A review on Data Mining for Indian Online Retail Industry

Ms. Pradnya Muley (Author)
Assistant Professor, MCA Department,
PES's Modern College Of Engineering,
Pune, India
pradnyamuley@gmail.com

Dr. Aniruddha Joshi(Author)
Department of Management Sciences,
Savitribai Phule Pune University,
Pune, India
joshiag@gmail.com

Abstract— Data mining, the technique of extracting hidden information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The Indian retail industry has emerged as one of the most dynamic and fast-paced industries due to the entry of several new players in the Indian Retail Sector. It increases the need of different knowledge extraction tools which can extract usable information form large data sets. In this paper we have discussed the Overview of an Indian retail sector, the previous study related to the retailing industry. Also the study presents the gaps identified in the previous work.

Keywords- Data Mining; Knowledge Discovery Data; Indian Retail Sector; Business Intelligence

_____*****_____

I. INTRODUCTION

Data Mining (DM) is the process of analyzing data from different perspectives and summarizing it into useful information. It is also known as Knowledge Discovery Data & Knowledge Mining (Chen et al., 2006). Furthermore, it extracts the hidden predictive information from large databases Ramageri and Desai (2013). The DM is a powerful technology with great potential to help organizations focus on the most important information in their data ware houses (Berry & Linoff, 1997; Fayyad et al., 1996; Han & Kamber, 2006).By the use of DM, nature and behavior of any kind of data can be predicted. In the retail industry, Data mining collects usage quantity of information from on purchase/ sales, customer purchasing history, consumption, and services. Data mining in there tail industry helps in identifying customers buying patterns and trends that lead to be improved the quality of customer services and satisfaction (Vohra, 2016).

II. BACKGROUND OF THE STUDY

Retailing is defined as business activities involved in selling goods and services to consumers for their personal, family or household use (Berman & Evans, 2001).The emergence of new formats and the evolution of modern retail in India have attracted attention in recent years (Mulky & Nargundkar, 2003). The business press in India has carried several articles and news items in the last three years about the modern formats (Bhattacharjee, 2001).Retailing in India is one of the biggest sectors witnessing tremendous transformation (NCERT, 2012).Now a day's retailer is facing dynamic and competitive environment, with an increase in globalization and competitiveness retailers are seeking better market campaign (Onestepretail, 2015). The retailer is collecting a large amount of customer daily transaction details. This data collected requires proper mechanisms to convert it into knowledge, using this knowledge retailer can make better business decisionRamageri and Desai (2013). However, analyzing such large amount of data is a challenging and time-consuming process as the analyst often needs to divide data into multiple segments based on certain criteria and also to combine objectives into groups based on needs. It is not possible to determine these data volumes in a conventional way; thus, DM has been introduced, which is commonly used to segment effectively as it enables to differentiate customers' markets to meet their various needs; therefore, marketing efforts are manageable (Bulysheva & Bulyshev, 2012; Tan et al., 2006). It is also apparent that the benefits of analytics will continue to expand and span a variety of dimensions, including an overall improvement in the quality and speed of decision, better alignment of resources to strategies, increased revenue, and improvement in cost efficiencies (Muley & Joshi, 2015)

III. OVERVIEW OF INDIAN RETAIL SECTOR

The Indian retail industry has experienced high growth over the last decade with a noticeable shift towards organized retailing formats. The industry is moving towards a modern concept of retailing. The size of India's retail market was estimated at US$ 435 billion in 2010. Of this, US$ 414 billion was traditional retail and US$ 21 billion was organized retail. India's retail market is expected to grow at 7% over the next 10 years, reaching a size of US$ 850 billion by 2020. Traditional retail is expected to grow at 5% and reach a size of US$ 650 billion (76%) while organized retail is expected to grow at 25% and reach a size of US$ 200 billion by 2020. The US-based global management consulting firm, A T Kearney, in its Global Retail Development Index (GRDI) 2011, has ranked India as the fourth most attractive nation for retail investment, among 30 emerging markets. As India's retail industry is aggressively expanding itself, the great demand for real estate is being created. The cumulative retail demand for real estate across India is expected to reach 43 million square feet by 2013. Around 46 per cent of the total estimated demand between 2009 and 2013 will become from Tier-1 cities. For instance, Pantaloon Retail added 2.26 million square feet (sq. ft.) of retail space during the fiscal 2011 and booked over 9 million sq. ft of retail space to fructify its expansion plans in future. Some of the key players in the Indian retail market, with a dominant share are:

