Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Application of K-Means Clustering Algorithm for E-Commerce Data Analysis

Laila Ali Putri [1*], Mazayah Tsaqofah[2], Dea Syahfira Hasibuan[3], Hasti Fadillah[4], Maria Ulfa[5], Mhd.Furqan[6]

[1,2,3,4,5,6] Ilmu Komputer, Universitas Islam Negri Sumatera Utara
lailap774@gmail.com[1], tsaqofahmazayah@gmail.com[2], deasyahfira16@gmail.com[3], hastifadillah9838@gmail.com[4], mulfa7903@gmail.com[5], mfurqan@uinsu.ac.id[6]*

Abstract

The development of information technology has driven significant changes in consumer behavior, especially in online shopping transactions through e-commerce platforms. Increasingly fierce business competition requires companies to not only focus on the product, but also understand the characteristics and needs of customers in order to maintain their loyalty. This research aims to identify customer behavior patterns so that segmentation can be carried out that is useful for a more personalized, effective, and efficient marketing strategy. The results of the analysis show that there is a segmentation of customers into several groups based on different transaction intensity and value. This segmentation can be used as a basis for strategic decision-making, especially in marketing planning and customer relationship management. By understanding customer behavior patterns through the clustering process, companies can develop a more personalized and effective service strategy to increase loyalty and business profitability.

Keywords: K-Means, E-Commerce, Clustering

1. Introduction

The development of information technology is increasing, of course, in line with the condition of human civilization, including in the business field. In business, competition requires companies to maximize their capabilities as best as possible in order to survive and compete among other companies. Companies must be able to understand and incorporate customer characteristics which is one of the important things to consid [1].

One of the growing businesses is shopping Online through e-commerce which shows how the behavior of customer purchases changes dynamically. As competition in the business industry becomes more intense, companies are required to shift their focus not only to the product, but also to the customer [2].

In maintaining and expanding the business, companies need to realize that retaining existing customers is just as important as finding new ones. Customer service is also one of the important things to provide Customer Experience good. Cause Customer Experience bad ones such as companies that are unable to provide solutions to problems faced by customers or have difficulties in identifying customers will lead to the risk of losing customers and reducing the number of transactions [3].
Not only being a social media at once marketplace who can only display products, but sellers in Facebook can also take advantage of the live streaming to show off their products interactively, promote their products/companies, and conduct tutorials on the use of certain products [4].

Setyawan, et al proved that the average student who has a smartphone accesses more than 6 hours per day for applications equipped with features live streaming. No wonder social media users with the live streaming has gained an increase in user engagement caused by direct interaction and two-way communication in activities live streaming aforementioned. Based on the increase in sales data through Facebook, an analysis will be carried out using one of the roles of data mining, namely clustering. In the clustering method itself, there are several algorithms that can be applied, including Hierarchical, K Means, and DBSCA [5].
In the sales industry Online, strengthening the relationship between customers is the main goal of the company to gain significant advantages in the market competition. Distribution companies need to identify the best customers and improve their understanding of customer needs in order to maintain customer loyalty. In this all-digital era, sales Online has changed the way we shop, to companies e commerce need to understand their customers deeply to provide a relevant and satisfying experience. In recent years, the business realm Online has experienced rapid increases on a global scale, and the market atmosphere is slowly but surely reaching a level of maturity. [6].

The K-means algorithm will group the data units in a dataset into various clusters based on the closest distance to the randomly selected initial centroid value, which is the initial central point. All of this data will be used to calculate the distance using the Euclidean Distance formula. Data that is close to the centroid will create clusters and this process will continue until there are no changes to each cluster [7].
The K-means clustering algorithm works by dividing the data into different groups based on the similarity of the attributes they have. In the context of sales OnlineK means clustering can be used to identify groups of customers with product preferences, purchasing behavior, pricing preferences, and other relevant factors. This method allows companies to better understand customer groups and provide services that suit the needs and preferences of each group [8].
The K-means algorithm, a non-hierarchical grouping method, shows relatively fast computational times. Comparative analysis conducted by Ghosh & Kumsar between K-Means and Fuzzy C-Means (FCM) [9]. shows that the former algorithm proved to be faster, with a time of 0.433755 seconds elapsed, in contrast to the latter's algorithm, FCM, which took 0.781679 seconds to complete [10].

2. Research Methodology

One of the effective methods in customer segmentation is the K-means clustering method. This method has been widely used in various research and business applications. K-Means is one of the most popular and frequently used grouping methods in various fields due to its simple nature as well as its ease of implementation. The results show that customer segmentation can help companies e-commerce in developing a more effective marketing strategy [11].

