Improving Customer Segmentation in E-Commerce using Predictive Neural Network

Himanshu Singh[1], Dr. Shubhangi Neware[2]
[1]M.Tech. student, Department of Computer Science and Engineering,
Shri Ramdeobaba College of Engineering and Management, Nagpur, India, singhhh@rknec.edu
[2]Assistant Professor, Department of Computer Science and Engineering,
Shri Ramdeobaba College of Engineering and Management, Nagpur, India, newares@rknec.edu

ABSTRACT

Segmentation is a critical enabler to achieve business objectives and realize benefits in order to satisfy and meet the needs of the customers. The main aim of the customer segmentation is to divide the target market into subsets that share similar characteristics, needs and priorities. In this paper, we have performed predictive neural network approach based on the parameters like product reviews, products, buying pattern, viewing pattern and time based segments and clustering techniques in the sector of electronic commerce. We have found the distinct brands per customer and evaluation of unigrams, bigrams and trigrams have been done as a part in the feature extraction. Also, neural network is used for finding out the most preferred brands which the customer prefers and thereafter the statistical analysis is done from the predicted data. As we know that the sentiment mining and classification plays a major role in predicting what the customer thinks about the products, brands, etc. Finally we will be evaluating the accuracy of the predictions to find out the best suitable brands from the given set of the input data.

Key words : Cluster segmentation, Customer analysis, Neural network, Segmentation

1. INTRODUCTION

The first step in the customer segmentation is that knowing the customer, because if you know the complete profile of the customer – psychological, geographic, preferred channels and sales – then you can build and apply more effective marketing strategies. The main goal of the customer segmentation is to better satisfy customer needs or wants. Segmentation, targeting and positioning (STP) may be a customer analytics methodology helps in identifying and choosing groups of potential customers. Many researchers now aim on clustering customers from the transactional data although there is an intense increase in the collection of customer behavior data [1]. The market segmentation may be defined as the technique which involves the identification of major groups within the general market that shares unique characteristics and consuming habits so as to provide the authorities to customize the products or services to satisfy their needs [4]. Creating detailed customer profiles can help us in identifying new sub-segments for existing products, or tailor existing products to meet the needs of a particular customer segment. You want to regularly review and refine your customer segments, since customer behaviors, perceptions and demographics change over time. As we know that the e-commerce industries and organizations depends on the customer segmentation to target particular customer groups with the data and products that customers inside a segment would likely find pertinent [3]. The customer purchases database consists of four sets of customers who have purchased the various products from the company's online website. The profile for each customer includes the unique id, product name, product URL, retail and discounted price, brand and product specifications. The results that we have obtained by performing the segmentation are the best suitable brands out of the several brands which the customer likes to prefer. The best preferred brands are taken as the result in our findings of the overall segmentation. Unique challenges for customer segmentation efforts attempting to depend on online customer data are presented by the major online social media platforms which are used for the products present and distributing content [3]. In good strategy, we are always looking on growing organizations whether looking for more profits or we can say greater social impact and it can be done by providing values. Therefore, in order to grow; targeting the appropriate and right customer group is necessary. The segmentation solution is created through a rigorous and iterative process which is very much essential from the customer's point of view. It includes the data processing or factor analysis of the data of various customers. Then the cluster analyses are done in order to segment the data according to various parameters. Finally, the review and refinement process takes place which ensures that the formed and selected clusters are best chosen. A good

customer service provides an experience that meets the customer expectations rather it produces the satisfied customers. However a bad customer service can generate complaints which can result in lost sales, because the customers might take their business to a competitor. Therefore by analyzing research conducted in this particular domain, this paper provides an overall picture of the current situation of the online market contributing to a better understanding of the scenario.

