Association Rule Generation analyzed that **e**xisting Association Rules Mining (ARM) algorithms basically use multiple scans to extract a rule from a transaction database. Sometime ARM algorithms exit without a rule in the desktop environment due to the high volume of transactions. Matrix Algorithm (MA) is proposed to minimize this issue. However, it is a computational expensive solution.

Association Rules

To generate all association rules from the given dataset, first find all frequent item sets and then generate strong association rules from the frequent item sets.

Support: It is the measure of how often the collections of items in an association occur together as percentage of all transactions. Support = (containing the item combination) / (total number of record.).

Confidence: It is the measure of uncertainty or trust worthiness associated with

each discovered pattern. Confidence of a rule = the support for the combination / the support for the condition.

Frequent Item Set: If an item set satisfies minimum support, then it is a frequent item set.

Strong Association Rules: Rules that satisfy both a minimum support threshold and a minimum confidence threshold.

Overall Objective

To identify the shopping patterns of the customers visiting the retail outlet and devise cross marketing and merchandising strategies to increase the profitably of the retail outlet.

Specific Objectives

To understand the Shopping behavior of the customer in retail outlet.

To identify the sales trend of the products

To identify the shopping patterns of consumers.

To develop methods of cross selling and merchandising for the retail outlet.

To provide suggestions and feed back to improve the retail s performance in terms of visual merchandising.

Sample Design

Sample design is a defined plan determined before any data are actually collected for obtaining a sample from a given population. Samples can be classified into probability samples and non-probability samples. In this study has used systematic sampling in the group of probability sampling.

Data

Point of sales data is the data collected where money transaction and billing takes place in a retail shop which is a secondary data. In this study point of data was used which has collected form super market The data structure is consists of transaction no, name of the product, no of units bought, cost and size of the product which gives information about the product.

Size of Sample

600 transactions in the Month of April 2014 to May 2014.

Tools Used For Prediction

In this study analysis part was used the SQL Server 2008 and Data mining ADD-INS in MS Office 2007. Using the Table Analysis Tools for Excel, first create a connection to an instance of SQL Server 2008 Analysis Services (SSAS). This connection gives to access the Microsoft data mining algorithms that are used to analyze.

Analysis and interpretation

Market basket bundle item

Bundle of items: It is the products which occur together.

Bundle size: It is the number of items in a bundle.

Number of sales: It is the number of times the bundle being bought.

Average value per sales: It is the average cost of items together.

