Customer Behavior Analysis Towards Online Shopping using Data Mining

Conference Paper · April 2018

DOI: 10.1109/IMTIC.2018.8467262

2 authors:

Muqaddas Gull

8 PUBLICATIONS **42** CITATIONS

SEE PROFILE

Arshi Pervaiz

University of Lahore

1 PUBLICATION **10** CITATIONS

SEE PROFILE

Customer Behavior Analysis Towards Online Shopping using Data Mining

Muqaddas Gull and Arshi Pervaiz
Department of Computer Science
The University of Lahore,
Sargodha, Pakistan
Muqaddasgull@yahoo.com, Aries_42008@yahoo.com

Abstract—**Study of customer behavior in online shopping usually deals with identification of customers and their buying behavior patterns. The aim of such studies is to make certain who buys where, what, when and how. The results of these studies are useful in the solution of marketing problems. Various studies on customer purchasing behaviors have been presented and used in real problems. For analysis of customer behaviors data mining techniques are consider more effective. The target of this paper is to analyze behavior of such people who are visiting the online shopping sites and spending their time there, surfing for different stuff. It would also be taken into account that how many people are there and how many of them are actually shopping. In this paper, different queries are applied to mine the database of a specified site which results in analysis of customer behavior towards online shopping.**

Keywords—*data mining; market strategies, association rule; frequent item set; priori algorithm*

I. INTRODUCTION

As human behavior is hard to predict but it is usually expected to know anyone by the way we get in contact with them. Humans communicate through many different ways. These ways can be face to face meeting, calling on phone or by sending an email or a text message. Sometimes people do show to others how they are feeling by merely not answering in with words rather showing just gestures. Behavior of human being keep changing constantly and it's hard for anyone to always remain in the same mood [1]. It is commonly known remedy that people who are sad and getting bored usually go for shopping. So, shopping is greatly considered a good idea for changing moods.

Now days with a rapid increase in E-commerce, people do not need to go somewhere for shopping. A person usually at home is spending hours and hours on just checking out the stuff for sale over the internet, even on which he/she is not interested to buy [2]. Such shopping habits of the customer are fabulously taking the online shopping industry to much higher profitable place.

A. Customer Behavior

Buyer conduct is the investigation of people, gatherings or associations about their procedure of choosing, securing, utilizing and arranging the items, administrations, encounters or thoughts to fulfill needs and the effect of this procedure on the shopper and the general public [3]. A customer can behave in many different ways while shopping. It can never be expected that every type of customer is same in personality, decision making and intellectually. A buyer can be affected by personal, psychological and social and cultural factors. There are numerous other factors that influence customer shopping habits. Few of these can be:

- Customer's financial conditions
- Customer's desire to buy
- Customer's need

Along with these factors there also exists a fear by the customer to share his/her personal data online. Many people find it unsafe and unsecure. Despite these behavior issues there are millions and millions of people who are shopping everyday online. The target of this paper is to analyze behavior of such people who are visiting the online shopping sites and spending their time there, checking different stuff. It would also be taken into account that how many people are there and how many of them are actually shopping.

B. Customer behavior influencing market strategies

The market now a day is getting very competitive. Organizations are spending lot of money in planning to sell their products in the most effective way. This has tested advertisers to dissect changing customer shopping conduct and novel choice making forms inside another, quickly creating medium. Web based showcasing research proposes that the Web is changing the path in which individuals utilize shopping channels to buy items [4].

Marketing strategy can be described as the combination of all most suited features of cost, product, distribution and sales that can attract the customer. It is the way of identifying some unique features in the grouping of products in such a way that the product is considered best sold [5]. We often encounter in our daily life with various selling deals. It usually happens that the things that are less sold are given in a discount rate or free with the most selling items. Like a bed linen is the most sold item and floor cushions are not sold much. The company plans to give one cushion on half prize to people who are buying bed linen. In such away customer gets more attracted towards the deals and even the company gets benefit from the unsold stock

978-1-5386-6594-7/18/$31.00 ©2018 IEEE

C. Data mining and customer behavior

The Data mining framework is helpful to Business house to discover the relationship of the clients with various items. Also, how clients are switching from one brand then onto the next in order to fulfill their need on the grounds that their prior purchasing propensities are legitimately contemplated by the Data Mining System [3]. Data mining (DM) is a process of discovering useful patterns of data for identifying current and predicting future trends. Almost all of the operational data of an online system is stored in the data warehouses for further use. Such data is then analyzed either regularly (large organizations) or whenever there is a probability that in the coming days, sales would likely be growing. Through DM it can be analyzed how different products are contributing in the total sales and how they are affecting the shopping habits of customers.

