Journal of Information Systems and Informatics
Vol. 6, No. 3, September 2024 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v6i3.799 Published By DRPM-UBD

Optimizing Motorcycle Sales: Enhancing Customer Segmentation with K-Means Clustering and Data Mining Techniques

Luis Fernando[1], Melissa Indah Fianty[2*]

[1,2] Department of Information System, Faculty of Engineering & Informatics, Universitas Multimedia Nusantara, Indonesia
Email: [1]luis.fernando@student.umn.ac.id, [2*]melissa.indah@umn.ac.id

Abstract

Information plays a crucial role in the sustainability of company operations. The development of information technology, especially in the industry 4.0 era, affects various fields including economics, social, and education. The company faces challenges in declining motorcycle sales due to intense competition and ineffective customer segmentation. To address these issues, this study proposes the use of the K-Means algorithm with Python tools for better customer segmentation. The study aims to identify diverse customer groups and tailor marketing strategies accordingly. By utilizing the Elbow method and Silhouette score, the analysis of customer data is simplified. This study also employs data mining techniques to uncover hidden patterns in motorcycle sales data, aiding companies in improving operational efficiency and decision-making.

Keywords: Clustering, Customer Segmentation, Data Mining, K-Means Algorithm

1. INTRODUCTION

In today's rapidly evolving business environment, the effective use of information is critical for the sustainability and growth of companies. Information technology has become a cornerstone of organizational success, impacting virtually every sector, from economic and social fields to education and defense, particularly during the Industry 4.0 era [1], [2], [3]. For companies operating in competitive markets, such as the motorcycle industry, leveraging technology to gain insights from data is essential for maintaining market relevance and driving strategic decision-making. Despite the global recognition of the Company's advanced technologies—tested both on public roads and racing circuits—and its pioneering work in developing fuel-efficient engines, it faces challenges in sustaining customer engagement and sales due to fierce competition and a lack of effective customer segmentation [4]-[7].

The Company currently struggles with declining motorcycle sales, which is exacerbated by ineffective customer targeting strategies in an environment

crowded with competitors offering similar products [7]. This challenge highlights the need for a more sophisticated approach to segmenting customers and personalizing marketing strategies. While information systems are recognized for their role in transforming raw data into valuable insights by providing timely, accurate, and relevant information to stakeholders [5], there remains a gap in the Company's utilization of these systems to their full potential. Current efforts, such as using the K-Means algorithm with Python tools for customer segmentation, show promise but lack comprehensive integration of advanced data analytics techniques that could further refine these strategies [8].

Data mining techniques, particularly clustering, offer powerful tools for uncovering hidden patterns in data, which can drive more informed decision-making in marketing and sales strategies [9], [10]. However, the Company's existing use of these techniques does not fully leverage the potential of data mining to address its specific challenges. Additionally, the analysis of unstructured data, which can provide deeper insights into customer behaviors and preferences, remains underdeveloped [14], [15], [16]. This underutilization creates an opportunity for the Company to enhance its approach by incorporating advanced data analytics tools, such as the Elbow method and Silhouette score, to simplify data analysis and improve decision-making processes.

Addressing this gap is crucial, not only for optimizing marketing and production strategies but also for enhancing overall operational efficiency. A more robust application of data analytics could enable the Company to better understand customer behavior, reduce costs, and increase productivity [17], [18], [19]. Moreover, by sharing these insights, the Company could contribute valuable knowledge to the entire industry, fostering broader improvements and innovation [20], [21]. Therefore, this study aims to explore how the integration of advanced data mining techniques can enhance the Company's customer segmentation and market targeting strategies, thereby strengthening its competitive position in a highly dynamic market.

2. METHODS

The methodology for this study, as illustrated in the provided Figure 1, involves several key stages: problem identification, data preprocessing, Elbow method, Silhouette score, modeling, and the application of the K-Means algorithm. Each stage is designed to enhance customer segmentation and optimize marketing strategies for the Company.

