Unveiling Banking Customer Pattern Through K-Means and DBSCAN Cluster Evaluation

Wilsen Grivin Mokodaser*[1], Fanny Soewignyo[2], Tonny Irianto Soewignyo[3]
[1]Informatika, Fakultas Ilmu Komputer, Universitas Klabat, Airmadidi
[2,3]Management, Fakultas Ekonomi dan Bisnis, Universitas Klabat, Airmadidi
e-mail: *[1]**wilsenm@unklab.ac.id**, [2]f.soewignyo@unklab.ac.id, [3]tonnysoewignyo@unklab.ac.id

Abstract

The banking sector's Customer segmentation plays a crucial role in understanding diverse customer behaviors and designing targeted financial strategies. This study employs K-Means and DBSCAN algorithms to analyze banking data and uncover customer behavior patterns. The clustering process with K-Means (k=5) produced five distinct groups of customers, each characterized by differences in loan amount, credit limit, account balance, and rewards points. These clusters demonstrate meaningful differentiation, such as customers with high loan exposure and moderate credit limits compared to groups with smaller loans but higher rewards accumulation. In contrast, DBSCAN (eps=1.9, min_sample=5) produced three clusters. Most data points were concentrated in one large cluster, while a few formed small groups with considerable noise. Evaluation metrics further confirmed the superiority of K-Means over DBSCAN. K-Means achieved better performance (Silhouette = 0.09, DBI = 2.8, CHI =820), indicating moderate separation and interpretability. In contrast, DBSCAN showed a negative Silhouette Score (-0.354) and a low Calinski-Harabasz Index (3.608), indicating poorly defined clusters. These results suggest that K-Means is more effective for banking customer segmentation, providing clearer profiling insights, whereas DBSCAN is less suitable due to the dataset's homogeneity and distribution.

Keywords— Banking, K-Means, DBSCAN, Evaluation.

1. INTRODUCTION

The development of the modern banking industry presents new challenges in understanding the increasingly diverse behaviors and needs of customers [1]. With the growing number of digital transactions, loans, credit card usage, and reward systems, banks need to have the right strategy to group customers based on their behavioral patterns [2]. Customer segmentation has become one of the key approaches to support marketing strategies, risk management, and the enhancement of customer satisfaction [3]. In the context of the banking industry, paying attention to customer activities is not merely about recording financial transactions, but also about understanding their behaviors, preferences, and patterns of interaction with the available services [4]. In-depth analysis of customer activities can help banks identify the varying needs across segments, design more relevant products, and provide more personalized services [5][6]. Moreover, a comprehensive understanding of customer activities also plays an important role in detecting potential risks, such as loan defaults or indications of suspicious activities related to financial security [7][8]. Thus, serious attention to customer activities not only contributes to improving customer satisfaction and loyalty [9] but also strengthens the competitiveness and sustainability of banks in facing increasingly competitive market dynamics [10].

To achieve a deeper understanding of customer activities, data-driven analysis methods such as clustering serve as an effective approach [11]. Clustering enables banks to group customers based on similarities in behavior and financial characteristics [12], without the need

for predefined labels or categories. Algorithms such as K-Means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise) provide different perspectives in customer segmentation: K-Means is effective for identifying general patterns in data with a defined number of clusters, while DBSCAN can identify groups of customers with unique patterns, including those with extreme or irregular characteristics [13][14]. By leveraging both approaches, banks can gain a more comprehensive overview of customer profiles, allowing marketing strategies, product offerings, and risk management to be carried out more accurately and in a data-driven manner. Among clustering techniques, K-Means is widely used for segmentation due to its simplicity and efficiency, although it performs poorly when data are unevenly distributed or contain complex shapes [15]. To address this, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) serves as an alternative, as it can detect clusters with irregular shapes and identify outliers that frequently appear in financial data [16].

