

Discovery of Customer Behavior with Data Mining in E-Commerce

Dr. J. Umamaheswari

Faculty on contract, School of Maritime Management, IMU, Chennai

Dr. R. Gokulakrishnan

Additional Director/Scientist-E, STPI Chennai, MeitY, Government of India

Article Info

Volume 82

Page Number: 5986 - 5995

Publication Issue:

January-February 2020

Abstract:

Data Mining is one of the basic techniques of extracting the required data from the huge collection of datasets produced in the various fields. This involves business area too whereas the significance of data mining benefit in business application is by discovering associative data and knowledge that can be utilized to produce good decision making over business area. However, the current e-commerce has been developing firmly which produces the service and resources over internet consistently. Simultaneously, the circumstance of fraudulent proceeding with users has reached the e-commerce system. In this period, e-commerce is deliberated as a killer-domain to succeeded mining data that provide appropriate factors from case to case but the most traditional method which e-commerce is able to process is Customer Relationship Management (CRM). Hence, this paper discusses about the data analysis that deals with the design of predicting customer behavior and relationship to manage the e-commerce business in a systematic manner. This proposed model has integrated K-Means clustering technique and classification using Neural Network (NN) as the data mining techniques for dealing both attributes of discrete and continuous to extract the weight of 5 major factors with customer credit card datasets for hidden knowledge. Thus, it efficiently supports in discovering the actual customer behavior and even classify recent customer in future with less time to improve CRM recommendation to customer in improving e-commerce business prediction performance.

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 29 January 2020

Keywords: Data mining, e-commerce, Customer Relationship Management (CRM), Naïve Bayes, Neural Network classification algorithm.

1. Introduction

In worldwide, the e-commerce has become more familiar over past decade and online shopping is also become a current trend which provide enormous useful for customers by creating foreign goods availability at minimum cost with immense options [1]. However, the accelerated generation of personal electronic devices, internet implementation and globalization are the major factors considered for driving the e-commerce diffusion. The main development in assurance of

online security and internet communication for commercial usage has provided the e-commerce growth [2]. Moreover, e-commerce is an activity of international trading that massively depends upon logistics for ensuring the product of online getting transported to the exact customer's location at right time. Hence, it assists in maintaining the measure with its increase of demand in world economy [3]. There are various technologies integrated with e-commerce namely internet marketing, electronic fund transfer, online transaction process, automate data collection

method and Electronic Data Interchange (EDI). The current utilization of WWW in e-commerce has produced minimum over transactional life cycle while it gets involved a large technologies range namely mobile devices, e-mail and internet through wired telephone. The electronically organized trade amount has developed unusually with usage of wide spread internet. Nowadays, the major participants of e-commerce in the world are Amazon, Flip kart and eBay which have huge product database and consumer count of utilizing those services whereas the policy, product and pricing has been varying by country wise. In different websites, mining the e-commerce data contains the available products and corresponding prices has been offered. These can be useful for companies at various pricing levels who offer the products and the choice is expanded for consumers. The companies can offer the various products and prices for consumers by the classification of e-commerce data. These will give the understanding in the market by which products are popular and which customers are eager to pay for definite products. The classification based on automation for both companies and customers when it comes to product type, reducing cost of price and timing. With the help of this classification data, companies can offer best services whereas customers can get the best deal for their products which have to raise their customer base. In recent times, the essential way to understand customer base by growing rapidly with the e-commerce business so that company is more attractive. The important way of e-commerce in data mining had moved for supporting the enterprise in e-commerce companies with the essential information regarding the business. Nowadays, many companies in possession of bid data which have to adopt e-commerce in their data repositories. This is the only way to increase the decision making by getting the data to mine or to enable the business intelligence [4]. Hence, when compared with traditional e-commerce business, there has large

number of customers are available as both business customers and individuals. The managing customer requirement has becomes very critical. In e-commerce industry, CRM is the essential supporting for constructing profitable relationships to support data analysis by designing CRM in a systematic way.

