



ISSN: 0975-833X

**RESEARCH ARTICLE**

**UTILIZING DATA MINING APPLICATION TECHNIQUES TO INVOLVE A BUSINESS ANALYTICS FOR RETAIL OUTLETS**

**<sup>1</sup>Pappu Rajan, A. and <sup>2</sup>Victor, S. P.**

<sup>1</sup>Department of Computer Science and Research Center, St.Xavier's College(Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India

<sup>2</sup>HOD, Department of Computer Science & Director of Computer Science and Research Center, St.Xavier's College(Autonomous), Palayamkottai, Tirunelveli, Tamil Nadu, India

**ARTICLE INFO**

**Article History:**

Received 29<sup>th</sup> August, 2012  
Received in revised form  
25<sup>th</sup>September, 2012  
Accepted 13<sup>th</sup> October, 2012  
Published online 23<sup>th</sup> November, 2012

**Key words:**

Data Mining , Business analytics , Market basket analysis, Cluster Analysis, Association rule.

**ABSTRACT**

This is a study on application of Retail Analytics regarding to the retail outlet in Tiruchirappallai, Tamil Nadu , South India using data mining technique. The scope of this paper is using mining technique 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 is 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 is also identifies 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.

*Copy Right, IJCR, 2012, Academic Journals. All rights reserved.*

**INTRODUCTION**

Business intelligence mainly refers to computer-based techniques used in identifying, extracting and analyzing business data, such as sales revenue by products and/or departments, or by associated costs and incomes. These technologies provide historical, current and predictive views of business operations. Common functions of business intelligence technologies are reporting, online analytical processing, analytics, data, process mining, complex event processing, business performance management, benchmarking, text mining and predictive analytics. Data mining is a concept of business intelligence has the power to harness the hidden knowledge present in the huge data which is got at point of sales. Generally dictionary is data about data ,here 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[16]. 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. The complete range of merchandise available at the retail store is divided into separate product categories consisting of related products. Category management is a retailing and supply management concept in which the range of products purchased by a business organization or sold by a retailer is broken down into discrete groups of similar or related products, these groups are known as product categories. It is a systematic, disciplined approach to managing a product category as a strategic business unit. The phrase Category Management was coined by Brian F. Harris [1]. 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 their as well as the end-user convenience.

**ANALYTICS IN RETAIL INDUSTRY**

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

\*Corresponding author: [ap\\_rajjan2001@yahoo.com](mailto:ap_rajjan2001@yahoo.com)

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 [3].

### NEED FOR THE STUDY

This study is important in the sense as we can measure the position of the product or retail shop after analyzing the data that we have processed data with the help data mining technique using business analytics. This study helps in devising cross selling and visual merchandising methodologies. Helps in identifying Key performance Indicators. Reduce cost and increase the profitability of the retail outlet. Maintaining efficient shelf space.

### OVERALL OBJECTIVE

To identify the shopping patterns of the customers visiting the retail outlet and devise cross marketing and merchandising strategies to increase the profitability of the retail outlet. In the data mining, web Usage Mining focuses on several techniques that could learning or predict user behavior and navigation pattern because user using the web round the clock. To specify the functions for clustering, prediction, and other techniques in the data mining. To identify the Market Basket analysis, Association rule related functions where Mining can be applied effectively.

### SPECIFIC OBJECTIVES

- ✚ To understand the behavior of shoppers.
- ✚ To identify the product's sales trend.
- ✚ To identify the shopping patterns of consumers.
- ✚ To identify the average spent dispersion.
- ✚ 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.
- ✚ To give suggestions and recommendation about the problem

### RESEARCH METHODOLOGY

Research can be defined as the scientific search for knowledge, or as any systematic investigation, to establish novel facts, solve new or existing problems, prove new ideas, or develop new theories, usually using a scientific method. The primary purpose for basic research is discovering, interpreting, and the development of methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe [7].

### TYPE OF RESEARCH AND APPROACHES USED ANALYTICAL RESEARCH

In analytical research, the researcher has to use facts or information already available, and analyze these to make a critical evaluation of the material. As all the data found in the survey is totally numerically so the type of analysis was statistical.