Nevertheless, clustering results need to be evaluated objectively to ensure the quality and relevance of the formed segments. One commonly used evaluation metric is the Silhouette Score, which measures both intra-cluster cohesion and inter-cluster separation [17]. Thus, the combination of K-Means, DBSCAN, and evaluation using the Silhouette Score can provide more comprehensive insights into banking customer patterns [18]. This study focuses on uncovering banking customer behavior patterns through a comparison of the K-Means and DBSCAN algorithms, complemented by cluster profiling and Silhouette evaluation. The findings are expected to contribute to the development of more effective customer segmentation strategies and support data-driven decision-making in the banking sector.

2. RESEARCH METHODS

The research methodology is designed to provide a systematic overview of the steps taken in analyzing banking customer patterns through the application of K-Means and DBSCAN algorithms. Each stage in this methodology plays an important role in ensuring that the data used is clean, the analysis process follows scientific principles, and the results obtained can provide meaningful interpretation. Thus, this methodology not only explains the technical procedures but also serves as a conceptual foundation for achieving the research objectives.

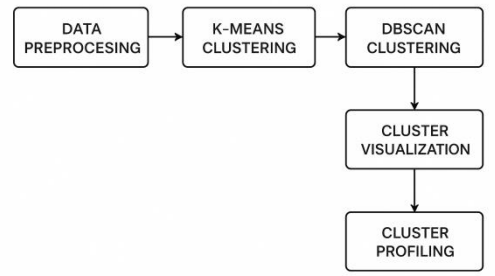

Figure 1. Research steps

2.1. Data Preprocessing

In this stage, raw data is cleaned by removing duplicate values, handling missing values, and performing normalization using StandardScaler to ensure all features are on the same scale [19]. Raw data is often not immediately ready for use. Normalization is necessary so that distance-based algorithms (such as K-Means and DBSCAN) are not biased toward features with larger scales.

2.2. K-Means Clustering

In this stage, the implementation of the K-Means algorithm will be carried out following the formula shown below.

$$d(xi, \mu j) = \sqrt{\sum(xi - \mu j)^2} \quad (1)$$

where:
$d(xi, \mu j)$ = Distance between the *i*-th data point and the *j*-th centroid
xi = The *i*-th data point
μj = The centroid of the *j*-th cluster
$(xi - \mu j)^2$ = The squared difference

In this stage, the K-Means algorithm is executed by testing several values of k (the number of clusters), and the best k is selected using the Elbow Method or Silhouette Score [20]. This process divides the data into groups based on minimizing the distance to the cluster centers. K-Means is a popular algorithm that is simple yet effective for segmentation. It can provide a good baseline for understanding the data structure, although it is only optimal for spherical-shaped clusters.

2.3. DBSCAN Clustering

In this stage, the implementation of the DBSCAN algorithm will be carried out following the formula shown below.

$$d(P, C) = \sqrt{\sum_{i-1}^{n}(x_{pi} - y_{ci})^2} \quad (2)$$

Where:
d(P,C) = The Euclidean distance from data point P to data point C (cluster center)
x_{pi} = The value of the *i*-th feature of data point P
x_{ci} = The value of the *i*-th feature of data point C (cluster center)
n = The dimension of the data

In this stage, DBSCAN is executed with the parameters *eps* (the maximum distance between points in a cluster) and *min_samples* (the minimum number of members in a cluster) [21]. Parameter tuning is performed to find the optimal combination. DBSCAN is capable of detecting clusters with irregular shapes and automatically identifying outliers. This is particularly important for banking data, as customer behavior is not always distributed in simple patterns.

2.4. Cluster Visualization

In this stage, the clustering results are visualized using PCA or t-SNE to reduce the dimensions into 2D/3D, and then displayed as a scatter plot with different colors for each cluster [22]. Visualization makes it easier to understand the data distribution and the patterns among clusters. This helps to see whether the formed clusters are truly separated or overlapping.

