e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12199

Unlocking Pharma Market Segmentation for Strategic Growth Through Advanced Data Intelligence

Alfi Purwaningrum[1], Amalia Nur Alifah[2*], Dwi Bagus Dermawan[3], Galuh Andini[4]
[1-4] Departement of Data Science, School of Computing,
Telkom University,
Bandung, Indonesia 40257
alfipurwaningrum@student.telkomuniversity.ac.id; amalialifah@telkomuniversity.ac.id; dwibagusd@student.telkomuniversity.ac.id; galuhandini@student.telkomuniversity.ac.id

*Correspondence: amalialifah@telkomuniversity.ac.id

Abstract – *Business competition compels companies to understand customer characteristics in order to maintain and enhance their competitiveness, especially in the pharmaceutical industry, which involves various customer segments such as hospitals, pharmacies, patients, and end consumers with diverse needs. Customer segmentation becomes crucial in developing effective strategies, with K-Means algorithm being one of the commonly used methods because it is easy to use and effective at clustering big datasets. This study combines the K-Means Clustering and the Elbow method to identify the ideal number of clusters in segmenting the customer profiles of a pharmaceutical company. The analysis results reveal two main clusters: the first cluster is dominated by hospitals with higher medication purchase volumes and longer delivery distances, ranging from 8 to 131 km, while the second cluster is dominated by pharmacies with smaller purchase volumes and shorter delivery distances. These findings enable the pharmaceutical company to better understand customer characteristics and design more effective strategies to compete in the market. It is recommended that the company adjusts its marketing strategies and products based on the needs of each cluster, enhances customer relationships through loyalty programs, and optimizes distribution routes to improve operational efficiency.*

Keywords: Clustering, Customer Profiling, Elbow, K-Means, Pharmaceutical Market, Segmentation.

I. INTRODUCTION

In the business world, competition is a common occurrence, and companies must optimize their capabilities as much as possible to survive and compete with other companies. The global pharmaceutical market is projected to reach $1.5 trillion by 2023, with a compound annual growth rate (CAGR) of 4-5%, driven by increasing competition and the need for targeted marketing strategies (IQVIA, 2021).

Understanding customer characteristics and behaviors is essential for companies to effectively segment their markets and maintain competitiveness. For example, a recent case study from Intermix, a women's retail company, revealed that optimized customer segmentation strategies led to a 15% increase in annual revenue through targeted email marketing campaigns (Econsultancy, 2021). These examples highlight the tangible benefits of effective segmentation approaches, particularly in highly competitive industries like pharmaceuticals.

One important aspect that companies need to understand for consideration is customer characteristics (Febrianty et al., 2023). In a pharmaceutical company, customers not only include hospitals and pharmacies but also patients and end consumers who use the pharmaceutical products (Wibowo et al., 2020). Thus, pharmaceutical companies must understand that each customer group has different needs and preferences. For instance, hospitals may look for cost-efficient medications.

The strategy to address this is customer segmentation, which aims to divide the target market into groups based on specific characteristics, in order to help the company grow (Suharti et al., 2022a). One of the methods that can be used for segmentation is the K-Means algorithm, which is one of the simplest and most used clustering methods. This method can efficiently group a lot of data in a short amount of time (Suharti et al., 2022a). To determine the number of clusters, the elbow method can be used, which is a technique for selecting the optimal number of clusters in the K-Means algorithm. This is done by plotting the number of clusters against the inertia value and finding the elbow point on the plot curve (Ramadhan, 2023).

Previous research has provided an important foundation in this field. A study conducted by Savitri, A. and colleagues showed that the use of the K-Means method in customer segmentation for Belle Crown Malang was initially suboptimal, as the elbow technique resulted in $k = 3$, while performance tests recommended 2 or 5 clusters. After a re-approach, the study ultimately settled on $k = 2$, which significantly improved segmentation accuracy. Performance tests such as Connectivity and Silhouette demonstrated that the choice of $k = 2$ provided the best fit (Febriani & Putri, 2020). Further research by Ika Murpratiwi, S. and colleagues demonstrated that the combination of the RFM model with K-Medoids resulted in optimal customer segmentation for UD XYZ. The Davies-Bouldin Index (DBI) indicated that $k = 5$ was sufficient to represent data variation. Customized marketing strategies can be implemented based on the customer groups formed (Ika Murpratiwi et al., 2021). Another study by Suhati, P. and team stated that the K-Means method was effective for customer segmentation based on transaction data. Three clusters were formed, helping the company better understand customer needs to design more precise marketing strategies (Suharti et al., 2022).

