Analysis and Prediction of E-Customers' Behavior by Mining Clickstream Data

Gökhan SİLAHTAROĞLU
Istanbul Medipol University
MIS Department
Istanbul, Turkey
gsilahtaroglu@medipol.edu.tr

Hale DÖNERTAŞLI
Beykent University
Institute of Science and Engineering
Istanbul, Turkey
hale@google.com

Abstract- **In a regular retail shop the behavior of customers may yield a lot to the shop assistant. However, when it comes to online shopping it is not possible to see and analyze customer behavior such as facial mimics, products they check or touch etc. In this case, clickstreams or the mouse movements of e-customers may provide some hints about their buying behavior. In this study, we have presented a model to analyze clickstreams of e-customers and extract information and make predictions about their shopping behavior on a digital market place. After collecting data from an e-commerce market in Turkey, we performed a data mining application and extracted online customers' behavior patterns about buying or not. The model we present predicts whether customers will or will not buy their items added to shopping baskets on a digital market place. For the analysis, decision tree and multi-layer neural network prediction data mining models have been used. Findings have been discussed in the conclusion.**

Keywords: data mining, clickstream, e- customer, customer behavior, digital market.

I. Introduction

As a new channel of marketing the Internet is providing consumers with different kind of products and services. An electronic market place has the advantage of offering more choices, lower prices, easy search and access to online customers. That is why the internet market share is expanding every other day. So, consumers' behavior patterns are also gaining importance in terms of buying or not buying [1]. For an electronic market place owner or manager, the probability that customers will buy or not is crucial. There have been few researches about examining online customer behavior with modern data mining techniques rather than examining with existing customer behavior theories [2]. Using modern data mining methods such as artificial neural networks and decision tree algorithms we can examine online consumer behavior.

Data mining is a technology to extract useful information from large data sets. It is an application area of machine learning. Classification, clustering and association rule discovery are the main models of data mining [3]. In order to perform data mining application a special data set is needed. This data set must be built in accordance with the aim of the data mining application. This special data set is called data warehouse or data mart [4]. For example, to perform a supermarket data mining application, the data warehouse must be mostly converted from the supermarket's transactional database; if the mining will be about production then data warehouse should be converted from production databases [5]. However, these days, both users and businesses use web environment in their daily lives when they do business, shopping or just window shopping. So, web based data have gained importance recently.

Text mining is another leg of data mining; it is an extension of data mining to text data [6]. All three models of data mining may be applied to text mining as well. For text mining data consist of text or words. If this text data are converted or extracted from a website, this application is usually called as web mining. Data collected from a website may be content of the site or mouse movements of online users. So, web mining is focused on extracting useful, interesting and valid patterns or rules from unstructured documents such as HTML files, emails or chat messages [7]. With the increase of web usage, clickstream data have become another source for information extraction. Mouse movements of an online user may reveal information related to the behavior of the clicker. This study introduces a new model and an application on prediction of customer behavior using clickstream data. A corporate web site clickstream data have been converted into proposed clickstream data warehouse model and analyzed with data mining decision trees and artificial neural network models.

II. Motivation

Clickstream is a record of a user's activity on the internet; these mouse clicks a user makes when he or she is surfing may tell us a lot about the behavior of the user if it is analyzed in an appropriate way [8]. These movements are sort of the behavior of online customers [9]. We can name this analysis as web mining or web farming approach to discover patterns in the navigation of websites and web contents [10], [11]. By analyzing users' navigation patterns and their relation with web content one can redesign a website, portal or e-business along with the behavior of the online users [12]. There are a number of studies on collecting and analyzing web portal [13], web content, text mining and also clickstream data analysis in literature [14]. We all know that the quality of web portals is crucial to the web administrator and management. In addition to this, we also know that a proper web mining may yield useful results about the quality of a web page [15]. Issues like web site performance, its quality and online marketing intelligence may be carried out with data mining techniques [16]. Companies may gain a lot by analyzing the relation between customers and products if they establish a proper platform, especially a web based platform which is suitable for web mining. Both Clustering and classification models may be used for this purpose [17]. This kind of analysis may be held both for products and services. For example a study has been carried out for tourism sector by Schaefer et. al. [16]. The study introduces a new design for web mining system which combines web usage and content mining. As it is mentioned above web usage and clickstream data may reveal the behavior of the users. This is also true for other software usage if it is connected to the internet and a corporate server. So, beside customer behavior, a detailed customer profile may also be extracted through such an analysis [18]. Nevertheless, before starting a web farming, web mining or clickstream analysis, analysts need to build a model with a proper data warehouse. [19]. The data warehouse will be in the heart of a web mining model. A detailed survey has been studied by Wang et. al. The study covers some important works on data mining technology applied to e-commerce [20].

