*International Journal of Engineering & Technology*, 7 (4.6) (2018) 161-164

International Journal of Engineering & Technology

*Website: www.sciencepubco.com/index.php/IJET*

***Research paper***

# Predicting User Behaviour on E-Commerce Site Using Ann

**Aditya Sai Srinivas[1]*, Somula Ramasubbareddy[2], Manivannan .S.S[3], Govinda.K[4]**

*[1,2]Research Scholar,SCOPE School ,VIT University,Vellore,India,632014.*
*[3,4]Associate Professor,SCOPE School ,VIT University,Vellore,India,632014.*
**Corresponding author E-mail:svramasubbareddy1219@gmail.com*

**Abstract**

One of the major tasks for the modern business is learning to use all of the data available to them in a way that is both expressive and actionable. However, the potential for using data produced by a website is often left unfamiliar, and as a result, the objectives and feedbacks of individual digital customers can be unheeded. On the average website, there is a plenty of data to be collected about who interrelates with your site and how. By leveraging all of this statistics, we can gain perceptions into customer actions. **Machine learning methods** can be used to define which customers may be fascinated in achieving a result on your site.

***Keywords***: *Back Propagation, Layers, Machine Learning, SOM, Regression..*

## 1. Introduction

Due to massive growth of product and availability of e-commerce website has revolutionized the world in the way of buying the products. Definitely the evolution of internet and cloud has played a major role in the way of buying products. Now a days we have different e-commerce websites like Amazon, flipkart, e-bay and many more where a customer can buy and compare his/ her desired product based on his/her needs. Due to competitiveness in the market we have seen a significant decline in the cost of products. Due to presence of many e-commerce websites on the important, it becomes very important for shopping sites to know the current mood of the customers in order to increase the stock of that product so that they don't run out of stock and it also helps them in meeting the customer's satisfaction. This helps the e-commerce company in attracting more customers to use their service or product. The details of the products are stored on cloud where a user does login or signup and then looks for different products. Every action of user is tracked by the e-commerce company after which they analyze the user's behavior and provide suggestion to look at different products too which the user may find attractive and will help the user for his cause.
Here in this paper we will be analyzing the mood of customers when they visit some e-commerce site. We will be taking into account five different factors which includes type of device used by customer, time duration the customer is spending time on the app, at which time slot he is using the site, weather the customer is returning user or not, and how many product is he viewing in that duration. According to our previous datasets we will try to predict the user behavior of buying a product from a particular website. I am using artificial neural network which contains two hidden layers with back-propagation method to predict the user behavior. We will also be comparing the efficiency of the output with the efficiency of logistic regression without training, and after softmax training.

## 2. Background

TAM was introduced by Davis[1] , and his motive was to understand and predict customer's IT approval over his workstation. According to Davis, observed usefulness "is the probable user's subjective possibility that using a particular Presentation system will rise his or her job show within an administrative framework" while comfort of practice "Refers to the degree to which the probable user assumes the system to be open from strength". In this literature survey TAM has been proposed to understand and predict the user behavior on e-commerce shopping websites. Here an important factor has been added, the perceived risk. Online shopping acceptance is influenced by perceived risk. However here perceived risk is divided into two sub category they are the transactional security and the product delivery and service.
One of the biggest hurdle adoption and online purchase is perceived risk. Jacoby and kaplyn [2] discovered seven different types of risk : psychological, social, time, financial, performance and opportunity cost risk. However there were only three types of risks which were observed in internet shopping. Those are financial, information and product risks. Customer feels risky while performing shopping through any online as compared to traditional way of shopping. Customers are reluctant in providing their personal information on internet. According to Armando [3], in his research paper he compared traditional machine learning with the most advanced deep learning technology. In e-commerce, user search different products and based on those searches website gets to know the intention of the user and shows him related products. He used probabilistic generative process to understand the behavior of human intentions in order to know what the user is going to search next. PEOU[4] is defined as a degree where user thinks that using a system is free from effort. PEOU is measured on basis of three scale according to venkatesh [5], that is learning to use my favorite website, my favorite website is simple and easy to use, and my interaction with them are crystal and clear. He says that PEOU and PU both influence the customer in positive way. For predicting someone's next step, we need to understand his behav-

**Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**

ior. Novak elat[6] indicated in his paper a correlation between recreational internet use and flow. He defined flow as a state which we see during network navigation often characterized by series of responses. According to Chwenley[7], customer's satisfaction is an important aspect for the user to return on the website to use it again to buy product. He also proposed the website manager to make the websites more artistic and should have better user interface so that user feel convenient to use it. Which will definitely increase the tally of visiting user. And once that there is in increase in the number of customer, then there is a great chance of their products getting sold. Caroline[8] in his paper focuses on the dataset of interest user and check the pattern in purchased data by the user. He aims to create the dynamic of purchase intent by studying the user's present and previous intent. He divided his paper into three levels user activity level, types of user action and content user interacts with. He conducted a data driven cross platform analysis of the customer's behavior and buying intentions of the product. Statistical analysis was performed in Grazyna's[9] paper from where he tried to access the login session timing of each and every user. Then he performs preliminary data analysis and then he compared the buying and non-buying session. These analysis were performed on the book store. In this paper, Rafel and Jerzy [10] examines how graphical factors influence the online working e-commerce website in terms of searching a particular product. His study mainly focuses on three different factors on two different levels are two location of main menu, presentation on product and different type of search task..re asked to work fully fast and with high precision of accuracy. The obtained results were analyzed and a standard method of variance was used to understand that weather the independent variable affect the mean search time or not. An intelligent SOM network[11] is proposed to perform web mining and predict the next visiting page. The intention is to use the web perfecting to improve the web access. For perfecting, it is required to know the users current status at which page he is currently at and then predict the next visiting page. SOM analysis is performed to know the user to know the relation between the continuous user requests. Idea is based on the requirement to provide flexible product recommendation system suitable for work driven websites[12] which is indeed able to personalize the web experience for each unknown visitor. The principle engine which he used is neural networks. His work uses unsupervised model to create and train special type of neural network. Which is called self-organizing map.

## 3. Methodology

There are 500 user data with 6 different attribute. We are gathering information weather customer is using mobile or laptop, how many product that user is viewing, how much time a user is spending time on our website, whether the user is returning user or not, their timing slot when they are using our website, their final action that weather they are buying our product or not. We are increasing the efficiency of our prediction using ANN as shown in Figure1 (back propagation method) which contains two hidden layers. We have used **Tanh** as activation function. Which is a hyperbolic function.

Initially the all the bias values are taken as zero. And the learning rate is taken to be 0.001.

Activation function is given by:

$$f(x) = 1— \exp(-2x) / 1 + \exp(-2x) \quad (3)$$

### 3.1. ANN architecture

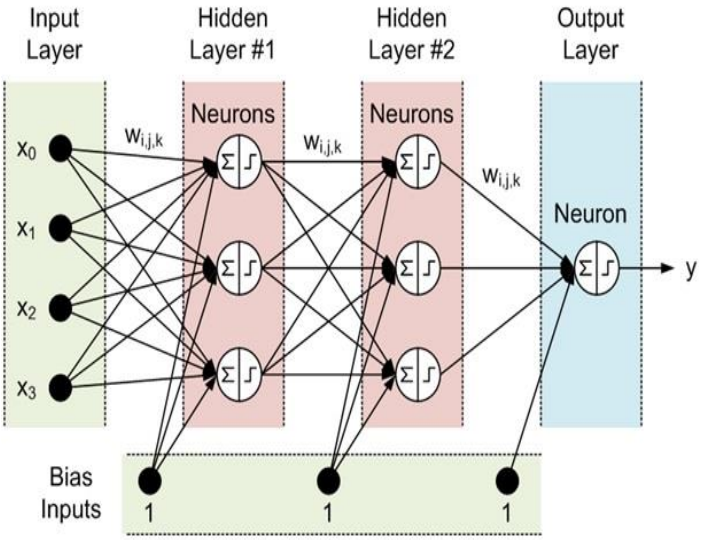



**Fig.1**: ANN Architecture

Feed Forward formula is given by:

$$Z = \text{Tanh}\ (\ (X_i * W_i) + b) \quad (1)$$

Where,

Xi is input matrix value.
$W_i$ is corresponding weight for the input.
And $b_i$ is the biased value for that node.

Bias update and weight update is given by:

$$W_1 = W_1 - \text{learning_rate} * (DE_{total} / DW_1) \quad (2)$$

Backward pass is given in Figure2.