Unlike previous studies that focused on general clustering techniques, this research uniquely combines the K-Means algorithm and the elbow method to identify the optimal number of clusters specifically within the pharmaceutical market context. This approach allows for more precise segmentation, tailored to the unique characteristics of this industry. Furthermore, the application of these methods to real-world pharmaceutical market data highlights the practical relevance of the study (Wicaksana et al., 2023). Thus, this research aims to identify the optimal k value as a basis for clustering, using the available data. This will enable a deeper understanding of customer characteristics in the pharmaceutical sector, providing actionable insights for market segmentation and targeted marketing strategies.

II. METHODS

This study eployed the Elbow method, K-Means Clustering, and pre-processing, which will be implemented on the pharmaceutical company's distributor data. The stages of this research are illustrated in the flowchart in Figure 1.

The flowchart in figure 1 outlines the process of customer segmentation using data collection, pre-processing, and clustering algorithms. The process begins with data collection, followed by pre-processing to prepare and clean the raw data. After that, relevant variables are selected for analysis, leading to the determination of how many clusters exist. Here, the Elbow Method and the Silhouette Score are two approaches that are utilized, both of which help to identify the optimal number of clusters for segmentation.

The data is divided into several clusters using the K-Means method after the ideal number of clusters is established. The results are then

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12199

presented through visualization, providing clear insights into the distribution of clusters and customer segments. The process concludes after the visualization step, offering a comprehensive view of customer profiles based on the clustering.

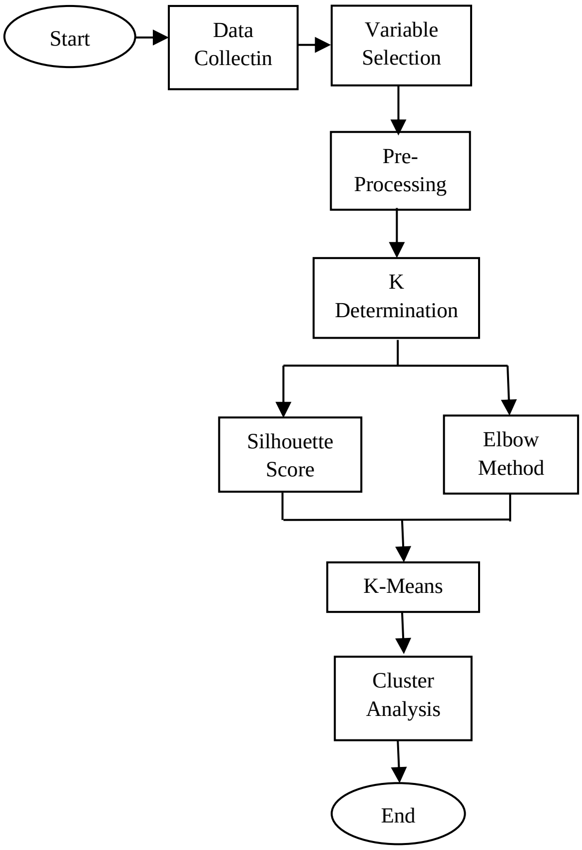

Figure 1 Flowchart of Data Analysis

1 Data Collection and Pre-Processing

The dataset used in this study consists of transactional records from a pharmaceutical company, covering the period from January to December 2023. It includes variables such as transaction quantity, total customer spending, and the distance between customer addresses and the company's nearest distribution center. The data was anonymized to ensure privacy compliance, and the collection process adhered to ethical guidelines.

