

Data Mining Approach in Retail Knowledge Discovery and Internet Technologies

Simranjit Kaur¹ and Seema Baghla²

¹M. Tech. Student, ²Assistant Professor, Computer Science Engineering, Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo, Bathinda, Punjab, India
E-Mail: simranjitnathan@gmail.com, garg_seema238@yahoo.co.in

Abstract - Online shopping has a shopping channel or purchasing various items through online medium. Data mining is defined as a process used to extract usable data from a larger set of any raw data. The data set extraction from the demographic profiles and Questionnaire to investigate the gathered based by association. The method for shopping was totally changed with the happening to internet Technology. Association rule mining is one of the important problems of data mining has been used here. The goal of the association rule mining is to detect relationships or associations between specific values of categorical variables in large data sets.

Keywords: Data mining, online shopping, SPSS, association technique

I. INTRODUCTION

It is the process used to extract usable data from a larger set of any raw data. It implies analyzing data patterns in large batches of data using one or more software [1]. Data mining involves effective data collection and warehousing as well as computer processing. For segmenting the data evaluating the probability of future events, data mining uses sophisticated mathematical algorithm [1]. Data mining is also known as knowledge discovery in data. Step by step the new advancements are assuming its own significant position which meets the cost of substitute way to deal with advertising techniques. Because of the improvement in PCs, the clients utilize PCs and access all data of any business endeavor from their home [2].

II. LITERATURE REVIEW

Bejju [1] discussed, e-commerce websites for sales analysis using DM techniques to evaluate, propose & improve strategies by using web mining techniques. Decision trees have been used for classification of probabilistic business model. ID3 algorithm has been used for create decision tree of gives data set. Ramageri [2] discussed some data mining techniques such as classification, clustering, prediction, association rule and neural network. Authors concluded that data mining techniques and algorithm are helpful in finding the patterns to decide upon the future trends in businesses to grow. Pahwa et al. [3] discussed a review of consumer online buying behavior to study the consumer buying behavior concerns with the individual or the group using internet live stats. Demographic profile of India online users has been also shown. Kalaivani et al. [4] discussed customer behavior predictive modeling for online shopping using

tuned decision tree method. The combination of two techniques has been used which are classification and prediction. Novel tuned decision tree predictive model has been proposed and evaluated in this study. This proposed technique has been compared and validated with other data sets of online shopping.

Arora [5] discussed the prospect of e-retailing in India to study the benefits and challenges associated with the e-retailing business setup. Secondary data has been collected through various books, business magazines, journals, newspapers and internet websites. The result of this study is that the best method to reach out to people is to create medium through blogs and discussion forums for them. Gerrikagoitiaa et al. [6] discussed new trends in e-marketing on the web mining. The main aim of this study is that to test SMEs of active users on web mining solutions. This analysis has been observing the significant of the digital footprint at the beginning of the web mining processes. Singh [7] discussed present scenario, emerging opportunities and challenges in retail sector in India to analyze the present structure of Indian retail sector. Collected and compiled data has been calculated and analyzed through relevant ratios and percentages. The result of this study is that the Indian retail sector is evolving rapidly. Furrer and Sudharshan [8] in this paper, identify many opportunities and its problems has be used Internet. The main aim of this study is that Internet has richer support for a questionnaire and allow the accorded automatically data.

Qinghai *et al.* [9] discussed the online shopping changes the retail business model. A survey of the people who use online shopping in China, has been made. The result of this study is that the low price, discount, feedback from previous shoppers and quality of product and information are also important factors in online shopping. Sana [10] discussed a study on the impact of online shopping upon retail trade business. The primary data has been collected through direct personal interview and investigation through schedules. The result of this study is that retail stores are more engaged to customer satisfaction. Consumers become more comfortable with online purchasing for its convenience and home delivery. Dawn and Kar [11] discussed the various issues, opportunities and effective strategies for growth & development of e-retaility in India. The main aim of this study is to draw an effective strategy based on the detailed

survey of e-retailed companies in India. The primary data has been collected based on interviews e-retailed consletants and 30 shoppers in kolkata. At the end of this study, industry should ensure that costumers must feel safe & secure while transality online. Yassir and Nayak [12] discussed the various issues in data mining and information retrieval. A qualitative and quantitative method has been employed with secondary data and analysis.

