IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-12 E-ISSN: 2347-2693

Analysis of Customer Behaviour using Modern Data Mining Techniques

S.J. Nasti[1*], M. Asgar[2], M.A. Butt [3]

[1*]Department of Computer Sciences, BGSB University, Rajouri, India
[2]School of Mathematical and Computer Sciences, BGSB University, Rajouri, India
[3] Department of Computer Sciences, University of Kashmir, Hazratbal, India

[*]*Corresponding Author: samiah.mushtaq14@gmail.com, Tel.: +91-8491027772*

Available online at: www.ijcseonline.org

Received: 17/Nov/2017, Revised: 29/Nov/2017, Accepted: 15/Dec/2017, Published: 31/Dec/2017

***Abstract*—** As we know enormous amount of data is present on the Internet and in order to get value out of this data and present the information to the user in a very simple form, researchers are working hard to collate this data.This colossal size of data on Internet is the most important source for decision making and marketing now-a-days. The paper presents a proposed model to understand online customer's buying behaviour based on decision tree and artificial neural network models.Our model is comparatively good at predicting the precision of customer's buying behaviour.

***Keywords*—** *Artificial Neural Network ,Buying Behaviour, Confusion Matrix, Data Mining, Decision Tree*

I. INTRODUCTION

An e-market is a place where customers can buy different kinds of products and services online with lower prices and this is the reason e-market is growing every other day. Today, online retailers hire analysts to build models in order to understand their customers in a good way. In traditional shopping, the behaviour of customers can be analysed by his gestures while as in online shopping the buying behavior of customers is predicted and analysed by clickstream data. With the proliferation of web usage, data recorded through clickstream is another source for information extraction. Indeed e-commerce is growing very fast.So analysing customers behavior is also gaining importance.Customer behavior reveals that if a user is likely to buy a product or not and is of great interest to online retailers [1].In order to predict the customer behavior with high accuracy, data mining is a desirable tool to transform the data into the information.Predicting customer behavior using data mining is not a magic, it actually uses neural networks and decision trees to build models that predict the customer behavior. In this chapter, we have presented a model to extract customer's buying behavior from weblogs and clickstream data. For this purpose, we use both decision tree and artificial neural network models.The paper is organized as follows:Section II contains the related work in the area, Section III presents the proposed approach,Section IV contains the results and Section V concludes the paper.

II. RELATED WORK

Ansari [2] proposed an integrated architecture for an e-commerce system with data mining.Their system can dramatically reduce the pre-processing, cleaning and data understanding, in knowledge discovery projects. Qing and his co-workers [3] proposed an architecture according to users needs and preferences by extending an e-shop into an intelligent e-marketing and selling platform.Their method uses two types of data mining techniques, namely classification and clustering. In e-commerce sites, to predict the users behaviour,a new approach was proposed by Vallamkondu and Gruenwald.In order to predict the purchase and traversal behaviour of future users, the proposed approach involves extracting of information from integrated data of past users [4]. Lee and his colleagues proposed a model for e-commerce known as integrated path traversal patterns (IPA) and association rules for web usage mining. This IPA model takes into consideration both the traversing and purchasing behaviour of the customers at the same time. This model not only take the traversal forward information of the user, but also takes into account the users backward traversal information, which makes the model accurate and correct for capturing users purchasing and traversal behaviours[5]. Satokar and Gawali presented a personalization system, which depends on features extracted from hyperlinks, for web search.Their personalization system which uses a weighted URL rank algorithm can help users not only to get relevant webpages, but also domains the user is interested in it [6]. Kiruthika and her co-workers [7] discusses the use of association rules in discovering patterns in web usage mining. They propose a system in which they preprocess the web server logs, cluster the closely related transactions, and discover the association rules. Their proposed system can help the website designers to improve their website design. Todi et al. [8] developed an application

which extracts information from e-commerce websites for classification of data in order to benefit both customers as well as companies. They use naïve bayes and decision trees, the two most popular supervised algorithms for classification and compared them.The results show that decision trees perform better than naïve bayes.Using such kind of applications, customers can understand the qualities of the available products and a competitor can understand how their competitors are priced.

III. PROPOSED MODEL

We have proposed a model to predict the behaviour of online customers in terms of buying and not buying by analysing the clickstreams of online customers.Initially clickstream data, demographic information about customers and the products they added to the shopping bag have been collected. This collected data from multiple sources is transformed and stored into a single place known as data cube. After that both decision tree as well as artificial neural network models are being used to extract customers buying behaviour patterns. The overall model is shown in figure 1.1.