2. LITERATURE REVIEW

The key to meaningful segmentation is to define customer variables and attributes that are relevant to your unique business. The customers are becoming more concerned and sophisticated in how they navigate their shopping choices, and the online retailers are discovering the one-size-fits-all marketing approaches that aren't so effective any more. Xiaojun Chen, Yixiang Fang, Min Yang, Feiping Nie, Zhou Zhao and Joshua Zhexue Huang suggested a partitioned clustering algorithm named "PurTreeClust" for the faster clustering of customer's transaction records where one of the major key element in achieving successful modern marketing and customer relationship management is the customer segmentation or the clusters of customers [1]. There is both a science and an "art" to sculpting and processing a successful segmentation. Since, it is the foundation for distinctive and sustainable competitive advantage. A customer segmentation can led to new value propositions by using the different customer segments to develop tailored value propositions. The segmentation area profiles will help you to develop new audiences by pinpointing where you are most likely to find the group of people from particular segment. Rachid Ait daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb in their paper propose that the LRFM (length, recency, frequency and monetary) model and clustering techniques helps in segmenting the customers and describes that the mode of payment is an essential factor in the ecommerce business [4]. Since, the insight is not meant to replace made to measure segmentation of audiences, participants. Rather it provides the missing link that allows you to put this internal view into an external context so that we can analyze the potential of the growing customers. Also, Shahriar Akter & Samuel Fosso Wamba propose that the big data analytics provides improvement in the online industry by making right use of people and modern technologies to modify the data into the perception for decision making and the major components of e-commerce include product development, operations, marketing, finance, human resource management and information systems which were studied and proposed in business value and big data analytics (BDA) [8]. Segments are not necessarily predictive in nature, although they are generally descriptive and serve as a type of classification that can be used to aid in understanding the future behaviors and the needs of customer.

3. METHODOLOGY

In this section the proposed model to evaluate the accuracy of the predictions is described out of various brands. The main purpose of this model is to performing customer segmentation using predictive neural network and statistical analysis to evaluate the accuracy measurement by plotting a graph of number of recommendations v/s accuracy and number of recommendations v/s delay. The plotted graph gives us an idea about the statistical analysis of the number of recommendations of a particular brand which a customer prefers and the most suitable and preferred brands a customer wishes to purchase.

3.1 Understanding Data

The dataset used in this model is of an online store which is obtained from its e-commerce website. With the help of this data of various customers we will be able to analyze variety of brands, detailed information regarding the customer and its transactional history. Also, for a particular customer the individual dataset contains unique id, product URL, product name, retail price, discounted price, description, rating, brands, etc. We have performed segmentation on the parameters like product reviews, products, buying pattern, viewing pattern and time based segments.

Table 1: Online shopping details for particular customer

uniq_id	crawl_timestamp	crawl_timestamp2	product_url	product_name	product_category_tree	pid	retail_price	d
c2d766ca9{	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2FF!	999	
7f7036a6d!	3/25/2016 22:59	68024	http://www.fli	FabHomeDecor Fa	["Furniture >> Living Roon	SBEEH3Q{	32157	
f449ec65d{	3/25/2016 22:59	68024	http://www.fli	AW Bellies	["Footwear >> Women's F{	SHOEH4GI	999	
0973b37ac{	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2F6	699	
bc940ea42{	3/25/2016 22:59	68024	http://www.fli	Sicons All Purpose	["Pet Supplies >> Groomir	PSOEH3ZY	220	
c2a173139!	3/25/2016 22:59	68024	http://www.fli	Eternal Gandhi Su	["Eternal Gandhi Super Se	PWTEB7H;	430	
ce5a6818f7	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2FV	1199	
8542703ca{	3/25/2016 22:59	68024	http://www.fli	FabHomeDecor Fa	["Furniture >> Living Roon	SBEEH3Q{	32157	
29c8d290c{	3/25/2016 22:59	68024	http://www.fli	dilli bazaaar Belli	["Footwear >> Women's F{	SHOEH3D;	699	
4044c0ac5{	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2FV	1199	
e54bc0a7c{	3/25/2016 22:59	68024	http://www.fli	Ladela Bellies	["Footwear >> Women's F{	SHOEH4KI	1724	
c73e78fb4{	3/25/2016 22:59	68024	http://www.fli	Carrel Printed Wo	["Clothing >> Women's Cl{	SWIEHF3E	2299	
c29af37837	3/25/2016 22:59	68024	http://www.fli	Sicons All Purpose	["Pet Supplies >> Groomir	PSOEH3ZYYFETGCCD		
9aacdeccel	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2FG	999	
83c53f8948	3/25/2016 22:59	68024	http://www.fli	Freelance Vacuun	["Pens & Stationery >> Sch	BOTEGYTZ	699	
d95b0456a	3/25/2016 22:59	68024	http://www.fli	Alisha Solid Wom	["Clothing >> Women's Cl{	SRTEH2FE	999	
849ab0569:	3/25/2016 22:59	68024	http://www.fli	FabHomeDecor Fa	["Furniture >> Living Roon	SBEEH3Q{	32157	