The data pre-processing stage involves several important steps that need to be carried out in data processing. First, data from various sources must be combined so that they can be processed together, resulting in a total of 489 rows. Next, missing values are handled by deleting rows with missing data, leaving 488 rows. Afterward, a check for duplicate data is performed, and any duplicates are removed to ensure that the analysis results are not affected by repeated data (Syahira & Arianto, 2024). After that, a new column, Address Distance (km), was added by calculating the distance between the existing addresses. Next, an outlier check was performed to detect extreme values that could affect the analysis results (Pratamawati et al., 2021), leaving 380 rows of data. Finally, data standardization was necessary to transform the values in the dataset into a uniform scale, using methods such as z-score or min-max scaling (Azzahra Nasution et al., 2019). All these steps are crucial to ensure the data is ready for further analysis.

2 Variable Selection and Cluster Number Determination

The dataset used in this study consists of transactional records from a pharmaceutical company, covering the period from January to December 2023, It includes variables such as a number of initial variables available, such as branch name, date, document number, product code, product name, price, quantity, total, bonus, discount value, customer group name, customer name, customer address, and phone number.

The selected variables in this research are the quantity and total purchase variables as the relevant ones for clustering. In addition, we created a new column, customer address distance, as an additional variable in the clustering process. The selection of these variables was based on considerations of market penetration and the potential for more accurate customer segmentation to enhance marketing and service strategies. Thus, it is expected that the clustering analysis results will provide a deeper understanding of customer purchasing patterns and preferences, which can be optimized for business purposes.

3 *K-Means Algorithm*

K-Means is a straightforward iterative clustering technique that divides data into k cluster using distance as a measure. Reducing the sum of squared distances between the data points and the centroids of each cluster is its aim (Yuan & Yang, 2019). The objective function being minimized is shown in Equation (1).

$$d = \sum_{k=1}^{K}\sum_{i=1}^{n}\|x_i - u_k\|^2 \qquad …..(1)$$

With K being the number of clusters, u_k is the centroid of the k-th cluster, x_i is the i-th data point, and n is the number of data points within the cluster (Yuan & Yang, 2019). The cluster centroid u_k is calculated using the formula shown in Equation (2).

$$u_k = \frac{1}{2}\sum_{i=1}^{n} x_i \qquad …..(2)$$

The final outcome of the K-Means algorithm is highly influenced by the initial selection of cluster centroids, which can lead to unstable results. Therefore, selecting the appropriate initial cluster centroids and the correct value of K is crucial for achieving optimal clustering results.

4 Elbow Method and Silhouette Score

The approach to determine the ideal number of clusters in K-Means Clustering is the Elbow method. This method identifies the point where the decrease in the sum of squared errors (SSE) begins to slow down, resembling an elbow shape on the graph (Sholeh & Aeni, 2023). SSE is calculated using the formula shown in Equation (3).

$$SSE = \sum_{k=1}^{K}\sum_{x_i \in S_K}\|X_i - C_k\|_2^2 \qquad …..(3)$$

With k being the number of clusters, C_k is the k-th cluster, and x represents the data inside each cluster.

Silhouette Score evaluates the clarity of cluster formations by measuring how closely a data point aligns with the cluster it belongs to in contrast to its relationship with other clusters (Alifah et al., 2022; Januzaj et al., 2023). A score around 0 denotes overlapping clusters, whereas a value between -1 and 1 shows that the data points are well-clusters and isolated from neighboring clusters. Data points that contain negative values may have been allocated to the incorrect cluster. The following formula is used to determine a data point i is Silhouette Score:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad …..(3)$$

Where:

$a(i)$: the intra-cluster distance, or average distance, between data point i and every other point in the same cluster.

$b(i)$: the average distance between the data point i and all points in the nearest cluster that is not the point's own cluster (inter-cluster distance).

$$Silhouette\ Score = \frac{1}{n}\sum_{i=1}^{n} s(i) \qquad …..(4)$$

Equation (4) shows the average Silhouette Scores of each data point, with n denoting the total count of data points. This results in the overall Silhouette Score for the clustering outcome.

