

A Review of various Machine Learning Techniques for Customer Segmentation in Ecommerce

¹Priyanka V Galagali, ²Dr.Jalaja G, ³Raghavendra CK

¹Student, ²Associate Professor, ³Assistant Professor

¹Computer Science and Technology,
¹^{BNM}Institute of Technology,
Bangalore, India

Abstract: The Ecommerce field has been a rapidly emerging technology and has been playing an important role in digital life. Customer segmentation mainly refers to the division of customers into certain groups and categories. This involves dividing the customers who visit ecommerce website depending on various characteristics or traits of customers into various or multiple variety of subdivisions of communities according to the characteristics of customers based on their shopping behavior or their online habits. In the previous studies, the many algorithms have been used for clustering and classification of data and the problem of segmenting of customers. what so ever it is very hard in these times in division for segmenting customers clearly with clarity while facing a real-time customer data that have the habits and lifestyles of the customers that purchased activities and much more. This paper mainly focuses on the different types of clustering algorithms are the different types of techniques which have been used in creating an algorithm for customer segmentation. by comparing all the techniques which have been used in the e-commerce field to analyze the real-time data of the customers that shopping habits have been taken into consideration to segment customers and to increase or maximize the profits of the company. The robustness and efficiency of every algorithm is discussed along with the limitations of the algorithms. Finally, the customer segmentation can be derived from the results of clustering result. types of experiments have been conducted for real-time data of the customers who have been shopping in varied e-commerce websites and the result taken an output by various methods is in accordant with the customer segmentation. The robustness and efficiency of every algorithm is discussed along with the limitations of the algorithms.

Index Terms— Ecommerce, RFMT, Customer Segmentation, Customer Analysis, k-Means algorithm, Clustering methods, Online Retail, E-commerce system, Mean shift algorithm, Agglomerative algorithm, Machine learning, Python, customer Relationship Management, Customer Segmentation, Data Mining, RFM.

1. INTRODUCTION

A good understanding of customers is extremely important in order to understand who are your loyal customers and who are the customers you want to retain for a longer time. Nowadays a lot of Enterprises have been emerged and are competing against each other for personalizing ecommerce products and customer services in order to ensure maximum satisfaction for different types of customers. The types of strategy which are taken for the development of the enterprise it's very essential has it determine the customer-oriented market behaviour. That also means that every enterprise should be based on a customer which means the values of the enterprise should be centered around the customer's value and their interests. The enterprise or any organization should primarily focus on the requirement of the customer or what a specific customer requirement from that organization. Therefore, In order to understand the key primary importance for the diversity of the requirements of the customers customer segmentation has become a topic which can never be avoided. generally the operators of the business from various companies collect the requirements of the customer in form of data they are either do it with the help of questionnaires to target specific customers. the different types of methods by which the business operators collect the information required to run their business in form of certain methods with the help of questionnaires interview one-on-one calls with the core customers. the customer retention intention can be extracted from the data collected from the customers. The basic premise of segmentation analysis is to divide the customers up for the purpose of customer in two different markets or statements based on their behaviour and other criteria when they are tense or even hundreds of thousands of customers we can't analyse them individually we need to categorise them and group them by using data related to how customers interact with the business of the company it's important to use multiple metrics or criteria because by using just one we don't get an accurate picture of what that customer is. With the emergence of interconnected devices which are intelligent the real-time customer retention data collection from the mobile terminal is all the possible and these help the customers are the Enterprises to better capture customer retention intention. Any enterprise or organisation will still faith the following two important issues house no matter what kind of advanced methods are used for capturing the customer retention data they use. Certain values measured between certain intervals are widely being used in collection of the customer retention data that helps them in the better expression of PR intention. sometimes it might also lead to full of ambiguity. collection of such first data Will aggravate the difficulty of capturing the actual customer retention data. When the data is seen over the past few years with its touch in the customer retention of a large scale products which have been personalized CR are added from customer retention data and shown to be more diverse.