1. Pantaloon Retail Ltd, a Future group venture: Over 12 mn sq. ft. of retail space spread over 1,000 stores, across 71 cities in India.
2. Shoppers Stop Ltd: Over 1.82 mn sq. ft. of retail space spread over 35 stores, in 15 cities.

401

3. Spencer's Retail, RPG Enterprises: Retail footage of over 1.1 mn sq. ft. with approx 250 stores, across 66 cities.
4. Lifestyle Retail, Landmark group venture: Has approximately 15 lifestyle stores and 8 Home centers.

Indian retailing is undergoing a process of evolution and is poised to undergo a dramatic transformation. The retail sector employs over 8% of the national workforce but is characterized by a high degree of fragmentation with over 5 million outlets, 96% of whom are very small with an area of less than 50 m 2 (Aggarwal, 2000).For the urban sector alone, the shop density increased from 4 per 1000 people in 1978 to 7.6 per 1000 people in 1996 (Venugopal, 2001). Because of their small size, Indian retailers have very little bargaining power with manufacturers and perform only a few of the flows in marketing channels unlike in the case of retailers in developed countries, (Sarma, 2000).

In recent years, there has been a slow spread of retail chains in some formats like supermarkets, department stores, malls and discount stores. Factors facilitating the spread of chains are the availability of quality products at lower prices, improved shopping standards, convenient shopping and display, and blending of shopping with 6 entertainment, and the entry of industrial houses like Goenkas, Rahejas, Piramals and Tatas into retailing (Ramaswamy & Namakumari, 2002). However, formats are not easily scalable across the country. Several companies have found that it is not easy to expand beyond some regions and cities as evident from the examples of Margin Free Market and Foodworld, which are active only in a few states or cities. Affordable real estate prices and availability of a sufficient number of economically well-off households in the catchment area are critical requirements that will determine new store viability and thus the possibility of further expansion(Mulky & Nargundkar, 2003).

A. Growth of Indian retail market

- 2010: continued growth expected to grow retail market to $427.00 billion by 2010
- 2008: The Indian retail boom is at its peak, and this phase has been termed the high retail gold rush.
- 2006-2007: Maintaining its #1 position as the market with the most opportunity for retail growth, India's retail market grew to $330.00 billion
- 2005-2006: The beginning of the Indian retail boom, India is also ranked as the #1 market for global retailers to enter accounting to global retail development Indices.
- 2003-2004: Growth: standing at $230.00 billion, India's retail market enters the growth phase, characterized by the entry of new domestic and international participants and expansion by existing retailers in India.
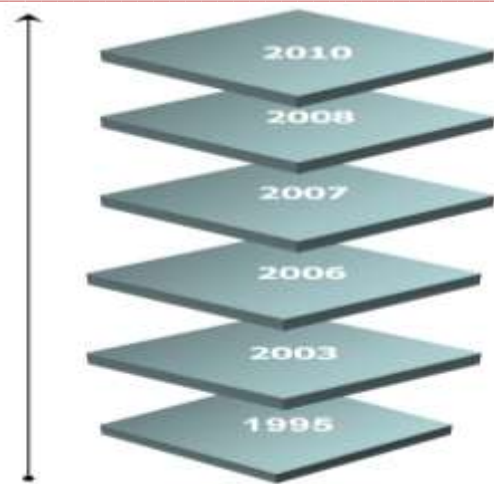- 1995: Emerging: the Indian retail market is classified as an emerging market.