The mentioned Table 1 describes about the online shopping details for a particular customer. The dataset is taken from the online website store in order to perform the processing on it. For a particular customer the shopping details are unique id, product URL, product name, retail price, discounted price, description, rating, brands, etc. regarding the purchase. On the given dataset the pre processing is done in order to remove the missing entries. Also, feature extraction is done using the calculation of unigrams, bigrams and trigrams. The dataset contains unique values which are used further in the prediction purpose of the research.

3.2 Data Preprocessing for Segmentation and Methods

The data preprocessing may be defined as an important step in the data mining process. The representation and quality of data is necessary and foremost before starting an analysis. In the phase of machine learning project, the data preprocessing seems to be important perspective. The data preparation and filtering steps can take adequate amount of processing time. The data preprocessing includes cleaning, normalization, selection, feature extraction, instance selection, etc. While the customer table contains fields such as unique id, product URL, product name, retail price, discounted price, description, rating, brands, etc. The market consists of various customers and the customers differ in one or more ways. Through the market segmentation organizations separate large, heterogeneous buyers into smaller fragments that can be reached more productive and fruitful with products and services that fulfill their respective needs. The different levels of markets are being proposed which includes mass marketing, segment marketing, niche marketing, and micro marketing. Since, the customers have unique needs and wants we may consider that each customer is potentially a separate market.

Figure 1: Framework for customer segmentation using predictive neural network

Figure 1 describes about the framework for performing the customer segmentation using the predictive neural network approach. For starting our research, we take the input as the ecommerce dataset which is available through online medium. Pre-processing is performed on the data in order to remove the missing entries in the dataset. The dataset should not contain any of the wrong or missing entries so that the error can be minimized. After that feature extraction is performed with the help of unigrams, bigrams and trigrams calculation. The calculation is based on the count of unique brands a particular customer chooses based on his preference. Then neural network is used for the finding out the most

preferred brands. The neural network prediction includes the calculation of score by adding the unigram, bigram and trigram values. Therefore, we then find out the price patterns and discount patterns for the brands of the particular customers. Finally, we recommend the best brands as our outputs by performing the statistical analysis. The recommendations of the best brands are based on the calculation of accuracy measurement which can be calculated by finding out the correct recommendations and dividing them by the total number of recommendations. Suitable accuracy graphs are prepared in order to depict the accuracy measurement for the particular customers.

A neural network may be defined as a network or circuit composed of artificial neurons or nodes, for solving artificial intelligence (AI). We can suggest that the artificial networks can be used for determining predictive modeling, adaptive control and the services where they might be trained through a dataset. The neural networks can be used in different fields such as function approximation, regression analysis which includes the modeling and time series prediction. Also, classifications including pattern and sequence recognition, etc. are its related field. Recent surveys have revealed that customer relationship management (CRM) systems are taking the place of traditional mass marketing strategies by personalized marketing practices [10].