This data makes the enterprise impossible to provide a response for all the customer retention queries with the limited resources they possess. When seen from the point of mining which is required for supply-chain is essentially factored on clustering of Data Analytics. And ambiguity in the data causes the clustering operation to be performed on the customer retention data with fuzzy or

hazy intervals these intervals are not timed properly. Besides when the data is being considered for diverse range of people who have shopped.

2. LITERATURE SURVEY

Customer segmentation is mainly aimed at the division of customers into multiple groups or subsets which are done according to the intentions of requirements of customers after the process of dividing the customers all the customers who are found to be grouped under the same subset will eventually have the same character or traits and where are the customers who are lying in different groups or subsets will have some of years differences between them. When the findings of past research are taken into consideration many of the words will have focused on the segmentation of customers depending on a users which are based on the segmentation theories of traditional market segmentation theories are proposed novel multiple segmentation attributes which will help take organisation to contribute to customer requirement data. On the flipside, incorrect consumer segmentation would result in an illogical market movement design that will be unable to meet customer expectations. Consider the influence of customer profiling on market activity design. To answer the customer segmentation challenge, several analysis methodologies have been applied. Clustering analyses was widely investigated as a data-driven approach among various analytic methodologies.

Customer segmentation could prove to be effective and it could help any critical for the decision-making by the company. The accuracy and preciseness of customer segmentation helps us to fully understand what are the requirements of the customer and what do they expect from the market conditions and therefore it will help us to design the market activities in order to fully satisfy the customers.

3. LITERATURE SURVEY

Zhao-Hui[1] has come up with the concept of Customer requirement Data to analyze the dataset and for the retention of customers. It uses boxplot which is a plotting method to plot the clusters which Clustering of customer requirement data generates. This paper mainly focuses on using GPHC [1] method for clustering of customers. Depending on the conservation of certain plastic with have been generated for the customer requirement data are you respect lasting method which is also termed has groschen peak heuristic clustering is used which can be analysed for fuzzy customer requirement data. which is term has GPHC.Each of the CR item is often labelled with a score. The GPHC algorithm which is used mainly consists of two types of targets. The First Target provides the decision makers of organisation with the clear understanding of customer segmentation result thereby accurately helping them understand the customer intention. The second target mainly uses intelligence of computational power in order to reduce drastically or lessen the human work on classifying and customer data clustering.