Figure 1.   Growth of Indian Retail Market. Source (Sensormatic, 2008)

IV.   PREVIOUS STUDY

The previous study related to the retailing industry is discussed as follows.

Mulky and Nargundkar (2003) this paper analyzes the developments in retailing in India. A literature survey of retailing in India and some newly industrialized countries is carried out. Factors stimulating modernization are discussed, and some implications for managerial action and policy are derived.

Sabitha et al. (2011) a framework for Decision tree and frequent item set is developed for the analysis of wholesale data. The wholesale customer dataset is taken and analyzed to know the session at which the product can be sold more using decision tree algorithm like ID3 and C4.5. The data in the dataset is preprocessed to make it suitable for classification. The preprocessed data is used for classification, and we obtained high classification accuracy.

Rajagopal (2011) aimed to identify the high-profit, high-value and low-risk customers by mining technique - customer clustering. First, they cleansing the data and developed the patterns through a demographic clustering algorithm using IBM I-Miner. Second, profiling the data develop the clusters and identify the high-value low-risk customers. However, this study planned to extend the detailed behavioral data and opportunity identification using association algorithms within the segments discovered.

Velu and Kashwan (2012) Proposed intelligent models for market classification of customers into suitable classes in order to enhance the business decision-making process. The models are validated with rigorous experimental training and then subsequently tested for effectiveness. Three models of the decision tree, ensemble model, and SOM model are applied and tested for customer classification. Results are quite encouraging. Classifications by models quite closely match with that of statistical results. The models are put into practice in real-time business analysis and again validated for accuracy and effectiveness. However, this study doesn't concentrate on more application areas such as health monitoring and field management for testing and validating models.

Kashwan and Velu (2013) proposed a K-means Clustering technique for segmenting applications in the market forecasting. Furthermore, develop a real-time and online system using SPSS Tool for a particular supermarket to predict

sales in various annual seasonal cycles. The model developed was an intelligent tool which received inputs directly from sales data records and automatically updated segmentation statistics at the end of day's business. The model was successfully implemented and tested over a period of three months. The experimental results were shown that quite encouraging and had shown high accuracy. However, this study focused only towards supermarket and didn't involve more trials with automation of market forecasting and planning.

Singh and Rana (2013) proposed a new Customer Centric Analytical model using data mining, computing active customer retention value and the parameters affecting this value. The data analysis done using data mining techniques helps the Organization for better decision making and strategy planning. An empirical study of Automobile industry has been discussed, data analysis, result interpretation thus providing a business solution to the Organization to improve prospect customer. Furthermore, this study has planned to extend to know the significant and influential features affecting the sales of the Organization, keeping customer choices in mind.

Raju et al. (2014) analysis the concept of Data mining and Customer Relationship Management in organized Banking and Retail industries. It also discusses standard tasks involved in data mining; evaluate various data mining applications in different sectors. They use data warehousing to combine various data from databases into an acceptable format so that the data can be mined. The data is then analyzed, and the information that is captured is used throughout the organization to support decision-making. However, this study doesn't focus the decision-making process. This will reap the immense benefit and derive considerable competitive advantage to withstand competition.

Agarwal (2014) Different customer profiles were segregated by making use clustering algorithms. The results were reviewed, and they were instrumental in identifying different groups. The six consumer profiles have been described with the Minitab results as a reference point. The proposed method not only increases profit for companies but also poses a challenge in the field of data mining about how probably "Recommended for You" items are chosen by the customers and how likable is the platform or store according to the customer preferences.

Nagpal (2014) discussed some of the key challenges facing the online retail industry and the use of data mining techniques in customer-centric business intelligence for online retailers. Some aspects of Flipkart, the online Indian giant, have also been discussed with regard to the online retailing challenges. Also, focused the use of business intelligence techniques for online retail brands to stay competitive in the cyber market.