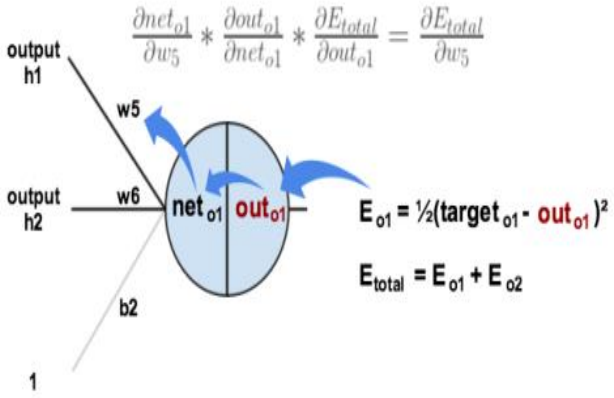



**Fig. 2:** Backward Pass

This paper comparing the efficiency of our neural network with logistic regression. **Logistic regression** is defined to be a statistical method used for analyzing a dataset where there are one or more independent variables that determine an outcome. The training of logistic regression is done using softmax function which is also known as normalization function as shown in Figure3.

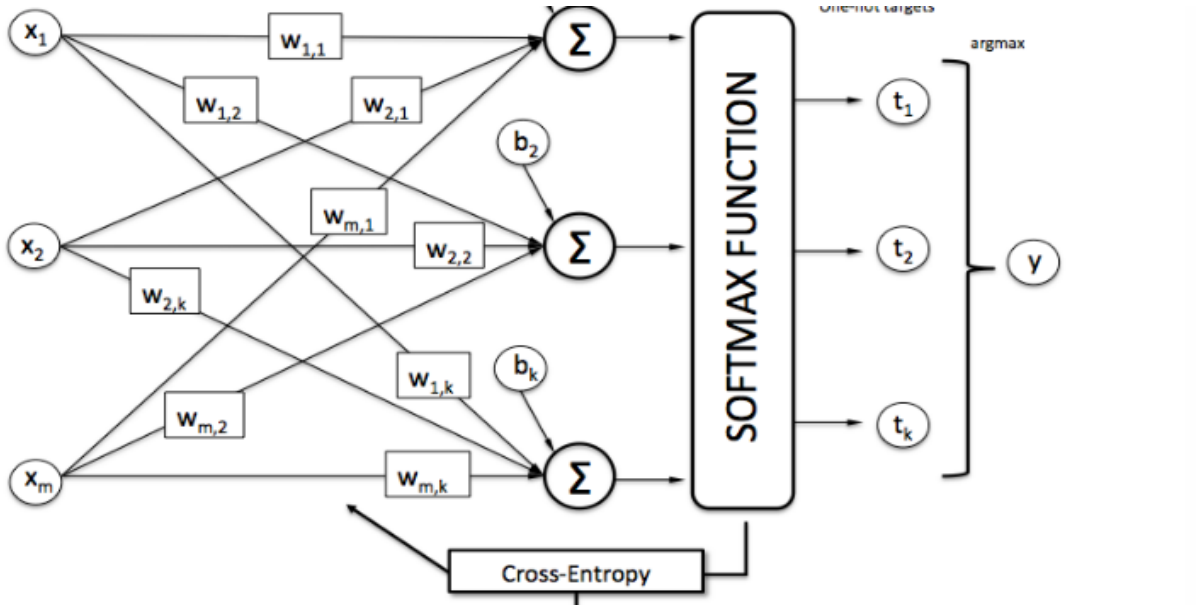



**Fig. 3:** Logistic Regression

Where cross entropy is given by the mean of change in the observed value and actual value and is given by

$$H(p,q)=-\sum \forall \exp(x)\log(q(x)) \quad (4)$$

We are calculating efficiency of an algorithm by (error in (observed output and actual output) /total number of dataset checked)*100.

## 4. Experimental Results

It is implemented using python and panda math's library. We first of all created a neural network structure. Then the dataset is passed to input node. As our neural network has two hidden layer so our neural network gets appropriately trained and is able to easily recognize the features of the dataset. Although two hidden layer is not enough to achieve most efficient. However we are still able to achieve close to **95%** efficiency. Our work is an implementation of problem available at kaggle for which the highest efficiency is 98%. Although we tried to compare the most common algorithm used to train and compare the data set out of which the first is logistic regression

**Logistic regression** is a statistical method for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes).
We have got efficiency about **87% for logistic regression.**

The method implemented by us is artificial neural network which is yet another machine learning algorithm which gives us more efficient result as compared to other algorithm because of availability of hidden layers to identify the features in a data set quite well.