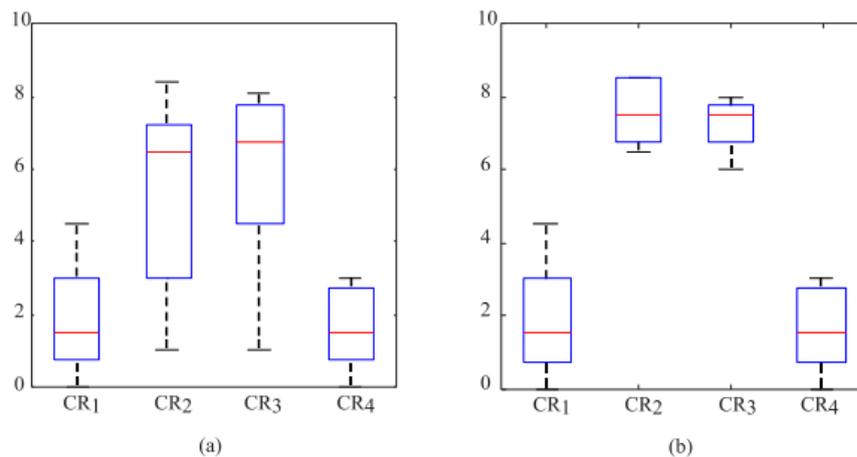


Fig 1: Example of plotting box plots from CRD clustering

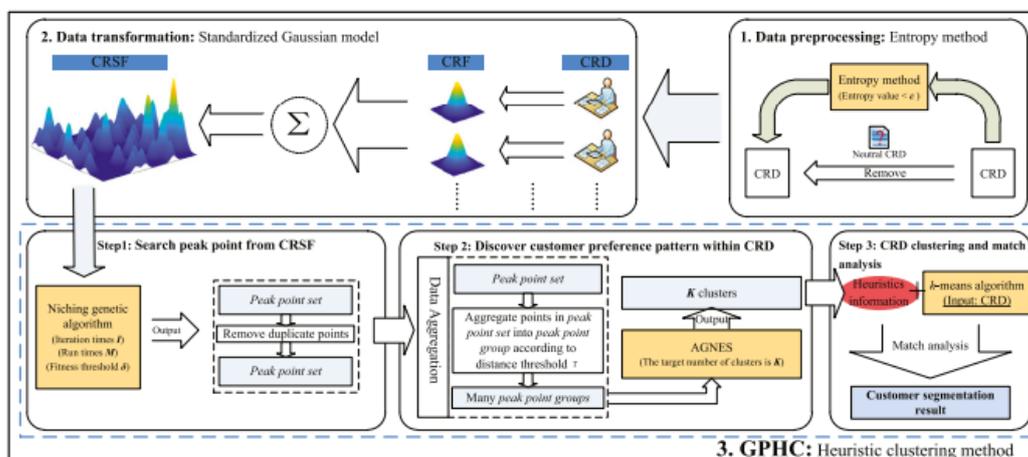


Fig 2: Methodology Followed with Clustering of customers

Joy [2] uses a process called RFM ranking to describe the behaviour of customers. The clients of any enterprise or organization are divided into groups by the process of segmenatation of same or habits based on how their purchasing habits are. The transactions of the company are taken as a data is evaluated over a certain time period. Segmentation provides a thorough grasp of the customers

needs and looks at how well the customers' needs are met. The company's potential customers are identified. Dividing customers into subgroups helps in rapidly increasing the profits and revenue of the organization. It is thought that keeping clients are more valued and salient than getting new ones. When taken an example, any corporation or organization can make use of the variety of methods which can be helpful in marketing and modified to a certain target nook people are used to retain customers. This study of identification begins with an RFM[2] analysing of transferable data and then moves on to cluster it. This study first does an RFM analysis on the transactional data before clustering it using regular data.

It is commonly used to rate consumers based on previous purchases. RFM analysis is used in a variety of applications that involve a large number of clients, such as online purchasing, retailing, and so on. They use Kmeans algorithms and go through a set of process for analysing the dataset which contains n number of variables or instances and would want to output k number of clusters. The customers are partitioned depending on the initial centroids. They also calculate the value of distance of each point taken from the chosen centroids and are eventually evaluated using Euclidian distance and slater these distances are compared and clusters are obtained.

Jun Wu[3] bases their research which is dealt with a problem in an organization. Using sales data of online transactions or purchases, a varied approach. Finally, the clusters of customers are obtained by applying the coefficients by loading the principal component matrix and factor matrix and eigen values.

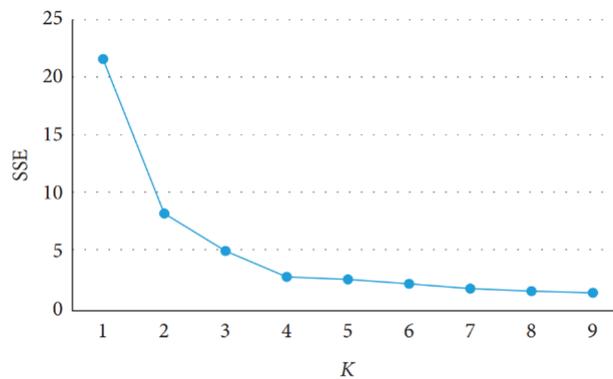


Fig 4: User cluster number against elbow method

T Kansal [4] explains the essence of the modern period is innovation in shopping and method for increasing customers through their lifestyle. The application of segmentation of customers clarifies the unclear idea of which section to target.

