A Review on Consumer Behavior Analysis using Association Rule Mining for Business Intelligence Applications

Ms. Kareena, ME Student, University Institute of Engineering (CSE), Chandigarh University Gharuan, Mohali, Punjab, India, kareenasardana3993@gmail.com

Er. Raj Kumar, Assistant Professor, University Institute of Engineering (CSE), Chandigarh University Gharuan, Mohali, Punjab, India, rajshira@gmail.com

Abstract — Data mining is the technique used for finding out information from large amount of data. Now a day's business mostly rely on consumer behavior and data mining is the most important way to get information about the consumer that what they want from companies. The data mining process that we carry out from finding out information to have some knowledge provide many benefits and due to the boom in technology sector, large amount of data is created every moment so this made the organizations to deal with new challenges, allowing them to understand the information beyond system. So basically data mining has emerged to help to understand the massive amount of content available and to produce the summary of the information in useful way so that the information further can be used effectively by business analyst and information technology professionals. In this paper Association rule mining is explained to see that how it can be made useful to predict consumer behavior and other related applications.

Keywords — Data Mining, Consumer Behavior, Business Intelligence, Association Rule, Mining.

I. INTRODUCTION TO DATA MINING

Availability of large amount of data has created lot of trouble for companies when they need to find out any kind of information from available data and it is no use until it is converted into useful information so that further it can be utilized for effective use. It is essential to examine this vast amount of data and gather useful information from it. Gathering information is not the only process that we need to or we should perform rather it also includes other processes such as removing redundancy, storing at one place, transforming data so that it become consistent, pattern evaluation and at the end presenting the data in such a way so that it can be understood by user. Data mining has emerged due to many reasons like creation of data at large amount and also not having proper technology to examine the data and to make decisions in effective and practical way. This is the reason why we are starving for knowledge and drowning in data, the solution to this problem is data warehouse and data mining. We have data everywhere created every moment we can't find, can't get, can't understand, can't use that available data but with the use of data mining we can have useful information and we can perform all these tasks very easily and we can use it for our own respective purposes.

- Data: Lowest level of abstraction.

- Information: Data that is ordered in some proper format and that data is valuable and understandable.
- Knowledge: The concept of understanding information based on recognized patterns in a way that provides insight to information.

Steps involved in data mining:

- Data Integration: This is the initial phase of whole process in which we collect data from various sources into a single data source with the help of some technology.
- Data Selection: Main focus in step is on that data which is required for our research/hypothesis or on that data samples on which discovery is to be performed.
- Data Cleaning: When we obtain data from different sources it may vary in formats so to obtain consistent or homogeneous data cleaning is performed using some cleaning algorithms.
- Data Transformation: Data is pre-processed into standard format using dimensionality reduction or transformation methods.
- Data Mining: In this process first we identify the data mining algorithm and then we apply it on selected data to obtain patterns.
- Pattern Evaluation: Then the pattern that we obtain from data mining is interpreted and

evaluated to have some kind of knowledge out of it.

- Knowledge Representation: This is the main aim of data mining to some knowledge which is taken into consideration so that business decisions can be taken.
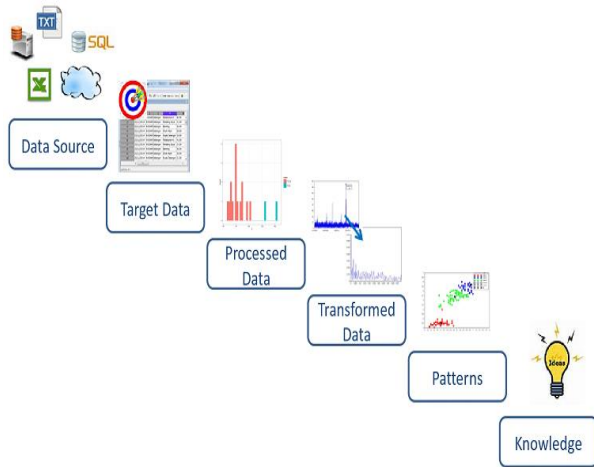
Fig. 1. Phases of Data Mining Process

Classification of data mining system:

- On the basis of database mined: It means that on the basis of type of data and database models from which we are trying to extract data. Like data can be categorical and numerical same way database models can be relational, transactional and object relational.
- On the basis of kind of knowledge mined: It means that what actually we are doing with data, what kind of knowledge we are getting from it. For example we can do prediction, association, correlation analysis.
- On the basis of technique utilized: Technique utilized depend on the kind of knowledge that we are mining like if we are doing association then we will use techniques that are used for association.
- On the basis of application adapted: It depends on the purpose for which we are using data mining like finance, telecommunication, stock market, healthcare.