The clustering results were validated by assessing both intra-cluster similarity and inter-cluster differences using two key metrics: the Sum of Squared Errors (SSE) and silhouette scores. SSE was employed to measure the compactness of clusters by calculating the sum of squared distances between data points and their respective cluster centroids, with lower values indicating more cohesive clusters. On the other hand, silhouette scores provided a quantitative evaluation of how well each data point was assigned to its cluster by considering

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12199

the distance between the point and points in other clusters. A higher silhouette score, closer to 1, signifies well-defined clusters, while a score near 0 or negative values suggests potential overlap or misclassification. Together, these metrics offered a comprehensive evaluation of clustering performance, ensuring that the identified clusters were both internally coherent and externally distinct.

III. RESULTS AND DISCUSSION

At this stage, the data analyzed by the authors consists of the columns Quantity, Total, and Address Distance. To identify the most suitable quantity of clusters for the K-Means model, we employed the Elbow method. This method involves calculating the Within-Cluster Sum of Squares (WCSS) for various numbers of clusters and then plotting these values (Naghizadeh & Metaxas, 2020). The graphic of the elbow indicates the “elbow” -the point at which the decline in WCSS starts to slow down. This elbow point indicates the optimal number of clusters (Khairati et al., 2019). Figure 2 displays the results obtained from the analysis of the Elbow method.

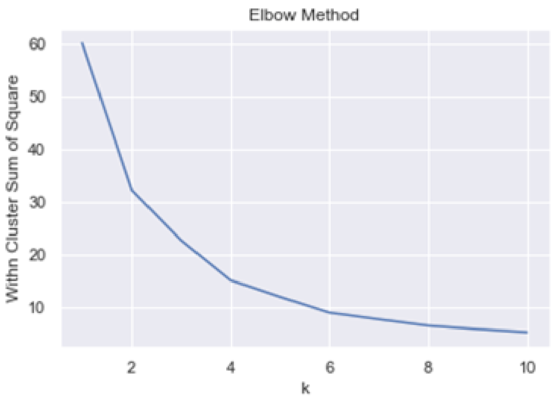

Figure 2 Elbow Method

The graph in Figure 2 shows the change in WCSS values against the number of clusters. When the number of clusters increases, especially from k=1 to k=2, the WCSS value falls noticeably. This decrease reflects that adding clusters up to $k = 2$ provides a substantial reduction in within-cluster variation. However, after $k = 2$, the decrease in WCSS slows down, indicating that adding more clusters offers diminishing returns. This suggests an elbow point at $k = 2$. The elbow point indicates that two clusters are optimal, as beyond this point, the decrease in WCSS becomes more gradual and less significant. Based on the findings we have gathered, it is clear that the most effective configuration for the K-Means model in thin dataset is to utilize two clusters.

Table 1 Silhouette Score

No	K	Silhouette Score
1.	2	0.4485
2.	3	0.4688
3.	4	0.4916
4.	5	0.4826

Additionally, based on the analysis, the silhouette score for different values of k was evaluated, as shown in table 1. While the silhouette score for $k = 2$ is 0.4485, which is lower compared to $k = 3$ or $k = 4$, k=2 is chosen because the characteristics of the data for each cluster are more clearly identifiable with this number of clusters. Despite the slightly lower silhouette score, using $k = 2$ allows for a better understanding and interpretation of the data distribution within each cluster. Therefore, $k = 2$ is opted to achieve more meaningful segmentation, even though other k values yield higher silhouette scores.
The results of the data clustering visualization using two clusters are shown in Figure 3. The scatter plot illustrates the results of the K-Means clustering algorithm, where the data points are grouped into two distinct clusters. Blue data points indicate Cluster 1, while purple data points indicate Cluster 2. The red dots indicate the centroids of each cluster, which represent the central point of each group. The centroids

are calculated as the average position of all the data points within their respective clusters. As seen in the plot, Cluster 1 tends to have data points concentrated towards the lower-left corner, whereas Cluster 2 has more dispersed data points across the graph. The centroids are positioned to minimize the distance between the data points and the center of their respective clusters, demonstrating a clear distinction between the two clusters.