2.5. Cluster Profiling

In this stage, the K-Means algorithm is executed by testing several values of k (the number of clusters), with the best k selected using the Elbow Method or Silhouette Score. This process divides the data into groups based on minimizing the distance to the cluster centers. K-Means is a popular algorithm that is simple yet effective for segmentation. It can provide a good baseline for understanding the data structure, although it is only optimal for spherical-shaped clusters.

2.6. Model Evaluation

In this stage, evaluation is carried out by calculating the Silhouette Score (to measure the quality of cluster separation) and examining the distribution of members in each cluster [23]. Evaluation is necessary to ensure that the clustering results are not only formed mathematically

but also truly possess good separation quality. The Silhouette Score provides insight into whether the formed clusters are compact and well-separated. The Davies-Bouldin Index is useful for assessing clustering quality, where lower values indicate more compact and better-separated clusters. The Calinski-Harabasz Index is useful for evaluating clustering quality by comparing within-cluster density and between-cluster separation, where higher values indicate better clustering results.

3. RESULT AND DISCUSSION

3.1 Data Preprocessing

	Customer ID	First Name	Last Name	Age	Gender	Address	City	Contact Number	Email	Account Type	...	Minimum Payment Due	Payment Due Date	Last Credit Card Payment Date	Rewards Points	Feedback ID	Feedback Date	Feedback Type	Resolution Status	Resolution Dat
0	1	Joshua	Hall	45	Male	Address_1	Fort Worth	19458794854	joshua.hall@kag.com	Current	...	226.22	11/26/2023	3/20/2023	8142	1	10/6/2023	Suggestion	Resolved	1/22/202
1	2	Mark	Taylor	47	Female	Address_2	Louisville	19458794855	mark.taylor@kag.com	Current	...	42.84	11/5/2023	6/16/2023	4306	2	4/7/2023	Complaint	Resolved	8/27/202
2	3	Joseph	Flores	25	Female	Address_3	Philadelphia	19458794856	joseph.flores@kag.com	Current	...	162.12	1/8/2023	3/20/2023	4842	3	9/7/2023	Praise	Pending	5/11/202
3	4	Kevin	Lee	52	Other	Address_4	Oklahoma City	19458794857	kevin.lee@kag.com	Savings	...	216.46	9/8/2023	10/15/2023	9463	4	5/28/2023	Complaint	Resolved	7/5/202
4	5	Linda	Johnson	68	Other	Address_5	Phoenix	19458794858	linda.johnson@kag.com	Savings	...	1.29	3/4/2023	7/27/2023	2209	5	2/12/2023	Complaint	Resolved	11/21/202

5 rows × 40 columns

Figure 2. Banking data structure

Figure 2 above shows the result of loading data in Google Colab, illustrating the structure of the dataset to be used in this study. The dataset consists of 40 columns and 5,000 records. The preprocessing stage is a crucial initial step in this research, as the quality of the analysis results strongly depends on the cleanliness and relevance of the data used. At this stage, customer data obtained from the file Comprehensive_Banking_Database.csv is first loaded using pandas and then validated for its structure and data types. Next, data cleaning is performed by removing sensitive or irrelevant columns such as Customer ID, Feedback ID, Contact Number, and Email, so that the analysis focuses on customers' financial behavior without involving personal identity attributes. From the available data, only numeric columns are selected for analysis using the select_dtypes function, resulting in a numeric-only dataset (df_numeric). Missing values are then checked and handled using imputation methods, such as applying the median value, which aligns with the characteristics of financial data. Feature cleaning is also conducted, including handling extreme outliers and adjusting unit scales if necessary to improve interpretability. To ensure each variable has balanced weight in the analysis, the data is transformed using StandardScaler, producing a standardized dataset (X_scaled). All processed results, including df_numeric, X_scaled, and scaler.pkl, are stored as artifacts to facilitate replication and interpretation of the research findings. With these steps, the data becomes more structured, free from irrelevant attributes, and ready for the modeling stage.