II. LITERATURE REVIEW

The online shopping has become the regular part of today's world. The factors associated with this shopping trend are Convenience, Better Product Selection and Useful Delivery Mode. There exist many pros and cons of online shopping and through their study the present and future trends of online shopping can be analyzed [2].

Data mining is also an efficient field to analyze consumer behavior in online shopping. The buying patterns of customer are somehow interlinked. Association rule mining helps identifying such patterns and helping the business decision making process [3].

The behavior of customers shopping over the internet has been analyzed through five factors. These factors are time, privacy, trust, convenience and product variety. The research conducted for this purpose was in the form of questionnaire and the results were analyzed statistically. Trust was claimed as such a human trait that affects their buying habits [6].

The shopping behavior of people in Pakistan gets affected by their psychological and emotional attitudes. Privacy is also considered as the most prominent factor in this online shopping trend in the way that people may sometimes not feel secure sharing their personal information over the internet. On the other hand, fascinating prices of various stuff may also attract individual attention and aid to urge online shopping. So, trust over the source is the issue that affects people buying behaviors [7].

To identify consumer behavior a research has been conducted on 7-Eleven convenience store. The factors considered in this regard are cultural, social, personal and psychological factor. After applying multiple regression analysis and hypothesis testing the coefficient of determination (R2) is derived that describes the influence of all the factors that affects consumer inclination to buy the products. All these factors are also independently discussed and analyzed statistically [8].

The online reviews help the customers establishing an opinion regarding online shopping. These reviews vary in both quality and quantity. They can have both positive and negative kind of effect on customers and as well as on business. The data in this regard is collected via questionnaire and the results are compiled after various complicated statistical methods. These results interpret that reviews do contribute to decision making in online shopping [9].

There have been various techniques of data mining for the identification of frequent item sets. As the data retrieved after processing is very large and requires some efficient technique to discover some useful pattern. The paper discusses those techniques that can aid in the formation of any such pattern. Association rule has been considered as one of the basic data mining tools. There exist many algorithms like Apriori algorithm, AIS, SETM, Apriorihybrid, FP-growth for pattern discovery. Apart from the pros and cons of these algorithms, any of these can be used along with association rule mining for data analysis [12].

III. METHODOLOGY

In this paper, to exemplify our research work we used an online shopping home decoration website's database as a sample database. The site is about multiple items like Duvet covers, Bedspread, Curtains, Cushions, pillows and bathroom and kitchen accessories. The site contains thousands of records that can give a little insight about customer purchasing habits in home decoration. The reason for choosing this database was that, it is a working site and has so many numbers of customers visiting every day that provides a huge dataset to analyze customer behavior as Fig. 1 presents a methodology of our work.

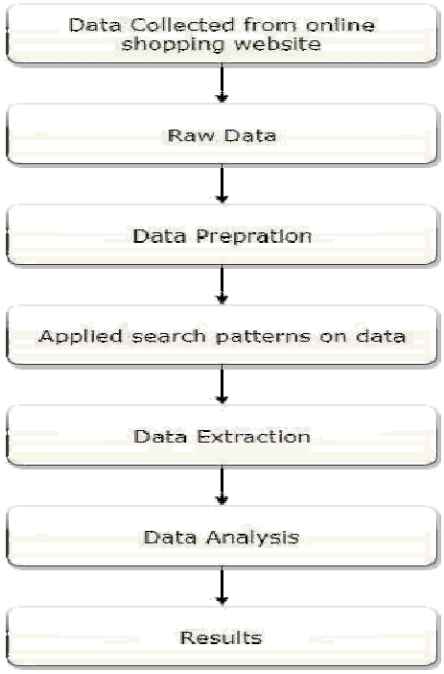

Fig. 1. Flow Chart of Methodology

To extract some useful pattern, it is here to place some questions that would be answered by applying queries using SQLyog (a complete MySQL management tool).

The questions to analyze shopping behavior are:

1) How many people are paying in the home decoration and how much money they are spending in this regard??
2) What is the most sold product?
3) What is the least sold product?
4) Which majority of cities are contributing to sales?