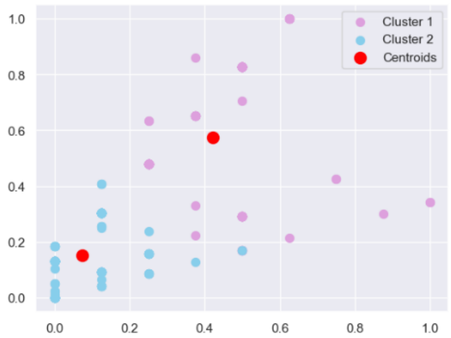

Figure 3 Scatter Plot Data After Clustering

Additionally, to better understand the characteristics of the data in each cluster, a boxplot is used to visualize the distribution of each feature within the clusters. The boxplot provides a clear representation of the variation and central tendencies of the data points in each cluster, helping to compare the characteristics of the clusters across the features used in the analysis. This allows for a more detailed examination of how the features contribute to the differentiation between clusters.

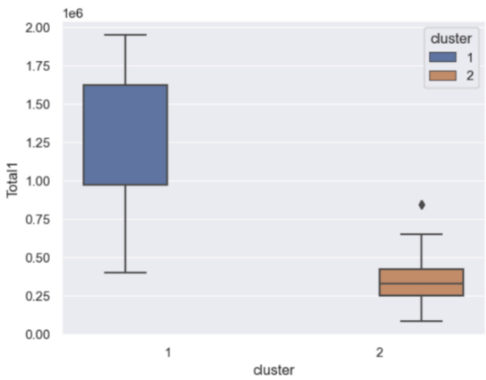

Figure 4 Boxplot Cluster vs Total

The boxplot in figure 4 illustrates the distribution of the variable "Total1," which represents the total amount paid by each customer, in millions of Rupiah, between the two clusters. The blue box, which represents Cluster 1, displays a greater range of total payments, with values ranging from approximately 400 thousand to nearly 2 million Rupiah. The median total payment in Cluster 1 is around 1.25 million Rupiah, indicating a higher spending pattern among customers in this group. Cluster 2, represented by the orange box, has a lower range of total payments, with values ranging from approximately 100 thousand to 800 thousand Rupiah, and a median of around 300 thousand Rupiah. Additionally, an outlier can be seen in Cluster 2, indicating a customer who has made a significantly higher payment compared to others in the same cluster. The different spending patterns of the two clusters are illustrated in this boxplot, with Cluster 1 typically having a larger total payment than Cluster 2.

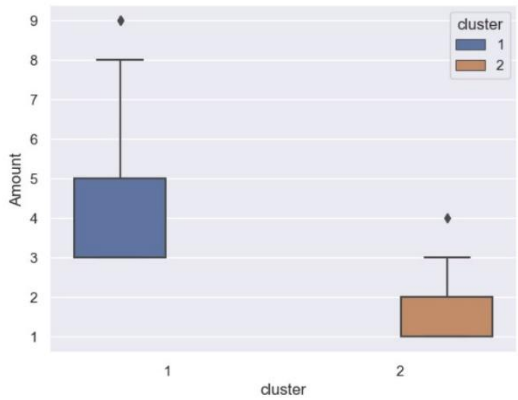

Figure 5 Boxplot Cluster vs Amount

The boxplot in figure 5 illustrates the distribution of the variable "Amount," which represents the total number of medicine items purchased by each customer, across the two clusters. The blue box, which represents Cluster 1 has a wider range of purchases, with the number of items ranging from approximately 3 to 8, and a median of around 4 items. There is also an outlier in Cluster 1, indicating a customer who purchased a significantly higher number of items. On the other hand, Cluster 2,

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12199

represented by the orange box, shows a narrower range of purchases, with the number of items purchased ranging from 1 to 3, and a median of approximately 2 items. An outlier is also observed in Cluster 2. Overall, the boxplot shows that customers in Cluster 1 tend to purchase a higher number of items compared to those in Cluster 2, highlighting distinct purchasing behaviors between the two clusters.