Vidhate (2014)aimed of this research work is to know what are the observations of different experts and researchers regarding development in the super bazaar with its impact on consumer response, economic development as well as competitive changes that have taken place in marketing processes of different super bazaars. It is also used to know how knowledge mining is linked to consumer buying behavior analysis for effective decision making in super bazaars.

V. Gaps identified

Rajagopal (2011) planned to extend the detailed behavioral data and opportunity identification using association algorithms

within the segments discovered. Velu and Kashwan (2012) this study doesn't concentrate on more application areas such as health monitoring and field management for testing and validating models. Kashwan and Velu (2013) focused only towards supermarket and didn't involve more trials with automation of market forecasting and planning. Raju et al. (2014) this study doesn't focus the decision-making process. This will reap the immense benefit and derive considerable competitive advantage to withstand competition.

References

[1] Agarwal, P. (2014). Benefits and Issues Surrounding Data Mining and its Application in the Retail Industry. International Journal of Scientific and Research Publications. [Online]. 4 (7). pp. 1–5. Available from: http://www.ijsrp.org/research-paper-0714/ijsrp-p3146.pdf.

[2] Aggarwal, A. (2000). Current issues in Indian retailing. European Retail Digest. 25. pp. 70–71.

[3] Berman, B. & Evans, J.. (2001). Retail Management. 8th Ed. Upper Saddle River, N.J: Prentice Hall.

[4] Berry, M.J.A. & Linoff, G. (1997). Data mining techniques for marketing, sales and customer support. USA: John Wiley and Sons.

[5] Bhattacharjee, P. (2001). Retail's new Godzilla Business World. In: A. Coughlan, E. Anderson, L. Stern, & A. El-Ansary (eds.). Marketing Channels. New Delhi: Prentice Hall of India Pvt. Ltd, pp. 44–46.

[6] Bulysheva, L. & Bulyshev, A. (2012). Segmentation Modeling Algorithm: A Novel Algorithm in Data Mining. Information Technology & Management. 13 (4). pp. 263–271.

[7] Chen, Y.-L., Chen, J.-M. & Tung, C.-W. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. Decision Support Systems. [Online]. 42 (3). pp. 1503–1520. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0167923606000030.

[8] Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (1996). Advances in Knowledge Discovery and Data Mining. Menlo Park, Calif: AAAI Press.

[9] Han, J. & Kamber, M. (2006). Data Mining Concepts and Techniques. 2nd Ed. Burlington: Morgan Kauffman.

[10] Kashwan, K.R. & Velu, C.M. (2013). Customer Segmentation Using Clustering and Data Mining Techniques. International Journal of Computer Theory and Engineering. [Online]. 5 (6). pp. 856–861. Available from: http://www.ijcte.org/index.php?m=content&c=index&a=show&catid=52&id=943.

[11] Muley, P. & Joshi, A. (2015). Application of Data Mining Techniques for Customer Segmentation in Real-Time Business Intelligence. International Journal of Innovative Research in Advanced Engineering (IJIRAE). [Online]. 2 (4). pp. 107–109. Available from: http://www.ijirae.com/volumes/Vol2/iss4/19.APAE10122.pdf.

[12] Mulky, A. & Nargundkar, R. (2003). Modernisation in Indian retailing: Managerial and policy perspectives. Mulky and Nargundkar. [Online]. 27 (2). pp. 1–8. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.546.8393&rep=rep1&type=pdf.

[13] Nagpal, S. (2014). Improving Online Retail Effectiveness through Business Intelligence. International Journal of Engineering, Business and EnterpriseApplications. [Online]. 10 (1). pp. 35–40. Available from: http://iasir.net/IJEBEApapers/IJEBEA14-416.pdf.