### SIMULATION APPROACH

Simulation approach involves the construction of artificial environment within which relevant information and data can be generated. This permits an observation of dynamic behavior system under controlled conditions. The term simulation in the context of business and social science applications refers to the operation of a numerical model that represents the structure of a dynamic process. Given the values of initial conditions, parameters and exogenous variables, a simulation is run to represent the behavior of the process over time. Simulation approach can also be useful in building models for understanding future conditions [7].

### RESEARCH PROCESS

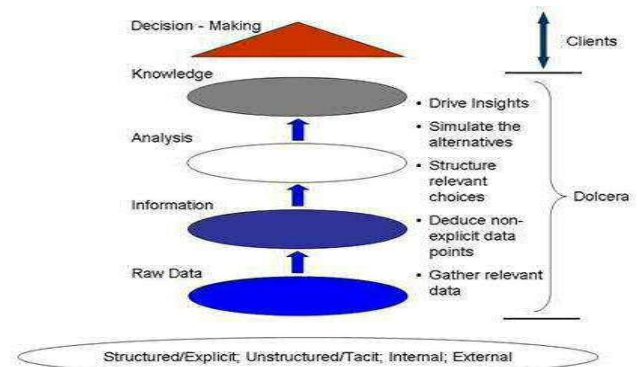


Fig. 1.0

Step 1: Gather relevant data. Using primary and secondary research techniques, we assemble both structured and unstructured data and information. We have access to a wide array of leading secondary data sources and also to a pool of well-regarded industry, academic, and functional experts. For Web-based research, our proprietary software tools and protocols efficiently and effectively collect the most relevant data. For primary research, we utilize comprehensive survey and data collection methodologies. Finally, we apply multiple filters to arrive at data that is relevant, consistent, and reliable.

Step 2: Deduce non-explicit data points. Data and information are not always available via primary and secondary research. In cases where there are gaps in the data, we employ assumptive methodologies such as devising intelligent proxies and statistical techniques.

Step 3: Structure relevant choices. Once all data has been gathered and filtered, we use advanced analytical techniques to structure and analyze data for patterns, trends, and outliers. Using our proprietary multidimensional databases, we can intelligently observe structured data sets in comparative and cross-relational formats.

Step 4: Simulate the alternatives. Leveraging our significant modeling expertise, we input structured data into customized predictive models. The models are constructed to serve as point and click web tools for our clients to simulate scenario-based business case outcomes. Simulated outcomes are driven by leading indicators such as changes in industry/sector dynamics, shifts in innovation and business model cycles, changes in operating paradigms, movements of macroeconomic drivers, etc.

Step 5: Drive insights. Our aim is to provide complete and error-free knowledge and insight that support rapid, effective decision making. This insight helps to stay ahead of the curve in a knowledge-intensive business environment

#### NATURE OF RESEARCH

Quantitative type of research has been used. As all the data was based on numerical figures obtained in the survey.

#### SOURCE OF DATA

The type of data used is Secondary data. The data which is already collected by someone else and which have been through the statistical process is known as the secondary data. Common sources of secondary data for Social science include censuses, surveys, organizational records and data collected through qualitative methodologies or qualitative research. Point of sales data (POS) is the data collected where money transaction and billing takes place in a retail shop which is a secondary data. POS Data Structure: Transaction Number, Name of the product, No. of units bought, cost and size of the product.

#### SIZE OF SAMPLE AND POPULATION

The sample size of the population was 150 respondents and 600 Transactions. Survey of people who either purchase or plan to purchase material from stores near in their home

#### TOOLS - THE SQL SERVER 2008 DATA MINING ADD-INS FOR OFFICE 2007

The SQL Server 2008 Data Mining Add-ins for Office 2007 provides wizards and tools that make it easier to extract meaningful information from data. They help to derive patterns and trends hidden in complex data, visualize those patterns in charts and interactive viewers, and then generate rich, colorful summaries for visual presentation and for business analytics. It can analyze correlations and run predictions against data stored in Microsoft Office Excel tables, or can create and modify data mining models that are stored on an instance of in SQL Server 2008 Analysis Services (SSAS), and graph the results in Microsoft Office Visio [8].