III. METHODOLOGY

To recognize the issue, to create and approach the issue and to plan a proper examine outline; essential and auxiliary information has been utilized [5]. To gather data for the investigation from clients, essential research is utilized. A few individuals who were web clients were recognized and essential information were gathered from respondents utilizing organized survey.

The methodology starts by recognizing the opinion of the shoppers through questionnaires to test the comprehensive purpose of the business. The customer’s opinion is then analyzed [4, 6, 8]. Thereafter, the data mining techniques is applied to analyze customer’s opinion related to customers, e-customers, shoppers and e-shoppers. Finally, the reasons for the users towards online or retail shopping are analyzed. Figure 1 shows the flowchart of the proposed methodology.

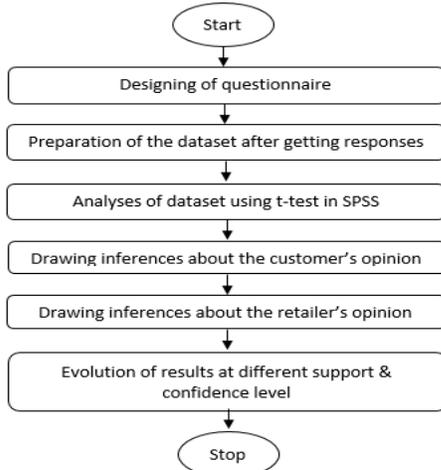


Fig.1 Flow chart of proposed work

Figure 1 represents flow of the proposed work that has been used for classification process.

1. How have the utilizations of web based business influenced customary retailers?
2. How to depict the market condition and the promoting procedures that should be changed.
3. The absence of leaning of more up to data online business destination.
4. To assemble form internet looking from the clients, empowering online business to best the homogenous customary business.
5. How it because more easily far the retailer’s batters always to pull in the consideration of the clients with respect to the buying exercises.

6. The online directs are affirmed in a few territories it can’t be across the board for the entire masses.
7. Make online channels and towards more current innovations adjusted by new age.
8. Make a security channels at high rates so costumers can more trust on online shopping.

IV. RESULTS

Figure 2 shows the SPSS data source opening dialogue box by clicking the "More files ..." option and locate the dataset.

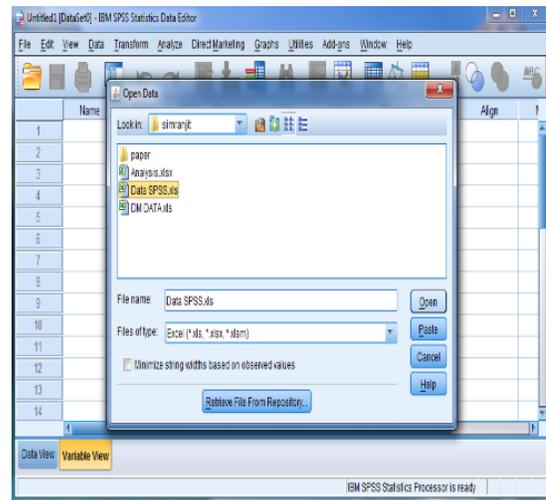


Fig. 2 SPSS excel data

Figure 3 shows the view of actual data in our dataset for each variable including the variable name, type, various properties of the data, label(s) for the variable and its values.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Format
1	GROUP	String	3	0		None	None	3	Left	
2	V1	Numeric	11	0		None	None	11	Right	
3	V2	Numeric	11	0		None	None	11	Right	
4	V3	Numeric	11	0		None	None	11	Right	
5	V4	Numeric	11	0		None	None	11	Right	
6	V5	Numeric	11	0		None	None	11	Right	
7	V5	Numeric	11	0		None	None	11	Right	
8	V7	Numeric	11	0		None	None	11	Right	
9	V8	Numeric	11	0		None	None	11	Right	
10	V9	Numeric	11	0		None	None	11	Right	
11	V10	Numeric	11	0		None	None	11	Right	
12	V11	Numeric	11	0		None	None	11	Right	
13	V12	Numeric	11	0		None	None	11	Right	
14	V13	Numeric	11	0		None	None	11	Right	

Fig. 3 Data view

Figure 4 and 5 shows the output of the t-test and independent samples test on the dataset respectively.