2. Literature Review

E-commerce has been developing worldwide business trend in recent years with the emergence of communication technology and personal computers, also rise in the consumer use of the Internet. In 2017, worldwide e-commerce retail sales had been exceeded US\$ 2.3 trillion and are estimated to continue over the years for further hitting US\$ 4.88 trillion in 2021[5]. Laudon and Traver[6], discusses e-commerce using internet and web which refers to professional transaction between individuals and businesses. There are various categories namely B2B and B2C e-commerce operations between business institutions of goods and services over the electronic media which refers to the direct exchange between an enterprise and the customer respectively [7].

There are large amount of e-commerce transactions has dispersed when compared to traditional logistics which leads to a numerous e-commerce clients. It is essential to accomplish customer relationships in an organized manner with the large consumer base [8]. CRM has described as a process chain and supporting processes to establish and maintain customer relationships, customer retention, improve customer acquisition, business profitability and customer loyalty [9][10]. There are four different stages in customer life cycle namely customer recognition, customer engagement, customer retention and customer development [11]. The estimation of customer behavior becomes a strategically significant and difficult problem due to the wide-ranging of customer orders, high

volatility and expectations based business in the e-commerce [12]. CRM in e-commerce logistics have important of customer behavior in decision-making and the companies are able to recognize in improving competitiveness [13][14]. Pan et al. [15] described to refine home delivery strategies for e-grocery by predicting customer behavior using customer-related data to minimize failure delivery levels. Wong and Wei [16] built an online customer behavior analysis platform by segmenting high-value consumers, eventually forecasting their next purchase and evaluating their online buying activity to optimize customer relationships.

Data mining methods were introduced to find the hidden customer information and customer behavior in order to learn more about customers [17]. The process of data mining is used for hidden patterns and discovering information through the convergence of artificial intelligence, analytics and machine learning from large data sets [18]. The traditional computational data mining methods are clustering and decision trees where clustering is based on similarity for defining the intrinsic structure of data and grouping the data artifacts into clusters [19]. In order to evaluate hidden data definitions, it is an unsupervised method for sorting data items into relevant clusters. In exploratory the pattern analysis, alliance, decision-making based machine-learning circumstances, and data mining is an essential and useful [20]. Amine et al. [21] proposed a clustering method to establish current marketing approaches for assessing the preferences of consumers on Morocco's e-commerce websites. Cao et al. [22] examined consumer purchasing data, assessed consumers based on the organized weighting average and the k-means algorithm in terms of Regency-

Frequency-Monetary (RFM) Value. Wang et al. [23] described to devise different types of consumer relations and marketing strategies which are grouped customers using the k-means method. Another important method in data mining is the decision tree, which is primarily used for model grouping and estimation. According to Tan et al. [24] input attribute set mapped to task of learning model classification into predefined class name. Classification of decision trees have allows exploring and generating a series of models created on the study of a collection of training data that divides the data into groups. The training data contains the test items which categorized as a class is also known supervised learning. Ma [25] directed by customer segmentation for improving customer loyalty of e-commerce website using decision trees. On the basis of the decision tree method, Guo and Qin [26] examined the attributes of consumer churns in e-commerce. Xu et al. [27] established a clustering decision-making tree-based model to enable mobile CRM facilities.

3. Design of predicting customer status model

One of the traditional features in e-commerce is customer management which has altered from the period of small societies and even the trend of e-commerce or e-business is to maintain the customer satisfaction from the shopkeepers that represent relationship management as an objective. Data mining can be used to foresee the kind of relationship apart from providing contacts, billing and retentions for the particular people. In order to predict the customer behavior, CRM systems has the most essential use of data mining. It involves value assessment, capturing appropriate customer information and behavior modeling. The proposed model in figure.1 illustrated the appropriate predictive model where the historical data are collected from the customer.