#### VISUALIZATION AND PRESENTATION

After analyzing the data, charts and underlying analytical data can be copied to Excel, or can build diagrams interactively by using the custom data mining shapes and data mining templates and shapes in Visio

#### LIMITATIONS OF STUDY

Collecting the point of sales data is highly difficult as it is a confidential data of any retail shops. Huge volume of data is required to find the association patterns. This study is limited to finding the frequent pattern of products that are bought together and does not relate with the type of customer or customer segmentation buying the products. For the time concerns this study focuses on only few retail outlet in Tiruchirappallai, Tamil Nadu, South India. Retail managers are reluctant to the suggestions provided. Research after implementing the suggested strategies is required to find the effectiveness of the decision made.

#### BUSINESS INTELLIGENCE

Business intelligence is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. Business analytics is part of BI which refers to the skills, technologies, applications and practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Business analytics focuses on developing new insights and understanding of business performance based on data and statistical methods.[11]

#### DATA MINING

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining [8]. This, data mining should have been more appropriately named knowledge mining from data, which is unfortunately somewhat long. Knowledge mining a shorter term, may not reflect the emphasis on mining from large amounts of data. Nevertheless, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material. Thus, such a misnomer that carries both data and mining became a popular choice. Many other terms carry a similar or lightly different meaning of data mining, such as knowledge mining from data, knowledge extraction, data pattern analysis, data archaeology, and data dredging. Many people treat data mining as a synonym for another popularly used term, knowledge discovery from data, or KDD. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery.

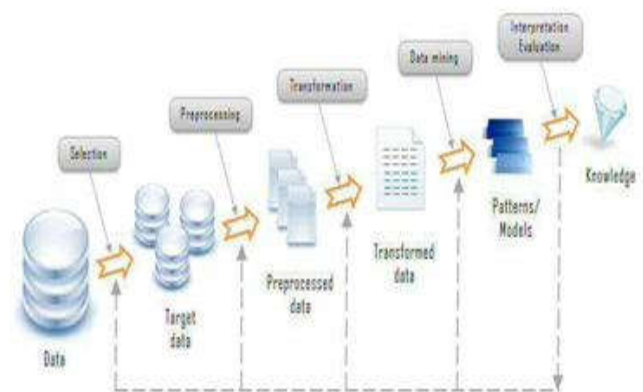


Fig.2.0

Steps 1 to 4 are different forms of data preprocessing, where the data are prepared for mining. The data mining step may interact with the user or a knowledge base. The interesting patterns are presented to the user and may be stored as new knowledge in the knowledge base. Note the according to this view, data mining is only one step in the entire process, albeit an essential one because it uncovers hidden patterns for evaluation.[7]

We agree that the data mining is a step in the knowledge discovery process. However, in industry, in media, and in

the database research milieu, the term data mining is becoming more popular than the longer term of knowledge discovery from data. Therefore, in this book, we choose to use term data mining. We adopt a broad view of data mining functionality; data mining is process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories

### KNOWLEDGE DISCOVERY PROCESS IN DATAMINIG

1. Data Cleaning is to remove noise and inconsistent data.
2. Data Integration where multiple data sources may be combined.
3. Data Selection where data relevant to the analysis task are retrieved from the database.
4. Data Transformation where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
5. Data Mining an essential process where intelligent methods are applied in order to extract data patterns.
6. Pattern Evaluation to identify the truly interesting patterns representing knowledge based on some interestingness measures.
7. Knowledge Presentation where visualization and knowledge representation techniques are used to present the mined knowledge to the user

### RETAIL ANALYTICS

Retail Analytics in-line analytics allows retailer organizations to get the most out of their existing management information systems and information sources without the risk of committing to significant investments which enable today's global retailers to make smarter decisions and manage their businesses more effectively [4].

### TECHNIQUES IN DATA MINING ASSOCIATION

Frequent patterns, as the name suggests, are patterns that occur frequently in data. There are many kinds of frequent patterns, including item sets, subsequences, and sub structures. A frequent item set typically refers to as set of items that frequently appear together in a transactional data set. A frequently occurring subsequence, such as the pattern that customers tend to purchase first a PC, followed by a digital camera, and then a memory card, is a frequent sequential pattern. A substructure can refer to different structural forms, such as graphs, trees or lattices, which may be combined with item sets or subsequence. If a substructure occurs frequently, it is called a frequent structured pattern. Mining frequent patterns leads to the discovery of interesting associations and correlations within data also called as market basket analysis.[10]

### CLASSIFICATION

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we make the software that can learn how to classify the data

items into groups. For example, user can apply classification in application that given all past records of employees who left the company, predict which current employees are probably to leave in the future. In this case, we divide the employee s records into two groups that are leave and stay . And then user can ask our data mining software to classify the employees into each group.