Kwak [5] provides a unique method for extracting consumer groups from online customer data for items delivered through online sources. Matrix is factorised is used to discover behavioural customer categories first, followed by demographic customer groups. The methodology to connect the two groups in order to create integrated and comprehensive consumer segments known as personas. Customer interactions with internet material produce behavioural segments. These clients' gender, age, and location are used to create demographic groupings. They examined the persona-generating approach and discovered that it produces genuine, stable,.

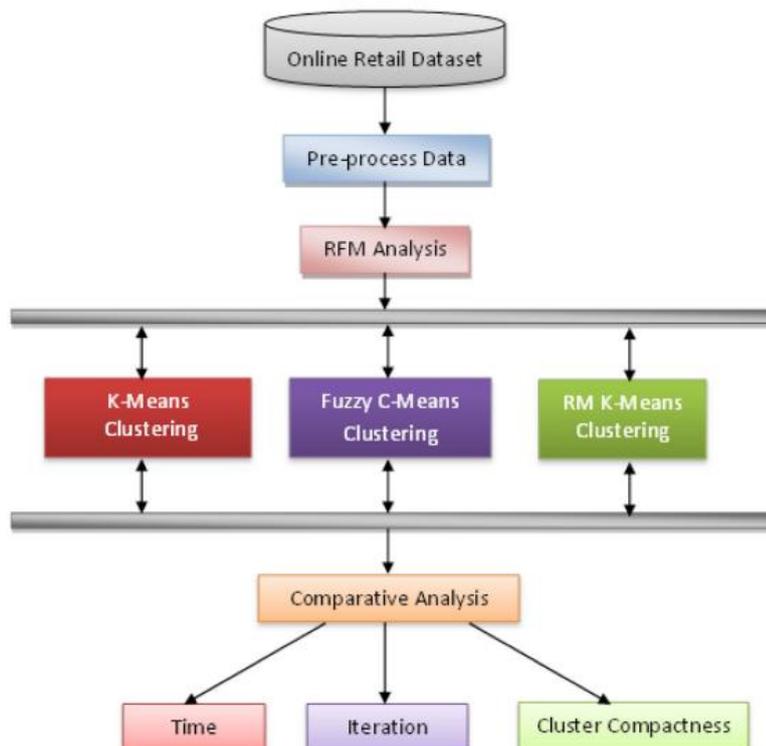


Fig 3: The framework used by [2] for RFM analysis

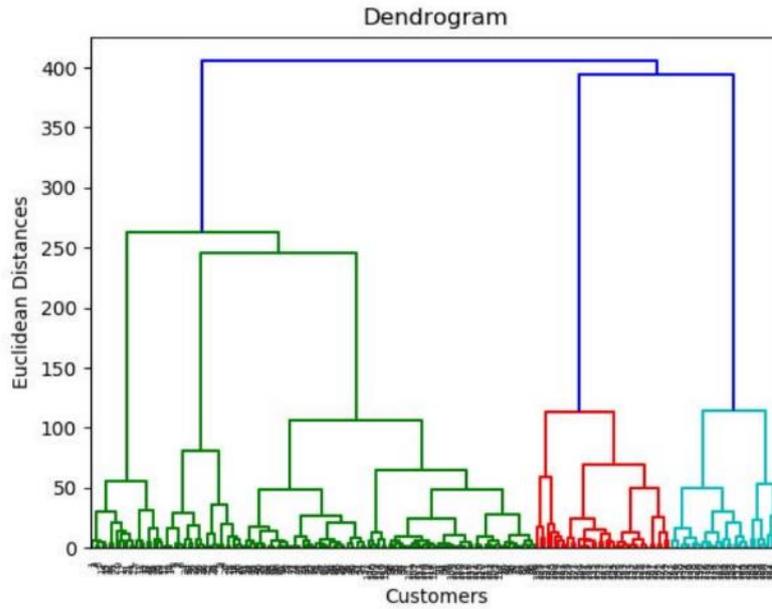


Fig 5: Structure of the dendrogram of the dataset used

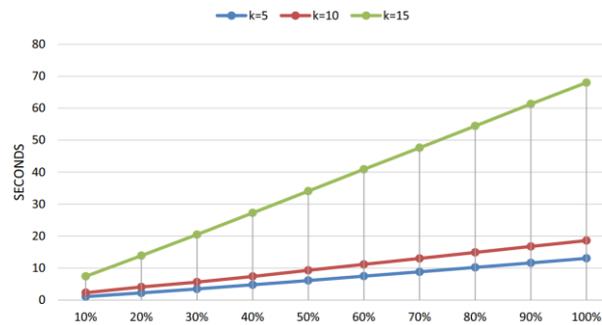


Fig 6: Proportion of dataset

The results indicate that utilising genuine online consumer data, we can reliably identify both behavioural and demographical customer segments from which we may construct personas depicting real groups of individuals. In this study, we explain how to construct personas quickly and automatically using behavioural data that mirror actual individuals and formed from massive data amounts that allow for quantitative analysis