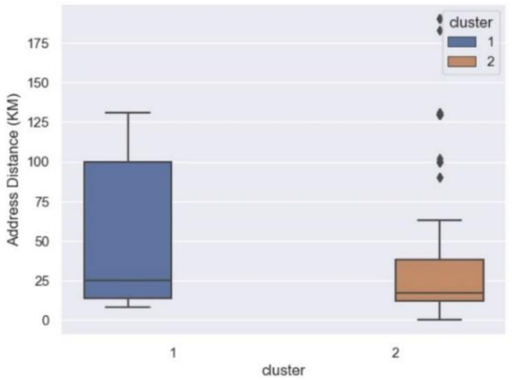

Figure 6 Boxplot Cluster vs Distance

The boxplot in figure 6 depicts the distribution of the variable "Address Distance (KM)," which represents the distance between each customer and the pharmaceutical market, across the two clusters. The blue box, which represents Cluster 1, shows a wider range of distances, with values ranging from approximately 8 to 131 km, and a median of around 30 km. This indicates that customers in Cluster 1 tend to be located farther from the market, with a few outliers reaching even greater distances. In contrast, Cluster 2, represented by the orange box, has a narrower range of distances, with values ranging from 0 to around 60 km, and a median of about 20 km. Several outliers can be observed in Cluster 2, indicating customers who are located farther away than the majority in this group. Overall, the boxplot suggests that customers in Cluster 1 are generally farther from the market compared to those in Cluster 2, highlighting geographical differences between the two clusters.

The segmentation into two clusters offers valuable insights for marketing and operational strategies. Cluster 1, characterized by larger spending and broader geographic dispersion, suggests an opportunity to prioritize high-value contracts and optimize distribution logistics to improve efficiency. Conversely, Cluster 2, with lower spending and more localized presence, could benefit from targeted promotional campaigns aimed at smaller pharmacies or less frequent buyers. These insights enable tailored strategies that maximize resource allocation and customer engagement.

Additionally, figure 7 and figure 8 show the types of customers within the pharmaceutical company for each cluster. This analysis will help to further differentiate the clusters by identifying the customer profiles associated with each group. By understanding the types of customers in each cluster, the pharmaceutical company can tailor its marketing strategies and services to better meet the needs of specific customer segments, optimizing both engagement and sales outcomes. This step is crucial for gaining deeper insights into the behavioral and demographic characteristics that define each cluster.

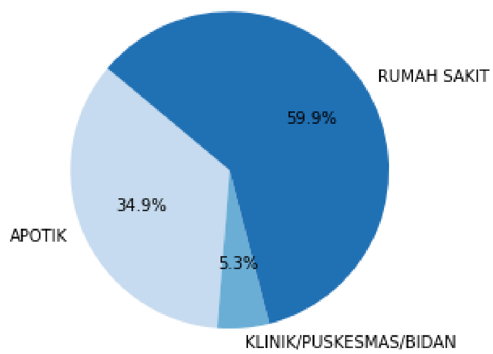

Figure 7 Pie Chart of Customer Group in Cluster 1

In total, Cluster 1 has 152 customers, which represents 40% of the entire dataset. The pie chart in figure 7 illustrates the distribution of customer groups within Cluster 1. The majority of customers in this cluster, accounting for 59.9%, are hospitals, suggesting that hospitals form the largest segment in this group. Pharmacies make up 34.9% of the customers, indicating a substantial secondary segment. Meanwhile, clinics, community health centers (puskesmas), and midwives account for 5.3% of the customers, representing a smaller but notable segment. This distribution shows that hospitals dominate Cluster 1, followed by pharmacies, with clinics and other health service providers forming a minority. These insights can help the company tailor its marketing strategies to focus on the needs and behaviors of the hospitals and pharmacies within this cluster.

