Engineering and Technology Journal
Journal homepage: engtechjournal.org

A Study about E-Commerce Based on Customer Behaviors

Zainab N. Mohammed [a*], **Suhad M. Kadum** [b]

[a] Computer Science Department, University Of Technology, Baghdad, Iraq, 110984@student.uotechnology.edu.iq
[b] Computer Science Department, University Of Technology, Baghdad, Iraq, 110102@uotechnology.edu.iq

Submitted: 04/03/2020 Accepted: 25/07/2020 Published: 25/07/2021

K E Y W O R D S

E-commerce, Purchases, Customers behaviors, Machine Learning, Prediction.

A B S T R A C T

E-commerce is one of the new virtual technologies that make life simpler for both traders, marketers, and customers. However, the main problem for the seller was how to know customer's intentions when they enter the website. This paper proposed to predict whether customers make a purchase or not from their previous behaviors. Therefore, this paper aims to predict the intentions of users that using online -shopping. the main aim of this study is to highlight customer behaviors to predict purchases and make a compression between the works that are related to the subject of this paper to conclude and suggest the best method to predict purchasing in e-commerce treading that depends on customer behaviors.

How to cite this article Z. N. Mohammed and S. M. Kadum," A study about E-Commerce Based on Customer Behaviors,", Engineering and Technology Journal, Vol. 39, No. 07, pp. 1060-1068, 2021.
DOI: https://doi.org/10.30684/etj.v39i7.1631

1. INTRODUCTION

Digital analytics growing up by using social media in shopping, it becomes very popular at a global level. These connectivity applications are having over million active users the interact to share and digest information helping marketing activities such as enhancing, ads and customer behavior from the information investigation of publication buying behavior [1].

A large amount of data has been used by users, with no specific target to their real need's users have used many applications they use the social media for communication, treading, and other purposes. Amazon, Alibaba, net flex and other websites that have a determined goal and other that have more than one goal like Facebook, tweeter and other new application that the user can make friends and communicate with the other users [2].

E-commerce is online shopping and trading by using electronic devices through the internet, with the development of e-commerce still needing good online marketing [3]. Many data mining techniques are proposed to forecasting customers that are most likely to be churn [4]. Many ways can be reached to the customer behaviors some of them are numerical or textual, so the first step to choose a good data set depend on best features used in the training to make the learning very comfortable, there are different

type of machine learning used and give different results some of them are in the next part of this study, this paper study e-commerce with customer behaviors for five last years when it's grown up [4].

This paper is the first study for reviewing previous works in predicting customer purchasing based on customer behavior.

The main aim of the paper is to thoroughly review the literature discuss the previous works in the field of decision support system in e-commerce purchasing prediction based on customers previous behaviors, it is a useful paper for researchers to know all the methodologies used in this field and its accuracies with compressions between them than improving the best method in prediction with the best method in feature selection for example. It has another benefit for summarizing the previous researches to facilitate the research process.

The paper contains: firstly, defined e-commerce and the effect of AI on it, secondly, make a review to the researches and presented customer behavior to have a thought about data collections and items and describing methods and algorithms that are used in each research, thirdly, make a discussion for the results of the related works and presented it in a table, finally, a conclusion of these study.

2. E-COMMERCE

A Commercial that is automated by Artificial Intelligence (AI) can be seen in our lives, it is used in working, buying, and so on. AI changes marketing to an automation system seen in social media [5].

E-commerce improving the tools online treatments of marketers in purchasing and selling. It's the modern way of commerce that becomes a huge tool between seller and consumers, the internet is growing up the effects of the overall world economy highly, it raised the rate of Exchanging a foreign currency [3]. In the last few years, AI has grown-up marketing, it helps to promotion strategies of marketing, It helps marketer's hands by summarized their workloads manually [5]. There are many types to deal with E-commerce, one of them is customer lifetime value that means depends on customer loyalty (how the customer will defend about the mark some product of it) [4].

E-commerce makes work strategies as simple as it was and it helping in making personal decisions by supporting them to make it faster and more effective, it makes smart seals that allow the sealer to know the products that the consumers favorite from that how is neglected [5].