III. THE DATASET USED IN THE STUDY

For any kind of web mining a special data warehouse is needed and this data warehouse is to be a multidimensional [21]. In this study, we have used a server side program to collect clickstream data from the company's web server; at the same time, another java script program has been used to collect data from client side. Data attributes we have collected and used in the study are as follows:

Special day: If it is one week or earlier than an official or religious day such as Christmas, Independence Day etc. this parameter takes true value as 1, otherwise it is 0.

Day: represents the day of week: Sunday, Monday, and Tuesday etc.

Period of day: We have four periods for this variable; morning, afternoon, evening and after midnight. So, morning: 1; afternoon: 2; evening: 3 and after midnight: 4.

Time spent on the side: This variable includes total time spent on the site. It is calculated as seconds.

Search: If the customer searches a certain product on the site by entering keywords, this variable takes true value labelled with 1; otherwise it is labelled 0 (false) by default.

Category of search: Products have been categorized into 4. Skirt, jeans, shorts and pants are one group labelled as 1; shoes, boots, sandals are grouped and labelled as 2; dress, jacket cardigan, overcoat, sweater etc. are 3 and underwear products are 4.

Number of items in basket: It shows the number of different items in the online shopping basket. If the customer has two identical products this is counted as one. There must be at least one difference, such as color or size between two products, to count them two separate items.

Discounted Item in the Basket: The e-commerce company makes promotion campaigns or applies discounts. This variable identifies if the item in the basket is a promotional or discounted one. If there is at least one discounted item in the basket this variable takes 1(True), otherwise 0 (False).

Product category of the item in the basket: There are five categories in this field: Female, male, unisex, child (girl), child (boy).

Item add time: It shows the time (in seconds) of the first item added to basket. If the basket is empty it takes 0 value.

Amount of clicks: Number of the items clicked. Menu item clicks are not counted. We counted only the click made on products.

Click No: This shows the order of the click made by the customer. It takes values as 1st, 2nd, 3rd, etc.

Clicked item: Items are labelled as in category of search: Skirt, jeans, shorts and pants are one group labelled as 1; shoes, boots, sandals are grouped and labelled as 2; dress, jacket cardigan, overcoat, sweater etc. are 3 and underwear products are 4.

Click time: It shows the time (in seconds) when the product is clicked. For example if the user clicked his/her first item to examine on the 100th seconds on his/her visit, 100 is attained to this variable.

Source: It represents the source where e-customers come to the site from. This may be a search engine, another site or promotional mails sent by the company. If the customer is coming from a search engine it takes 1, if s/he is coming from a promotional mailing, it takes 2 and all other sides or sources are labelled as 3.

Left without purchase: If the customer checks out properly by making payment this variable takes false value labelled with 0; otherwise it is 1 (true) by default.

As it was already mentioned above the data warehouse includes demographic data such as neighborhood, gender and age. Neighborhood has been categorized into seven different parts. If the customer neighborhood is labelled as 1, that means s/he lives in an area where estate prices are the highest; if it is labelled 7 that means the customer's address falls into an area where estate prices are the lowest.

All numeric data fields have been normalized in order to avoid any bias due to the value differences. For normalization, Equation (1) has been used. Categorical fields such as neighborhood, gender, period of the day, category of search and special day have also been normalized and converted into double values. This conversion is necessary for neural network learning process. However, for decision trees categorical values have not been normalized and they have been used as they are.

$$s' = \frac{s - \bar{x}}{\sigma} \quad (1)$$

Here, s represents the original value in the data set; $\bar{x}$ is the average of the field and σ is the standard deviation of the related data field and s' is the normalized value.