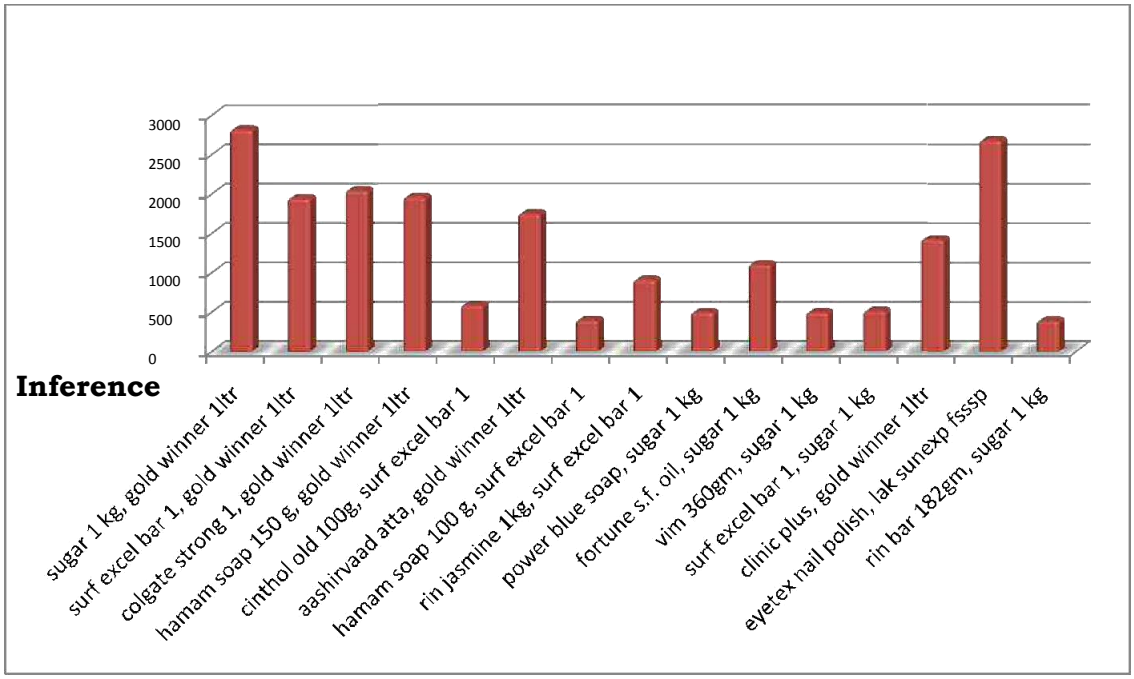
Table No. 1.0

Bundle of Items

Bundle of items	Bundle Size	Number of Sales	Average Value Per sale	Overall value of Bundle
Sugar 1 kg, gold winner 1ltr	2	25	113	2825
Surf excel bar 1, gold winner 1ltr	2	21	93	1953
Colgate strong 1, gold winner 1ltr	2	18	114	2052
Hamam soap 150 g, gold winner 1ltr	2	17	115	1963
Cinthol old 100g, surf excel bar 1	2	16	37	592
Aashirvaad atta, gold winner 1ltr	2	15	117	1755
Hamam soap 100 g, surf excel bar 1	2	13	31	403
Rin jasmine 1kg, surf excel bar 1	2	12	76	912
Power blue soap, sugar 1 kg	2	11	46	502
Fortune s.f. oil, sugar 1 kg	2	10	111	1110
Vim 360gm, sugar 1 kg	2	10	50	500
Surf excel bar 1, sugar 1 kg	2	13	40	520
Clinic plus, gold winner 1ltr	2	11	130	1430
Eyetex nail polish, lak sunexp fsssp	2	12	224	2688
Rin bar 182gm, sugar 1 kg	2	10	40	400
Vim 360gm, gold winner 1ltr	2	13	116	1505