### CLUSTERING

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. To make the concept clearer, user can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. By using clustering technique, user can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in a topic, he or she would only go to that shelf instead of looking the whole in the whole library [12].

### PREDICTION

The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, prediction analysis technique can be used in sale to predict profit for the future if user consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, user can draw a fitted regression curve that is used for profit prediction.

### SEQUENTIAL PATTERNS

Sequential patterns analysis in one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data.

### DECISIONS THAT CAN BE MADE USING RETAIL ANALYTICS

How wide should the ranges be? What ratios should we buy each color in? What ratios should we buy each size in? How should we allocate initial stock injections to stores to maximize sales? How accurately can we forecast store/stock keeping unit level sales to drive replenishment? When should we markdown slow-moving items and how deep should the markdown be? Business intelligence systems make a major contribution to improved decision making. For example, many retailers analyze historic sales by size and use last season s size ratios as a guide to this year s. But most stock outs occur in the best-selling sizes and this approach will perpetuate under buying those sizes for the next season. Analyzing past sales for the period of time that stock coverage was good will give a more accurate planning assumption [3,4].

In a similar way, analyzing the sales rate of an stock keeping unit may show that it is not selling at the planned rate and that action needs to be taken. Drilling down by store often shows that some stores are doing well with specific sizes or colors, some are doing good, and some doing badly. A retailer might then decide to do nothing in the stores where the product is selling well and apply a 30 per cent markdown in the stores where it is doing less well. In the stores where it is moving far too slowly for the markdown to make any real difference, retailers might transfer the stock to stores where it sells well. In this way, they can greatly reduce markdown as a per cent to sales and improve overall profitability from the merchandise.

### **IMPROVING PROCESSES ACROSS THE STORE**

With targeted business intelligence, retailers can perform detailed analysis in a range of areas including operations, customer management, events, trends, supply chain, and supplier relationships.

### **OPERATIONS ANALYSIS**

By frequently identifying purchased combinations of products, retailers can help improve store layout and assortment planning. Basket analysis can drive promotional campaigns and help measure the effectiveness of marketing initiatives. In addition, better analysis of in-store operations can improve staff planning to maximize efficiency and cut staffing costs.

### **CUSTOMER MANAGEMENT ANALYSIS**

Business intelligence can help retailers improve customer targeting, revenue management, customer knowledge management, and multichannel integration. By analyzing transactional data during the day, retailers can identify differences in customer profiles, while real-time tools help stores adapt prices to match demand patterns and improve profitability. Critically, by understanding how personal priorities affect shopping habits, retailers can also anticipate customer needs to provide more appropriate services and products.

### **EVENT ANALYSIS**

Customer theft and internal fraud need to be identified quickly. Retail analytics can help control shrinkage by pinpointing exceptional events quickly and help retailers sort any problems out rapidly. Business intelligence can also help with promotions management by determining whether or not to rush in extra stocks or mark down prices to cope with changes in consumer demand.

### **TREND ANALYSIS**

With accurate analysis of stock-outs, retailers can tailor demand forecasting to improve their in-stock levels. At the same time, by assessing the impact on customer demand patterns of removing slow-moving product lines, retailers can gain insights to optimize shopper satisfaction and stock turn.

### **SUPPLY CHAIN ANALYSIS**

Real-time traceability is essential for retailers who want to

respond quickly to changing demand. Business intelligence can help retailers attain a clearer view of their supply chain and even re-route global consignments to accelerate deliveries as necessary.

### **CATEGORY AND EVENT ANALYSIS**

Using business intelligence tools in event monitoring, category management, and new product introductions is vital. Retailers need to be able to share information between the point of sale and suppliers so that forecasts can be shared and performance accurately mapped to maximize sales opportunities. In addition, new product performance and the impact of promotions need to be quickly analyzed so that stores can maintain a competitive edge.

### **SUPPLIER RELATIONSHIP ANALYSIS**

Successful replenishment often depends on the efficiency of third-party logistics providers. Sharing accurate performance data with suppliers can help avert queries over service level agreements and provide a common basis for mutual improvement.