IV. METHOD

For the analysis we used KNIME program [22]. KNIME is an open source, Eclipse based program for data mining. It resides decision tree and artificial neural network algorithms and other open source data mining platforms such as WEKA and R. So, KNIME is quite useful to run different programs and algorithms on its own platform.

For learning, decision tree and artificial neural network algorithms have been used. Decision tree algorithms generate rules from data sets. When they run they create a tree structure and if-else-then rules. In this study we used two decision tree algorithms together. This technique is called bagging or bootstrapping. The algorithms we have used are C4.5 [23] and SPRINT (Scalable Parallelizable Induction of Decision Trees). C4.5 algorithm uses entropy function to determine the best attribute or data field to arch from. SPRINT algorithm uses gini function to choose the best data field to create branches. Neural networks learn through input, a set of hidden layers and an output layer. For each node an activation function is used. These activation functions are sigmoid, hyperbolic tangent and binary step activation function. They are used both for supervised and unsupervised learning. An objective or cost function is also used to assess how well the learning is. Sum of squares function as in (2) is one of them [24].

$$E = \frac{1}{NP}\sum_{j=1}^{P}\sum_{i=1}^{N}\left(t_{ji} - y_{ji}\right)^2 \qquad (2)$$

Overall model is depicted in the Figure 1.

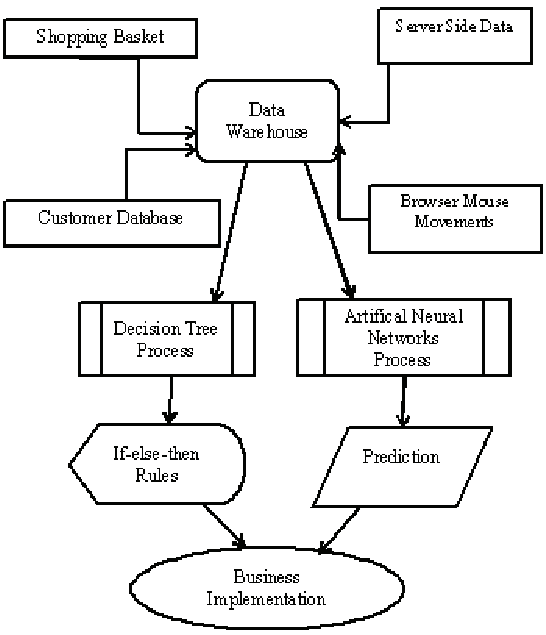

Fig. 1. Overall model applied for the analysis.

V. ANALYSIS AND FINDINGS

A. Online Analytical Process

As it is mentioned above the aim of the analysis is to predict whether an online customer will leave the site without a regular checkout at a certain time of his/her shopping. Before we applied artificial neural network and decision tree learning, we exercised some online analytical processes (OLAP). We may summarize our OLAP findings as follows.

9.367 visits out of 10.400 have been terminated without a regular checkout. Only 1.033 visits have ended up with a successful purchase. This is 9.9% of all visits. In other words 90.1% of all visits do not turn into a real sale. So, in the study we will find the reasons behind "***leaving the site without purchasing any item***", in other words the behavior patterns that show which customer will buy the item in his / her basket and which will not.

Most of the customers come from the 3rd neighborhood where estate prices are average or a little bit higher than average.

7.601 out of 10.400 customers are female. This is 73% of all visits.

There is a balanced distribution for source, so we cannot say that one **source** is superior to any other.

When we examine the attribute **period of day** we can say that a great majority of visits were paid either in the afternoon or in the evening hours.

Almost 50% of customers prefer searching a specific product on the site.

46.7% of customers (4865 records) put one or more discounted item in his/her basket; the others have nothing or no discounted items in their baskets during their visits.

Although only a couple of weeks are considered as **special day** during the data collection period, 2010 visits have been made during those days. So, 19.3% of all visits were during so called special days which are up to 7 days before a religious, national holidays or any day like Mother's Day, Saint Valentine's day etc.