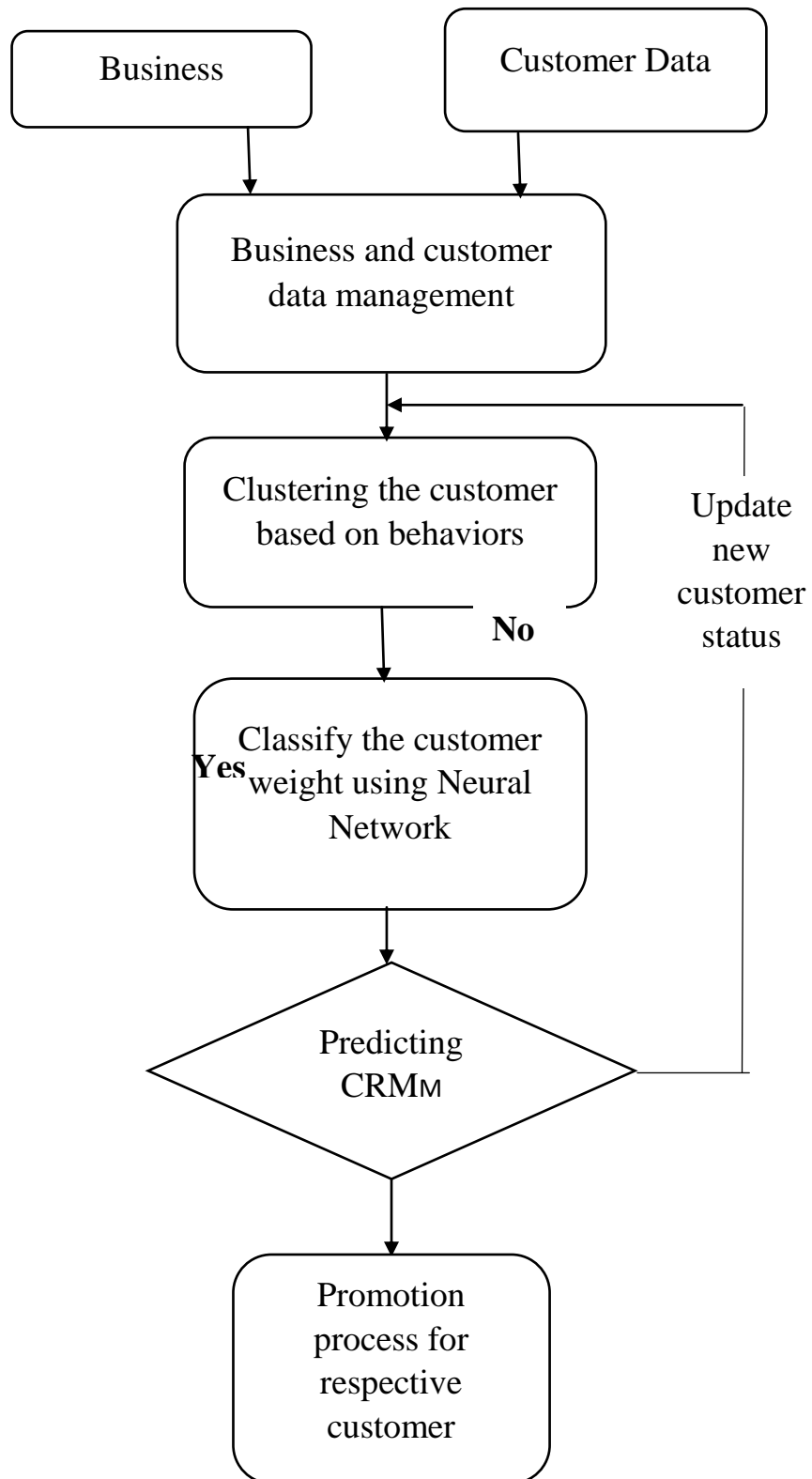


Figure 1 Proposed model of customer behavior classification in e-commerce using Neural Network

3.1 Business and customer data management

In earlier, customer management was spread across various departments throughout the business. Instead they had to be harmonized. In case of business and technical point of view this was necessary also customer may connect via the web or call centers. The customer data should therefore be kept up-to-date allowing staff to communicate easily and gather new information. The CRM approach has to build a centralized understanding of the customer with incorporates the information from multiple sources and then create this consumer knowledge base for numerous analyses of CRM applications which has been accessible as source data. Then create new perspectives by every customer dealings based on customer behavior.

3.2 Clustering the customer based on behavior

While a variety of techniques are being introduced as new generation techniques, clustering could be a common strategy that could be used in e-commerce business strategies. The fragments of methodology tracks sub-sets in a database based on a set of attributes and grouped together with related records is known as clustering whereas similar values of records are grouped together for the attribute. Because there are broad collections of customer data in ecommerce, the method of estimation is done by clustering k means.