There are several ways based on to predict the purchasing in electronic shopping some of them are listed below:

I. Social media

It is one of the e-commerce tools which have a large number of customers, this type of tread makes a path for people making communications with their needs of purchases online. Most customers of social media belong to the young or teenager's side rather than aging people [3].

II. E-commerce loyalty

It means the degree of loyalty of the customer to the product that it could be considered as a lifetime value, CLV means Customer Life value can be counted by the seller that each customer has an ID number and CLV (it is the Value of the relationship with existing customers by looking at the future cash flows of the customer relationship) [4].

III. Brand and its relate on CLN (Clint Loyalty Number)

Some people prefer the products with a specific brand, the loyalty that they make to their favorite brand is to make a purchase even if there are goods best than this treading mark products, and they also can be classified as CLV [4].

IV. Churn customer

Churn is a term that deals with a customer how is CLV customer and have a high (CLN) but then this customer decide to live this product that the seller or the system can denote the difference in the purchasing of that customer and the time spend to visit this market to do shopping is longer than the last time [4].

3. RELATED WORKS

Artificial Intelligence makes salespeople able to understand Customer behavior and there is no difficulty to maintain a personalized relationship with a customer. The following are studies that used deep learning to predicate customer behaviors depending on special features that affected the purchases of customers:

1) B. Zheng et al in (2013) [6] proposed a model based on users check-ins, it based on the location from mobile apps, and this check-in posted and can be shared with people these features make marketing more easer using information from social media users that received to them, it gave a higher result of accuracy using Artificial Neural Network (ANN) that most other researchers used Recurrent Neural Network (RNN), ANN was given a higher accuracy (93.13%) with its higher accuracy ANN have a problem with the time of training than other classifiers, support vector machine (SVM) can solve a problem of time but with very low accuracy that SVM with sigmoid function(54.00%) and SVM with radial basis function(53.89%), RNNs can train a big data with accepted time and accuracy.
2) J. Qiu et al in (2015) [7] proposed a new method called the "hot method". The model used two parts of the process is training the crowdsource with regression and then training it with (SVM). The data set that is used is JD dataset from a Chinese e-commerce website. Model is based on customer preferences brand, the result of precision was 75.3 percent.
3) H.Salehinejad et al in (2016) [4] proposed The RFM behaviors (Recency, Frequency, Monetary), a Recency means how many days the last purchasing was, Frequency means how many times the customer buy from that shop (total number of treading), monetary means total money every customer paid (How much money the customer paid during a limited time), using RNN algorithms. Data divided into three parts training, testing and validation, ReLU (Rectified Linear Units) activation function with RNN (Recurrent Neural Network) they also used LSTM-RNN (Long Short Term Memory) but it gives higher performance for monitory feature than ReLU, this paper also learning the data with SRNN (Simple Recurrent Neural Network). results were 67 percent for the SRNN and 77 percent when using LSTM while the ratio of accuracy raised to 80 percent when using ReLU-RNN.
4) K. Attanudi et al in (2018) [8] have integrated a greedy algorithm with neural that forecasts the behaviors of the customers to save the time of computations using online behaviors as a dataset, the result was good for the aim of this research that the computational time reduced compared with the method without it, but the accuracy also decreased, it helps a machine to make the decision using a greedy algorithm. The accuracy is 63 and 64 percent.
5) C. Okan Sakar et al in (2018) [9] applied association rule mining on the dataset to evaluate and predicate the visitors of a page that make purchases of the visitors and extract valuable features of the behavior from different profiles of users to present the sales for a valued customer that have an intention to buy only using MLP(multilayer perceptron), SVM (Support Vector Machine), RRN, LSTM algorithms for training and testing. have results for MLP which is used for predicate if the customer offers the product or not 72 percent, and 70 percent for SVM classifier and 74 percent for LSTM.
6) A. Muto et al in (2019) [10] determined the truth of the classification in data mining a classification method is very good in data mining. The technique used is a classification for all shoppers who have paid their credit premium to the next marketing target and determining the targeted merchandise. Proposed ANN and backpropagation (BP) to make classification and prediction about the customers of the bank using the bank market dataset, the accuracy of this model was 80 percent.
7) I. Al-Zuabi et al in (2019) [11] has proposed a system to predict the gender and the age of customers in communication networks from their behaviors, information's and services. both call details records (CDRs) and customer relationship management (CRM) has used to analytics the behaviors. SyrialTel Telecom company relabel dataset has used with 18000 users. The accuracy from the improved SVM method was 85.6% for gender prediction, and 65.5% for age prediction.
8) C. Giri and et al in (2019) [12]. has proposed a system for customer analyses in fashion, which described that the system can predict the customers' Requirements and can meet and understand