Average customer **age** is 46.52 with a 14.32 standard deviation.

We have collected visits which lasted at least 90 seconds. We considered visits shorter than 90 seconds as accidental visits or outliers. They share 2.1 % of all visits and that share has been excluded. Also, considering them as outliers, we have ignored visits which lasted more than 1.500 seconds no matter it has ended up with a sale or not. Indeed, they make only 1.5% of all visits. Thus, we applied outlier reduction considering '**Time spent on the site**'.

B. Decision tree analysis

For this part of the analysis we applied bootstrapping operation with the algorithms which apply both gini index and entropy function. Data are divided in two parts. 70% of data have been used for decision tree learning and 30% data have been used for prediction. A confusion matrix scorer has been applied to calculate overall accuracy. Table 1 presents the confusion matrixes and accuracy statistics for both decision tree and artificial neural network analysis. Table 1 reads that overall accuracy for prediction is 90.42%.

Accuracy at predicting whether customer will leave without buying is 96.3%. When it comes to predict whether customer will pay and leave then accuracy is 40.2%. Since, F measure is 0.947 for predicting whether a customer will leave without paying; the rules generated by decision tree analysis may be used for some business purposes.

TABLE 1. CONFUSION MATRIX AND ACCURACY STATISTICS FOR DECISION TREE ANALYSIS.

Left Before Paying \ Prediction (Decision Tree)	yes	no	Accuracy	Precision	Specifity	F-Measure
yes	2690	103	96,30%	93.2%	40.2%	0.947
no	196	132	40,20%	56.2%	96.3%	0.469
Correct classified:2822						
Wrong classified:299						
Accuracy:90.42%						
Error:9.58%						
Left Before Paying \ Prediction (Neural Network)	yes	no	Accuracy	Precision	Specifity	F-Measure
yes	2720	104	96,30%	96%	62%	96.2%
no	113	184	62,00%	63.9%	96.3	62.9%
Correct classified: 2.904						
Wrong classified: 217						
Accuracy: 93.047%						
Error :6.953%						

Decision tree analysis may be summarized as follows:
The most important factor to affect 'leaving the site with or without buying at least one item' is the variable "**Special Day**". As it is depicted in Figure 2 (a), when the visit is paid on a special day, probability of buying an item increases from 9.9% to 46%. However, on other days only 1.3% of all visits turn into a real purchase.

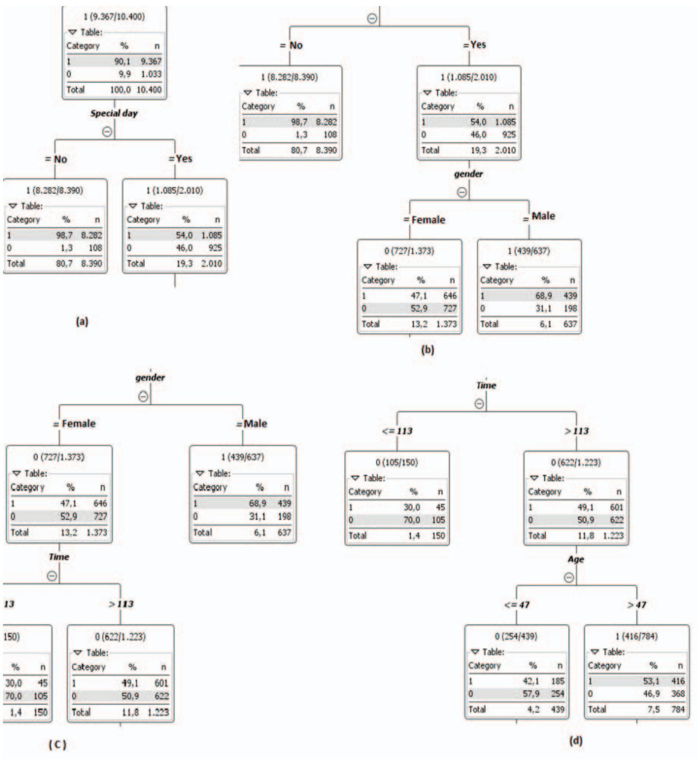

Fig. 2.Special day, gender and time play an important role on buying or not.