In this section, we will explain the flow of the research method carried out. The following can be seen in figure 1 is a diagram of the flow of the method in the research.

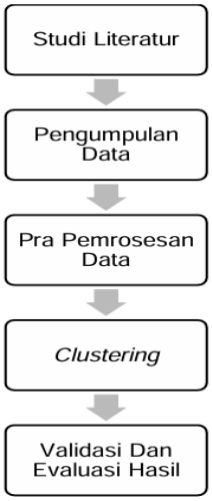

Fig. 1: Flowchart

3. Results and Discussion

This study uses a descriptive quantitative approach, where this study is devoted to using the K-Means Clustering algorithm, to group e-commerce customers based on purchasing behavior. The data used is in the form of customer transaction data which includes recency (last time of purchase), frequency (number of purchases), and monetary (total spending value), which is then processed through the process of cleaning, transforming, and normalizing data. The number of clusters is determined using the Elbow and Silhouette Score methods to obtain optimal segmentation. Furthermore, the K-Means algorithm is applied to form customer groups based on common characteristics, and the results are evaluated and visualized to provide useful insights for marketing strategies and business decision-making. The analysis process is carried out using the Python programming language with the help of libraries such as Pandas, Scikit-learn, and Matplotlib.

Contoh hasil klasterisasi:

	ID Pelanggan	Recency	Frequency	Monetary	Cluster
0	1	186	1	275031	1
1	2	95	1	21199	2
2	3	166	1	39819	2
3	4	293	1	485896	0
4	5	163	1	72706	2

Fig. 2: Cluster result example

The table above is the result of grouping e-commerce customers based on RFM (Recency, Frequency, and Monetary) analysis using the K-Means algorithm. This table shows the first five customers with each attribute value. The Recency attribute shows the time since the last purchase was made by a customer, where the smaller the value means that the customer has just made a transaction. Frequency indicates the number of transactions made by a customer, and in this example all customers have a value of 1 indicating that they have only made one purchase. Monetary reflects the total value of customer spending, which ranges from tens of thousands to hundreds of thousands of rupiah. The Cluster column shows the results of the grouping, which divides customers into clusters based on their shopping behavior patterns. For example, customers with ID 4 who have high recency (293) and large monetary (458,896) are included in Cluster 0, while customers with lower total purchases such as ID 5 (72,726) are classified in Cluster 2.