Personalization will be beneficial and has become a solution to the problem of overloading Personalization may be used in marketing to entice potential consumers.

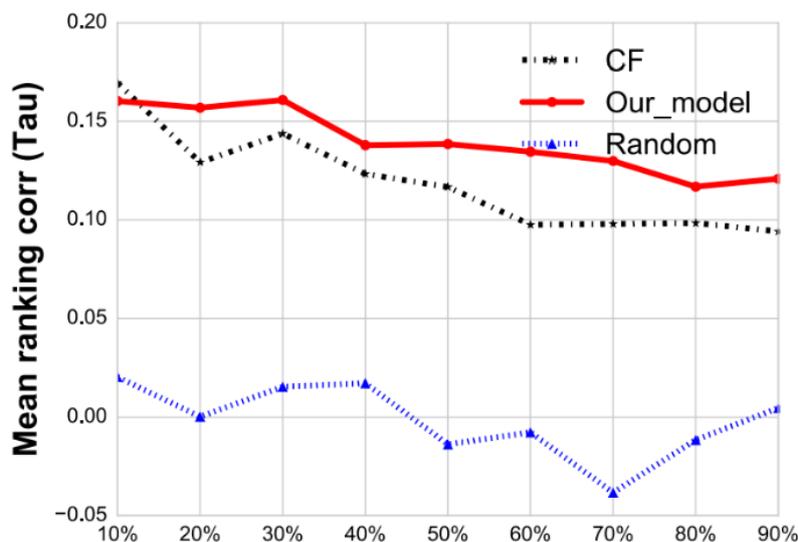


Fig 8: Size of training dataset against the algorithm

Pondel [6] examines the application and advantages of systems which are used for helping the objectives and provides recommendations based on data mining processes aimed at the e-commerce business. On real-world marketing datasets, the good commercial results of collective clustering are proven.

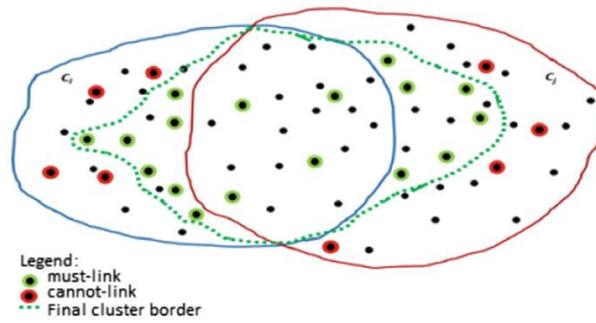


Fig 9: Example of space of clusters merged

Nurma Sari[7] explains about Ecommerce transactions as no longer being novel. Many individuals purchase online, and many businesses utilise it to promote and sell their products. As a result, buyers are subjected to an overabundance of information. This method uses the approach of operationalizing variables which are used for the usefulness of the perceived content from customers. It mainly considers three things which are variables on which the customer behaviour is calculated, the dimensions of variables are also included for indicating the segmentation process. In identifying the aim of business, collecting the required raw data, preparing or preprocessing of data and variable selection, the approach and procedure were generalised and data analysis, processing, and performance evaluation.