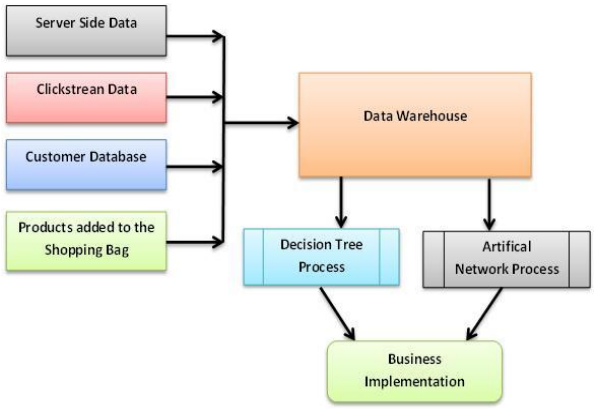

Figure 1.1 Proposed Model

a) The Data Set

The essential data attributes used in the study are briefly discussed below:

Time Stats: This variable is calculated in seconds and is the total time spent on the particular website.

Search Category: Products are categorized as Women , Men, Kids and Personal Care. Women being labelled as 1, Men labelled as 2 , Kids as 3 and Personal care as 4.

Day / Special Day: Day represents any day of the week and special day represents religious or any official day. If it is a special day, this variable is set to 1 otherwise it is set to 0.

Shopping Bag:This variable denotes the number of items customer had chosen for purchasing.

Product Add Time: This variable is calculated in seconds and denotes the first product added to the shopping bag. It is set to 0, if the shopping bag is empty.

Ordered Items: This variable is set to 0, if the Customer purchases the product as it is then removed from the shopping bag, otherwise its value is 1.

b) Decision Tree Analysis

Decision tree algorithms create if-else-then rules like tree structure when they run. We use bootstrapping technique, i.e. we combine C4.5, a statistical classifier and scalable parallelizable induction of decision trees (SPRINT) together. In order to determine the best attribute to arch from, we use C4.5 algorithm and SPRINT algorithm's Gini function to select the best attribute to create branches. For decision tree analysis, we use Konstanz Information Miner (KNIME) program.KNIME is an eclipse based open source program for data mining. KNIME is quite friendly to run different kinds of algorithms and programs on its own platform.Confusion matrix for decision tree analysis is listed in table 1.2 of results section.Confusion matrix is a table that is used to describe the performance of a classification model on a set of test data for which true values are known.Various measures with their formulas are listed in table 1.1 as shown below :

Table 1.1: Various Measures

S.No.	Measure	Formula
01	Accuracy	ACC = (True positive +) True Negative / (positive + Negative)
02	Precision	PPV = True positive / (True positive + False positive)
03	Recall	TPR = True positive / (True positive + False Negative)
04	Specificity	SPC = True Negative / (False positive + True Negative)
05	F1- Score	F1 = 2 True positive / (2 True positive + False positive + False Negative)

c) Artificial Neural Network Analysis

After decision tree analysis, we apply the Artificial neural network model as shown in figure 1.2. As can be seen from figure 1.2, data is fetched from the database, then normalization process is carried out for all data fields. After normalization partitioning is done in two parts, one being 70% and the other as 30%. Which means 70% of data is used for Artificial neural network analysis and remaining 30% for prediction and scoring. Every neural network has three types of layers: input, hidden and output. For learning a special type of artificial neural network, namely multilayer perceptron learner (MLP) having an input layer with ten nodes, ten hidden layers each having ten nodes and an output layer are used. After learning is complete, data is sent to decision tree predictor for the process of prediction. Scorer mode is used to create both confusion matrix and accuracy statistics for increasing readability.As artificial neural networks do not indicate rules like decision trees do. So in this context, we only present confusion matrix. ANN model can be used to predict whether a customer will buy at least

one item or will leave without buying. Table 1.3 shows that confusion matrix and accuracy statistics for neural network analysis with a 0.968 F-score, accuracy and precision values as 94.14% and 96.97% respectively, we can say that a neural network model works well for predicting whether a customer will buy at least one product.