[14] NCERT (2012). Organized Retailing. [Online]. Bhopal. Available from: http://educationhp.org/Files/RS101-RetailBusiness.pdf11_54_2013_12_07_01.pdf.

[15] Onestepretail (2015). 6 Biggest Challenges Retailers Face Today. [Online]. 2015. Onestepretail. Available from: https://onestepretail.com/blog/white-paper-6-biggest-challenges-retailers-face-today-sneak-peek/. [Accessed: 21 April 2016].

[16] Rajagopal, S. (2011). Customer Data Clustering Using Data Mining Technique. International Journal of Database Management Systems. 3 (4). pp. 1–11.

[17] Raju, P.S., Bai, V.R. & Chaitanya, G.K. (2014). Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries. International Journal of Innovative Research in Computer and Communication Engineering. 2 (1). pp. 2650–2657.

[18] Ramageri, B.M. & Desai, B.L. (2013). Role Of Data Mining In Retail Sector. International Journal of Computer Science and Engineering. [Online]. 5 (1). pp. 47–50. Available from: http://www.enggjournals.com/ijcse/doc/IJCSE13-05-01-051.pdf.

[19] Ramaswamy, V.. & Namakumari, S. (2002). Marketing Management. Delhi: Macmillan India Ltd.

[20] Sabitha, Amma, N.G.B., G.Annapoorani & P.Balasubramanian (2011). Implementation of Data Mining Techniques to Perform Market Analysis. International Journal of Innovative Research in Computer and Communication Engineering International Journal of Innovative Research in Computer and Communication Engineering. [Online]. 2 (11). pp. 7003–7008. Available from: http://www.rroij.com/open-access/implementation-of-data-mining-techniques-toperform-market-analysis.pdf.

[21] Sarma, M.. (2000). Some issues in retail management in India. Vision. 4 (1). pp. 35–40.

[22] Sensormatic (2008). Overview of the Indian Retail Market.

[23] Singh, A. & Rana, A. (2013). Mining of Customer data in an Automobile Industry using Clustering Techniques. International Journal of Emerging Technologies in Computational and Applied Sciences. 5 (3). pp. 251–258.

[24] Tan, Steinbach, M. & Vipin Kumar (2006). Introduction to Data Mining. [Online]. Pearson Addison-Wesley. Available from: https://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=8&cad=rja&uact=8&ved=0CE8QFjAHahUKEwjM1_T41YfGAhUneaYKHb3qAKk&url=http://pencilji.com/download/37729661.pdf&ei=_4J5VYzcFqfymQW91YPICg&usg=AFQjCNF2vo06UKTaSoeEFUkaH6aRYsnbMw&sig2=X9XrUgg-gfYK-hTDiFp_5Q&bvm=bv.95277229,d.dGY.

[25] Velu, C.. & Kashwan, K.R. (2012). Performance Analysis for Visual Data Mining Classification Techniques of Decision Tree, Ensemble, and SOM.International Journal of Computer Applications. [Online]. 57 (22). pp. 65–71. Available from: http://research.ijcaonline.org/volume57/number22/pxc3883874.pdf.

[26] Venugopal, P. (2001). Marketing Channel Management : A Customer-Centric Approach. New Delhi: Response Books.

[27] Vidhate, D.R. (2014). A Conceptual Study of Consumer Behavior Analysis in Super Bazaar using Knowledge Mining. In: Sinhgad Institute of Management and Computer Application. [Online]. 2014, SIMCA, pp. 70–75. Available from: http://nci2tm.sinhgad.edu/NCIT2M2014_P/data/NCI2TM_16.pdf.

[28] Vohra, J. (2016). Data Mining Approach for Retail Knowledge Discovery. International Journal of Advanced Research in Computer Science and Software Engineering. [Online]. 6 (3). pp. 740–742. Available from: http://www.ijarcsse.com/docs/papers/Volume_6/3_March2016/V6I3-0388.pdf.

[29]