This paper uses a web crawling system to fetch the data available publicly in the system such as the customer data including the reviews of the customers and ratings given by the customers after they have purchased the products. They also embed URL manager to take care of the quantity of pages along with the scheduler, parser, database.

All the records which included shopping and purchasing of products were originally obtained through the source of website which belonged to an American store using a specialised web crawling technology. Companies can use this knowledge to decide which customers are the most profitable.

With technological advancement, the collaboration between businesses and customers is becoming increasingly important. This connection must be managed for the company's future success. Customer Relationship Management refers to the process of communicating between businesses and their consumers (CRM) [10].

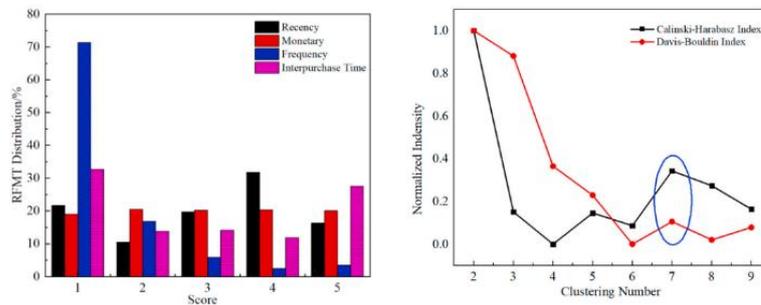


Fig 10: RFMT distributions on a grades scale.

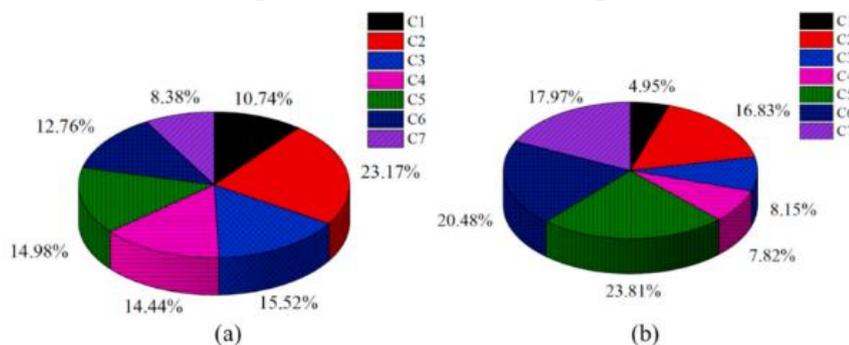


Fig 11: Cluster size and sale contributions

4. CONCLUSION

There are several approaches involved in the segmentation process. Companies utilise to determine the intended customer audience. It guarantees that marketers will focus their marketing efforts on the most likely to attract customers. Additionally, organisations may implement more effective marketing methods as a result of successful consumer segmentation. The procedure helps in decreasing the problems associated with the unreliability which comes with investment. The k-means algorithm was utilised as an algorithm in this case. Here we have reviewed numerous techniques which are used in Machine Learning which can

be used to segment customers in Ecommerce field. There a number of approaches which give an average result .but throughout the review it can be concluded that Kmeans with RFM analysis is the best approaches for segenation.