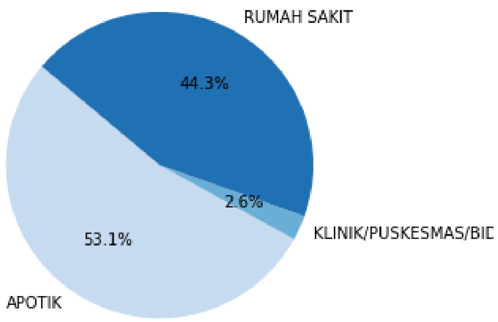

Figure 8 Pie Chart of Customer Group in Cluster 2

The pie chart represents the distribution of customer groups within Cluster 2, which contains 228 customers, accounting for 60% of the total dataset. In this cluster, 53.1% of the customers are pharmacies, making them the largest segment. Hospitals make up 44.3% of the customers, representing a significant secondary group. A smaller group, consisting of clinics, community health centers (puskesmas), and midwives, accounts for 2.6% of the customers. This distribution shows that Cluster 2 is primarily dominated by pharmacies, followed closely by hospitals, with a minor portion of clinics and other health service providers. These insights are crucial for developing tailored marketing strategies aimed at pharmacies and hospitals, which constitute the majority of customers in this cluster.

Despite the robustness of our findings, certain limitations should be acknowledged. The exclusion of variables such as customer demographics or product-specific preferences may introduce potential biases in the segmentation results. Additionally, data preprocessing steps, including the removal of outliers, could impact the generalizability of the findings to the broader customer base. Future research could incorporate these variables to provide a more comprehensive segmentation framework

IV. CONCLUSION

This study successfully identified two main customer clusters for the pharmaceutical company using the K-Means algorithm and the elbow method. Cluster 1 is dominated by hospitals that purchase large quantities of medicine and have high total purchases. Additionally, customers in Cluster 1 tend to be located farther from the pharmaceutical company. In contrast, Cluster 2 consists mainly of pharmacies with lower medicine quantities and total purchases, and these customers are generally located closer to the pharmaceutical company compared to those in Cluster 1. This knowledge provides the pharmaceutical company with valuable insights into the specific needs, preferences, and geographic patterns of each customer group.

Based on the findings, the pharmaceutical company is advised to tailor marketing strategies by focusing on different approaches for hospitals and pharmacies, taking into account the distinct needs and preferences of each group; enhance product and service offerings by providing cost-efficient products for hospitals and competitively priced products

e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v7i1.12199

for pharmacies; strengthen customer relationships by improving engagement with key customers in each cluster through loyalty programs and personalized customer service.

While this study provides valuable insights into customer segmentation, certain limitations should be acknowledged. The dataset used in this analysis was limited to a specific geographic region, which may not fully capture the diversity of customer behaviors across different markets. Additionally, the exclusion of certain variables, such as customer demographics or product preferences, could have impacted the segmentation results. The preprocessing steps, particularly outlier removal, were necessary to improve model accuracy but may have excluded some extreme data points that could offer further insights.

Future research could explore a broader range of demographic or behavioral data to provide a more comprehensive understanding of customer segmentation. Additionally, testing advanced clustering algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) could offer more flexibility in identifying clusters of varying shapes and densities. Longitudinal studies could also be conducted to examine how customer segmentation evolves over time, providing deeper insights into customer behavior and retention strategies.

REFERENCES

Alifah, A. N., Fadhilah, H. N., & Sianipar, T. M. (2022). Klasterisasi KabupatenKota di Jawa Barat Berdasarkan Tingkat Kenyamanan dengan Metode K-Means Clustering. *PROSIDING SEMINAR NASIONAL SAINS DATA*, *2*(1), 30–38.

Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. JIEMS (Journal of Industrial Engineering and Management Systems), 13(2). https://doi.org/10.30813/jiems.v13i2.2274

Febrianty, E., Awalina, L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions. Jurnal Teknologi Dan Informasi (JATI), 13. https://doi.org/10.34010/jati.v13i2

Ika Murpratiwi, S., Gusti Agung Indrawan, I., & Aranta, A. (2021). ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL. Jurnal Pendidikan Teknologi Dan Kejuruan, 18(2).