A. Observations of Customers

The observations of the customers have been listed in table 1. It is clear from Table 1 that there is a significant difference between e-shopping and shopping in response to customers. According to them, parameters such as “All

Branded products are available on e-shopping”, “Timely delivery of products with Payment security”, “e-shopping has multiple payment options” and “e-shopping has Easy product selection” are highly significant as indicated by “***”. Other parameters such as “Website provides much information to take purchase decision”, “Convenient replace, return and refund policy” and “Order tracking system” were found to be significant as indicated by “**”.

These parameters indicate that there is a difference between online and traditional marketing these days and customers used to prefer e-shopping scenario.

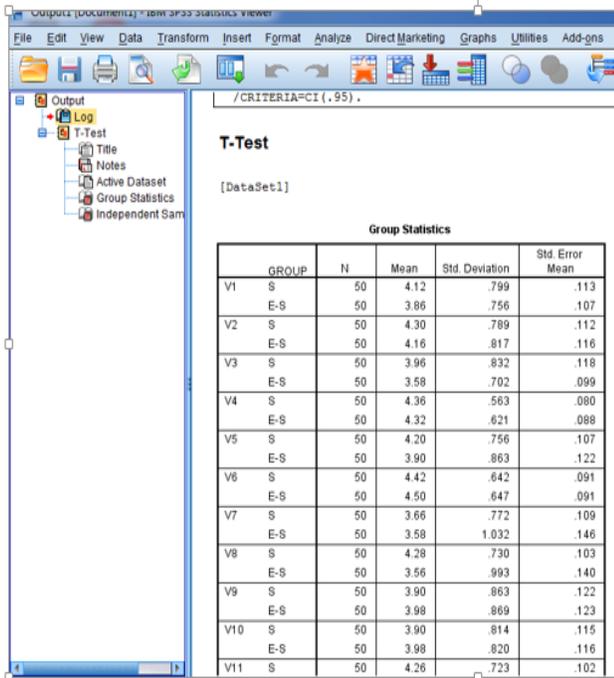


Fig. 4 Output of t-test

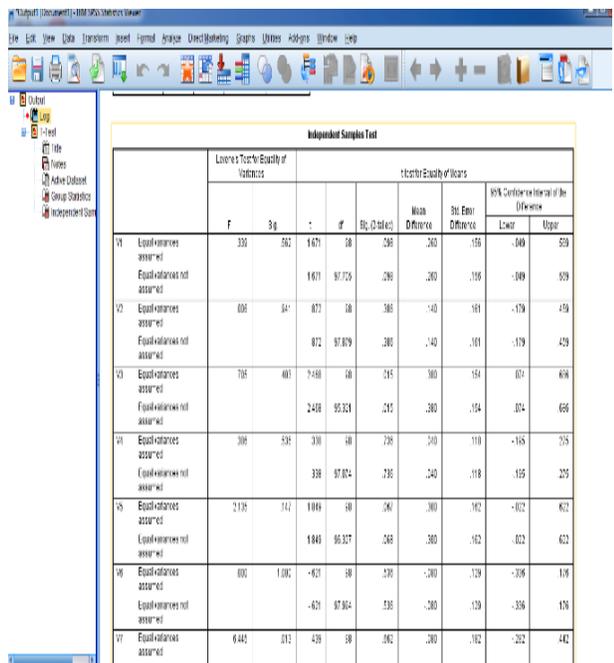


Fig. 5 Independent sample test

TABLE I OBSERVATIONS OF CUSTOMERS

Parameters	Customers		
	E-Shopping	Shopping	T-Test
Website easy to navigate and operate	4.12	3.86	1.67
24/7 websites service for shopping	4.3	4.16	0.87
Website provides much information to take purchase decision	3.96	3.58	2.47*
Required and Advertised products in stock	4.36	4.32	0.34
Gift any item to anywhere is easy via online	4.2	3.9	1.85
Accurate/Relevant information of products	4.42	4.5	0.62
Local stores have better prices & promotions than e-stores	3.66	3.58	0.44
All Branded products are available on e-shopping	4.28	3.56	4.13**
