

Comportement Analyse And Recommendation Using ClickStream Data

Khushboo Kumari¹, Aishwarya Kumar², Surabhi Umare³, Gaurav Phadtare⁴,
Harsh Khajgiwale⁵, Prof. Vishal Chaudhary⁶

^{1,2,3,4} Student, Computer Engineering, International Institute of Information Technology,
Maharashtra, India

⁵ Student, Information Technology, International Institute of Information Technology, Maharashtra, India

⁶ Assistant Professor, Dept. of, Information Technology, International Institute of Information Technology,
Maharashtra, India

ABSTRACT

Now-a-days, e-commerce is very much trending. People prefer to shop online on Internet than to go on a retail shop and purchase the items. The shopkeeper in a general marketplace or retail shop can easily analyze the customers' behavior and predict if he/she will buy a product or no. But in online shopping system it is difficult to predict the behavior of the customer as he/she is absent physically for shopping. Hence, in this study we propose a model for the analysis of e-customers behavior using clickstreams and extraction of data to predict the shopping behavior of customers on digital market. This model will predict whether the customer will buy the product added to the shopping basket or not.

Keyword: Web Mining, Data Mining, Behavior Analysis, E-Customer, Clickstream, Customer Behavior.

1. INTRODUCTION

People purchase wide range of commodities ranging from variety of costs. Initially people used to purchase things from the shop by physically visiting the store. But now, there is complete paradigm shift in the way people used to purchase commodities. Almost all the commodities are now purchased by the consumer online. Commodities ranging from low cost monthly provision to the high cost electronic gadgets and equipment. The markets have remarkably migrated from the physical world to the digital space. With the lots of services offered by the ecommerce retailer to the consumer like cash-on-delivery, EMIs and much more, the consumer has developed a significant interest in migrating to the digital market. While purchasing any commodity from the market physically, the retailer becomes friendly with the consumer by developing relationship with him/her, providing handsome discounts on the future purchases to the consumers who are regular customer to that retailer and satisfying service to the consumer. Since people are remarkably migrating to e-commerce, to ensure smooth business and increased purchase of commodity from digital market, trust and reliance of the people should be established. Hence, to achieve this goal various techniques are used by the e-commerce giants to make sure that stable, healthy and good relationship with consumer should be maintained. Few of these techniques are data mining, text mining, web mining, behavior analysis etc.

Data mining can be roughly defined as the process that analyzes data from different dimensions and perspectives. The analyzed data is then summarized into useful information. This information is then used to discover patterns, identify relationships based on the attributes and take effective measures to increase profit, revenue and so on. In brief, data mining is the process where correlations and/or patterns are found from the huge amount of relational databases available. Most of the data that comes in the internet is one way or the other in the form of text. Hence,

the consumer's data over the e-commerce is mostly in the text format. To analyze the user based on the text that comprises of user's data, technique of text mining is used.

Text mining, as the name suggests can be understood as data mining from the text, is the process of high-quality information extraction from the input text. This information is used to get knowledge and then further using it for promoting and recommending particular commodities to particular users.

Also the text data is extracted from the websites, processed and analyzed to understand the user by learning about the user's interest as he crawls over different websites. This is typically defined as Web mining. This data obtained from websites may be content of any site or user events over the site like clicking of mouse and much more. Based on the above mentioned three techniques of gaining the information about the user, behavior of the user can be understood and learnt in a better fashion. Hence by using above three techniques, namely web mining, data mining, and text mining, e-commerce service providers can get familiar to the user and hence relationship with user can be established over the digital space.

2. LITERATURE REVIEW

Silahtaroglu and Dönertaşlı (2015) used KNIME, an eclipse based open source program for data mining which consists of artificial neural network and decision tree algorithms and some other open source data mining tools like R and WEKA. Different programs and algorithms are run using KNIME, on its own platform. Use of decision tree and artificial neural networks is made for extracting customer's shopping patterns from their mouse movements and web logs. However the conclusions drawn from the study presented by the authors cannot be considered as general guidelines to understand the behavior of online customers [1].