### **COLLABORATIVE ANALYSIS**

Real-time analysis of customer demand improves order management and can help streamline production and cut lead times. With the seamless exchange of information between the store and the supplier, everyone in the supply chain has access to the same product data, at the same time, to reduce errors and duplicated effort. A common frame-work for discussion can be established. To keep pace with their competitors, retailers must ensure that their technology infrastructure can support initiatives to understand customers and consequently increase sales.

### **THE BIGGEST AREAS TO BENEFIT FROM RIGOROUS RETAIL ANALYSIS ARE:**

Merchandising decisions .Store operations. Identification and adoption of marketing strategies and campaigns Supply chain management

### **MARKET BASKET ANALYSIS**

A typical example of frequent item set mining is market basket analysis. This process analyzes customer buying habits by finding associations between the different items that customers place in their shopping baskets . The discovery of such associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased by customers. For instance if the customer are buying milk, how likely are they to also buy bread on the same trip to the supermarket? Such information can lead to increased sales by helping retailers do selective marketing and plan their self space [2].

### **LEADING RETAILERS ARE LEVERAGING MARKET BASKET ANALYSIS TO:**

Develop more profitable advertising and promotions. Target offers more precisely Improve loyalty card promotions with longitudinal analysis. Attract more traffic into stores Increase

the size and value of the basket of purchase. Test and learn by using the market place as laboratory. Empower planners and merchants to make smarter decisions. Determine the magic price point for individual stores. Match inventory to needs by customizing layouts, assortments and pricing to local demographic.

## ASSOCIATION RULES

To generate all association rules from the given dataset ,support greater than a specified minimum and confidence greater than a specified minimum

*Support* : It is the measure of how often the collection of items in an association occur together as percentage of all transactions  
 $\text{Support} = (\text{Containing the item combination}) / (\text{Total number of record})$   
 Support: 2% of Transaction has bread and jam.

*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)

*Confidence*: 60% of customers who have bought bread also buy jam.

*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 In Association rule mining, we first find all frequent item set and then generate strong association rules from the frequent item set  
 Rule Form\_LHS -> RHS [confidence, support] Bread -> Jam [60%, 2%]

## THE APRIORI ALGORITHM

Apriori algorithm is the most established algorithm for finding frequent item sets mining. Progressively identifies large item sets of different sizes. The basic principle of Apriori is any subset of a frequent itemset must be frequent . We use these frequent item set to generate association rules [12].

## REVIEW OF THE STUDIES

IBM Software Business Analytics White Paper Titled The Future of Market Research states 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. Data mining uses clustering techniques to find naturally occurring groups within the customer database. While each approach individually provides insight into basic customer groups, combining these approaches yields still deeper insights. [14]

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. The application of association mining to time series is very promising. The purpose of that article is to propose a new fast algorithm to discover the association that can exist between two time series. In that discretisation is to segment time series to

a number of shapes, and we classify these shapes to pre-defined shape classes to generate association rules using Genetic Algorithm [15]. IBM Software Business Analytics (2011) white Paper titled Delivering a Smarter Shopping Experience with Predictive Analytics. States that applying market basket analysis, algorithms can reveal associations between products that are typically bought together. Key insights from Predictive analytics might be: 1.Product A and B tend to be bought together. 2. If a customer buys products C and D then it is extremely likely they will buy product E. 3.Multiple purchases with category W are common, and customers who make such purchases also tend to buy Y. Armed with this insight, retailers can make calculated decisions about how to match offers with customers. Retailers can predict who is most likely to participate and they can forecast the ROI associated with customer response. By applying Market Basket Analysis retailers achieve higher sales, greater returns on marketing spend and a more targeted approach to future offerings. At the same time, customers feel that their needs are understood and met. Higher customer satisfaction and increased spend [14]. 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. There are different kinds of models for frequent items mining task. For general models such as cash register and Turnstile, we classify existing algorithms into sampling-based, counting-based, and hashing-based categories. Accordingly, as an extension of the general data stream model, four more specific models including time-sensitive model, distributed model, hierarchical and multi-dimensional model, and skewed data model are introduced [13].