Further analyzing from *special day = yes* node on, we see that gender has importance on buying as well, see Figure 2 (b). Women (labelled as female) buy at least one item with a ratio of 52.9%. This is 31.1% for men.
As it is in Figure 2 (c), 113rd second of the visit has great importance for women. 70% of female customers who shop on a special day complete their payments before the 113rd second of the visit. If the time exceeds this cut point, then only 50.9% of them buy, pay and check out. Since 113 seconds make about 2 minutes only it may be considered as a short time, so we can comment that, if a woman has got something in mind before shopping or if she likes an item at first sight she certainly buys it.
As it is depicted in Figure 2 (d) after the 113rd second of shopping node, age of the (female) customer is very important. The cut point for the age is 47. Our analysis shows that 57.9% of those who are younger than 47 pay and terminate the visit. 46.9% for the female customers who are older than 47 leave without paying even if they are shopping on a special day.
From Figure 3, we can conclude that down the node age 47 the only parameter that plays a role on whether a customer will or will not buy is the variable **time**. The break points for time are 259, 379 and 661. Each break point changes the percentage of purchasing. This change may be seen in Figure 3 below.

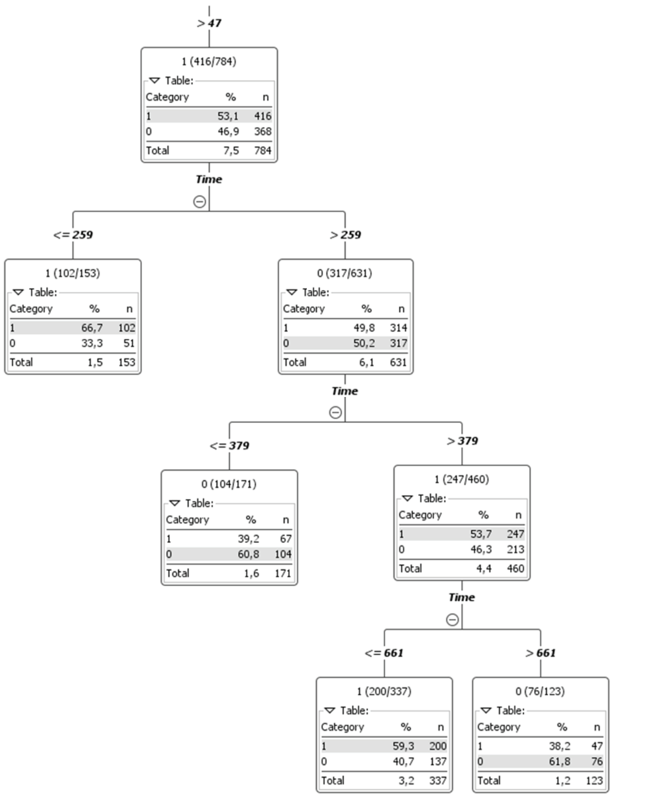

Fig. 3. 259th, 379th and 661st seconds of shopping are important for buying or not buying.

So far our analysis has yielded information about female customers only. So, in a further study we will exclude female customers and focus on male customers. We will do this because, majority of customers are women and decision tree cannot see or extract any rules for male customers when number of women customers are much more than that of men. Figure 4 depicts the results of our focus group (male customers). Figure 4 shows that the most important factor to determine if a male customer will leave with or without payment on a special day is the category of the product in the shopping basket. The rule suggests that if a male customer puts a product for women in his basket, there is a big probability that he will pay and check out. 71.2% of male customers did so. From Figure 4 we can conclude that if the product in the basket is any product such as a women's clothing or accessory then the most important factor to determine how the customer will leave the site is "item add time". 84.2% of those who add a women category item before the 295th second pay and leave. After the 295th seconds this drops to 44.1%.