them better than personal analytics. The proposed model has created value to the fashion purchasing, therefore the revenues have improved and invested from these techniques. It is also investigating the purchases and also it is keeping the old customers. Using transition matrix to predicting their future behaviors by using the higher revenues for many customers for 6 months ago including (RFM).

4. MACHINE LEARNING ALGORITHMS

Big data is too huge to store in the main memory of a system, and the need to process big data by parallel and distributed algorithms arise in internet search, network traffic monitoring, signal processing, machine learning, scientific computing [13]. Machine learning is about the reconnaissance and evolution of mathematical models and algorithms to learn from data. Its sample concentrate on classification objectives and consists of modeling an optimal mapping between the data domain and the knowledge set and developing the learning algorithms [14]. In this section the study will explain all the algorithms used in the related works prediction:

I. Artificial Neural Network (ANN)

Computer algorithms that are inspired by nature designed to mimic real neural networks in the process the information. ANNs gather their knowledge by detecting the patterns and relationships in data and train through experience, it is formed from hundreds of single neurons, processing elements (PE), connected with weights, which modeling the neural structure and are formed layers[15].

II. Recurrent Neural Network (RNN)

It is based on remove outliers from the data and then evaluates parameters from the filtered data. It filtered the data by removing outliers from both the target function and the inputs of the neural network. [16]. Bengio [17] has proposed (ANN) in statistical language modeling used feedforward neural networks with fixed-length context. Single model implements, better than enhancing sundry of other algorithms. neural network-based models supply considerable refinements for several tasks.[18].

III. Support Vector Machine (SVM)

SVM algorithm preserves a candidate Support Vector set. It begins the set with the adjacent pair of points from obverse classes. Support Vector may be prevented by other elect Support Vectors already present in the set. [19].

IV. Rectified Linear Units (ReLU)

It is a hidden layer activation in a deep neural network. It is finished by activation of the before the last layer in a neural network, then use to learn the weight parameters of the ReLU classification layer through backpropagation.[20].

V. Backpropagation (BP)

Is one method of ANN that has several units in one or more hidden layers, BP uses search methods to find the minimum point that helps to find the weights with the fewest errors. BP uses an error output to change the weight backward [10].

VI. Long Short-Term Memory Networks (LSTM)

The LSTM model has a memory cell instead of the hidden units that controlled inputs and outputs by gate [3]. It contains input, forgets, and output units and an activation cell [21]. for each time step LSTM has a range of repeated modules as in a standard RNN[22].

5. DATASETS SURVEY

Several kinds of datasets have been described in this section, all of them can deal with customer behaviors so all of them could help in prediction purchasing with different features has come from user information, session and so statistics of purchasing. All types discussed below:

I. Online shopper intention behaviors [9]

This dataset contains numerical features (Administrative, Administrative Duration, Informational, Informational Duration, Product related, Product-related duration, Bounce rate, Exit rate, Page value, and Special day**),** and categorical features(Operating Systems, Browser, Region, Traffic Type, Visitor Type, Weekend, Month, and Revenue). The feature vectors pertinence to 12,330 sessions within the dataset. Each session during this dataset would pertinence to a distinct user during a year to avoid tendency in session, big day, and so on. 10,422 from the label have a negative value and this means that visitors didn't make purchasing finally and just 1,908 were the positive class sample that ended with shopping. The training dataset has been observed and declared that the dataset was biased to the 'false' category from the label. This influence any model in training and predicting despite all the models have given better accuracy for predicting the 'false' category but the accuracy of predicting the 'true' label was much lower than that of zeros and this was a clear problem and very evident when the accuracy has been calculated as shown in Figure 1.