Applications of data mining:

- Computer hardware and software
- Science and engineering
- Government and science
- Travel industry
- Healthcare and medicine
- Entertainment industry
- Sports
- Business
- Education
- Smart appliances

II. BUSINESS INTELLIGENCE

Business intelligence describes the way and methods so that the organizations can lead in such a hectic environment when there is massive amount of data is created, above all they have to solve their problems and create more opportunities. So it refers to the use of latest technology and applications which are used to gather data and analyse it for the betterment of their company and business purposes. This can be done with the help of data mining technique. The main thing that companies want is that they have an effective predictive method for this the need historical data and current data for long term stability and competitive market advantage.

Common task of business intelligence:

- Data mining
- Predictive analysis
- Prescriptive analysis
- Text mining
- Online analytical processing
- Business performance management
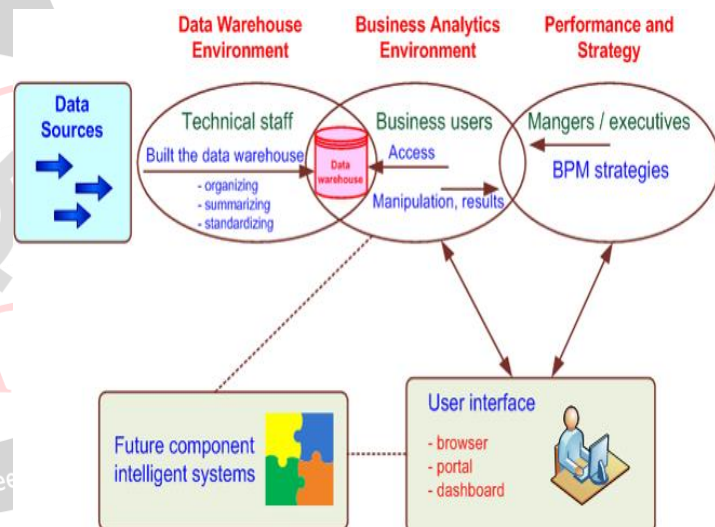
Fig. 2. Architecture of Business Intelligence

Business intelligence consists of four major parts:

- Data warehouse: It is the repository to store information from different sources of data and it transform into such a model so that it can be analysed easily. It can contain summary data, meta data, raw data.
- Business analytics: Business analytics contains the tools to manipulate the data warehouse in order to gain business insight and pre-plan the business to improve business performance.
- Business performance management: It is basically for measuring the performance of business which is done by business managers further make strategies and executes them.
- User interface: Through browser, portal or dashboard user interacts with computer system. It

is used for communication of user and system.

The main part of business intelligence consists of decision support system. The main roles of decision support system are:

- Analysis
- Prediction
- Decision

Whenever we talk of business we always have some kind of goals but it is not easy to achieve them because of technology that we use or that we have so it basically helps us to achieve our goals.

III. ASSOCIATION RULE MINING

Association rule mining is used for finding relationship among items. It is based on the theory that if you will buy a certain group of items then you are more or less likely to buy another item. Company mainly wants to see that how much their business is growing and this can be done with the help of algorithms. This has many applications in business world today. It can be used in various sectors like education, healthcare, market basket analysis etc. Market basket analysis is a technique used by retailers to examine the consumer behavior. It has wide variety of application in social media also like if you want to find out how consumer behaves in a particular situation. This thing is done with the help of association rule. For example if customer buy frozen pizza then he is more or less likely to buy ketchup. This thing is done by market analyst; market analyst thought that what are the items that are frequently bought by his customers.

Association rule:

- IF [BODY] THEN [HEAD]

Measures of rule quality:

- Item set: It state the manner in which products are purchased together by the consumer
- Support: It tells that how often A and B go together
- Confidence: It tells how often A is purchased when B is purchased.

For example if we 100 customers in a store and 20 of them bought bread, who bought bread bought butter also and they were 5.