3.2 K-Means Clustering

This stage uses the K-Means clustering algorithm to group customers based on similarities in their financial behavior. The process begins by determining the optimal number of clusters using the Elbow Method and Silhouette Score. After identifying the best number of clusters (for example, k=3), the standardized data (X_scaled) is used as input for K-Means. The result of this stage is a cluster label for each customer, dividing the entire dataset into three major segments. Thus, K-Means provides an initial overview of how customer patterns can be grouped relatively homogeneously within their respective clusters.

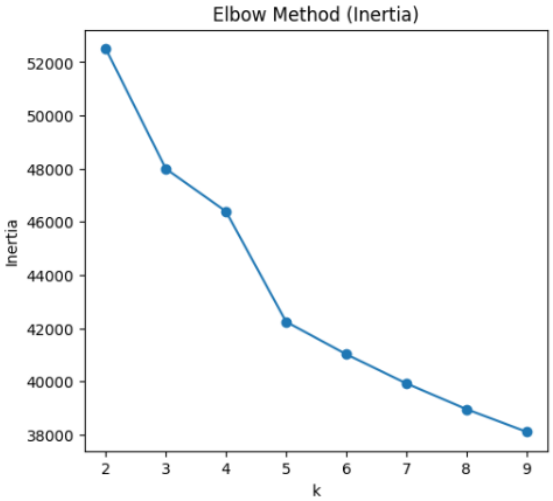

Figure 3. K-Means Elbow Method

The figure above shows the analysis results using the Elbow Method to determine the optimal number of clusters in the K-Means algorithm. The horizontal axis (X) represents the number of clusters (k), while the vertical axis (Y) represents the inertia value, or the total squared distance between each data point and its nearest cluster center. It can be seen that the inertia decreases sharply from k=2 to k=5, but after k=5 the reduction becomes more gradual and less significant. This pattern indicates that the optimal point lies around k=5, as at this number of clusters the model is able to partition the data effectively without losing efficiency. Therefore, selecting five clusters is the most appropriate choice to represent data segmentation in this case, banking customer profiles, so that the resulting groups are more representative and easier to analyze.