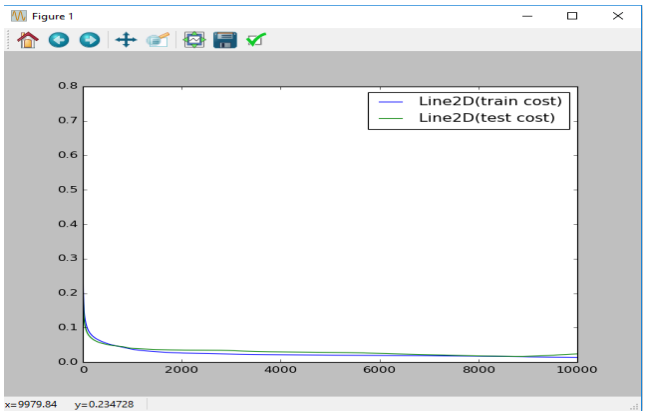



**Fig. 4:** Train Cost and Test Cost for ANN

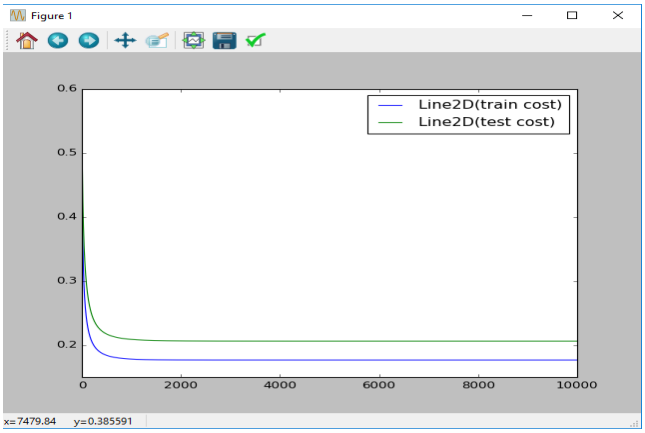



**Fig. 5:** Train Cost and Test Cost for Logistic

## 5. Conclusion

We have successfully trained our neural network using 250 datasets by back propagation method and then for next 250 dataset we are predicting the outcome using our trained neural network which shows around 95% efficiency. We are able to compare the efficiencies between two type of methods used for predicting the users behaviors those are ANN using Back propagation method and other is logistic regression. ANN is observed to be more efficient than logistic regression because ANN has around 95% efficiency whereas logistic regression has around 85% efficiency. However we can increase the efficiency of our neural network by adding more hidden layer. Here we have restricted our hidden layer to two because of the computation power of the computer. Thus increase in the number of hidden layer will consume extra time for computation but will result better efficiency. Thus we are able to successfully predict the customer's behavior using artificial neural network model using back propagation method. And we can say that ANN is better method among the two methods for predicting user behavior on e-commerce site.

## References


[1] Davis F.D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13:318-339.

[2] Moon, J., Kim, Y. Extending the TAM for a world-wide-web context. Information and Management 2000;38:217-230.

[3] Vidushi, D., Singh, Y.: SOM improved neural network approach for next page prediction. IJCSMC **4**(6), 175–181 (2015)

[4] Maditinos, D. I., Sevic, Z., Theriou, N. G., & Tsinani, A. V. (2007). Individual investors' perceptions towards dividends: the case of Greece. *International journal of monetary economics and finance*, *1*(1), 18-31.

[5] Zhang, X., Edwards, J., & Harding, J. (2007). Personalised online sales using web usage data mining. *Computers in Industry*, *58*(8-9), 772-782.

[6] Lin, C. Y., Fang, K., & Tu, C. C. (2010). Predicting consumer repurchase intentions to shop online. *JCP*, *5*(10), 1527-1533.

[7] Dinwiddy, C. L., & Teal, F. J. (1996). *Principles of cost-benefit analysis for developing countries*. Cambridge University Press.

[8] Suchacka, G. (2013). Statistical analysis of buying and non-buying user sessions in a Web store. *Information Systems Architecture and Technology—Network Architecture and Applications*, 163-172.

[9] Suchacka, G., & Chodak, G. (2013, June). Practical aspects of log file analysis for e-commerce. In *International Conference on Computer Networks* (pp. 562-572). Springer, Berlin, Heidelberg.

[10] Abramowicz, W., & Mayr, H. C. (2006, May). Business information systems. In *9th International Conference on Business Information Systems, BIS*.

[11] Shih, H. P. (2004). An empirical study on predicting user acceptance of e-shopping on the Web. *Information & Management*, *41*(3), 351-368.

[12] Bhatnagar, A., Misra, S., & Rao, H. R. (2000). On risk, convenience, and Internet shopping behavior. *Communications of the ACM*, *43*(11), 98-105.