A very broad classification of web mining into three categories is done by Wei et al (2015) namely web content Mining, web structure mining and web usage mining. Also the stages involved in web mining namely source data collection, data pre-processing, pattern discovery and pattern analysis. The authors conclude that web mining plays a very significant role in E-commerce that has major influence over merchant, customer and company [2].

Jiang and Yu (2008) used data mining principles for clustering segments of customer with the help of K –Means algorithm and web log data from e-commerce websites. In business media, satisfying and increasing the expectations of customers simultaneously with decrease in costs is required. Additionally with fascinating new customers, great deal of resources are devoted to delight and retain the already existing customers for maintaining a long term and close relationship with e-customers. How to analyze the collected consumer information is of major interest to the marketers and analyzers. According to authors, traditional methods were mainly focused for predicting quantity of similar products rather than finding the relationship between the products purchased and customer. Authors created clusters of customers visiting their website based on different parameters like gender, age, time, etc. and based on these clusters they analyzed if the customer bought the product or not. The business strategy is totally changed by E-commerce [3].

Tamimi and Mohammadpourzarandi (2013) discussed broadly over the threats possessed by user in e-commerce. Few threats which include XSS and CSRF specifically are targeted towards application's users and web application are targeted by all other attacks. For detecting the attacks over e-commerce transaction authors explicitly mentioned two strategies namely Rule-based attack detection and anomaly based attack detection. Authors focused on web mining to unearth hidden relationships among various data and based on the various ways to obtain information, divided web mining broadly into three categories namely, web content mining, web structure mining and web usage mining [4].

Liu et al (2011) proposed a snippet click model for the task of query recommendation. User can easily describe their information needs through Query recommendation. On closely observing into the search behavior logs obtained

from a famous commercial search engine, authors concluded that 15.36% query sessions contained clicks based on the query recommendation links. Based on the review conducted by the authors, it is found that major focus is on locating the popular queries those are similar with the current queries either in content or in click context. Finally based on the analysis into user's interaction process with search engine, authors conclude that user's information need to be depicted in the clicked snippets [5].

Chen et al (2014) focused on web services for improving the accuracy of recommendation. Existing application can be transformed into Web-application using Web services. Web services are based on web applications that interact with other web applications for the purpose of exchanging data. Authors employed the use of location information and Quality-of-Service (QoS) to represent Web Services characteristics. The properties like response time, availability, throughput etc. is defined as QoS. Based on historical records, authors predicted QoS values and came out with the best recommendation for active users. It is found that effective than a single method result [6].

Dziczkowski and Wegrzyn-Wolska (2013) described a system for e-commerce client's behavior analysis so that the system can interact with web site and extract the information regarding the client's behavior and need. Authors used support vector machine, clustering methods and hybrid evolutionary algorithm for product characterization, market segmentation and forecasting the sales related problem respectively. Authors' proposed system is based on results obtained from both Web Usage Mining and Opinion Mining. Authors' tried to distinguish between the web users as well as the individuals using the same connection because the two different individuals can log in from the same machine or same individual can log in from two different machines. Required data for the analysis can be collected from Log Files, Cookies, Query Data, Registration Data, client side Data, Session Identification etc. It is concluded that as it is difficult to get accurate results only from analyzing the client's side data, the combined data from different sources considering all the aspects can improve the behavior analysis of user [7].

3. PROPOSED SYSTEM

Based on the review done from different available literatures, it is learnt that different modules like recommendation system, behavioral analysis, web mining, text mining etc. are suggested for improving the interaction of customer with e-commerce service provider. In our proposed system, we intend to create a system which will predict the customer's behavior on the basis of the clickstream data. Clickstream data is the data that is obtained from the number of mouse clicks that a user does while surfing through different products. On the basis of number of clicks over particular products, similar products will be recommended to the customer. However, if the customer thinks to buy the product, proceeds till the end but does not purchase the product. Hence, the product remains in the cart. Now in this context, excited offers will be provided to the customer that will attract the user to buy the product.