Reducing Number of Scans for Association Rule Generation analyzed is that existing Association Rules Mining algorithms basically use multiple scans to extract a rule from a transaction database. Sometime these algorithms exit without a rule in the desktop environment due to the high volume of transactions. Matrix Algorithm is proposed to minimize this issue. However, it is a computational expensive solution. Advanced Matrix Algorithm is used to generate an efficient rule by a single scan using the Boolean matrix concept. This is comparatively effective and efficient than traditional approaches in terms of computational cost for database scan and frequently candidate sets generation [9].

Neuro-marketing is a relatively new field that studies consumers sensory-motor, cognitive and affective response to marketing stimuli. In neuro-marketing, researchers use technology to learn why consumers make the decisions they do and what part of their brain is telling them to do it. Market researchers can use neuro-marketing to better measure consumers preferences, since the verbal response to questions such as Do you like this product? may not always be the true answer due to cognitive bias. By fitting respondents with devices that measure their physical response to certain stimuli such as ads, commercials, color and packaging researchers can learn what the consumer reacts to [10].

## DATA ANALYSIS AND INTERPRETATION

Point of sale or Point of purchase or checkout is the location where a transaction occurs. A "checkout" refers to a POS terminal or more generally to the hardware and software used for checkouts, the equivalent of an electronic cash register. Point of sales data is the data collected at point of purchase. The data collected is having 12 basic categories containing 27 companies, 50 Brands and 350 products of different stock keeping units.

**INFERENCE OF OVERALL SALES PROFILE**

The high performing products are from the category of Food, Fabric Wash and Personal Wash. And the moderate performers are from the category of snacks, beverages, oral care and skin care and low performers are sanitary napkins, hair care, house hold care, cosmetics and deodorants.

**ASSOCIATION RULES**

Association rules are formed with minimum support and minimum confidence. Minimum Support = 10%  
Minimum Confidence / Probability = 50%

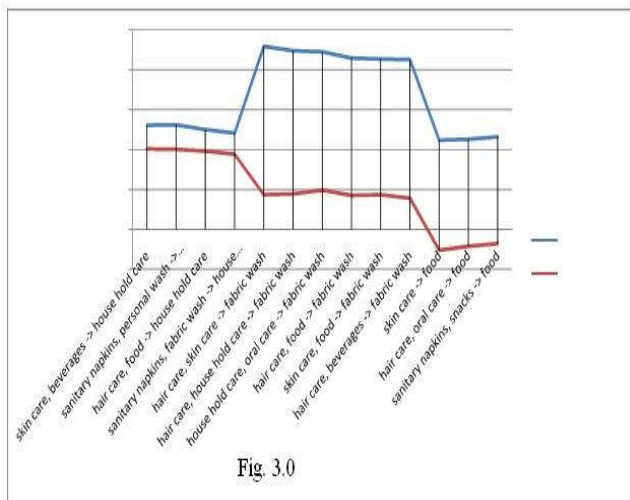


Fig. 3.0

**ASSOCIATION RULES FOR CATEGORY**

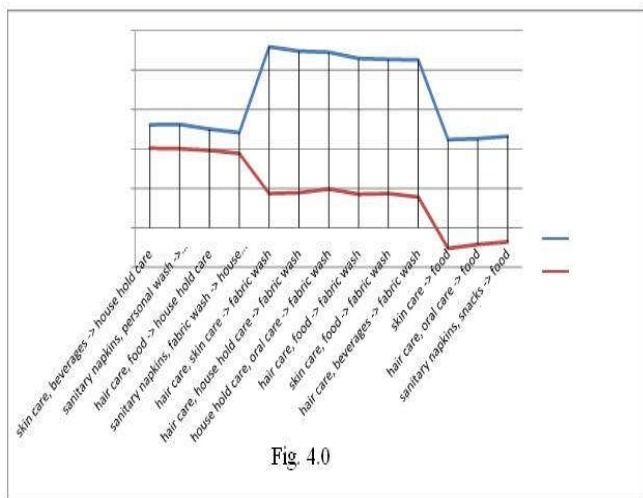


Fig. 4.0

**ASSOCIATION RULES FOR BRAND**

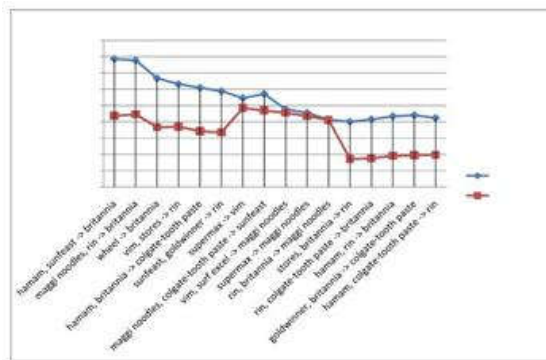


Fig. 5.0

**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 beingbought. Average value per sales: It is the average cost of items together. The bundle Gold Winner and Sugar 1Kg is more profitable than other bundle item set Eytex nail polish and Lakme Sun exp 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