3.2.1 K-means clustering for Segmentation

K-means clustering is used by a distance calculation, the Euclidean Distance, to classify data items into k clusters. In this paper, clustering k-means allows extract hidden customer data interactions and the collection of large customer data is converted into classes based on the company's size. It also produces updated table with changed data after producing the effects of the k-means clustering which has been offering arbitrary segmentation that allows the data more relevant to the study of the decision tree. For

instance, it is possible to identify a group of customers in the business with large monetary value for each month. Even the same amount of money in the same month if a customer spent with two different businesses, it may have their business scales and models are different by means of different things to the two businesses. The grouped results vary according to the company's size. Range to Euclidean, as shown in Eq. (1) and (2) shall be used to calculate the distance to the centroid cluster between each point.

$$d(x_e - m_j) = \sqrt{(x_e - m_j)^2} = |x_e - m_j| \quad \dots\dots\dots(1)$$

$$m_j = \frac{\text{total value of data objects in the cluster}}{\text{total number of data objects in the cluster}} \quad \dots\dots\dots(2)$$

3.3 Classify the customer behavior weights using Neural Network

The data mining technique of NN classifier has been chosen and implemented in the customer classification analysis tier to analyze the existing customers by extracting the valuable customer behavior and knowledge. One of the major issues in this e-commerce business is building the customer relationship as success. As an instance, e-commerce business can be considered for both existing and recent customers. In the current scenario of e-commerce, the business has send verification code and e-mail messages to the customers based on their status which have taken from the historical data available in social media, other e-commerce business and cibil score from banking sector.

The current issues in the e-commerce business are predicting the customers who may cause endanger by overestimating and underestimating them has produced a loss in both short-term and long-term. This doesn't assist in building the relationship with each customer whereas CRM implementation in an organization has created high costs. Moreover, the organization need to endure the

costs for building relationship with customer which involves both discount cost for product and renewed products for customer. Hence, it is complex in identifying the benefits of customer is crucial. Thus, the frequency of identifying customer is complex due to the information that refer CRM concept to both monetary benefit and positive word of mouth about the product of e-commerce business on the consideration of business sorting. This research work has proposed Neural Network as a classifier for providing evidence in assist of decision making whether the business is profitable and essential to improve the relationship with customer or not.

The output for this model has represented solution to the e-commerce business. "Is the contribution from the business in the relationship with this customer is profitable?" is the question raised to analyze the business. Hence, the solution is in form of binomial probability based on the weighted vector generated from the network has

been converted as criterion function for taking binary decision. Thus, the criterion function is based on decision maker and considered as following equation 3 and 4:

$$Y_{dec} = 1 \text{ for } Y \geq Y_{cus} \quad \dots\dots\dots(3)$$

$$Y_{dec} = 0 \text{ for } Y < Y_{cus} \quad \dots\dots\dots(4)$$

Where,

Y_{dec} = Binomial probability of decision variable

Y = Decision value

Y_{cus} = Value of the customer generated by Neural Network

This proposed model architecture has utilized for accessing the profitability in building relationship of customer based on the CRM concept over e-commerce. The network structure has shown in figure-2.

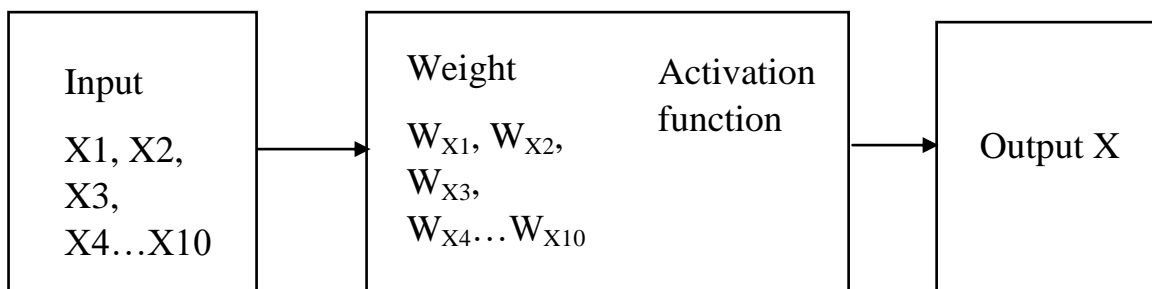


Figure 2. NN model to process an assessing profitability of building customer relationship in e-commerce