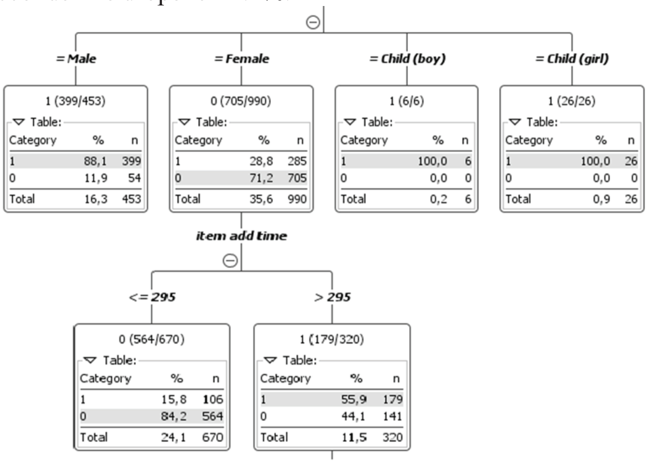

Fig. 4. The category of the product and item add time play an important role on male customers.

The most important rules extracted from decision tree analysis may be summarized as follows:

If it is a special day and the customer is a woman and the shopping duration is not more than 113 seconds yet, then she will buy the item in her basket (70%). Analysis shows that 105 customers draw this pattern. After the 113rd second possibility that the visit will turn into a real purchase is getting lower.

Another rule says: If it is a special day and the customer is a woman who is older than 47 and the shopping duration is between 259 and 379 then she will buy and leave the site (60.8%). 104 customers have followed this pattern.

Following the same pattern, if the duration exceeds 661 seconds, with a percentage of 61.8 she will buy the items in her shopping basket. If the time is between 379 and 661 seconds then purchase likelihood will be lower. 137 customers of those 337 have bought the items in their shopping basket (40.7%).

If the customer is male and he is shopping on a special day and he has at least one female product in his basket he buys the item(s) in his basket: 85.2%. 564 male customers followed this pattern in 90 shopping days. Otherwise, namely, if he has added another category of product rather than men's, the probability that he will leave without buying is higher: 55.9%. 179 male customers drew this pattern when shopping.

C. Neural Network Analysis

After performing decision tree analysis, we applied neural network multilayer perceptron learner. In the Figure 5 below, data processing steps are shown. As it can be easily seen in the figure, firstly data are read from the database, after that a serial of z-score normalization processes have been implemented for all data fields. After normalization the dataset is partitioned in two parts with a 70% to 30% ratio. That means 70% of data are used for neural network learning and 30% are used for prediction and scoring. Learning has been achieved using a multilayer perceptron learner (MLP). For MLP, a ten-node input layer, 10 hidden layers, each has 10 nodes and an output layer have been used. In the system, after the learning is completed 30% of data and learning parameters are sent to multilayer perceptron predictor node for prediction process. Scorer node creates a confusion matrix and accuracy statistics to be used to assess the quality of learning.

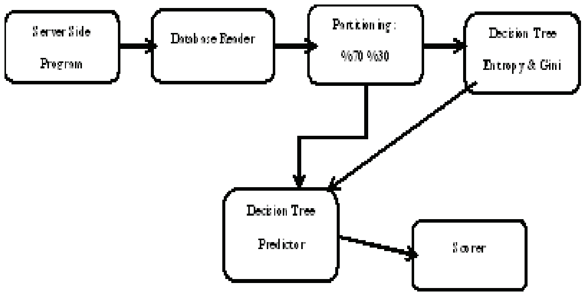

Fig. 5.Artifical Neural Network model used for analysis.

Since neural networks do not reveal rules like decision trees do, in this part of the study only confusion matrix will be presented. Table 1 shows error and accuracy statistics. With a 96.2% F measure we can say that neural network model may be used to predict whether a customer will leave without paying or buying at least one item. Indeed, accuracy and prediction values are high for guessing "yes". F-measure for predicting "*if a customer will purchase at least one item*" is relatively low: 62.9%. So, it is the risk of the business to use this, namely predicting and accepting that a customer will buy at least one item, because accuracy and precision values are 62.0% and 63.9%.