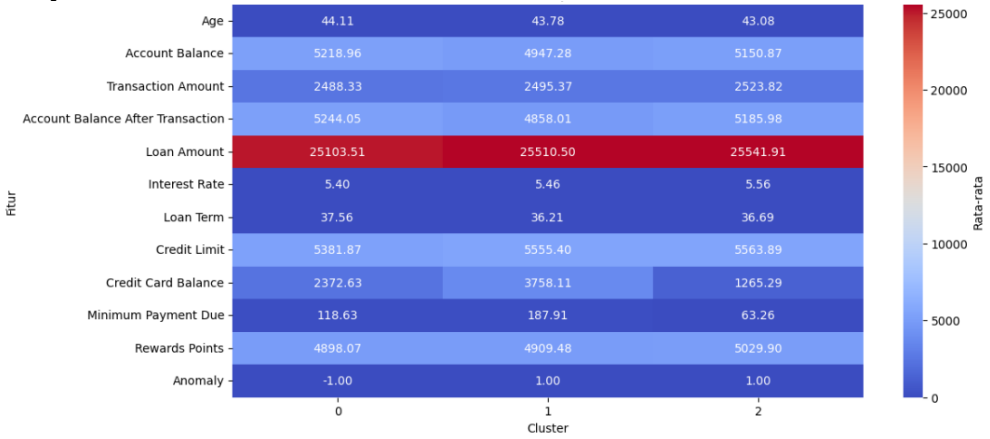

Figure 4. K-Means Heatmap cluster average

The figure above shows the results of cluster profiling, which reveal distinct characteristics for each customer group. Cluster 0 consists of 300 customers, characterized by a relatively high Loan Amount (25,103.51), a Credit Limit of 5,381.87, and an Account Balance After Transaction of 5,244.05. However, this cluster has an Anomaly value of -1, a low Interest Rate (5.40), and the shortest Loan Term (37.56). Cluster 1, with 2,318 customers, is marked by a high Loan Amount (25,510.50), Credit Limit (5,555.40), and Account Balance (4,947.28), but tends to have an Anomaly value of 1, an Interest Rate of 5.46, and a relatively short Loan Term (36.21). Meanwhile, Cluster 2 consists of 2,382 customers with strong indicators in Loan Amount (25,541.91), Credit Limit (5,563.89), and Account Balance After Transaction (5,185.98). Similar to Cluster 1, this group also has an Anomaly value of 1, an Interest Rate of 5.56, and a relatively short Loan Term (36.69). Overall, all three clusters are characterized by high loan amounts and

credit limits, but they are differentiated by variations in ending transaction balances, interest rates, and loan terms

3.3 DBSCAN Clustering

At this stage, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is applied. Unlike K-Means, DBSCAN does not require the number of clusters to be predetermined; instead, it relies on two key parameters: *eps* (the neighborhood radius) and *min_samples* (the minimum number of points required to form a dense region).

In this study, the tuning of these parameters was conducted systematically by evaluating several *eps* and *min_samples* combinations, supported by visual inspection of the k-distance graph and cluster separation quality. The final configuration (*eps = 1.9* and *min_samples = 10*) was selected because it produced the most interpretable grouping pattern with a reasonable number of clusters and a manageable amount of noise data. This process ensures that parameter tuning is not arbitrary but rather based on data-driven assessment of density and separability.

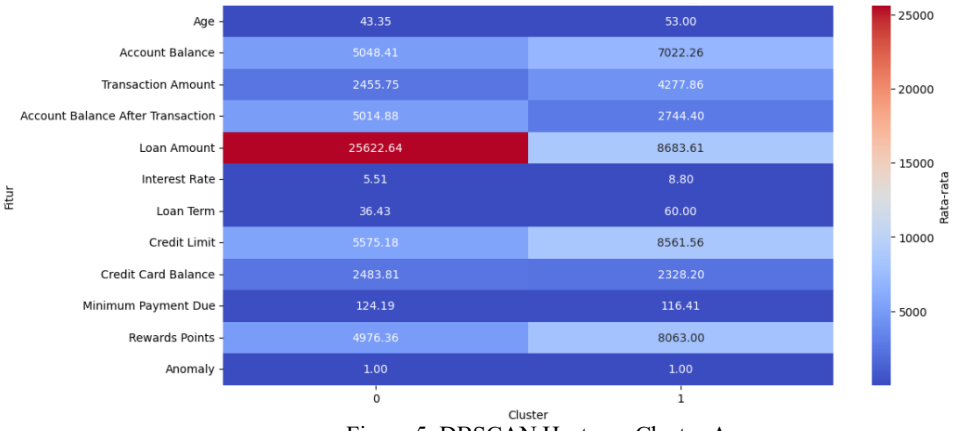

Figure 5. DBSCAN Heatmap Cluster Average

The image above shows that the clustering results generated three valid clusters and a significant number of noise points (1,103 customers). Cluster 0 emerged as the dominant segment (4,477 customers) with high Loan Amounts (25,622.64), moderate Credit Limits (5,575.18), and stable Account Balances (around 5,048.41). In contrast, low values were observed in Interest Rate (5.51%) and Loan Term (approximately 36 months), indicating a segment of moderate-risk borrowers with large but stable credit exposure.

Although Cluster 1 and Cluster 2 contain only 4 and 3 customers, respectively, they are analytically significant. Cluster 1 represents older customers (average age 53) with lower loan amounts but exceptionally high Reward Points (8,063) and Credit Limits (8,561.56), suggesting premium clients with high spending capacity but conservative borrowing behavior. Cluster 2, showing extreme deviations in several variables, likely represents outlier or exceptional customer profiles.

While the Silhouette Score of -0.0875 indicates that overall cluster separation is weak, the presence of these small and distinct clusters highlights DBSCAN's ability to capture niche financial behaviors and potential anomalies that might be overlooked by partition-based methods like K-Means. Therefore, these clusters, although small, provide valuable analytical insights for identifying unique customer patterns, potential fraud risks, or targeted financial strategies.