4. Results and Discussion

In this research, the proposed model has utilized for better regression in predicting the beneficial customers from the existing and analyzing the new customers. In this implementation, the customer credit card details as an input data with various attributes namely Customer ID (cust_ID),

Balance (Bal), purchase, purchase frequency (pur_Frq), Tenure, payments (Pay), credit Limit (CL), oneoff_purchase (OO_Pur), oneoff_purchase_frequency (OO_Pur_Frq), cash_Advance_TRX (Cash_adv_TRX), Purchase_TRX (Pur_TRX), Minimum payments (Min_pay), purchases_installments_frequency (Pur_Inst_Frq), are shown in the Table 1.

Table 1 Datasets for credit card details for E-Business

Cust_ID	Bal	Purchase	Pay	Min_Pay	CL	Tenure	Pur_Frq	Pur_Inst_Frq	OO_Pur	Oo_Pur_Frq	Inst_Pur
C101	40.90	95.40	201.80	139.51	1000.00	12	0.17	0.08	0.00	0.00	95.40
C102	3202.47	0.00	4103.03	1072.34	7000.00	12	0.00	0.00	0.00	0.00	0.00
C103	2495.15	773.17	622.07	627.28	7500.00	10	1.00	0.00	773.17	1.00	0.00
C104	1666.67	1499.00	0.00	0.0	7500.00	12	0.08	0.00	1499.00	0.08	0.00
C105	817.71	16.00	678.33	244.79	1200.00	8	0.08	0.00	16.00	0.08	0.00
C106	1809.83	1333.28	1400.06	2407.25	1800.00	12	0.67	0.58	0.00	0.00	1333.28
C1007	627.26	7091.01	6354.31	198.07	13500.00	12	1.00	1.00	6402.63	1.00	688.38
C1008	1823.65	436.20	679.07	532.03	2300.00	9	1.00	1.00	0.00	0.00	436.20
C1009	1014.93	861.49	688.28	311.96	7000.00	12	0.33	0.25	661.49	0.08	200.00
C110	152.23	1281.60	1164.77	100.30	11000.00	8	0.17	0.00	1281.60	0.17	0.00
C111	1293.12	920.12	1083.30	2172.70	1200.00	12	1.00	1.00	0.00	0.00	920.12
C112	630.79	1492.18	705.62	155.55	2000.00	6	0.25	0.00	1492.18	0.25	0.00
C113	1516.93	3217.99	608.26	490.21	3000.00	12	1.00	0.92	2500.23	0.25	717.76
C114	921.69	2137.93	1655.89	251.14	7500.00	7	0.75	0.75	419.96	0.17	1717.97
C115	2772.77	0.00	805.65	989.96	3000.00	11	0.00	0.00	0.00	0.00	0.00
C116	6886.21	1611.70	1993.44	2109.91	8000.00	12	0.50	0.50	0.00	0.00	1611.70
C117	2072.07	0.00	391.97	376.58	3000.00	12	0.00	0.00	0.00	0.00	0.00
C118	41.09	519.00	254.59	73.20	2500.00	12	0.42	0.33	0.00	0.00	519.00
C119	1989.07	504.35	1720.84	744.61	13000.00	10	0.67	0.58	166.00	0.08	338.35
C120	40.90	398.64	1053.98	12465.55	4000.00	11	1.00	1.00	0.00	0.00	398.64
C121	3202.47	176.68	223.07	13557.30	2000.00	9	0.67	0.67	0.00	0.00	176.68

The major and signified attributes are considered in this database whereas there are 1500 available records. However, the NN model has assist in manipulating the variance, z-score of each attributes and evaluates the weight of each attributes that are segregated in term of 5 factors by using of clustering method which is an advantage of this hybrid proposed model. The tenure, credit limit attribute weight is interpreted with factor (A), attributes purchase, payment and minimum_payment are weighted get interpreted with factor (B), attributes oneoff_purchase_frequency and purchase frequency are used for weighting factor (C), the attributes balance and cash_advance_TRX are utilized for weighted the factor (D) and the

attributes Purchase_Installation and Installation_Purchase_Frequency are interpreted for weighted factor (E).

In order to determine the profitable in building the customer relationship by analyzing the behavior factor in a CRM system for developing the business of e-commerce.