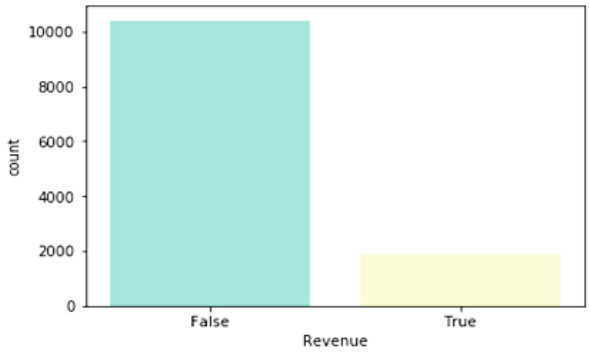

Figure 1: Unbalance data biased to zero revenues

II. Ta-Feng dataset [4]

The dataset has collected from a Chinese grocery e-commerce website, it contains 817,741 transactions belonging to 32,266 users and 23,812 items, transaction data of 4 months from November 2000 to February 2001. The dataset contained 9 attributes as it is described in Table I.

TABLE I: Ta-Feng dataset features description

Class name	Description
Transaction _DT	Transaction date –timestamp not available
Customer _ID	Unique Customer identifier
Age _Group	10 possible values - <25,25-29,30-34,35-39,40-44,45-49,50-54,55-59,60-64,65>
Pin _Code	8 possible values - 105,106,110,114,115,221, others, Unknown
Product _Subclass	product subclass
Product _ID	unique product identifier
Amount	quantity
Asset	asset
Sales _Price	billed price

III. Foursquare restaurant check-ins dataset [6]

The dataset is collected from the Foursquare.com website for user check-ins in their locations, the data contains three types of data: text(comments), timestamps (the date and time for chick-ins), and hyperlinks(tweets links). The features are used in this data (User ID, Timestamp, longitude, latitude, and tweet). The dataset contained 120,825 rows for chick-ins in restaurant locations from the Foursquare.com website.

IV. JD dataset [7]

It was collected from Jing Dong(JD website) it is known as B2C e-marketing website, see Figure 2, the features depend on was (price, average rating, and some reviews, on-shelf date, rating, and date of the most recent review). The features are divided into three parts of products features: static-like (price, brand, and so on), dynamic (average rating and seals), and certain product features like (size, color, and so on).

Figure 2: JD chines e-commerce website

V. Bank direct marketing data [10]

The data has been collected from a contact center of a Portuguese bank, the data has collected is contained 17 features which were recorded between May 2008 to November 2010 and it has contained 79354 contacts. The features were variate in its types there are three types which are numerical features (age, balance, duration, day, p-days, and previous), polynomial features (job, materials, education, contact, month and p-outcome), and binomial features (Default: means the customer has credit in default, housing, loan, and the label Y).

VI. Mobile phone data [11]

The dataset represents human and social behaviors, it's combined some databases to gather which are:

1) CDRs: it contains all features that deal with customers it contains numerical features which are (call types, GSM (A), GSM (B), Direction, Cell identifier, and Duration Date).
2) Customer's services: it deals with all service features, and it's classified manually according to the type of features like political news, sports news, and horoscopes and so on which handle as customers' services.
3) Customers contracts information's: this dataset deal with all basic information about customers like(age, gender, location, and so on).

VII. Apparel Industry Dataset[11]

This dataset is Specializes fashion, it contains two-period time, the first one is 01-06-2016 to 01-12-2016, and the other is from 01-12- 2015 to 01-06-2016. It contains Recency value for all customers as the main feature, creating a piece of clothes is based on a recency value that is calculated for each customer for six months. The customers are classified into "Inactive," "Less Active," "Active," "Highly Active," and "New Customers." According to their evaluated recency. Customers' features are Purchase amount, Date of purchase, Days since, and the Recency Frequency Amount data that calculated for each customer to distinguish between customers.