As mentioned above, the recommendations to the customers purchasing product/commodity will be given on the basis of clickstream data. Data includes log files, days (special days like weekends, festive days etc.), time (how much time spent on a product), likes/dislikes, products added to the cart etc. Now once the data is obtained from clickstreams, it will be preprocessed, analyzed and appropriate knowledge will be extracted. The wide variety of data will be extracted and the same will be stored into the database.

In the proposed system, method of Naïve Bayes classifier is used to classify the products on the grounds of above mentioned criteria like time, days, number of views, gender, etc. Once the data is classified into multiple classes based on different criteria, predictions will be made about the intention of customer towards purchasing the product. Predictions will determine whether customer will purchase the product or not.

For instance, customers surfs over the e-commerce website and clicks on a particular product A and product B. Suppose, customer clicks on product A for ten times and on product B five times. Since, we are generating the data from the clickstreams, product A will be awarded first priority as number of clicks over product A is greater than the number of clicks of product B. Since number of clicks over A are greater than B, the probability of A being sold is greater as compared to B. Hence, on the basis of above mentioned and calculated probabilities, recommendations

will be made to the customer for product A in the form of different offers like discount, free home delivery, exchange offers, try and buy, etc. which are appealing to user. Based on the above predictions, technique of association rule mining will be applied to recommend the similar products to the customer. Additionally, the proposed system will check whether the customer has purchased the product or left it in cart without purchasing and checked out.