After gathering such information, we select the reasonable strategy from several different choices. For analyzing customer shopping behavior there exist many old techniques such as conducting surveys through questionnaires and polling. Even various algorithms have been applied on online data for this kind of information. We here by apply association rule for checking the association between products which are bought by the customers. The basic idea here is to provide a mechanism of creating a market strategy out of some information retrieved from online purchasing.

A. Data Analysis

As described earlier the sample database chosen in this regard has thousands of records saved. The result for questions raised is in thousands and are as follow:

1) How many people are paying in the home decoration and much money they are spending in this regard?

All the customers who did shopping over this site support the answer i.e. total customers, representing number of people that are spending over home decoration and total sales, representing the amount of money they are spending in this regard.

TABLE I.

Total Customers	Total Sales
26586	$1810044.98

2) What is the most sold product?

TABLE II.

Qty	Product Name	Product Price	Tax Applied	Discount	Price (Including Tax)
826	Stag Natural Duvet Cover Set - King Size	18.32	3.67	0.00	21.99

The product named “Stag Natural Duvet Cover Set – King Size” here extracted from the database is the most sold product. With efficient query processing most sold product can be retrieved, which can further serve as a base in offering many purchase deals.

3) What is the least sold product?

TABLE III.

Qty	Product Name	Product Price	Tax Applied	Discount	Price (Including Tax)
2	Love Grey - Bed in a Bag: Double	14.30	2.86	0.00	17.16

The product named “Love Grey- Bed in a Bag: Double” here extracted from the database is the least sold product. This type of information can help in introducing either discounted deals or offering the product in combination with any other product.

4) Which majority of cities are contributing to sales?

Cities like London, Leeds and Bradford are mostly contributing in the sales.

The name of the cities is extracted via query processing over the database of the related site. Similar information can be extracted to discover more useful information that can further boost the sales process and help in analyzing the overall customer behavior in shopping.

The answers to such questions can help extracting some patters like how most unsold item can be added in the deal with the most sold item or if people are spending much money in the home décor every now and then, then price of such items can either be increased or decreased. So many such profitable business ideas can be mined in this way.

Although many such data mining techniques exist that can be applied over huge datasets for patterns extraction. The idea behind implementing Apriori algorithm of association rule mining is that Apriori algorithm has least memory consumption. It has easier implementation and it filters out the undesirable patterns in an easy and most efficient manner [9]. Data mining techniques are being widely used in the field of risk and market management, telecommunication networks, inventory control and medical diagnosis.

B. Association rule of mining

Association data mining finds fascinating relationship connections among the vast information, the analyst is turned out to be more intrigued by mining association rules from database. The disclosure of fascinating relationship among tremendous measures of client exchange records could strengthen decision power [10].

Association rule mining is also sometimes termed as market basket analysis. It finds the relationship between different kinds of product in the “shopping basket” the customer places while shopping. [3] It is usually witnessed that the kind of things that are used in combination are mostly offered together. For example, in our sample data, in the home decoration site, all such household item that contribute in home decoration are offered, whether these products belong to bedroom, kitchen or bathroom. A customer when visit the site for any of the item like curtain or bed linen, starts getting

interested in the kitchen stuff also and it's a close proximity that he may order more than one product.
To further elaborate the above discussion and explaining the association rule an example is derived from the data we possess in our database of customers, it can be quoted as: If a person is purchasing the Duvet then it's more likely that he will purchase pillow.

Association rule applied on output i.e. in this database no. of Duvet sold is 36286, no. of Pillow sold is 44 would be like:

TABLE IV.

Customers who bought this item	Customers who also bought this item
Sheet	Pillowcase
Sheet, Cushions	Pillowcase

The association rule will have the following form X→ Y that form has meaning that people who bought items of set x are often also bought items on set Y [3].

{Sheet} → {Pillowcase}

{Sheet, Cushions) → {Pillowcase}

The two basic rules of association are support and confidence. Support is the frequency of transaction to have all the items on both sets X and Y are bought together.

To calculate the support and confidence we here apply a well-known algorithm i.e. Apriori algorithm of association rule mining. For Apriori algorithm we need some item sets to extract a pattern of most frequent item sets.

Frequent Item set: An item set whose support is greater than or equal to a min_sup threshold. Frequent item sets can help support marketing decisions and provide a very useful pattern. In association rule mining task from a set of transactions T, the goal of association rule mining is to find all rules having Support >= min_sup threshold and Confidence>= min_conf threshold [11].