3.4 Cluster Visualization

This step aims to provide a visual representation of the clustering results from both K-Means and DBSCAN. By using dimensionality reduction methods such as PCA (Principal Component Analysis), data with many features are projected into two dimensions for visualization. The resulting graph shows the distribution of customers as points, colored according

to their respective clusters. From this visualization, it is apparent that K-Means tends to form clusters with more regular boundaries, whereas DBSCAN produces more small clusters along with a large number of points categorized as noise. This visualization helps to enhance the understanding of the complex structure of customer data.

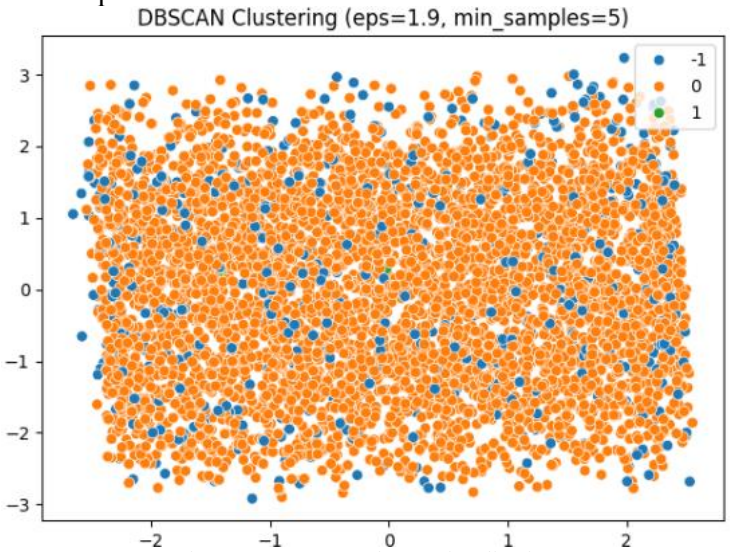

Figure 6 DBSCAN cluster visualization

The image above shows that, based on the summary of average feature profiles, Cluster 0 is dominated by customers with an average age of 43 years, an Account Balance of 5,048.40, and a Transaction Amount of around 2,455.75. The average post-transaction balance reaches 5,014.88, with a relatively high Loan Amount of 25,622.64 and a Credit Limit of 5,575.18. The average Loan Term is 36 months, while the Credit Card Balance is 2,483.81, with a Minimum Payment Due of approximately 124.19. Customers in this cluster also have a substantial amount of Reward Points, totaling 4,976.36, and a consistent Anomaly value of 1.

Meanwhile, Cluster 1 shows characteristics of older customers, with an average age of 53 years. They have a higher Account Balance of 7,022.26 and a Transaction Amount of 4,277.87. However, the post-transaction balance is much lower at only 2,744.40. Their Loan Amount is smaller at 8,683.62, but with a longer Loan Term of 60 months. Additionally, their Credit Limit is higher (8,561.56) compared to Cluster 0, with a Credit Card Balance of 2,328.20 and a Minimum Payment Due of 116.41. This cluster also has higher Reward Points, totaling 8,063. Both clusters share an Anomaly value of 1, but the main differences lie in age, post-transaction balance, loan amount, and loan term