1. A - Solvency with weighted 0.2
2. B – Loyalty with weighted 0.25
3. C – Purchasing Frequency with weighted 0.15
4. D – Positive word of mouth over market with weighted 0.25
5. E –Money in single transaction with weighted 0.15

The NN Classifier model consists of randomly 10 inputs as random from the dataset get applied to the network that has determined in the earlier list. The value of above 5 factors is classified based on below segregation

- Very Low as 1
- Low as 2
- Moderate as 3
- High as 4
- Very High as 5

Moreover, the activated simulation function is calculated as bias along with weighted function as the concept of proposed NN model. The suitability of model has been verified based on accomplished data from random 10 clients as input are consider for classifying the customer with the trained data. Therefore, the simulation results is shown in Table 2 which consist of rated value from the clients and weighted value get calculated with decision value mentioned in term of “W” whereas the simulated value is mentioned in term of “S”. Thus, the result assist in promoting the customer whose Y_{Dec} is 1 and for the Y_{Dec} is 0 will be negligible.

Table 2. Simulation results for the cost-effectiveness analysis process in building customer relationships by NN model.

Customer ID	Customer Features					S	W
	A	B	C	D	E		
1	5	4	5	3	2	1	1
2	3	1	4	2	2	1	1
3	2	3	1	1	2	0	0
4	4	5	3	4	3	1	1
5	5	3	2	3	5	1	1
6	1	2	3	2	3	0	0
7	5	3	4	4	4	1	1
8	4	3	5	4	2	1	1
9	1	2	4	3	1	0	0
10	2	1	3	2	3	0	0

Moreover, the prompted customers may be updating e-commerce business customer about their status. If the business uses the weighted value of NN and predicts a customer who has already purchased the product frequently and loyalty to the business as the benefits. Based on the classification of NN using weighted and activation function the e-banking / e-commerce business can be able to identify the customer behavior and the relation in order to prompt customer with several offer and benefits. Such systems should have rating and feedback of the e-banking / e-commerce business about the old customer behavior and the check for the relationship of the new customer to prompt their product using the historical data. In addition, the requirement of human intervention from case to case performs a successful data mining with definite predictive models for preventing frauds in e-commerce.

4. Conclusion

The several activities in the business can be beneficial from mining e-commerce data even though they are not performed by statistician over economy and financial plan for future. However, the country is facing a huge conclusion involving recessions. In addition, data mining pattern assist in analyzing the data and submit statistics whereas the knowledge can able to reflect strategy of e-commerce systems. The proposed NN model has predicted the customer behavior based on the factors weighted in order to determine the profitability in building relationship of customer for CRM concept in e-commerce. This result of mining has assisted e-commerce businesses by the weighted factors to simulate. It can be interpreted hidden knowledge in performing international business with continual guidance for the regulations of ever changing business. Moreover, the proposed model has presented a model for e-commerce system to manage the large customer bases, and building long-term and profitable relationships. Hence, it can efficiently support in

discovering the actual customer behavior and even classify recent customer in future with less time to improve CRM recommendation for the customers in improving e-commerce business prediction performance. Thus data mining has created astonishingly impressive sales force in ecommerce.