Then

- confidence will be 5/20 = 25%
- Support will be 5/100 = 5%

Popular algorithm that [6] comes under association rule mining is APRIORI algorithm. Features of APRIORI algorithm:

- This algorithm employs unsupervised learning.
- This algorithm uses bottom-up approach.
- In this algorithm item subsets increases from one item-set to two item-set, from two item-set to three item-set and so on. Further in those item-sets some are selected based on support count. For example if the support count is 3 then the subsets which has

support count less than 3 will be eliminated and the subsets which has support count equal to or more than 3 will be selected.

For example: Min support= 60%

Support= 60*5/100=3

Min confidence= 80%

Transaction ID	Itemset
1	M,O,N,K,E,Y
2	D,O,N,K,E,Y
3	M,A,K,E
4	M,U,C,K,Y
5	C,O,O,K,E

Itemset	Support count
M	3
O	4
N	2
E	4
Y	3
D	1
A	1
U	1
C	2
K	5

Itemset	Support count
M	3
O	4
E	4
Y	3
K	5

Itemset	Support count
MO	1
ME	2
MY	2
MK	3
OE	3
OY	2
OK	3
EY	2
EK	4
KY	3

Itemset	Support count
M,K	3
O,K	3
O,E	3
K,E	4
K,Y	3

Itemset	Support count
M,K,O	1
M,K,E	2
M,K,Y	2
O,K,E	3
O,K,Y	2

Association rule	Support	Confidence	Confidence%
O and K gives E	3	3/3=1	100%
O and E gives K	3	3/3=1	100%
K and E gives O	3	3/4=.75	75%

Fig. 3.Example of apriori algorithm

Applications of Association Mining:

- **Consumer segmentation** – It help us to make clusters of consumers on the basis of their interest and need.
- **Finding consumer needs** – It help us to examine the needs of consumer so that further they can have new ideas of attracting consumers and build new strategies
- **Analyzing the market basket** – In this it find out the relation between products so that they can be placed together to improve   business.
- **Finding pattern** – It help us to find out the pattern among consumers so that further business plan can be made out.
- **Examining consumer buying behavior**- It examines the buying behavior of consumer so that company can send them different offers according to their buying behavior.
- **Summary of consumers and company** – Overall

It helps to maintain the record of both company and consumer.

IV. LITERATURE REVIEW

In order to understand the behaviour of consumer it is very necessary for the e-commerce companies to improve their user friendly experience. Web access log is one the source used for the extraction of consumer behaviour. Many research works have been proposed on the web log analysis

which benefits the e-commerce industries. But these logs are very big in size.

Ru Jia, et.al (2017) proposed [1] a Bayesian classification technique which involved the clicking behaviour features to predict the purchasing strategy of users. There is no need to include rating data in this proposed approach unlike the traditional purchase prediction techniques. Further, the Collaborative Filtering algorithm also does not face issues such as sparse data within the proposed approach. Due to its simplicity and high speed, the Naïve Bayes rule is included here through which the user clicking behaviour features are applied to perform classification. By making comparisons against the User-Based Collaborative Filtering algorithm, the performance of classification of proposed method is evaluated in real-time applications. The effectiveness of proposed technique is shown by the results achieved at the end. Thus, several ecommerce purchase prediction applications can further implement this approach.

M. R. Zakiullah, et.al (2017) proposed [2] a novel mechanism known as Command Line Interface for Consumer Behaviour for eliminating the issue of not being able to access Hadoop tools without using proper tools. The implementation of proposed model is easy and highly economical. Irrespective of the presence of any other tools or the type of Hadoop platform involved, this approach can be applied. The consumer behaviour of small scale businesses or huge enterprises can be known with the help of this approach. Therefore, the revenue can be increased and the global competition of ecommerce applications can be won. A virtual environment is used to implement the proposed model. A usual web server is the source from which the data is accessed since the confidentiality of e-commerce data is very high. For consumer identification, the IP or Hostname is used which is replaced by username here. With the help of using data from real e-commerce sire, the new patterns can be extracted by applying this proposed model.

Liping Zhang, et.al (2017) presented [3] that there is huge increase in the diversity of electricity consumption behaviour in different regions of China as per the increase in growth of its economy. The data is partitioned into relevant clusters traditionally, using the K-means algorithm. However, this algorithm categorizes the data into the best clusters that already exist even through there is not much correlation found amongst that data. Thus, for choosing the outliers from each cluster, DBSCAN algorithm is applied in this proposed technique. The cluster-self and sub-cluster are created by dividing the clusters of K-means. The electricity behaviour analysis of the proposed approach is highly detailed which results in providing better affects. On the basis of complementation of K-means and DBSCAN, the power consumer behaviour is analysed by considering the properties of electricity behaviour. Improvement is seen within the achieved results as per the experiments conducted by implementing combined model. Highly similar clustering can be achieved for the behavioural properties by implementing the new model since new clusters are generated after selecting the outliers of each cluster.