1) Problem Identification

The first phase aims to identify the challenges faced by the Company's sales department, particularly regarding customer segmentation for motorcycle buyers.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

The primary issue stems from the limitations in data effectiveness, which impedes the Company's ability to understand its customer base thoroughly. A comprehensive assessment is conducted to identify specific problems in the existing segmentation approach, ensuring that potential motorcycle buyers are accurately targeted based on various relevant factors, such as demographics and purchasing behavior.

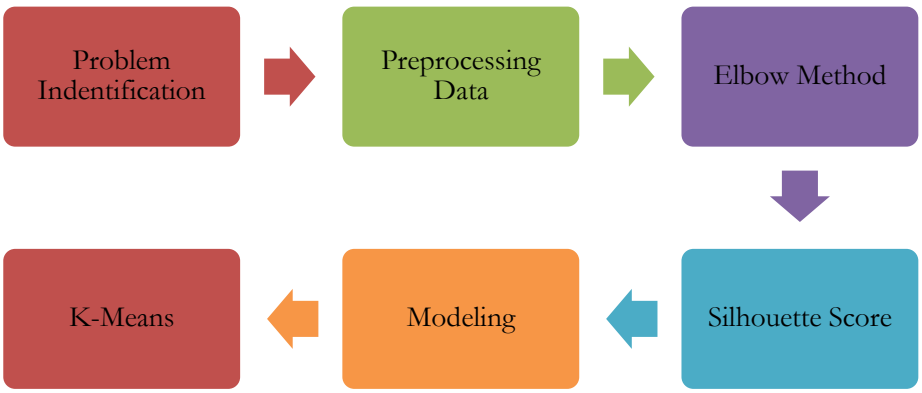

Figure 1. Research Flow

2) Preprocessing Data

In this phase, data preparation is conducted to ensure the dataset is suitable for the subsequent modeling stage. This involves cleaning and organizing the data, including handling missing values and standardizing data formats. The preprocessing phase ensures that all data attributes are correctly aligned, and any inconsistencies, such as missing data points, are identified and addressed [22][23]. This step is crucial to improve the quality and reliability of the dataset, which will be used for clustering and analysis.

3) Elbow Method

The Elbow method is applied to determine the optimal number of clusters for the K-Means clustering analysis. This process involves plotting the number of clusters (K) against the Within-Cluster Sum of Squares (WCSS) using Python tools. By analyzing the curve of the graph, the "elbow point"—where the WCSS starts to diminish at a slower rate—indicates the most appropriate number of clusters for the dataset. In this study, the Elbow method is specifically used to analyze the types of motorcycles and provincial data, providing insight into the natural grouping within the data [24].

4) Silhouette Score

The Silhouette Score method is used to evaluate the quality of the clusters generated by the K-Means algorithm. This method calculates the Silhouette score

for each data point, which measures how similar the data point is to its cluster compared to other clusters. A higher Silhouette score indicates better-defined clusters with greater separation and homogeneity. The Silhouette score is particularly useful for validating the effectiveness of the clustering approach. In this study, it is applied to assess clusters based on motorcycle types and customer age, ensuring meaningful segmentation that aligns with the Company's marketing objectives [25].

5) Modeling

The modeling stage involves visualizing the classification model using the K-Means clustering method. The K-Means algorithm is implemented using the sklearn library in Python to apply clustering models to the Company's dataset. During this stage, the results from the Elbow method and Silhouette score are used to refine the clustering process and ensure optimal data grouping. The visualization helps in understanding how different data points are grouped, aiding in the interpretation of clustering results and supporting strategic decision-making [26].

6) K-Means

The final stage utilizes the K-Means algorithm to classify the data into distinct clusters based on similar characteristics. This algorithm iteratively assigns data points to clusters by minimizing the variance within each cluster. The K-Means analysis is performed to provide insights into customer segmentation, specifically focusing on motorcycle types and geographical provinces. The results are presented in visual formats, such as scatter plots, to illustrate the cluster points clearly and reveal patterns in the data that are useful for targeted marketing strategies. The visualization of these clusters helps the Company understand the different customer groups and tailor marketing efforts accordingly.