With these high accuracy values, we may conclude that by using this model and data, artificial neural network model works well for predicting customer behavior. In this study we aimed at predicting whether a customer will or will not buy the items s/he added to his/her basket.

D. Practical issues

This analysis generates rules to determine if a customer will pay and buy the item which is added to his/her shopping basket. If e-commerce firms have got rules like these they may use it for promotions. For example, when those firms know if a customer will buy the item(s) in his/her basket they can make instant promotions or discounts tailored for each customer on the spot. These rules may be turned into program codes and may be embedded in web pages. The codes will check each customer's clickstream pattern, compare it with the pattern found by decision tree and neural network and finally decide the possibility of the shopper to buy the item(s) added to basket or not. Firms may decide on the percentage of possibility to buy or not. It may be any value from 0% to 100% depending on the needs of the firm. After that, another piece of code may be embedded in web page and server to offer automatic discounts or promotions. For example if the possibility to buy the items is less than , say 30% then 10% discount may be offered to the customer individually. These examples and ideas may be multiplied in practice.

Although we have chosen to buy or not buy as class to predict with data mining, another field could be chosen as class variable as well. For example, firms want to learn who their anonymous surfers are, surfing on their websites. They might want to learn their age or gender. If the one has not signed in and surfing on an e-commerce web site, through data mining and examining the clickstream pattern of the surfer we may predict the age or gender with a known error of margin. As it is done in this study, we can extract the clickstream pattern of known customers with decision trees and artificial neural network and use these patterns for prediction. Again all this information may be used by management for individual discount and promotions.

VI. Conclusion

In this study, we have introduced a novel approach so as to extract customers' shopping patterns from mouse movements and website logs in order to predict if a customer will buy the items s/he has added to his /her basket. In the study, we used number and order of mouse clicks. We also added the type of the product which is clicked by customers. Besides these, color and price of the product, menu items and mouse idle time may also be added to this dimension. In addition, the data fields to be used may differ from one study to another.

We used decision tree and artificial neural networks models together. In the first part of the study, data about customers' mouse movements (what and when they click), their demographic information (age, sex, and neighborhood) and the items they added to their shopping baskets have been collected. For this purpose, a program has been embedded in an e-commerce company's website to collect necessary data for ninety days in Turkey. All collected data are converted into a data cube. After that, artificial neural networks and decision tree data mining models have been used to extract online customers' behavior patterns. The most striking findings of our analysis are as follows:

Category of the product is important to men: They buy the items in the shopping basket with a higher probability, if they have added a female product to the basket.

Men are more likely to shop on special days such as Christmas, Mother's Day etc. than other days.

Category of the product is not important to women.

Women shop both on special days or any other ordinary day.

Item add time and shopping duration are important for women to buy or not.

If women spend more than 113 seconds on the site and have added an item to the basket their buying probability is high.

Those who are younger than 47 are more likely to buy.

259, 379 and 661st seconds of shopping are very important for buying or not.

If women have added the first item to the basket before the 295th seconds they are more likely to buy it. Buying likelihood is less between the 259th and 379th seconds of shopping. However, after the 661st second of shopping probability to buy the items in the basket is rising.

Although all the findings listed above depend on the limited data collected from an e-retailer operating in Turkey, the same model may be used with different data to spot some other patterns which may be turned into business decisions by managements. Using the introduced data warehouse and the model with a further analysis we may predict issues like age and sex of the current customer [21] (If the customer has not been signed in), how many items the customer will buy, what color of items this customer is interested in and even what item or the page the customer will click next.

Since decision trees produce If-Else-Then rules, they may be used in software on client side. All these predictions and rules may be turned into business decisions and implementations such as promotion, sale campaigns and personal discounts. Results of the analysis may be used for customer relation management (CRM) and Business Intelligence.

Nonetheless, findings of this study cannot be regarded as general rules for online customers or their behaviors, instead this study shows the "practicability" of such a work.

References

[1] Thompson S., H. Teo, 2006. To buy or not to buy online: adopters and non-adopters of online shopping in Singapore. Behaviour & Information Technology,25 (6), 497 – 509.