Takes hard time to search on Internet to find required products	3.9	3.98	0.46
Dislike the delivery problems & backorders of e- buying	3.9	3.98	0.49
Visual Appeal of product	4.26	4.36	0.71
Timely delivery of products with Payment security	4.08	3.56	2.84**
Prices are lower compared to local stores	4.44	4.28	1.12
Convenient replace, return and refund policy	4.32	3.94	2.52*
e-shopping has multiple payment options	4.32	3.82	2.79**
e-shopping has Easy product selection	3.64	4.12	2.99**
Order tracking system	4.3	3.94	2.39*
Sales people have a good knowledge of products	3.8	3.74	0.39
Automated emails & messages of new products & discounts	3.9	4.04	0.89
Stores are conveniently located	4.32	4.22	0.67

All the significant values have been represented in figure 6. The values that were non-significant have been removed.

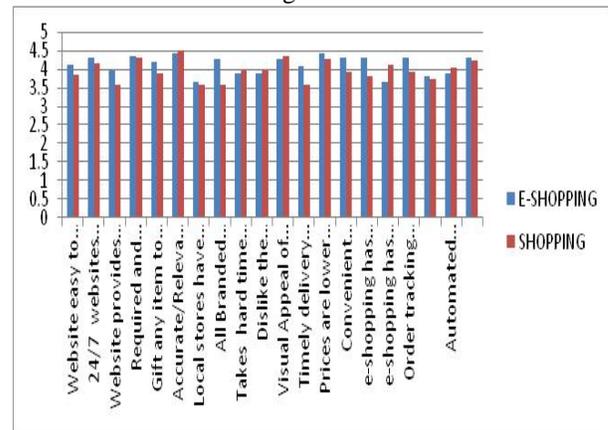


Fig. 6 Observations of customers

B. Observations of retailers

The observations of the retailers has been listed in table 1. It is clear from Table 2 that there is a significant difference between e-shopping and shopping marketing in response to retailers. According to them, parameters such as “Website

Figure 8 shows the dataset with number of rows, columns and attributes with minimum values for support as 0.3, confidence as 0.5 and maximum card item sets as 6.

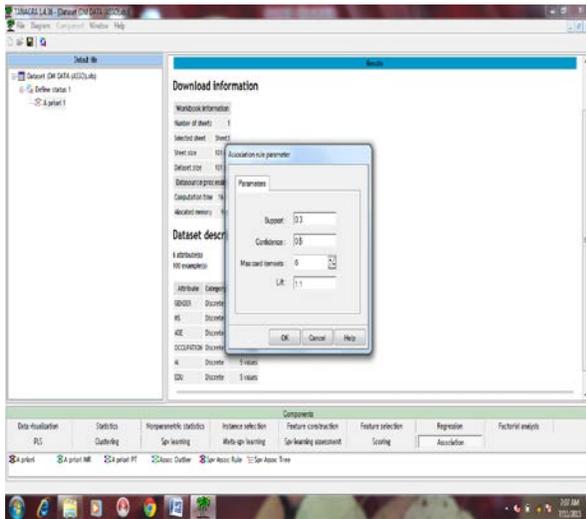


Fig. 8 Dataset with no. of rows, columns and attributes

Figure 9 and 10 shows the discrete and continuous attributes and attributes selected as input to ARM respectively.