3. RESULTS AND DISCUSSION

3.1 Problem Identification

Company is an operating in the motorcycle industry, that facilitate customer segmentation in motorcycle sales. One of the main issues in the sales sector is the lack of motorcycle purchases due to intense competition, which can be interpreted as a situation or event leading to reduced company revenue. In 2023, Company faced challenges with customer data in motorcycle sales, caused by data limitations being one of the main hurdles. The lack of detailed and thorough data on customer motorcycle purchasing behaviour makes it difficult to identify patterns and preferences that differentiate customer groups.

3.2 Preprocessing Data

In this stage and beyond, Python programming language will be used. The data obtained from customers in Company motorcycle sales is still raw and has not been analyzed yet. This research utilizes Jupyter Notebook for its Python language. First, import the data, then select only the data that will be used, followed by removing missing data.

Preprocessing Data

```
In [2]: missing_data = data.isnull().sum()
        print(missing_data)

id                0
frameno           0
sales_dt          0
tahun_sales       0
type_motor        0
warna_mtr         0
gender         1099
tgl_lahir         0
provinsi          0
kreditor       6089
downpayment    6022
cicilan        6022
main_dealer       0
dealer            0
dtype: int64
```

Figure 2. Displays the Amount of Data Lost

Figure 2, shows the display of missing data in each column of the imported data. After running data processing, there are missing data points in the gender column (1,099), creditor column (6,089), down payment column (6,022), and installment column (6,022). The aim is to modify or prepare the data to fit the requirements of the analysis or model to be built within the dataset.

3.3 Data Mining

This stage involves using techniques and algorithms to uncover patterns and convert them within the dataset. In this stage, the conversion format is first changed to numeric to enable the process.

	id	frameno	sales_dt	tahun_sales	type_motor	warna_mtr	gender	tgl_lahir	provinsi	kreditor	downpayment	cicilan	main_d
4	FH/BA /0074054/V	MH1JM8212NK523720	Juni	2023	BEAT STREET	HITAM	LAKI-LAKI	1999-06-24	Sumatera Utara	PT. Federal International Finance	3800000.0	790000.0	
6	FH/AC /0136707/V	MH1JBK311NK418426	Juni	2023	REVO CW FI	MERAH-HITAM	LAKI-LAKI	1973-10-21	Jawa Tengah	PT Adira Dinamika Multifinance Tbk	2700000.0	678000.0	
7	FH/BB5 /0065386/V	MH1JM8211NK533848	Juni	2023	BEAT STREET	HITAM	LAKI-LAKI	1985-12-09	Riau	PT Mega Finance	1950000.0	1364000.0	
10	FH/AC /0151227/V	MH1JMB119NK019602	Juni	2023	GENIO CBS ISS	HITAM	LAKI-LAKI	1968-02-20	Jawa Tengah	PT. Federal International Finance	10000000.0	1256000.0	
11	FH/DA3 /0055950/V	MH1JMB119NK019468	Juni	2023	GENIO CBS ISS	COKLAT	LAKI-LAKI	1994-10-09	Sulawesi Tenggara	PT. Federal International Finance	2400000.0	1140000.0	

Figure 3. Converted Data to Datetime in Numerical Form

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Figure 3, shows data mining where the column has been converted. Also, to calculate age using .loc for data cleaning purposes, the output is displayed in tabular form. Before data cleaning, there were 10,996 rows, while after data cleaning, there were 4,885 data just showing 5 row.

3.4 Elbow Method

In this process, the K-Means clustering algorithm is used, employing the Elbow method to determine the optimal number of clusters to be formed. Here are the results of the data analysis conducted to understand customer preferences towards various types of motorcycles and related purchasing patterns:

3.4.1 Determining the Number of Customer Segmentation Clusters Based on Motorcycle Types Using the Elbow Method

Customer segmentation analysis based on motorcycle types is a crucial step in understanding customer purchasing behavior. At this stage, the Elbow method is used to determine the optimal number of clusters in customer segmentation, focusing on the types of motorcycles purchased by customers.

Figure 4. Number of Customer Segmentation Clusters Based on Motorcycle Type

Figure 4 illustrates a graph representing the WCSS (Within-Cluster Sum of Squares) values as the horizontal axis and the number of clusters as the vertical axis. WCSS serves as a metric in cluster analysis, measuring the distance of data points within a cluster from its center, spanning from k = 2 to k = 5. The WCSS value reflects the proximity of data points to the center of their respective clusters.