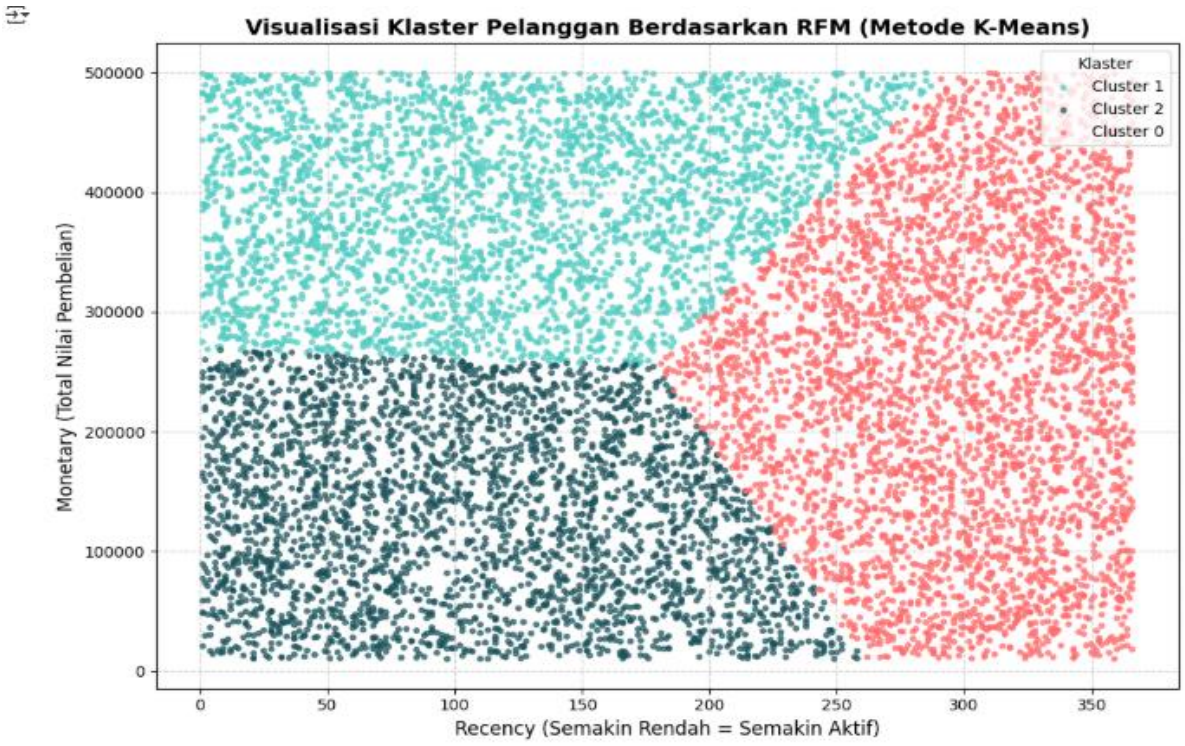

Fig. 3: Visualization of customer clusters based on RFM

The clustering results are depicted in the form of a two-dimensional scatter plot chart with the X axis representing Recency and the Y axis representing Monetary. Each point in the graph represents one customer, and its color indicates the cluster to which that customer belongs. There are three clusters that are differentiated by color: red for Cluster 0, dark blue for Cluster 1, and light blue for Cluster 2. From the visualization, it can be seen that customers with high recency and medium to high monetary are generally in Cluster 0 (red), new customers with low spending values are classified in Cluster 1 (dark blue), while new customers with large spending are classified in Cluster 2 (light blue). This visualization makes it easier for companies to understand customer segmentation and can be the basis for developing a more effective and personalized marketing strategy.

4. Conclusion

The application of the K-Means Clustering algorithm in e-commerce data analysis successfully categorized customers based on purchasing behavior using the RFM approach. This segmentation provides useful insights for companies in developing a more targeted and personalized marketing strategy. By understanding customer characteristics, companies can improve the effectiveness of marketing campaigns as well as long-term loyalty and profitability.

Reference

[1] I. S. B. E. Adiana, “No Title,” Anal. Segmentasi Pelangg. Menggunakan Komb. RFM Model dan Tek. Clust. JUTEI (Jurnal Terap. Teknol. Informasi), vol. 2, pp. 23–32, 2018, doi: 10.21460.
[2] J. W. et Al, “No Title,” “An Empir. Study Cust. Segmentation by Purch. Behav. Using a RFM Model K -Means Algorithm, vol. 2020, 2020, doi: 10.1155/2020/8884227.
[3] A. Fauzi, “No Title,” Data Min. dengan Tek. Clust. Menggunakan Algoritm. K-Means pada Data Transaksi Superst., 2019.
[4] Charlie, “No Title,” MMEMBANGUN KEPERCAYAAN DAN KETERLIBATAN Konsum. MELALUI LIVE STREAING Soc. MEDIA E-COMMERCE, 202AD.
[5] D. E. . Amin, “No Title,” pengaruh live streaming dan onlline Cust. Rev. terhadap pembelian Prod. Fash. muslim (Studi kasus Pelangg. tiktok shop di surabaya), vol. 07, 2023.
[6] Deng, “No Title,” Specif. Strateg. Innov. Mark. Model. E-commerce Enterp. Internet Era. Acad. J. Bus. Manag., pp. 22–26, 2023.
[7] R. K. Gupta, G. K., Agrawal, D., Singh, R. K., & Arya, “No Title,” Prevalence, Risk Factors Socio Demogr. Co-Relates Adolesc. Hypertens. Dist. Ghaziabad, pp. 293–298, 2020, [Online]. Available: https://www.iapsmupuk.org/journal/index.php/IJCH/article/view/331

[8] M. I. Istiana, “No Title,” Segmentasi Pelangg. Menggunakan Algoritm. K-Means Sebagai Dasar Strateg. Pemasar. pada LAROIBA Seluler, 2019, [Online]. Available: http://eprints.dinus.ac.id/12733/2/abstrak_12903.pdf

[9] K. Brahmana, R. W. S., Mohammed, F. A., & Chairuang, “No Title,” Cust. Segmentation Based RFM Model Using K-Means, K-Medoids, DBSCAN Methods, vol. 11, p. 32, 2020, [Online]. Available: https://ojs.unud.ac.id/index.php/lontar/article/view/58025

[10] D. Chandra, M. D., Irawan, E., Saragih, I. S., Windarto, A. P., & Suhendro, “No Title,” Penerapan Algoritm. K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi. BIOS J. Teknol. Inf. Dan Rekayasa Komput., vol. 13, pp. 30–38, 2021, [Online]. Available: https://francis-press.com/papers/11066

[11] G. Zhang, Z., Ni, “No Title,” Comp. Trajectory Clust. Methods based K means DBSCAN. Proc. 2020 IEEE Int. Conf. Inf. Technol. Big Data Artif. Intell., pp. 557–561, 2020.