In our model, the statistical and data mining techniques have been used to analyze the results from the dataset. The data mining techniques helps in discovering interesting patterns or the bonding in the data and predict the behavior by fitting a model based on that particular model [10]. This paper proposes a neural network (NN) based approach to predict the accuracy based on plotting of graphs. Using the regression approach to make predictions doesn't necessarily involve prediction of future. Instead, we predict the mean dependent variable given specific values of the dependent variables. We need to collect data for relevant sources, formulate a model and examine how well the model matches the data. The general procedure for using regression to make suitable predictions is starting the research in the subject area so that you can build on the work of others. Then collect the data for the relevant variables and specify your regression model. If you have a model that adequately fits the data use it to make predictions accordingly.

In this paper the proposed model works on the basis of neural network prediction and results the statistical analysis with the help of plotting graph. Our main focus will be on the following criteria based on the parameters like product reviews, products, buying pattern, viewing pattern and time based segments. The dataset is taken where the individual customer parameters are taken such as unique id, product URL, product name, retail price, discounted price, description, rating, brands, etc. We then calculate the number of brands and products according to their unigrams, bigrams and trigrams values. The n gram is a continuous sequence of n

items from a given sample of text. N-grams of texts are majorly used in text mining field and natural language processing tasks. As a result we get brand name and its unigrams, bigrams and trigrams values respectively. Thereafter, by using these unigrams, bigrams and trigrams values of particular brands which are a part of feature extraction we will recommend the best suitable brands for the prediction.

The result is shown by the prediction output which involves the brands along with their prediction index values respectively. Further the analysis is done of the particular brands in order to find out the average price and the discounted price. By taking the previous values along with the prediction index we now try to analyze the particular brands by showing the number of times user has purchased a particular brand, average price, average price after discount and the average overall discount price for that particular brand. Finally, by considering the previous values we calculate the delay time needed and the percentage of accuracy prediction with the help of neural network. The results are shown with the help of graph in which we will be plotting the graphs between the number of recommendations v/s accuracy and number of recommendations v/s delay.

A more detailed analysis regarding the accuracy measurement and the statistical analysis is shown with the help of graph. With the help of results we get an idea about the accuracy measurement that which are the most preferred brands the customer likes to purchase. With the help of statistical analysis it gives us a clear image about the processing we have performed on the dataset and what we obtain as a result. It helps in the overall customer segmentation process and building of various strategies which helps to enhance and nourish the ecommerce market which is growing larger day by day.

3.3 Result and Analysis

The statistical analysis was performed by plotting the graphs between the number of recommendations v/s accuracy and number of recommendations v/s delay. Here, we can observe that the table of four different users along with their number of recommendations or predictions and the accuracy percentage and delay time measured in seconds are formed for plotting the graph. For the particular user or customer respective numbers of recommendations are obtained along with accuracy percentage. The accuracy percentage denotes that how much accuracy we have achieved from the total number of mentioned recommendations present in the dataset. The accuracy measurement is done by finding out the correct recommendations and dividing them by the total number of recommendations.
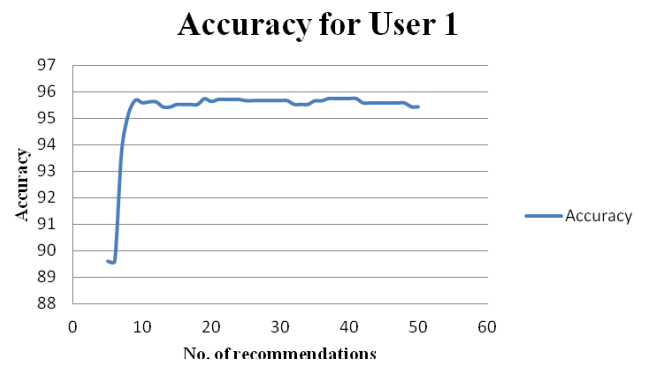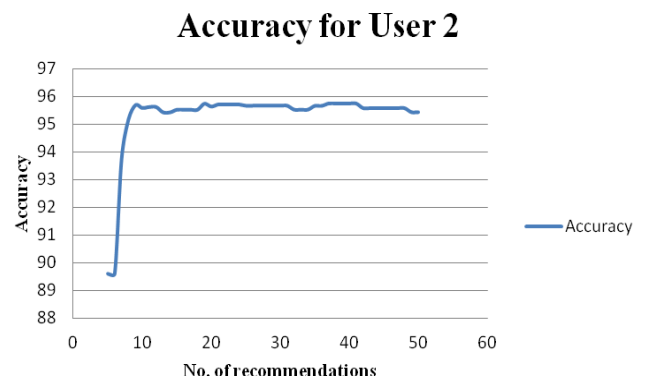