**CLUSTERS**

Cluster technique is applied for the Category and Brand combined from the data and there are seven clusters formed. The profiles of each Cluster are as follows: Cluster 1 is formed by the Fabric wash and its associated Brands such as Rin, Surf Excel, Power, Tide, Comfort, Wheel and Ariel. This shows that the retail shop should allot a separate shelf display with the brands specified above. From the POS data it is found that this display will increase the awareness of brands, its products and their promotions thereby increasing the sales of the products. Cluster 2 contains Snacks and its associated products such as Britannia, Sunfeast, Maggi noodles and Top ramen. This shows that the retail shop should allot a separate shelf display with the brands specified above. And visual merchandising for new products of Britannia will be induce impulse purchase of its products. Cluster 3 has Personal wash and related products. This shows that the retail shop should allot a separate shelf display with the brands specified above. Brands such as Hamam Cinthol and lifebuoy are well established in sales and people has to be educated about the other brands below lifebuoy with promotions like free samples and discounts which will increase the sales. Cluster 4 is formed with 83% of Food, 15% of Sanitary napkins and 1% of beverages with its related products. This shows that the cluster formed is heterogeneous where the retailer should allot more brands in these categories with less number of stocking units so that the sales would increase on impulse purchase. Cluster 5 is formed with 54% of Beverages, 42% of House hold Care, 3% of Deodorants, 1% of snacks and their related products. This shows that the cluster formed is heterogeneous where the retailer should allot more brands in these categories with less number of stocking units so that the sales would increase on impulse purchase. Cluster 6 is formed by 91% of Oral care and 9% of Cosmetics and their related products. The cluster formed is heterogeneous where the retailer should allot more brands in these categories Cosmetics with less number of stocking units so that the sales would increase.

Cluster 7 is formed by 61% of Skin Care, and 39% of Hair Care products and with their related products. The cluster formed is in these categories where the retailer should allot more brands in these categories with less number of stocking units so that the sales would increase.

**CLUSTER DIAGRAM**

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.

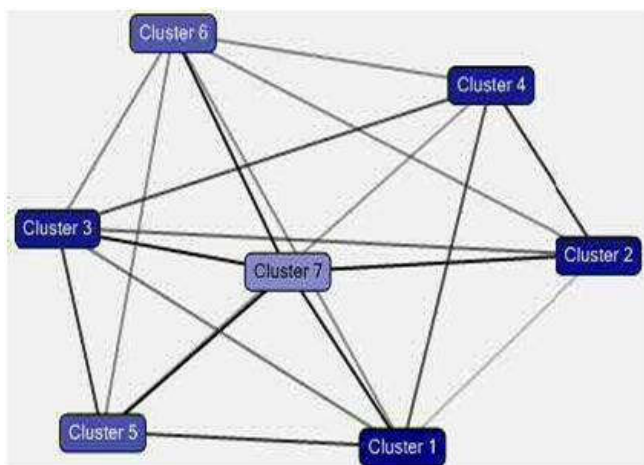


Fig 6.0

**FINDINGS**

**AVERAGE SPENT DISPERSION**

Spend dispersion is the percentage of spend dispersed across various categories of products like Hair care, personal care, skin care etc. by the shoppers in the retail outlet.

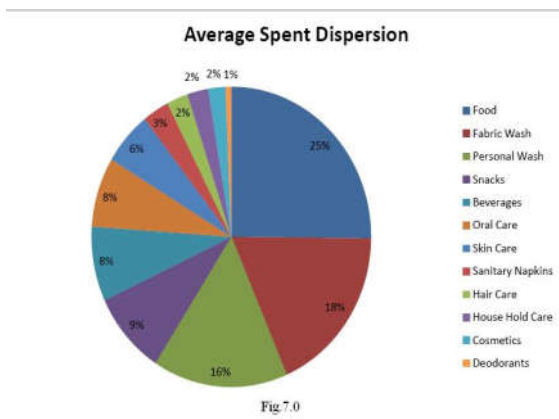


Fig 7.0

People who visit this retail shop spend 25% of their money on food related and it accounts first when compared to other product category. Next is the fabric wash, Hair care and other product categories which contributes to the spent dispersion.