A lower WCSS value indicates a more optimal cluster, signifying that data points within it are closely related. The graph displays a distinct trend, starting with a sharp decline as the number of clusters increases, peaking at two clusters. However, the plateau at the graph's end implies that adding clusters beyond a certain point may not significantly enhance our understanding of customer purchase patterns.

3.4.2 Determining the Number of Customer Segmentation Clusters Based on Provinces Using the Elbow Method

Customer segmentation based on province is an important step can look at each region to understand existing motorbike sales purchased by customers from customer segmentation based on these creditors.

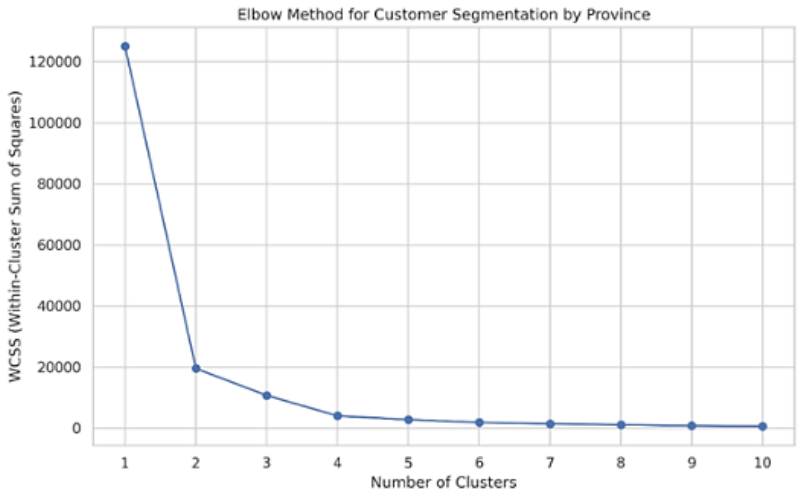

Figure 5. Number of Customer Segmentation Clusters Based on Province

Figure 5 is a graph shown by the Elbow Method for determine the optimal number of clusters in customer segmentation based on province. The elbow point is visible at k = 2 to k = 5. This shows that this cluster is the optimal cluster for customer segmentation based on creditors. The decrease in WCSS is not significant after k = 5. This is shows that adding clusters of more than 4 does not provide significant improvement in segmentation quality.

3.5 Silhouette Score

In this process, the K-Means clustering algorithm is used with the Silhouette Score method for evaluation, which is used to assess how well the data is divided into clusters generated by the K-Means model.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

3.5.1 Evaluation of the Quality of Customer Segmentation Clusters Based on Type Motorcycle Using Silhouette Score

After determining the optimal number of clusters using the Elbow method, the next step is to evaluate the quality of the formed clusters. In this section, we use the Silhouette score to evaluate the score value of data points fitting into their assigned clusters.

Table 1. Customer Silhouette Score Value Based on Motorcycle Type

Cluster	Silhouette Score
2	0 ,70423467
3	0,652810637
4	0,626157649
5	0,642754533
6	0,581682667
7	0,54501575
8	0,519861401
9	0,510085628
10	0,481896284

Based on Table 1, it is evident that the highest Silhouette Score occurs in cluster 2 with a value of 0.704. This value indicates that dividing the data into two clusters results in better separation among these groups compared to other cluster numbers. Although the Silhouette Score tends to decrease with an increase in the number of clusters, cluster 2 still maintains a significant value relative to other clusters.

3.5.2 Evaluation of Cluster Quality in Customer Segmentation Based on Province Using the Silhouette Score

Evaluating the quality of clusters in customer segmentation based on provinces using the Silhouette score can be optimal as it indicates better separation and homogeneity of data points within the clusters.

Table 2. Silhouette Score Value of Customer Segmentation Based on Province

Cluster	Silhouette Score
2	0,854869
3	0,64877
4	0,645809
5	0,613938
6	0,435797
7	0,459684
8	0,462472

Cluster	Silhouette Score
9	0,436727
10	0,403816

Based on Table 2, it's observed that the highest Silhouette Score is found in cluster 2 with a score of 0.854. This score indicates that dividing the data into two clusters provides better separation among the groups compared to other cluster numbers. Although the Silhouette Score typically decreases as the number of clusters increases, cluster 2 still maintains a relatively high score compared to the other clusters.