Figure 2: Accuracy graph for User 1
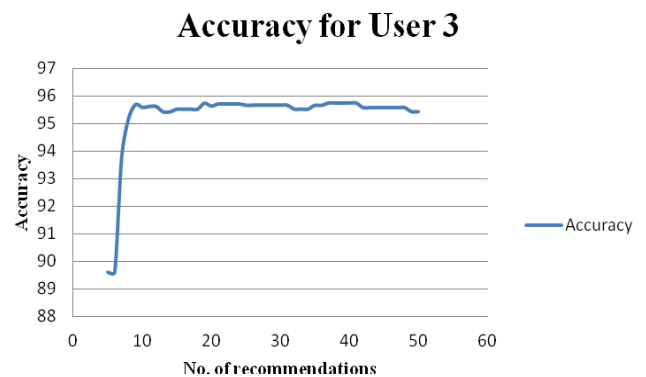

Figure 3: Accuracy graph for User 2

Figure 4: Accuracy graph for User 3
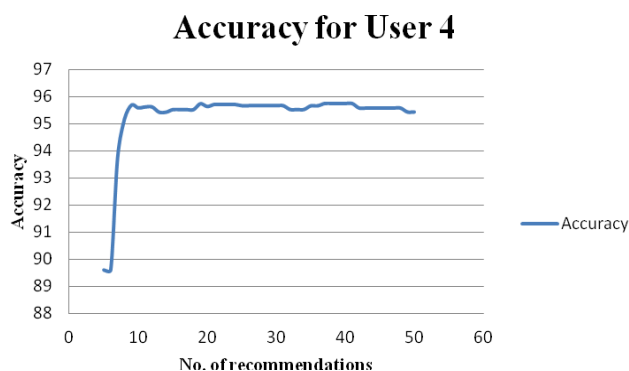
Accuracy for User 4

Figure 5: Accuracy graph for User 4

Therefore, by analyzing the graphs of the individual four users which we have taken we can clearly observe that the accuracy improves with the number of recommendations or predictions. As compared to the other papers and research work we can slightly say according to the results that the accuracy improves as we go on adding or multiplying the number of predictions. This may be very much helpful in the e-commerce industry as customer choices and their overall shopping strategy can be predicted. Figure 2 describes the accuracy chart for the user 1 where we can see the accuracy v/s number of recommendations. Clearly it is seen that as we go increasing the number of recommendations the accuracy increases gradually. The accuracy percentage is reaching up to 95% as we can see from the obtained graphs. This helps in the overall evaluation of the statistical analysis. Similarly, Figure 3, Figure 4 and Figure 5 respectively depict the graphs of the user 2, 3 and 4 of accuracy v/s the number of recommendations. Hence, by predictive neural network approach we can achieve the accuracy which can lead to better segmentation of customers.

4. CONCLUSION

Therefore, this paper aims to propose a model by segmenting a group of customers in online market by using the predictive neural network approach and statistical analysis. Here we have done the prediction of accuracy and found that the accuracy percentage increases as we increase the number of recommendations. The accuracy is measured by finding out the correct recommendations and dividing them by the total number of recommendations. After the analysis we conclude that the accuracy may be taken as an important factor in determining the prediction of best brands out of the several brands which we have taken and ensures that the recommendations also play a major in terms of individual customers.