**DEPENDENCY NETWORK FOR CATEGORY**

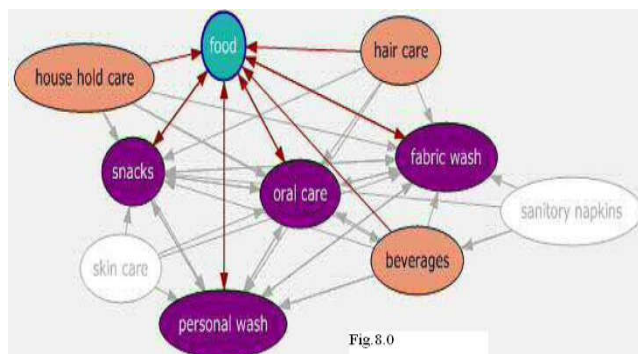


Fig. 8.0

Food, Snacks, Oral Care, Personal Wash, Fabric Wash are connected by double ended arrows which means that when user buys any of these product he is likely to buy other products connected. House hold care and hair care determines the occurrence of food by the direction of the arrows. There is no association between skin care and sanitary napkins with food.

**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 s 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 xample 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. Products for which there is less association can be improved by bundling those products on an offer. For example Clinic Plus and Lux can be bundled on an off price. 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.

**CONCLUSION**

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. Pradham, Swapna (2007), Retailing Management 2E, The McGraw Hill Companies. pp. 226-228. ISBN 978-0-07-062020-9.
2. David Willis, Market Basket Analysis (Cited on April 2012). Available from URL: [http://www.information-drivers.com/market\\_basket\\_analysis.php](http://www.information-drivers.com/market_basket_analysis.php)
3. Derick Jose, Retail Analytics Demystified(Cited on May 2012). Available from URL: <http://retailanalyticsdemystified.blogspot.com/2009/08/w-hich-behaviorial-variables-to-segment.html> (Cited on January 2012).
4. DMS Retail, KPI For Retail(Cited on November 2011). Available from URL: [http://www.dmsretail.com/interpreyou\\_kpi.htm](http://www.dmsretail.com/interpreyou_kpi.htm)
5. 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.
6. Jiawei Han and Micheline Kamber (2009), Data Mining: Concepts and Techniques), Second Edition, Morgan Kaufmann Publishers.
7. C.R.Kothari (2010), Research Methodology Methods and Techniques, Second Edition, New Age International Publishers.
8. Michael J.A. Berry and Gordon S. Linoff, Data Mining Techniques, Second Edition, Wiley Publishing, Inc., 2007.
9. 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.
10. Mourad Ykhlef (2011), Association mining of dependency between time series using Genetic Algorithm and discretisation, International Journal of Business Intelligenceand Data Mining 2011 - Vol. 6, No.1 pp. 55-70.
11. Thomas H. Davenport and Jeanne G. Harris, Competing on Analytics: The New Science of Winning, First Edition, Harvard Business School Press,2007.
12. Mourad Ykhlef: Association mining of dependency between time series using Genetic Algorithm and discretisation. IJBIDM 6(1): 55-70 (2011)
13. Journal Knowledge and Information Systems archive Volume 26 Issue 1, January 2011 Pages 1-30 , Hongyan Liu, Yuan Lin, Jiawei Han Hongyan Liu et al (2011)
14. IBM Software Business Analytics (2010) White Paper Titled The Future of Market –[www.ibm.com](http://www.ibm.com)
15. Advanced Matrix Algorithm (AMA): reducing number of scans for association rule generation Marghny H. Mohamed, Mohammed M. Darwieesh, A.B.M. Shawkat Ali DOI: 10.1504/11.39412 . International Journal of Business Intelligence and Data Mining 2011 Vol. 6 No. 2 .page 202-214
16. IIBA (March 2009). Kevin Brennan. ed. A Guide to the Business Analysis Body of

\*\*\*\*\*