3.6 Modeling

In the modeling stage, we will use the K-Means algorithm from the sklearn library to build a clustering model. K-Means algorithm is one of the clustering techniques used in data analysis and machine learning. Its main objective is to group data into several clusters based on similarity of specific attributes or characteristics found in the data visualization.

3.7 K-Means

3.7.1 Customer Cluster Results Based on Motorcycle Type

In these results, clusters of customers based on the type of motorcycle can be observed. In this analysis, the K-Means clustering algorithm is used with the number of clusters determined based on the highest Silhouette Score value.

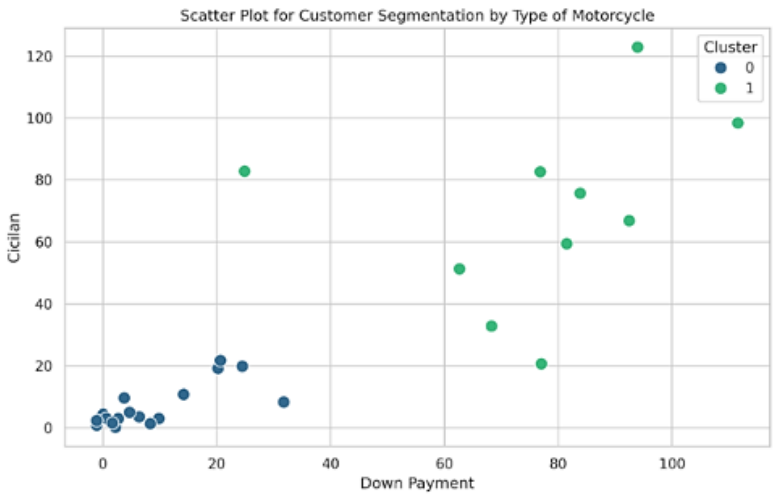

Figure 6. Scatter Plot of Customers Based on Motorcycle Type

Figure 6 shows the customer segmentation results based on the variables of down payment and installments, with each point representing a customer. The points are colored according to the cluster they belong to cluster 0 (turquoise) and cluster 1 (green). From the graph, it is evident that there are two clearly separated clusters. Cluster 1 has higher points. which is dominated by motorcycle types such as BEAT SPORTY CBS, PCX 160 CBS, and REVO FIT FI, indicates customers with higher payments. This might suggest that this group consists of buyers who are more financially capable or prefer motorcycle models with higher features and specifications, even with greater payments. While cluster 0 has lower points. This indicates differences in customer purchasing preferences based on the level of down payment and installments they choose.

3.7.2 Customer Cluster Results by Province

In this result, we can see customer clusters based on provinces. In this analysis, the K-Means clustering algorithm is used with the number of clusters determined based on the highest Silhouette Score value.

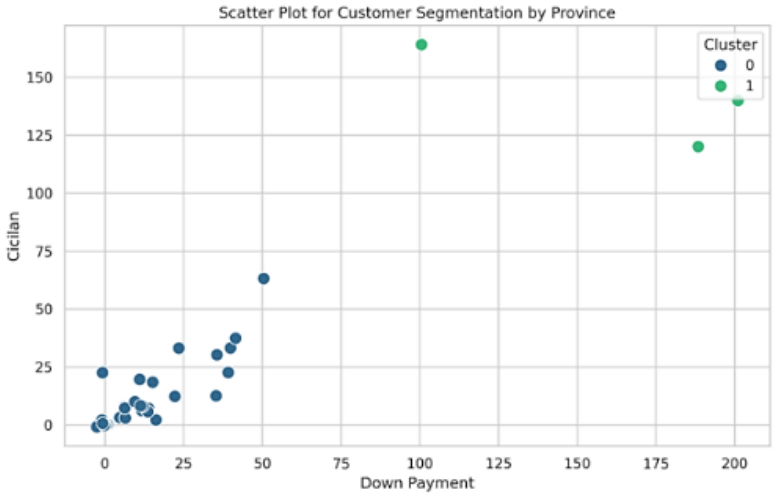

Figure 7. Scatter Plot of Customers by Province

Figure 7 illustrates the segmentation outcomes of customers based on downpayment and installment variables, with each data point color-coded according to its cluster: cluster 0 (teal blue) and cluster 1 (green). The graph reveals three distinct clusters, with cluster 1 containing data points with higher values and cluster 0 predominantly featuring data points with lower values. Moreover, the majority of data points belong to cluster 0, while cluster 1 comprises fewer data