From the database, it is here by extracted some patterns to be considered for calculating frequent item sets. The patterns are represented alphabetically to help in our working like Sheets are denoted as A, Pillowcase as B, Mattress as C, Duvet as D, and Curtains as E and Cushions as F. As items have very complex ids we represent item sets by item numbers.

TABLE V.

Items	Item sets
01	A, B
02	A, C
03	D, E, F
04	A, D, B

It is here considered a threshold of (Support, 50% AND Confidence, 50%)

Min occurrence = 50/100*4 =2

TABLE VI.

Items	Support
A	3
B	2
C	1
D	2
E	1
F	1

Considering only those item sets that are equal to or above min. occurrence. L1

TABLE VII.

Items	Support
A	3
B	2
D	2

Again, forming sets from the above items i.e. C2

TABLE VIII.

Item set	Support
A, B	2
A, D	1
B, D	1

AB has the highest support so taking this as a final pattern.

TABLE IX.

Items	Support	Confidence
A → B	2	2/3 *100= 66%
B → A	2	2/2*100= 100%

As above mentioned, A is used for pillowcase and B is used for the mattress. We applied the Apriori algorithm to

extract the confidence and support. Two association set are remained in the end which are A→B and B→A. Both have the support 2 and confidence of A→B is 66%, while B→A have confidence of 100%.

IV. Results

After applying multiple queries, the results obtained are shown in the tables above. These tables are quite self-explanatory in the sense that they provide the appropriate answer to the 4 questions raised by us, for the purpose of research. The Apriori algorithm further support the answers in this regard providing a tip to marketing strategy in the analysis of customer behavior in a way that the products that are closely related together, in terms of use or offered in a deal together are more of a chance to be bought together. Like in the example above the sheets and pillow case provide 100% confidence of being purchased together.

V. CONCLUSION & FUTURE WORK

As it is discussed above target of this paper is to analyze behavior of such people who are visiting the online shopping sites and spending their time there, checking different stuff. We have taken the database of a running website related to the online shopping. Number of products and categories related to the product is present in it. We mined the data from database, ultimate tool used and queries are applied to extract the data. Association rule mining and Apriori algorithm is applied for the customer behavior analysis. Support and confidence is the result of the Apriori algorithm which is implemented on the association of different product, like A→B and B→A, support and confidence is gained from this association.

In future Advance Apriori algorithm can be applied to observe the customer behavior analysis through support and confidence of product mining association rule.

References

[1] J. H. a. P. Gendall, "Understanding and predicting human behaviour," Wellington , 2008.

[2] S. M. Prof. Sanjeev Kumar, "Status and Scope of Online Shopping: An Interactive Analysis through Literature Review," International Journal of Advance Research in Computer Science and Management Studies , vol. 2, no. 12, 2014.

[3] R. Abhijit Raorane, "Data mining techniques: a source for consumer behavior analysis," 2011.

[4] D. S. Sven F. Crone, "Predicting customer online shopping adoption an evaluation of data mining and market modelling approaches".

[5] J. C. O. J. Paul Peter, "Consumer behavior and marketing strategy, McGraw-Hill", 2005.

[6] I. M. W. K. B. Dr Rizwana Bashir, "Effects of online shopping trends on consumer-buying behavior: an emparical study of pakistan," Journal of Management and Research, vol. 2, no. 2, 20\15.

[7] A. T. A. S. H. u. R. I. J. Sajjad Nazir, "How online shopping is affecting consumers buying behavior in pakistan?," International Journal of Computer Science Issues, vol. 9, no. 3, 2012.

[8] P. Gianie Abdu, "Analysis of consumer behavior affecting consumer willingness to buy in 7-eleven convenience store," Universal Journal of Management, vol. 1, no. 2, pp. 69-75, 2013.

[9] G. Kaur, "Association Rule Mining: A Survey," International Journal of Computer Science and Information Technologies, 2014.

[10] S. J.-b. Z.-z. Ming-ye, "The impact of online reviews on consumers' purchase decisions in online shopping," in management science & engineering (ICMSE), Helsinki, Finland, 2014.

[11] P. Y. C.C Aggarwal, "Mining association with the collective strength approach," vol. Volume: 13, no. Issue: 6, 2001.

[12] C. Kaur, "Association rule mining using apriori algorithm: a survey," vol. 2, no. 6, 2013.

[13] P. S. D. Pooja R. Harne, "Mining of assosiation rule- a review paper," International Journal of Engineering Technology, Management and Applied Sciences, vol. 3, no .4, 2015