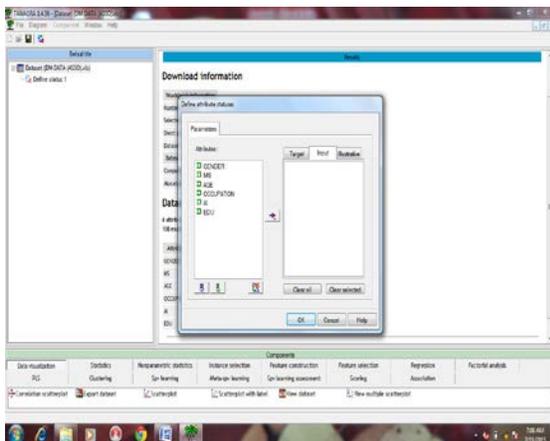


Fig. 9 Discrete and continuous attributes

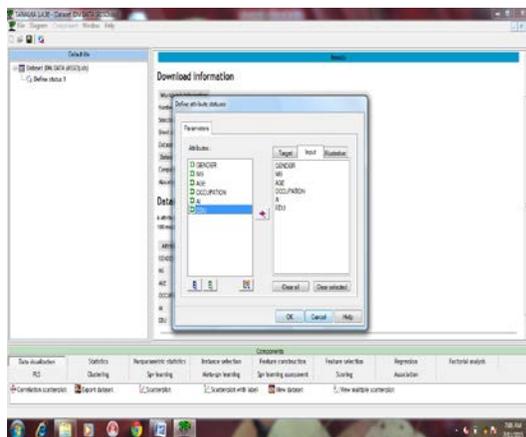


Fig. 10 Attributes selected as input to ARM

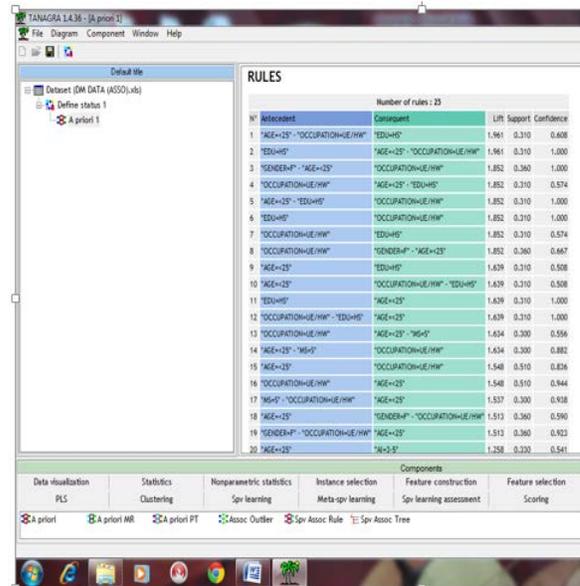


Fig. 11 Values of the parameters and results in 25 rules.

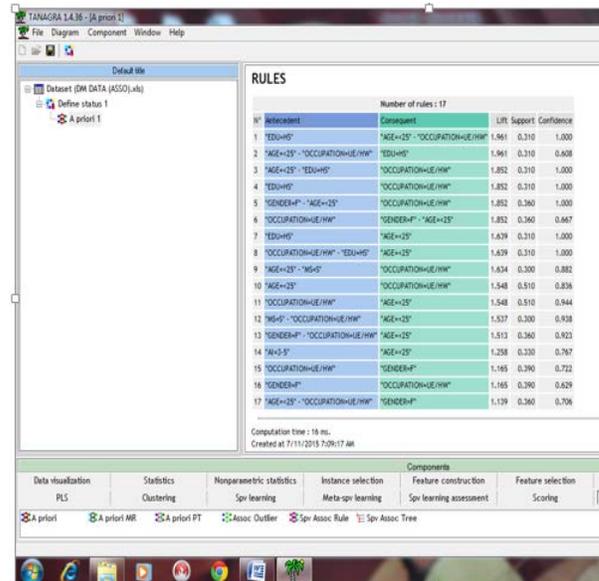


Fig. 12 17 rules left

Figure 11 shows the values of the parameters and results in 25 rules. It shows various rules such as Age <25 and Education is HS gives Support as 0.31 and confidence as 0.508. Similarly Gender as F, Age <25 and Occupation as UE/HW gives Support as 0.36 and confidence as 1.000.

Figure 12 shows the 17 rules left with minimum values for support as 0.3, confidence as 0.6 and maximum card item sets as 6.

Figure 13 shows the 06 rules left with minimum values for support as 0.35, confidence as 0.7 and maximum card item sets as 6.