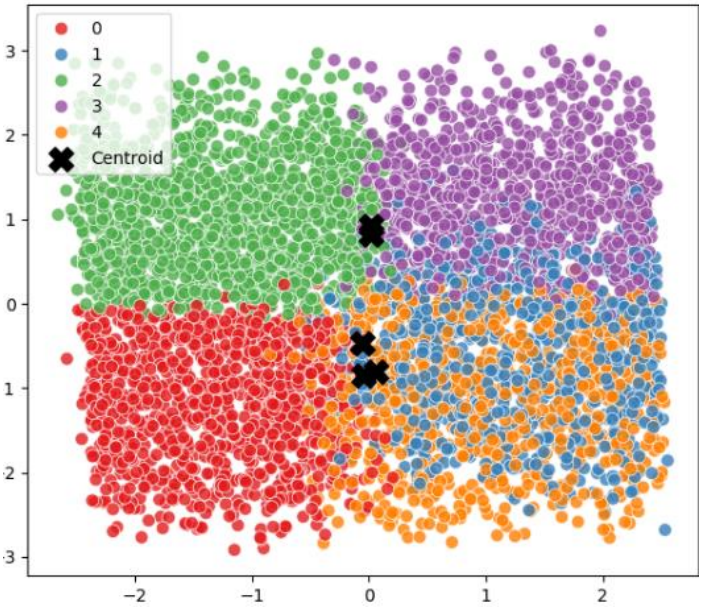

Figure 7. K-Means cluster visualization

The image above shows the results of applying the K-Means clustering algorithm with the number of clusters set to five (k=5). Each color represents a formed data group, while the black "X" marks indicate the positions of the centroids, serving as the center of gravity for each cluster, calculated from the average position of its members. This visualization shows that the data can be segmented into five main groups, although there are still overlapping areas indicating observations with similar characteristics at cluster boundaries, particularly among the blue, orange, and purple clusters. The presence of centroids provides an average representation of customer profiles within each group, facilitating the interpretation of segmentation results. Methodologically, these results confirm K-Means' ability to divide customer data into more structured groups, which can, in turn, be leveraged to gain deeper insights into banking customers' financial behavior patterns.

3.5 Cluster Profiling

This stage aims to interpret the clustering results through cluster profiling. The approach involves calculating the average values of key features (such as Loan Amount, Credit Limit, Rewards Points, and Account Balance) within each cluster. This profiling is visualized as a heatmap, allowing the dominant patterns in each customer group to be clearly observed. For example, some clusters may be characterized by high loan amounts, others by large account balances, or some may stand out for their high number of reward points. These results provide a concrete picture of the distinctive characteristics of each group, which is highly valuable for banks in developing marketing strategies and product offerings.

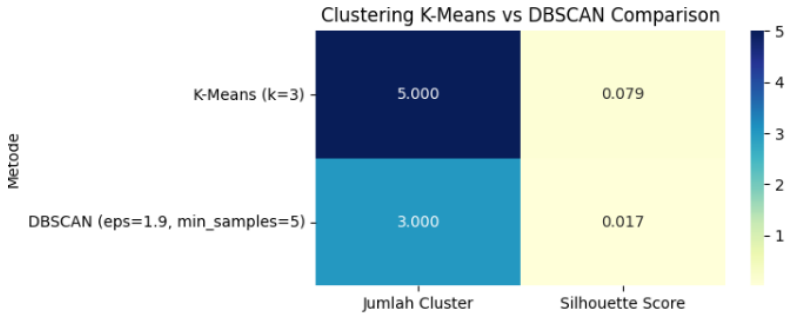

Figure 8. Cluster Comparison

The image above shows that the comparison between K-Means and DBSCAN reveals quite significant differences both in the number of clusters formed and in the Silhouette Score values. In K-Means with k=5, five clusters are formed with a relatively balanced distribution, ranging from 812 to 1,294 members per cluster. Nevertheless, the Silhouette Score of 0.078692 is still considered low, indicating that the separation between clusters is not entirely clear and there is overlap between segments.

In contrast, DBSCAN with parameters eps=1.9 and min_samples=5 produces only three clusters, where the majority of the data (4,477 observations) falls into one main cluster, while the other clusters contain only 4 members, and 519 data points are categorized as noise (-1). The Silhouette Score for DBSCAN is even lower at 0.017217, suggesting that the resulting cluster structure is less representative.

Overall, these results indicate that for the banking dataset used, K-Means is better able to divide the data into more balanced groups, even though the quality of cluster separation remains limited, whereas DBSCAN tends to be less effective in identifying meaningful cluster structures with the parameters applied.