Reference

- [1] M. Falk, E. Hagsten, "E-commerce trends and impacts across Europe" *International Journal of Production Economics*, vol. 170, pp. 357-369, 2015,.
- [2] R. Rahayu, J. Day, "Determinant factors of e-commerce adoption by SMEs in developing country:evidence from indonesia" *Procedia Soc Behav Sci.*, vol. 195, pp.142-150 ,2015.
- [3] C.C. Luk, K.L. Choy, and H.Y. Lam, "Design of an Intelligent Customer Identification Model in eCommerce Logistics Industry" , MATEC Web of Conferences, 2019.
- [4] Mustapha Ismail, Mohammed Mansur Ibrahim, Zayyan Mahmoud Sanusi, and Muesser Nat, "Data Mining in Electronic Commerce: Benefits and Challenges",*Int. J. Communications, Network and System Sciences*, vol.8,pp. 501-509,2015.
- [5] X. Chen, H. Lin, "Research on e-commerce logistics system informationization in Chain"*Procedia Soc Behav Sci.*, vol.96, pp.838-843, 2013.
- [6] K.C. Laudon, C.G. Traver, *E-commerce*. Pearson (2013).
- [7] R. Mangiaracina, G. Marchet, S. Perotti, A. Tumino, "A review of the environmental implications of B2C e-commerce: a logistics perspective" *Int. J. Phys. Distrib. Logist. Manag.*, 45(6), 565-591 ,2015.
- [8] J. Bhattacharjya, A. Ellison, S. Tripathi, "An exploration of logistics-related customer service provision on Twitter: The case of e-retailers " *Int. J. Phys. Distrib. Logist. Manag.*, vol.46(6/7),pp. 659-680 ,2016.
- [9] E.W. Ngai, L. Xiu, D.C. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification"*Expert Syst. Appl.*, vol.36(2), pp.2592-2602, 2009.
- [10] R.S. Hassan, A. Nawaz, M.N. Lashari, F. Zafar, "Effect of customer relationship manage satisfaction" *Procedia Econ. Financ.*, vol.23, pp.563-567, 2015.
- [11] N. Hill, J. Alexander, *The handbook of customer satisfaction and loyalty measurement*. Routledge ,2017.
- [12] P.A. Sarvari, A. Ustundag, H. Takci, *Kybernetes*, "Performance evaluation of different customer segmentation approaches based on RFM and demographics anaysis"vol.45(7), pp.1129-1157 ,2016.
- [13] S. Akter, S.F. Wamba, "Big data analytics in e-commerce: a systematic review and agenda for future research" *Electron. Mark.*, vol. 26(2), pp.173- 194, 2016.
- [14] I. Gamayanto, H. Christian, "The Development of Innovative CRM E-commerce: The case of Blibli.com",*Binus Bus. Rev.*, vol.9(1),pp. 29-38,2018.
- [15] S. Pan, V. Giannikas, Y. Han, E. Grover-Silva, B. Qiao, "Using Customer-related Data to Enhance E-grocery Home Delivery"*Ind. Manage. Data Syst.*, vol.117(9), pp.1917-1933, 2017.
- [16] E. Wong, Y. Wei, "Customer online shopping experience data analytics: Integrated customer segmentation and customised services prediction model"*Int. J. Retail Distrib. Manag.*, vol.46(4), pp. 406-420, 2018.
- [17] I.T. Afolabi, O. Oladipupo, R.E. Worlu, I.O. Akinyemi, "A Systematic Review of Consumer Behaviour Prediction Studies",*Covenant Journal of Business and Social Sciences*, vol.7(1),pp. 41-60, 2016.
- [18] I.H. Witten, E. Frank, M.A. Hall, C.J. Pal, "Data Mining: Practical machine learning tools and techniques", Morgan Kaufmann, 2016.
- [19] B. Liu, Y. Xia, P.S. Yu, "Foundations and advances in data mining" pp. 97-124, Springer Berlin Heidelberg ,2005.
- [20] X. Wu, X. Zhu, G.Q. Wu, W. Ding, "Data mining with big data",*IEEE Trans. Knowl. Data Eng.*, vol.26(1), pp.97-107 ,2014.
- [21] A. Amine, B. Bouikhalene, R. Lbibb, "Customer Segmentation Model in E-commerce Using

- Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco”, International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol.9 (8), pp.2000-2010, 2015.
- [22] J. Cao, X. Yu, Z. Zhang, “Integrating OWA and data mining for analyzing customers churn in E-commerce” J. Syst. Sci. Complex., vol.28(2), pp.381-392, 2015.
- [23] X. Wang, W. Qu, J. Dang, ICIDEL 2017, pp. 162-166 (2017).
- [24] P. N. Tan, M. Steinbach, V. Kumar, “Introduction to Data Mining”, vol.1, pp.145-205, 2006.
- [25] H. Ma, “A Study on Customer Segmentation for E-Commerce Using the Generalized Association Rules and Decision Tree”, American Journal of Industrial and Business Management”, vol.5 (12), pp.813-818, 2015.
- [26] F. Guo, H.L. Qin, “The Analysis of Customer Churns in E-commerce based on Decision Tree” International Conference on Computer Science and Applications (CSA), Wuhan, China, pp. 199-203, 2015.
- [27] L. Xu, H.C. Chu, J. Comput., 26(4), pp. 63-72 (2016).