REFERENCES

1. Zhao-Hui Sun, Tian-Yu Zuo, Di Liang, Xinguo Ming, Zhihua Chen, Siqi Qiu, GPHC: A heuristic clustering method to customer segmentation, *Applied Soft Computing*, Volume 111, 2021,107677,ISSN 1568-4946.
2. A. Joy Christy, A. Umamakeswari, L. Priyatharsini, A. Neyaa,RFM ranking – An effective approach to customer segmentation,*Journal of King Saud University - Computer and Information sciences*,Volume 33, Issue 10,2021,Pages 1251-1257,ISSN 1319-578,https://doi.org/10.1016/j.jksuci.2018.09.004.
3. Jun Wu, Li Shi, Wen-Pin Lin, Sang-Bing Tsai, Yuanyuan Li, Liping Yang, Guangshu Xu, "An Empirical Study on Customer Segmentation by Purchase Behaviors Using a RFM Model and K-Means Algorithm", *Mathematical Problems in Engineering*, vol. 2020, Article ID 8884227, 7 pages, 2020. <https://doi.org/10.1155/2020/8884227>
4. T. Kansal, S. Bahuguna, V. Singh and T. Choudhury, "Customer Segmentation using K-means Clustering," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2018, pp. 135-139, doi: 10.1109/CTEMS.2018.8769171.
5. An, J., Kwak, H., Jung, Sg. et al. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Soc. Netw. Anal. Min.* 8, 54 (2018). <https://doi.org/10.1007/s13278-018-0531-0>
6. Pondel, M., Korczak, J. (2019). Recommendations Based on Collective Intelligence – Case of Customer Segmentation. In: Ziemba, E. (eds) *Information Technology for Management: Emerging Research and Applications. AITM ISM 2018 2018. Lecture Notes in Business Information Processing*, vol 346. Springer, Cham. https://doi.org/10.1007/978-3-030-15154-6_5
7. Nurma Sari, Juni & Nugroho, Lukito & Ferdiana, Ridi & Santosa, Paulus. (2016). Review on Customer Segmentation Technique on Ecommerce. *Advanced Science Letters*. 22. 3018-3022. 10.1166/asl.2016.7985.
8. B. G. Muchardie, A. Gunawan and B. Aditya, "E-Commerce Market Segmentation Based On The Antecedents Of Customer Satisfaction and Customer Retention," 2019 International Conference on Information Management and Technology (ICIMTech), 2019, pp. 103-108, doi: 10.1109/ICIMTech.2019.8843792.
9. Jinfeng Zhou, Jinliang Wei, Bugao Xu, Customer segmentation by web content mining,*Journal of Retailing and Consumer Services*,Volume 61,2021,102588,ISSN 0969-6989,https://doi.org/10.1016/j.jretconser.2021.102588.
10. E. Y. L. Nandapala and K. P. N. Jayasena, "The practical approach in Customers segmentation by using the K-Means Algorithm," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 344-349, doi: 10.1109/ICIIS51140.2020.9342639.
11. Y. Chen and X. Li, "The effect of customer segmentation on an inventory system in the presence of supply disruptions," *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 2009, pp. 2343-2352, doi: 10.1109/WSC.2009.5429207.
12. Ezgi Akar (2021) Customers' Online Purchase Intentions and Customer Segmentation During the Period of COVID-19 Pandemic, *Journal of Internet Commerce*, 20:3, 371-401, DOI: 10.1080/15332861.2021.1927435
13. Acar, S., Köroğlu, F., Duyuler, B., Kaya, T., Özcan, T. (2022). Customer Segmentation Using RFM Model and Clustering Methods in Online Retail Industry. In: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A.C., Sari, I.U. (eds) *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation. INFUS 2021. Lecture Notes in Networks and Systems*, vol 307. Springer, Cham. https://doi.org/10.1007/978-3-030-85626-7_9
14. Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer Segmentation using RFM Model and K-Means Clustering. *International Journal of Scientific Research in Science and Technology*, 591–597. <https://doi.org/10.32628/IJSRST2183118>