The results have depicts that the accuracy percentage may be taken as an important parameter for the analysis purpose for the prediction. The research also tries to contribute towards the understanding of the customer behavior and the patterns regarding the analysis. Despite of the major advancement in this sector, in some areas more emphasis is given on the customer behavior and their prediction analytics. The analysis of parameters has also shown that the number of recommendations is certainly a major and key factor in defining the best suitable brands and also helps in the prediction purpose in the ecommerce market industry. Finally, we can analyze that provided with the mentioned strategies we can contribute towards the betterment and enhancement in the prediction analysis sector in the online store industries.

REFERENCES

1. Xiaojun Chen, Yixiang Fang, Min Yang, Feiping Nie, Zhou Zhao and Joshua Zhexue Huang, "**PurTreeClust: A Clustering Algorithm for Customer Segmentation from Massive Customer Transaction Data**" in *IEEE Transactions on Knowledge and Data Engineering*, IEEE, 2017, pp. 1-3.
2. Peter Kolarovszki*, Jirí Tengler, Margita Majercáková, "**The new model of customer segmentation in postal enterprises**", Elsevier, 2016, pp. 121-123. https://doi.org/10.1016/j.sbspro.2016.09.015
3. Jisun An, Haewoon Kwak, Soon‑gyo Jung, Joni Salminen, Bernard J. Jansen, "**Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data**", Springer, 2018, pp. 1-4.
4. Rachid Ait daoud, Abdellah Amine, Belaid Bouikhalene, Rachid Lbibb , "**Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco**", International Journal of Computer and Information Engineering , 2015, pp. 2000-2009.
5. María Teresa Ballestar, Pilar Grau-Carles, Jorge Sainz, "**Customer segmentation in e-commerce: Applications to the cashback business model**", [Elsevier] *Journal of Business Research*, 2017, pp. 1-4.
6. Pedro Quelhas Brito, Carlos Soares, Sérgio Almeidab, Ana Montea, Michel Byvoetc , "**Customer segmentation in a large database of an online customized fashion business**", [Elsevier] *Robotics and Computer-Integrated Manufacturing*, 2015, pp. 1-4.
7. Alexandra Amado, "**Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis**", [Elsevier] *European Research on Management and Business Economics*, 2017, pp. 1-3.
8. Shahriar Akter & Samuel Fosso Wamba, "**Big data analytics in E-commerce: a systematic review and agenda for future research**", Springer, 2016, pp. 173-176. https://doi.org/10.1007/s12525-016-0219-0

9. Er. Rupampreet Kaur, Er. Kiranbir Kaur, "**Data Mining on Customer Segmentation: A Review**", IJARCS, Vol 8, No. 5, 2017, pp. 857-859.

10. Anuj Sharma, Dr. Prabin Kumar Panigrahi, "**A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services**", *International Journal of Computer Applications,* Vol 27, No. 11, 2011, pp. 26-29.
https://doi.org/10.5120/3344-4605

11. Aaron Don M. Africa, Lourdes R. Bulda, Matthew Z. Marasigan, Isabel F. Navarro, "**A Study on Number Gesture Recognition using Neural Network**", *International Journal of Advanced Trends in Computer Science and Engineering* [IJATCSE], Vol 8, No. 4, 2019, pp. 1076-1078.
https://doi.org/10.30534/ijatcse/2019/14842019

12. Aaron Don M. Africa, Francis Xavier Asuncion, Janos Lance Tiberio, Raymund Miguel Franciso A. Munchua, "**Sensor-based Traffic Control Network with Neural Network Based Control System**", *International Journal of Advanced Trends in Computer Science and Engineering* [IJATCSE], Vol 8, No. 4, 2019, pp. 983-984.
https://doi.org/10.30534/ijatcse/2019/01842019