Figure 1.0

Bundle of items



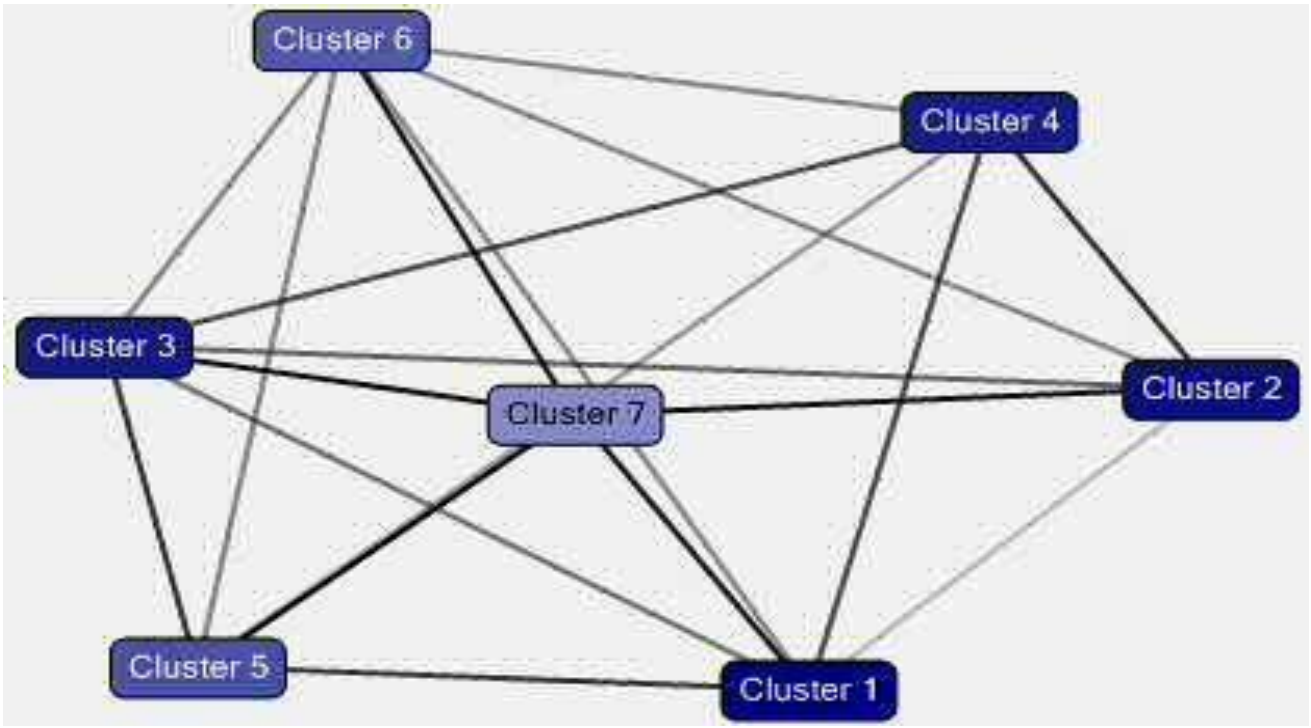
Inference

The bundle Gold Winner and Sugar 1Kg is more profitable than other bundles. And the item set Eyetex nail polish and Lakme Sunexp has the second highest value in terms of cost. The other value of bundles are less in value because they occur less number of time in the transaction data because there is no promotion for the products and so failed in cross selling those products.

Cluster Diagram

Figure 2.0

Association between Clusters



There is a strong association between cluster containing different brands and categories. Hence these relations prove that there is interdependence between the Brand and the Category of the products which ensures that associations are present between Brands of one cluster with the other. This helps the retailer to improve the visual merchandising design of the retail outlet.

Major Findings

Food, Fabric wash, Personal Wash and Snacks are the key performing categories.

Gold winner 1Ltr is the most frequently occurring product in every basket.

There is a strong association between the Fortune Sunflower oil, Sugar 1Kg and gold winner 1Ltr.

Eyetex nail polish and Lakme sunscreen is one of the highly associated products under cosmetics

The highest selling product under the category of fabric wash is Rin which is determined by the purchase of personal wash products such as Hamam, Pears etc. through impulse purchase.

The purchasing behavior identified is that when there is a product from food category there is at least one product from oral care, fabric wash, personal wash and snacks.

Clusters formed with brand and category had heterogeneous which proves that there is a strong influence of brand to choose a product from the category. It also had few homogeneous clusters which denote that they are independent.

Brand of the product influences the sales of a product in a particular category.

Britannia is been determined by many other brands which shows that there is an impulse purchase of Britannia biscuits.

Suggestions

The highly associated products that are found should be kept in such a way that the customers will feel easy to reach the products.

From the findings Fortune oil and sugar are highly associated products but the value is less than other bundles. This can be improved by keeping a branded sugar such as parry sugar near the fortune oil this will improve the value of sale and adds profits to the retail.

These associated frequent items should be located in such a way that the customer should reach it after passing through products which can be bought on impulse. For example customer should pass through the cosmetics section before reaching the sugar or oil.

Fabric Wash, Snacks and Personal wash should have a separate shelf display having more brands and more number of stock keeping units.

Beverages, Deodorants, House Hold care, oral Care and cosmetics should maintain more brands with less keeping units to increase the impulse purchase.

Product category which has no association such as snacks and fabric wash products should never be kept aside as this will de-motivate the shopper to buy any of those products.

Products are of poor in performance can be used to make bundles with high performers under the condition that both have a moderate association, which can be used to improve the sales.

Retail bulletin boards should be kept at the entrance having the bundled product display information to promote the products which will act like a road map for the customers to purchase the products.