Januzaj, Y., Beqiri, E., & Luma, A. (2023). Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique. International Journal of Online & Biomedical Engineering, 19(4).

Khairati, A. F., Adlina, A. A., Hertono, G. F., & Handari, B. D. (2019). Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA. PRISMA. Prosiding Seminar Nasional Matematika, 2, 161–170. https://journal.unnes.ac.id/sju/index.php/prisma/

Naghizadeh, A., & Metaxas, D. N. (2020). Condensed silhouette: An optimized filtering process for cluster selection in K-means. Procedia Computer Science, 176, 205–214.

https://doi.org/10.1016/j.procs.2020.08.022

Pratamawati, M. H. S., Hidayat, T., Ibrahim, M., & Hartatik, S. (2021). Hubungan Minat Belajar dengan Prestasi Belajar Matematika Siswa di Sekolah Dasar. Jurnal Basicedu, 5(5), 3270–3278. https://doi.org/10.31004/basicedu.v5i5.1331

Ramadhan, A. G. (2023). How to cite: Ade Guntur Ramadhan (2023) Data Mining Untuk Segmentasi Pelanggan dengan Algoritma K-Means: Studi Kasus pada Data Pelanggan di Toko Retail DATA MINING UNTUK SEGMENTASI PELANGGAN DENGAN ALGORITMA K-MEANS: STUDI KASUS PADA DATA PELANGGAN DI TOKO RETAIL. 8(10). https://doi.org/10.36418/syntax-literate.v6i6

Sholeh, M., & Aeni, K. (2023). PERBANDINGAN EVALUASI METODE DAVIES BOULDIN, ELBOW DAN SILHOUETTE PADA MODEL CLUSTERING DENGAN MENGGUNAKAN ALGORITMA K MEANS. https://archive.ics.uci.edu/ml/datasets/Trave

Suharti, P. H., Suryandari, A. S., & Amalia, R. N. (2022a). ANALISIS KINERJA MODUL PENGENDALI TEKANAN UDARA PCT-14 BERBASIS PLC DENGAN BERBAGAI METODA TUNING. Sebatik, 26(2), 420–427. https://doi.org/10.46984/sebatik.v26i2.2134

Suharti, P. H., Suryandari, A. S., & Amalia, R. N. (2022b). ANALISIS KINERJA MODUL PENGENDALI TEKANAN UDARA PCT-14 BERBASIS PLC DENGAN BERBAGAI METODA TUNING. Sebatik, 26(2), 420–427. https://doi.org/10.46984/sebatik.v26i2.2134

Syahira, N., & Arianto, D. B. (2024). PREDIKSI TINGKAT KUALITAS UDARA DENGAN PENDEKATAN ALGORITMA K-NEAREST NEIGHBOR. Jurnal Ilmiah Informatika Komputer, 29(1), 45–59. https://doi.org/10.35760/ik.2024.v29i1.10069

Wibowo, A., Handoko, A. R., & Korespondensi, P. (2020). SEGMENTASI PELANGGAN RITEL PRODUK FARMASI OBAT MENGGUNAKAN METODE DATA MINING KLASTERISASI DENGAN ANALISIS RECENCY FREQUENCY MONETARY (RFM) TERMODIFIKASI SEGMENTATION OF CUSTOMERS OF DRUG PHARMACEUTICAL PRODUCT RETAIL USING CLASTERIZATION MINING DATA METHOD USING MODIFIED MONETARY RECENCY FREQUENCY (RFM) ANALYSIS. 7(3). https://doi.org/10.25126/jtiik.202072925

Wicaksana, R. S., Heksaputra, D., Syah, T. A., & Nur'aini, F. F. (2023). Pendekatan K-Means Clustering Metode Elbow Pada Analisis Motivasi Pengunjung Festival Halal JHF#2. Jurnal Ilmiah Ekonomi Islam, 9(3), 4162. https://doi.org/10.29040/jiei.v9i3.10591

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. J, 2(2), 226–235. https://doi.org/10.3390/j2020016