points, mainly from provinces like West Java, Central Java, and East Java, suggesting customers with higher payment levels. This indicates varying purchasing behaviors among customers from these regions compared to others, potentially influenced by factors such as economic status, urbanization, and population density. These findings highlight notable differences in preferences for downpayment and installments between these two customer groups.

3.8. Discussion

This study provides critical insights into the use of data mining techniques for customer segmentation in the motorcycle industry, specifically for the Company under investigation. Through the systematic application of the K-Means clustering algorithm, alongside the Elbow method and Silhouette score, this research has identified the optimal number of clusters for segmenting customers based on motorcycle types and geographic regions. These findings offer a deeper understanding of customer purchasing behaviors and preferences, laying a foundation for more targeted marketing strategies.

A key challenge identified in this study is the Company's limited ability to effectively utilize customer data due to poor data quality and granularity. This deficiency, coupled with intense market competition, has contributed to a decline in motorcycle sales. Without detailed insights into customer purchasing behavior, it becomes difficult for the Company to pinpoint distinct customer groups, which hampers the development of targeted marketing strategies. To address these limitations, the data preprocessing phase was crucial. By cleaning and organizing the dataset, including handling missing data, we established a solid foundation for subsequent clustering analysis, enhancing the overall reliability and accuracy of the results.

Building on this foundation, the Elbow method was employed to determine the most suitable number of clusters for effective customer segmentation. For both segmentation by motorcycle types and provinces, the Elbow method identified an optimal range of 2 to 5 clusters. In the context of motorcycle types, the Within-Cluster Sum of Squares (WCSS) values showed a sharp decline up to 2 clusters, after which the reduction plateaued, suggesting that adding more clusters would yield diminishing returns in segmentation quality. A similar trend was observed for geographic segmentation, further confirming that a limited number of clusters can effectively capture key distinctions among customer groups. This insight is pivotal, as it allows the Company to streamline its marketing strategies, focusing on fewer, but more meaningful, customer segments without losing critical differentiation.

To further refine the clustering results, the Silhouette score was applied to evaluate the quality of the identified clusters. The analysis revealed that the highest Silhouette scores for both motorcycle types and geographic regions were achieved

with 2 clusters (0.704 and 0.854, respectively), indicating well-defined and distinct groupings. As the number of clusters increased, the Silhouette scores decreased, reinforcing the finding that two clusters provide the best balance between internal cohesion and external separation. This result underscores the value of dividing customers into two primary groups—based on motorcycle types and regions—to gain the most actionable insights for tailored marketing and customer engagement efforts.

The findings from the K-Means clustering analysis, supported by the Elbow method and Silhouette score, have several practical implications for the Company. The identification of optimal clustering enables the Company to segment its customers into two primary groups: those with higher purchasing power who prefer premium motorcycle models, and those with lower purchasing power who lean toward more economical options. Additionally, geographic segmentation suggests that customers from regions such as West Java, Central Java, and East Java typically exhibit higher purchasing power, influenced by various socio-economic factors. These insights empower the Company to more precisely target its marketing efforts, aligning promotional strategies with the distinct needs and preferences of each customer segment.

Given these findings, the Company should adopt a more customized marketing approach to effectively reach these distinct customer segments. For example, marketing campaigns promoting high-end motorcycle models should be concentrated in regions and among customer groups with higher purchasing power, while more affordable models should be marketed in areas where customers show a preference for lower payment options. Additionally, the Company would benefit from investing in improved data collection methods to capture more granular details about customer preferences, behaviors, and demographics. Such enhancements would allow for even more refined segmentation and targeted marketing, ultimately driving increased sales and strengthening the Company's market position.