Marina E. Tsoy, et.al (2016) proposed [4] a novel approach on the basis of application of RFM-analysis along with classification techniques to analyse the consumer's behaviour towards high-tech products. The analytical method of ABC-analysis is applied instead of standard partitioning of the customer base into quintiles. It is seen through the experimental results that for increasing the sales and profits of client groups, the RFM analysis can be used as an efficient segmentation tool. Three target groups of consumers are identified by adapting the improved RFM analysis on the wholesale and retail trade of high-tech products in this paper. The enhancement of marketing and promotion policy is recommended here. Further, this work can be extended in future by focusing on the particular product categories or markets for growing its versatility.

Shao Fan, et.al (2018) proposed [5] a novel model through which the behaviour of consumers' businesses can be analysed on the basis of Hadoop platform. The large distributed properties of E-commerce transaction data are considered to be the base of this proposed model. The demand of E-commerce data processing was viewed through the analysis of consumer behaviour through this paper. For achieving the distributed storage of data, Hadoop Distributed File System (HDFS) was applied. There was a rational distribution of file to each node and the system was seen to run normally without the loss of data. For achieving the parallel storage and data processing, the MapReduce model was applied. It became more obvious to achieve advantages for the data processing of clusters. The E-commerce data is processed effectively using the Hadoop cluster tool.

Made Ayu, et.al (2017) presented [6] that varieties of hedonic consumer behaviours can be achieved as advantages of using e-commerce within various fields. E-impulse buying is one of such behaviours. The impulse buying is of great importance it generates a profit of about 40% to 80% of all the purchased products. With the growing development of online shopping, buying of products on internet is increasing day by day. The objective of this study is to analyse three components; individual, product attributes, and web browsing activities as antecedents of consumer's impulse buying behaviour. The simulations were performed on the SmartPLS 3.2 in which the 124 respondents were analysed. Therefore, the results of

this research concluded that the important factor influencing the consumer's behaviour like shopping enjoyment tendency, impulsive buying tendency and most importantly their personality. But the attributes of products do not have any important effect on impulse buying behaviour. Moreover, web browsing activities show that the web browsing having great possibility to make the purchase real and leading to impulsive buying.

V.   CONCLUSION

  Data mining has wide variety of applications in various fields like the it has for predicting consumer behaviour for business intelligence but effective technique should be needed for predicting it because such algorithms cannot determine human sentiments, likes, dislikes etc and every human has different perspective towards everything. In this review paper association rule mining is used to predict the consumer behavior but it has some drawbacks like First is the complex candidate generation process which uses most of the time, space and memory. Another drawback is that it requires multiple scans of the database. So many other techniques like Naïve Bayes, SVM and Decision Trees etc can be used to predict consumer behavior because they can have high accuracy in predicting using different kind of data sets.

REFERENCES

[1] Ru Jia, Ru Li, Meiju Yu, Shanshan Wang, "E-commerce Purchase Prediction Approach By User Behavior Data", International Conference on Computer, Information and Telecommunication Systems (CITS), 2017

[2] M. R. Zakiullah, M. S. Islam, "Design of a Low Cost and Convenient Hadoop Application For Extracting Consumer Behavior", 20th International Conference of Computer and Information Technology (ICCIT), 2017

[3] Liping Zhang, Song Deng, Shiyue Li, "Analysis of power consumer behavior based on the complementation of K-means and DBSCAN", IEEE Conference on Energy Internet and Energy System Integration (EI2), 2017

[4] Marina E. Tsoy, Vladislav Yu. Shchekoldin, "RFM-analysis as a Tool for Segmentation of High-tech Products' Consumers", 13th International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE), 2016

[5] Shao Fan, Yao Jun, "The Establishment of Data Analysis Model about E-commerce's Behavior Based on Hadoop platform", International Conference on Intelligent Transportation, Big Data & Smart City, 2018

[6] Made Ayu, Aristyana Dewi, Isnaeni Nurrohmah, Nitto Sahadi, Dana Indra Sensuse, Handrie Noprison, "Analysing the Critical Factors Influencing Consumers' E-Impulse Buying Behavior", International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2017