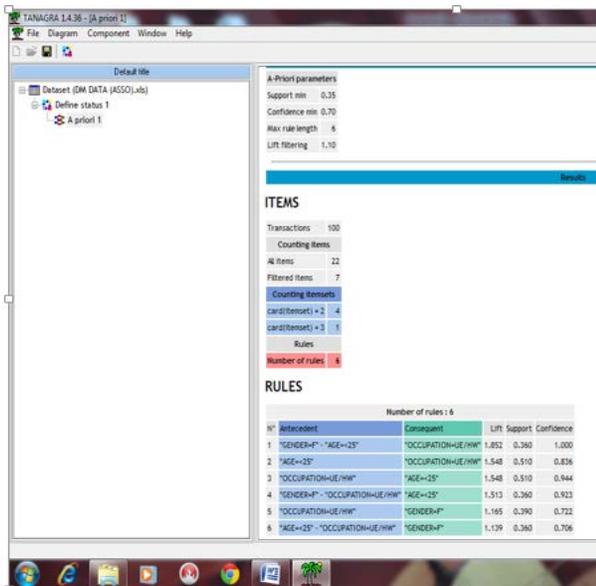


Fig. 13 6 rules left

V. CONCLUSION

1. Association predicts which group of customers used to shop online or retail more than others.
2. By setting the values of support and confidence, every marketer can check the particular group for their sales.
3. According to the data collection, it is predicted that females do e-shopping more than males.
4. Maximum customers and retailers belong to age group 25-60 years which depicts that customer are fond of shopping to spend some time with family and leisure activities.
5. Housewives do the e-shopping as they may not have time to visit retail store due to their household work.
6. Qualified customers used to shop than illiterate ones.

VI. FUTURE SCOPE

The responses gathered from customers and retailers are confined to the specific region. It can be extended with wider area such as state or different states. Difference

between various websites or states can be done. More powerful tools such as neural networks etc. or other data mining techniques can be applied for refined results. Sample size can be taken more in order to gather more response which may improve the results.

REFERENCES

- [1]. Bejju, "Sales analysis of e-commerce websites using data mining techniques", *International Journal of Computer Applications*, Vol. 133, No. 5, pp. 36-40, 2016.
- [2]. B. M. Ramageri, "Data mining techniques and application", *Indian Journal of Computer Science and Engineering*, Vol. 1, No. 4, pp. 301-305, 2014.
- [3]. B. Pahwa, S. Taruna and N. kasliwal, "Role of data mining in analysing consumer's online buying behaviour", *International Journal of Business and Management Invention*, Vol. 6, No. 11, pp. 45-51, 2017.
- [4]. D. Kalaivani, P. Sumathi and T. Arunkumar, "Customer behaviour predictive modelling for online shopping using tuned decision tree method", *Helix the Scientific Explorer*, Vol. 7, No. 5, pp. 1994-1999, 2017.
- [5]. J. Arora, "Prospect of e-retailing in India, the shopper's discernment in regards to shopping has been changed with the presentation of web media", *Journal of Computer Engineering*, Vol. 10, No. 3, pp. 11-15, 2013.
- [6]. J. K. Gerrikagoitia, I. Castander, F. Rebon and A. A. Sorzabal, "New trends of intelligent e-marketing based on web mining for e-shops", *International Conference on Strategic Innovative Marketing*, pp. 75-83, 2015.
- [7]. K. Singh, "Retail sector in India: present scenario, emerging opportunities and challenges", *Journal of Business and Management*, Vol. 16, No. 4, pp. 72-81, 2014.
- [8]. O. furrer and D. Sudharshan, "Internet marketing research: opportunities and problem", *Qualitative Market Research: International Journal*, Vol. 4, No. 3, pp. 123-129, 2001.
- [9]. Qinghai, C. Wenyuan and L. Kaiming, "The online shopping change the retail business model: a survey of the people use online shopping in china", *Journal of Business and Management*, Vol. 15, No. 5, pp. 77-110, 2014.
- [10]. Sana, "A study on the impact of online shopping upon retail trade business", *Journal of Business and Management*, pp. 74-78, 2015.
- [11]. S. K. Dawn and U. Kar, "E-tailing in India: its issues opportunities and effective strategies for growth and development", *International Journal of Multidisciplinary Research*, Vol. 1, No. 3, pp. 101-115, 2011.
- [12]. Yassir and S. Nayak, "Issues in data mining and information retrieval", *International Journal of Computer Science & Communication Networks*, Vol. 2, No. 1, pp. 93-98, 2014.