Conclusions

Thus the retail analytics is used to explore the point of sales data, identifying the customer pattern of buying the products, using the association rule mining with the help of market basket analysis in the retail outlet. The association rules defined is used to find the frequent item sets that occur in the transaction data and is prioritized in terms of value. This is then used to devise the cross selling strategies such as bundling the high value product with low performing product on an offer based on the suggestions of association rules that are defined. Dependency network and clustering is used to define relationship between the product category and brand, where the brand is the key influencer to the buying behavior of the shopper. Based on the findings and suggestion derived the visual merchandising setup or arrangement of various product category is drawn having in mind that the customer should reach the highly associated products easily and induce the impulse purchase of other products. Hence the retail analytics is used to mine the

potential information from the point of sales data which will be boosting the profitability of the retail outlet in term of customer satisfaction and sales revenue.

References

1. Derick Jose.(2011) *Retail Analytics Demystified.* Available from URL http://www.retailanalyticsdemystified.blogspot.com/2009/08/which-behavorial-variables-to-segment.html (Accessed on 06th December 2013).

2. DMS Retail.(2011) *KPI For Retail* . Available from URL: http://www.dmsretail.com/interpretyou_kpi.html (Accessed on 07th November 2013).

3. Gupta C K .(2006) *Introduction to Data Mining with Case Studies*,2nd Ed. PHI/Pearson Education.

4. Hongyan Liu, Yuan Lin, Jiawei Han. (2011) *Methods for mining frequent items in data streams: an overview,* Knowledge and Information Systems (KAIS), vol. 26, no. 1, pp. 1-30.

5. IBM Software Business Analytics. (2010) *The Future of Market Research* Available from URL: http://public.dhe.ibm.com/common/ecm/en/ytw03099usen.pdf (Cited on 1st April 2014).

6. IBM Software Business Analytics. (2011) *Delivering a Smarter Shopping Experience with Predictive Analytics.* Available from URL: http://public.dhe.ibm.com/common/ecm/en/ytw03162usen.pdf (Cited on 2nd June 2014)

7. Jaideep Srivasatava et al. (2000) *Web Usage Mining: Discovery and applications of usage, patterns from web data,* SIGKDO, Explorations, Vol.1.Issue 2.

8. Jiawei Han and Micheline Kamber. (2009) *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann Publishers.

9. Michael J.A. Berry and Gordon S. Linoff. (2007) *Data Mining Techniques*, Second Edition, Wiley Publishing, Inc.,

10. Maia-Intelligence. (2013) *KPI for Retail*. Available from URL: http://blog.maia-intelligence.com/2009/09/30/kpi-for-retail (Cited on June 2013).

11. Marghny H. Mohamed, Mohammed M. Darwieesh, A.B.M. Shawkat Ali. (2011) *Advanced Matrix Algorithm (AMA): reducing number of scans for association rule generation*, International Journal of Business Intelligence and Data Mining (IJBIDM), Vol. 6, No. 2.

12. Mourad Ykhlef . (2011) *Association mining of dependency between time series using Genetic Algorithm and discretisation* , International Journal of Business Intelligence and Data Mining 2011 - Vol. 6, No.1 pp. 55 70.

13.Pappu Rajan A , Victor S.P .(2012) *Features and Challenges of web mining systems in emerging technology* , International Journal of Current Research ,Vol.4, Issue ,07 ,pp.066-070.

14.Pappu Rajan,G.Prakash Raj,Rosario Vasantha Kumar .P.J. (2013) *A study on application o f data and web mining techniques to enrich user experience in libraries and on line book stores*, International Journal of Research in Computer Application & Management ,ISSN 2231-1009 , Volume No .3, Issue No.8 .

15. Roberto Battiti and Mauro Brunato. *Business Intelligence Tools*. Available from URL http://wiki.ask.com/Business_intelligence_tools (Cited on December 2013).

16.Suneetha, K. R. ,. Krishnamoorthi D R. (2009) *Identifying User Behavior by Analyzing Web Server Access Log File*, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.4.

17.The Fact Point Group. (2008) *Leading Practices in Market Basket Analysis.* Available from URL: http://www.plattretailinstitute.org/documents/jra/detail.phx (Accessed on 1st June, 2014).

18. Thomas H. Davenport, Jeanne G. Harris. (2007) *Competing on Analytics: The New Science of Winning*, First Edition, and Harvard Business School Press.