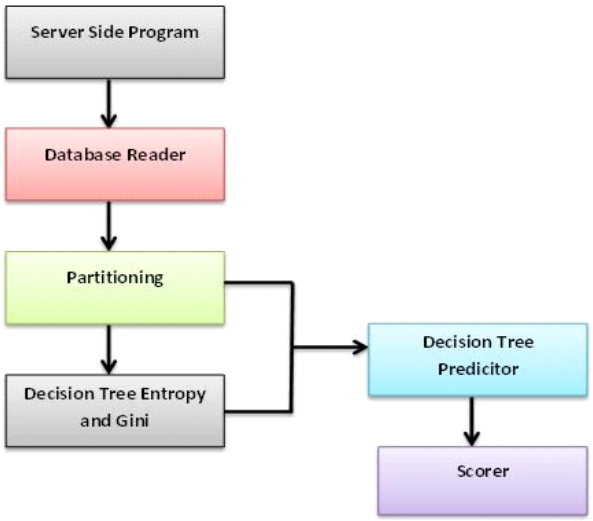

Figure 1.2: ANN Model for Analysis

IV. RESULTS AND DISCUSSION

The most vital rules summarized from decision tree analysis are as follows:

If it is a special day and if the customer is a woman she will buy the product. 115 customers draw this kind of pattern in 15 days.

If it is a special day and if the customer is a woman and is above 48 years, she will buy the product. 103 customers have followed this pattern in 20 shopping days.

If it is any day of the week and if the customer is a male and has at least two female products in his shopping bag, he will buy the product. 560 customers followed this pattern in 60 days while shopping.

Table 1.2 shows that the overall accuracy for prediction is 91.43% while as accuracy at predicting is 96.97% for a customer leaving without buying. In order to predict whether customer will pay and leave, then accuracy is 36.69% and F-score is 0.953 for predicting whether a customer will leave without paying. The rules generated from Decision Trees may be used for some business purposes.

Table 1.2: Confusion Matrix for Decision Tree Analysis

Left before paying/ Predection (Decision tree)	Yes	No	Sensitivity	Precision	Specificity	F1-Score
Yes	2660	83	96.97%	93.79%	36.69%	0.953
No	176	102	36.69%	55.13%	96.97%	0.440
Accuracy = 91.43% **Error Rate= 8.57%**						

Table 1.3: Confusion Matrix for Artificial Neural Network Analysis

Left before paying/ Predection (Decision tree)	Yes	No	Sensitivity	Precision	Specificity	F1-Score
Yes	2690	84	96.97%	96.65%	62.35%	0.968
No	93	154	62.35%	64.71%	96.97%	0.635
Accuracy = 94.14% **Error Rate= 5.86%**						

V. CONCLUSION

In this paper, we have presented a model for predicting the precision of customers buying behaviour. By predicting customers buying behaviour, we understand who are our best customers so that we can connect with them in a good way. Our proposed model predicts the behaviour of customer by using decision trees and artificial neural network models. The practical application of this model is that online retailers may use these rules generated from decision trees and can make instant discounts on the spot which in turn may dramatically increase their profit.

REFERENCES

[1] T. SH. Teo,*"To buy or not to buy online: adopters and non-adopters of online shopping in Singapore",* Behavior and Information Technology, Vol. 25(6) , pp.497-509,2006.

[2] S. L. Ansari, "*An integrated approach to control system design",* Accouting, Organizations and Society, Science Direct, Vol2 (2), pp. 101-112 , 1977.

[3] S. Qing, D. Gollmann, J. Zhou,*"Information and communications society",* 5th International Conference, ICICS Huhehaote, China, October, pp. 10-13, 2003.

[4] S. Vallamkondu and L. Gruenwald, "*Integrating purchase patterns and traversal patterns to predict HTTP requests in E-Commerce sites",* IEEE International Conference on E-Commerce, USA, pp. 256 – 263,2003.

[5] Y-S. Lee, S-J. Yen, G-H. Tu, M-C. Hsieh, *"Mining travelling and purchasing behaviors of customers in electronic commerce environment",* Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service,2004.

[6] K. D. Satokar , S. Z. Gawali,)"*Web search result personalizationusing web mining",* International Journal of Computer Applications (0975 – 8887) Vol. 2 (5), 2010.

[7] Kiruthika , R. Jadhav, D. Dixit, Rashmi, A. Nehete, T. Khodkar, T, "*Pattern Discovery Using Association Rules*", International Journal of Advanced Computer Science and Applications, Vol 2 (12), pp.69-74, 2011.

[8] A. Todi, A. Agrawal, A. Taparia, N. Lakhmani, R. Shettar,*"Classification Of E-Commerce Data Using Data Mining",*International Journal Of Engineering Science & Advanced Technology, Vol.2, (3), pp.550 – 554,2012.