[2] Aikaterini C. et.al., 2013. Online and mobile customer behaviour: a critical evaluation of Grounded Theory studies, Behaviour & Information Technology, 32(7), 655–667.

[3] Han J., & Kamber, M., 2011. Data Mining Concepts and Techniques, 3rd ed., Morgan Kaufman Publishers Academic Press.

[4] Watson H.J., Goodhue D.L., Wixom B.H., 2002. The Benefits Of Data Warehousing: Why Some

Organizations Realize Exceptional Payoffs, Information Management, 39 , 491–502.
[5] Inmon W.H.,2005. Building the Data Warehouse, Wiley.
[6] Wu HE, 2013. Improving user experience with case-based reasoning systems using text mining and Web 2.0, Expert Systems with Applications, 40(2), 500–507.
[7] Al-Zaidy, Rabeah; Fung, Benjamin C. M.; Youssef, Amr M.; et al., 2012. Mining criminal networks from unstructured text documents, Digital Investigation, 8(3-4), 147-160.
[8] Premchaiswadi, Wichian; Romsaiyud, Walisa, 2012 .Extracting WebLog of Siam University for Learning User Behavior on MapReduce, Proceedings of 4th International Conference on Intelligent and Advanced Systems (ICIAS) and A Conference of World Engineering, Science and Technology Congress.
[9] Senecal, S., Kalczynski, P. J., & Nantel, J., 2005. Consumers Decision making process and their online shopping behaviour: A clickstream analysis. Journal of Business Research, 58, 1599 – 1608.
[10] Hu, Jia, Zhong, Ning, 2008. Web farming with clickstream International Journal Of Information Technology & Decision Making, 7(2), 291-308.
[11] Ting, I-Hsien; Kimble, Chris; Kudenko, 2009. Daniel, Finding Unexpected Navigation Behaviour In Clickstream Data For Website Design Improvement, Journal Of Web Engineering, 8(1), 71-92.
[12] Detlor B.,2000. The corporate portal as information structure: Towards a framework for portal design. International Journal of Information Technology.
[13] Granić A., Mitrović I.& Marangunić N., 2011. Exploring the usability of web portals: A Croatian case study. International Journal of Information Technology.
[14] Lin, C.J. & Lu, H. ,2000. Towards an understanding of the behavioural intention to use a web site. . International Journal of Information Technology.
[15] Domingues, Marcos Aurelio; Soares, Carlos; Jorge, Alipio Mario.2013. Using Statistics, Visualization And Data Mining For Monitoring The Quality Of Meta-Data In Webportals, Information Systems And E-Business Management, 11(4), 569-595.
[16] Schaefer, Kerstin; Kummer, Tyge-F.,2013. Determining The Performance Of Website-Based Relationship Marketing, Expert Systems With Applications 40(18), 7571-7578.
[17] Vicari, Donatella; Alfo, Marco.2014. Model based clustering of customer choice data, Computational Statistics & Data Analysis,71 Special Issue, 3-13.
[18] Pachidi, Stella; Spruit, Marco; van de Weerd, Inge,2014. Understanding users' behaviour with software operation data mining, Computers In Human Behavior ,30, 583-594.
[19] Hu, XH; Cercone, N, A., 2004. Data Warehouse/Online Analytic Processing Framework For Web Usage Mining And Business Intelligence Reporting, International Journal Of Intelligent Systems,19(7), 585-606.
[20] Wang P, Zhang XY, Xu LB., Edited by: Xia, GP; Osaki H. 2002. Applications of data mining to electronic commerce: A survey, Proceedings Of The Sixth China-Japan International Conference On Industrial Management, 397-400 .
[21] Silahtaroglu G. 2015, Predicting Gender of Online Customer Using Artificial Neural Networks, Proceedings of International Conference on Management and Information Technology, New York, 28th August, 45- 50.
[22] Berthold MR. 2008. KNIME: the Konstanz information miner In Data analysis, Machine Learning And Applications, Springer-Verlag, 319-326.
[23] Quinlan, J. R. 1993, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.
[24] Hassoun Muhammed. 1995, Fundamentals of Artificial Neural Networks, Massachusetts Institute of Technology.