6. DISCUSSION

Generally speaking, various kinds of literature have experimented with different behaviors of customers in different applications like website e-commerce, banking, mobile, and communications, restaurants prefer, and fashion.

Literature papers for predicting customer purchasing from his previous behaviors (the behaviors depend on the dataset has used) during the e-commerce has used different methods with different

features, and the best-affected method was ANN it has given higher accuracy than other methods used but it takes a long time to learn many layers and of number neural in each layer.

Our findings suggest that the disparity between the accuracy of the methods depend on two things, first, the dataset selected and the features determining from it, second, the method used with the features as it shows when using ANN with Foursquare Check-Ins locations dataset it gave 93% accuracy, but when used backpropagation ANN with bank direct marketing dataset, it gave 80% as its accuracy.

There are different results of accuracy, they are between 63 and 93 percent, these differences in accuracy ratios came from the methods used in training and the features that determined the best and the global behaviors to make purchasing.

Table II presented review literature for the marketing with customer behaviors, it contains author name or names for each research, the year of the post, a research name, method or algorithm used to forecasting in each research, the best features that the learning based on it, a data set that contain the items to make training and testing on it, a result of accuracy that given from each research that comes from verification algorithm through a certain dataset.

TABLE II: Compression of literature reviews

Author Name	Year of Publish	Research Name	Best features	Dataset Used	Methods Used	Accuracy Ratio
B. Zheng, K. Thompson, Sarah S. Lam, Sang W. Yoon	2013	Customers' Behavior Prediction Using Artificial Neural Network	Check-in's features	Foursquare Check-Ins locations	ANN	93.13%
					SVM activation function	54.00%
					SVM- bias	53.89%
J. Qiu, Z. Lin,Y. Li	2015	Predicting customer purchase behavior in the e-commerce context	customers' preference (brand, price)	JD dataset	Association Rule Heat Model	Precision (75.3%)
H. Salehinejad, S. Rahnamayan	2016	Customer Shopping Pattern Prediction: A Recurrent Neural Network Approach	RFM, CLN	Ta-Feng dataset	SRNN	67%
					LSTM	77%
					ReLU RNN	80%
C. Okan Sakar, S. Olcay Polat, M. Katircioglu, Y. Kastro	2018	Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks	Numerical And Categorical features (user information and session)	Online shopper intention behaviors	MLP	72%
					SVM	70%
					LSTM	74%
K. A. Perichiapp	2018	Greedy Algorithm Based Deep Learning Strategy for User Behavior Prediction and Decision Making Support	Session of users, Product page, Item display with a session, Purchase	Online Behavior	Greedy	63-64%
A. Mutoi,, S. S. Faisal, H. Handayani, A. Jalaludin	2019	Classification Data for Direct Marketing using Deep Learning	Name, age, housing, loan	Bank direct Marketing data	Backpropagation	80%

7. CONCLUSION

In this study, several models are used in predicting purchasing based on customers previous behaviors, first, artificial neural network and perceptron backpropagation with classification features, and LSTM recurrent neural network, one new method has experimented with a greedy algorithm for predicting online behaviors and it gave a low accuracy compared with other researches, but it was more speed than other algorithms.

The main features are needed to know the behaviors of customers like human experts knowledge, these features are sentiment analyses (word "text features"), URL features (customer page, session, parches page), historical features and comments on tweeter, categorical features like name of months and special days, and numerical features like the binomial features that gave true or false(happen or not). These behaviors and features presented in this study must be training to gather to deal with all of the customers' behaviors. The suggestion is to experiment ANN with classification features by improving its speed by using one of the artificial intelligence algorithms (swarm, genetic, and so on).

References

[1] K. S. Mahajan, S.S. Jamsandekar, Dr. A M. Gurav, Machine Learning Approach for Marketing Intelligence: Managerial Application, Int. J.Eng. Comput. Sci.,2017.

[2] Yu, Fei, et al. Friend recommendation considering preference coverage in location-based social networks. Pacific-Asia conference on knowledge discovery and data mining. Springer, Cham, 2017.