While the current study provides valuable insights into customer segmentation using advanced data mining techniques, it is important to recognize its limitations. The findings are constrained by the scope of the dataset and the specific context of the Company. Future research should aim to expand the dataset to encompass a broader range of customer demographics and behaviors and explore additional clustering algorithms to validate and extend the results. Furthermore, conducting longitudinal studies could offer deeper insights into the long-term effects of these data-driven strategies on the Company's sales performance and market share, providing a more comprehensive understanding of their impact.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

4. CONCLUSION

Leveraging the K-Means algorithm for customer segmentation has empowered the Company to uncover critical insights into motorcycle purchasing behaviors and preferences, enabling the formulation of more precise and impactful marketing strategies. The use of the Elbow method and Silhouette score effectively identified the optimal number of clusters and evaluated their quality, facilitating a more precise analysis of customer data. The data preprocessing phase ensured the accuracy of these results by addressing incomplete or irrelevant data, and the application of data mining techniques revealed hidden patterns in motorcycle sales that were previously unrecognized. The findings show clear differences in purchasing preferences among customer groups, offering actionable insights to optimize the Company's marketing and production strategies. Overall, this study demonstrates the value of integrating advanced data mining methods in customer segmentation, contributing to enhanced operational efficiency and strategic decision-making not only for the Company but also providing a model for the broader motorcycle industry to adopt similar approaches for improved market positioning.

REFERENCES

[1] P. E. Sudjiman and L. S. Sudjiman, "Analisis Sistem Informasi Manajemen Berbasis Komputer Dalam Proses Pengambilan Keputusan," *TeIKa*, vol. 8, no. 2, 2020, doi: 10.36342/teika.v8i2.2327.

[2] W. Wahyudi, "Rancang Bangun Deteksi Warna Berbasis Machine Learning Arduino Mega Pro Mini Atmega2560-16AU," *Micronic J. Multidiscip. Electr. Electron. Eng.*, 2023, doi: 10.61220/micronic.v1i1.20235.

[3] S. Hasan and N. Muhammad, "Sistem Informasi Pembayaran Biaya Studi Berbasis Web Pada Politeknik Sains Dan Teknologi Wiratama Maluku Utara," *IJIS - Indones. J. Inf. Syst.*, vol. 5, no. 1, 2020, doi: 10.36549/ijis.v5i1.66.

[4] W. Gede Endra Bratha, "Literature Review Komponen Sistem Informasi Manajemen: Software, Database Dan Brainware," *J. Ekon. Manaj. Sist. Inf.*, vol. 3, no. 3, 2022, doi: 10.31933/jemsi.v3i3.824.

[5] Tarisno Amijoyo and A. K. Pasya, "Sistem Informasi Manajemen Seminar Pada Indonesia Gaming Arena," *J. Inform. Teknol. dan Sains*, vol. 4, no. 4, 2022, doi: 10.51401/jinteks.v4i4.2041.

[6] R. Siswanto, A. Susanto, and E. Saputra, "Sistem Informasi Helpdesk Ticketing di PT Tunas Artha Gardatama," *J. Ris. dan Apl. Mhs. Inform.*, vol. 1, no. 03, 2020, doi: 10.30998/jrami.v1i03.313.

[7] V. Plotnikova, M. Dumas, and F. Milani, "Adaptations of data mining methodologies: A systematic literature review," *PeerJ Comput. Sci.*, vol. 6, 2020, doi: 10.7717/PEERJ-CS.267.

[8] F. A. Setiawan, M. Sadikin, and E. R. Kaburuan, "Analisis Permasalahan

Perangkat Pencetak Menggunakan Metode Algoritma K-Means dan K-Medoids," *Teknika*, vol. 11, no. 2, 2022, doi: 10.34148/teknika.v11i2.471.

[9] M. I. Fianty, M. E. Johan, A. Aulia, and M. M. Veronica, "Application of Clustering-Based Data Mining for the Assessment of Nutritional Status in Toddlers at Community Health Centers," *J. Inf. Syst. Informatics*, vol. 5, no. 4, 2023, doi: 10.51519/journalisi.v5i4.586.