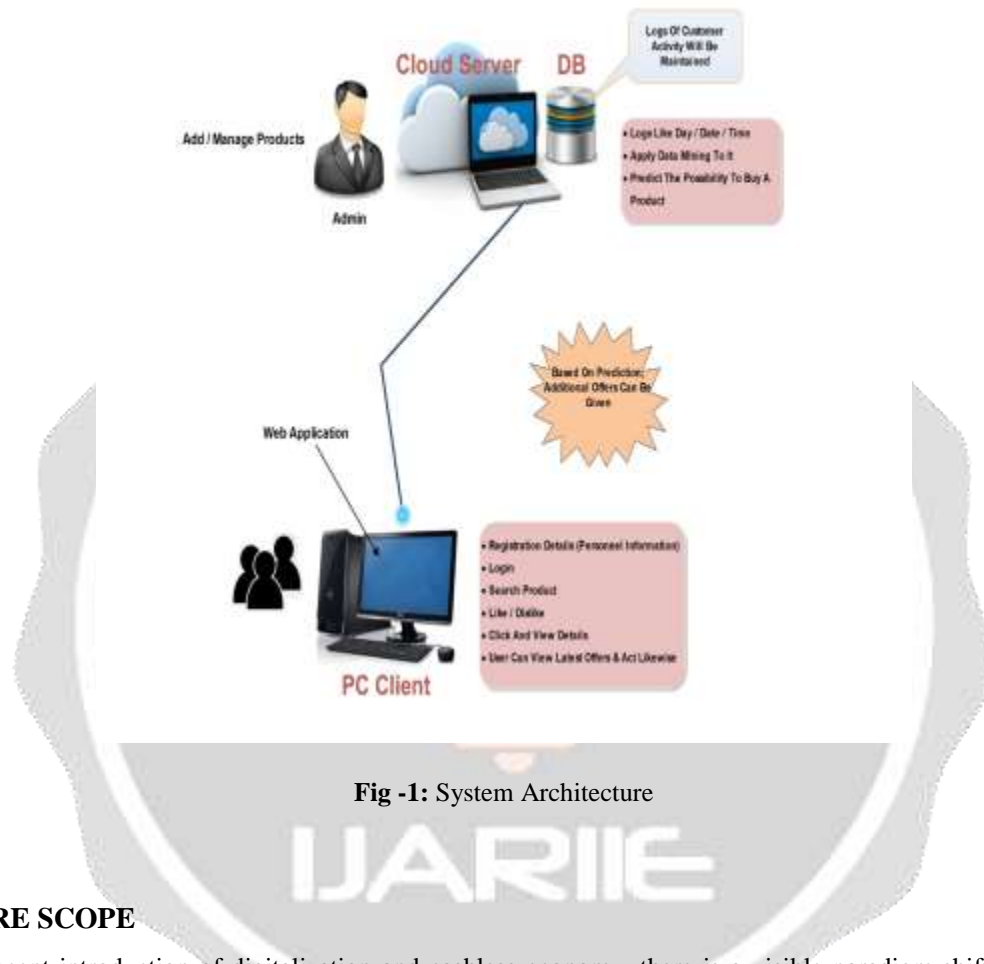


Fig -1: System Architecture

4. FUTURE SCOPE

With the recent introduction of digitalization and cashless economy, there is a visible paradigm shift in the ways people are purchasing products. There is a remarkable shift from the traditional purchase of products as people now prefer to purchase the products online rather than visiting the shops and purchasing. The proposed model can be extended to another dimension of machine learning techniques, where purchase patterns of different customers will be analyzed and accordingly the predictions will be made. Based on these predictions, recommendations to different customers purchasing the similar product will be made. This will eventually pave way to automation and also the customers will feel comfortable while purchasing the products from e-commerce platforms.

5. CONCLUSION

With the advent of technology and fast migration of business to digital platform, becoming familiar with the customer to meet their demands has become need of the hour. A wide range of literature survey is done to study and learn the technologies used till date to predict the behavior of the customer over the digital market. With customers changing their way of purchasing the commodity (from physical market to digital market), it is important to make customer feel comfortable over the internet and help him/her purchase the product by properly addressing the concerns and demands of the customer. This paper addresses the above mentioned aspects and makes the use of

Naïve Bayes classifier to predict the customer behavior and recommend different products as per his/her visit to the e-commerce website.

6. REFERENCES

- [1]. Silahtaroglu, G., & Dönertaşli, H. (2015, October). Analysis and prediction of E-customers' behavior by mining clickstream data. In *Big Data (Big Data), 2015 IEEE International Conference on* (pp. 1466-1472). IEEE.
- [2]. Wei, G. T., Kho, S., Husain, W., & Zainol, Z. A (2015, February). Study of Customer Behaviour Through Web Mining. In *Journal of Information Sciences and Computing Technologies*, 2(1), 103-107.
- [3]. Jiang, Y., & Yu, S., (2008). Mining the E-Commerce Data to Analyze the Target Customer Behavior, *Workshop on Knowledge Discovery and Data Mining*.
- [4]. Tamimi, R., & Mohammadpourzarandi, M. E. (2013, April). The application of web usage mining in E-commerce security. In *e-Commerce in Developing Countries: With Focus on e-Security (ECDC), 2013 7th International Conference on* (pp. 1-8). IEEE.
- [5]. Liu, Y., Miao, J., Zhang, M., Ma, S., & Ru, L. (2011). How do users describe their information need: Query recommendation based on snippet click model. *Expert Systems with Applications*, 38(11), 13847-13856.
- [6]. Chen, X., Zheng, Z., Yu, Q., & Lyu, M. R. (2014). Web service recommendation via exploiting location and QoS information. *IEEE transactions on parallel and distributed systems*, 25(7), 1913-1924.
- [7]. Dzikowski, G., Wegrzyn-Wolska, K., & Bougueroua, L. (2013, August). An opinion mining approach for web user identification and clients' behaviour analysis. In *Computational Aspects of Social Networks (CASoN), 2013 Fifth International Conference on* (pp. 79-84). IEEE.