[3] S. Lodhi, M. Shoaib, Impact of E-Marketing on Consumer Behavior: a Case of Karachi, Pakistan, IOSR J. Bus.Manage., 19(2017) 90-101. **https://doi.org/**10.9790/487X-19010590101

[4] H. Salehinejad and S. Rahnamayan, Customer Shopping Pattern Prediction: A Recurrent Neural Network Approach, University of Ontario Institute of Technology in 2016 IEEE.

[5] M. Avinash, Dr. R. Jayam, Dr. M.G.R. Educational, Artificial Intelligence - The Marketing Game Changer, Int. J. Pure App. Math., 2018. https://doi.org/10.4018/978-1-4666-8459-1.ch003

[6] B. Zheng, K. Thompson, Sarah S. Lam, Sang W. Yoon, Customers' Behavior Prediction Using Artificial Neural Network, Industrial and System Engineering Research Conference, 2013.

[7] J. Qiu, Z. Lin, Y. Li, Predicting customer purchase behavior in the e-commerce context, Springer,2015

[8] K. Attanudi, P. Perichapsson, Greedy Algorithm Based Deep Learning Strategy for User Behavior Prediction and Decision Making Support, J. Comput. Commun., 6(2018) 45-53. DOI: 10.4236/jcc.2018.66004

[9] C. Okan Sakar, S. Olcay Polat, M. Katircioglu, Y. Kastro, Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks, Springer, 2018.

[10] A. Mutoi, Siregar, S. Faisal, H. Hikmayanti, Handayani, A. Jalaudin, Classification Data for Direct Marketing using Deep Learning, Buana Perjuangan University,Sci. J. PPI-UKM, Indonesia, 2019.

[11] Al-Zuabi, Ibrahim Mousa, Assef Jafar, and Kadan Aljoumaa. Predicting customer's gender and age depending on mobile phone data. J. Big Data 6.1 (2019): 18. https://doi.org/10.1186/s40537-019-0180-9

[12] Giri, Chandadevi, Sebastien Thomassey, and Xianyi Zeng. Customer analytics in fashion retail industry. Functional Textiles and Clothing. Springer, Singapore, 2019. 349-361. https://doi.org/10.1007/978-981-13-7721-1_27

[13] G. Manogaran* and D. Lopez, A survey of big data architectures and machine learning algorithms in healthcare, J. Biomed. Eng. Technol.,2017. DOI:10.1504/IJBET.2017.087722

[14] Alpaydin, Ethem. Introduction to machine learning. MIT press, 2020.

[15] Bahrami, Mehrdad, et al. Develop 24 dissimilar ANNs by suitable architectures & training algorithms via sensitivity analysis to better statistical presentation: Measure MSEs between targets & ANN for Fe–CuO/Eg–Water nanofluid.Physica A: Stat. Mech. Appl., 519 (2019): 159-168. https://doi.org/10.1016/j.physa.2018.12.031

[16] Rubanova, Yulia, Ricky TQ Chen, and David K. Duvenaud. Latent ordinary differential equations for irregularly-sampled time series. Advances in Neural Information Processing Systems. 2019.

[17] Nasser, Ibrahim M., Mohammed Al-Shawwa, and Samy S. Abu-Naser. Artificial Neural Network for Diagnose Autism Spectrum Disorder. (2019).

[18] Karita, Shigeki, et al. A comparative study on transformer vs RNN in speech applications. 2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). IEEE, 2019.

[19] Ibrahim, Hadeel Tariq, et al. A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets, Neural Comput. Appl., 31 (2019) 5965-5974. DOI:10.1007/s00521-018-3414-4

[20] Abien Fred M. Agarap, Deep Learning using Rectified Linear Units (ReLU), arXiv.org, 2019.

[21] H. Salehinejad, Learning over long time lags, arXiv preprint arXiv:1602.04335, 2016.

[22] C. Zhou , C. Sun, Z. Liu and Francis C.M. Lau , A C-LSTM Neural Network for Text Classification,arXiv.org,2015