[10] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Comput. Sci. Rev.*, vol. 40, p. 100402, 2021.

[11] C. Wu, B. Yan, R. Yu, B. Yu, X. Zhou, Y. Yu, and N. Chen, "k-Means Clustering Algorithm and Its Simulation Based on Distributed Computing Platform," *Complexity*, vol. 2021, no. 1, p. 9446653, 2021.

[12] E. T. Naldy and A. Andri, "Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN," *J. Nas. Ilmu Komput.*, vol. 2, no. 2, pp. 89-101, 2021.

[13] R. D. Handayani and R. A. Aziz, "Framework Information Technology Infrastructure Library (Itil V3) : Audit Teknologi Informasi Sistem Informasi Akademik (Siakad) Perguruan Tinggi," *Explor. J. Sist. Inf. dan Telemat.*, vol. 11, no. 1, 2020, doi: 10.36448/jsit.v11i1.1456.

[14] I. Pratiwi, "Pengaruh Dukungan Manajemen Puncak, Kecanggihan Teknologi Informasi, Kualitas Sistem Informasi Akuntansi Dan Kinerja Individual Terhadap Efektifitas Sistem Informasi Akuntansi (Studi Pada Hotel Berbintang Tiga dan Empat Di Provinsi Banten)," *J. Ris. Akunt. Tirtayasa*, vol. 4, no. 1, 2019, doi: 10.48181/jratirtayasa.v4i1.5476.

[15] K. I. Jones and Swati Sah, "Application of Machine Learning With Big Data Analytics In The Insurance Industry," *Int. J. Data Informatics Intell. Comput.*, vol. 11, no. 12, 2022.

[16] Reymar, "Top 9 Big Data Challenges (And How You Can Solve Them Easily)," *SynchApps by Cazoomi*. 2023.

[17] A. I. Ramdhani, R. B. Agung, and W. Hermawan, "Rancang Bangun Aplikasi Persediaan Barang Pada Cv . Indoprima Motor Menggunakan Metode Fifo (First in First Out) Berbasis Web," *J. Gerbang STMIK Bani Saleh*, vol. 13, no. 1, 2023.

[18] R. Setiawan, "Menurunkan Claim Next Process Reject Plate R Cembung Pada Proses Perakitan Crankshaft Menggunakan Metode Eight Steps di PT XYZ," *Technologic*, vol. 13, no. 2, 2022, doi: 10.52453/t.v13i2.417.

[19] E. P. Yudha and R. A. Dina, "Pengembangan Potensi Wilayah Kawasan Perbatasan Negara Indonesia (Studi Kasus: Ranai-Natuna)," *Tata Loka*, vol. 22, no. 3, 2020.

[20] N. Nurhani and F. I. Sari, "Pengaruh Pelayanan Terdepan Terhadap Kepuasan Konsumen Pada PT Astra Internasional Tbk - Motor Di Makassar," *Nobel Manag. Rev.*, vol. 3, no. 2, 2022, doi:

10.37476/nmar.v3i2.2931.
[21] N. Nur'aeni and S. Supartono, "Pengaruh Kualitas Produk, Citra Merek, Dan Desain Produk Terhadap Keputusan Pembelian Sepeda Motor Honda Beat," *J. Dimens.*, vol. 11, no. 1, 2022, doi: 10.33373/dms.v11i1.3520.
[22] Salsabila MR, "Teknik Analisis Data Pengertian Hingga Contoh Penggunaan," *DQLab*. 2022.
[23] B. P. Wongso, M. E. Johan, and M. I. Fianty, "Empowering Pregnancy Risk Assessment: A Web-Based Classification Framework with K-Means Clustering Enhanced Models," *J. Inf. Syst. Informatics*, vol. 5, no. 4, 2023, doi: 10.51519/journalisi.v5i4.568.
[24] D. M. Saputra, D. Saputra, and L. D. Oswari, "Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method," 2020. doi: 10.2991/aisr.k.200424.051.
[25] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, 2023, doi: 10.30871/jaic.v7i1.4947.
[26] M. Sholeh and K. Aeni, "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 8, no. 1, 2023, doi: 10.30998/string.v8i1.16388.