3.6 Model Evaluation

```
=== K-Means Evaluation ===
Silhouette Score        : 0.090
Davies-Bouldin Index    : 2.801
Calinski-Harabasz Index : 820.416

=== DBSCAN Evaluation ===
Silhouette Score        : -0.354
Davies-Bouldin Index    : 1.814
Calinski-Harabasz Index : 3.608
```

Figure 9. Model evaluation result

The evaluation results show a significant difference in clustering quality between K-Means and DBSCAN. For K-Means, the Silhouette Score of 0.090 indicates that the separation between clusters is still weak, although some group structures are formed. The Davies-Bouldin Index (2.801) is relatively high, suggesting that the distances between clusters are not ideal and there is similarity between groups. Meanwhile, the Calinski-Harabasz Index (820.416) is fairly large, indicating some variation between clusters that can be considered, although the overall quality is not optimal.

In contrast, the DBSCAN evaluation results show much poorer performance. The negative Silhouette Score (-0.354) indicates that most data points do not fit well within the clusters formed and are closer to other clusters, making the separation highly suboptimal. The Davies-Bouldin Index (1.814) is indeed lower than K-Means, but in the context of DBSCAN, this occurs because most of the data is concentrated in a single large cluster with an imbalanced distribution. This is also supported by the very low Calinski-Harabasz Index (3.608), confirming that the cluster structure produced by DBSCAN provides almost no meaningful differentiation between groups.

Overall, although K-Means produces clusters of still weak quality, its results are more representative compared to DBSCAN for this banking dataset. DBSCAN appears unsuitable for this data, likely due to the dense distribution and lack of clear density patterns to separate.

4. CONCLUSION

The evaluation of clustering model performance shows a significant difference between the K-Means and DBSCAN algorithms in the context of banking customer segmentation. For K-Means, the Silhouette Score of 0.090, although relatively low, still indicates the presence of

distinguishable group separations. The Davies-Bouldin Index of 2.801 suggests that there are still similarities between clusters, yet the cluster distribution provides valuable information regarding customer characteristics. Additionally, the Calinski-Harabasz Index of 820.416 reflects a meaningful level of variation between clusters, indicating that this model can still be used to identify differences in customers' financial patterns.

In contrast, DBSCAN demonstrates much poorer performance. This is reflected in the negative Silhouette Score (-0.354), indicating that the formed clusters lack clear separation. Although DBSCAN's Davies-Bouldin Index (1.814) is lower than K-Means, this is primarily due to the majority of the data being concentrated in one large cluster, which does not reflect pattern diversity. The very low Calinski-Harabasz Index (3.608) further confirms that the cluster structure produced by DBSCAN is insufficient for meaningful analysis.

Overall, these findings suggest that K-Means is more suitable for banking customer segmentation, as it is capable of producing more stable and informative clusters. K-Means successfully forms five customer groups with distinct characteristics, such as a group with high loan amounts and moderate credit limits, and another group with higher account balances and significant reward points. Meanwhile, DBSCAN only forms three clusters with an unbalanced distribution, where most customers are concentrated in a single large cluster with many points classified as noise, reducing its practical usefulness.

Thus, although the K-Means evaluation results are not yet optimal, the model still provides a useful foundation for banks to understand customer financial patterns and behavior. The findings also indicate that DBSCAN is less effective for banking data, which tends to be relatively homogeneous, suggesting that its application should be reconsidered or combined with other, more adaptive methods.

REFERENCES

[1] B. Arthur and H. Omari, "Adaptation of Technology on Products and Services Related in the Banking Industry: Challenges and Solutions," *South Asian J. Soc. Stud. Econ.*, vol. 20, no. 4, pp. 78–89, 2023.

[2] R. T. Tambunan and M. I. P. Nasution, "Tantangan dan strategi perbankan dalam menghadapi perkembangan transformasi digitalisasi di era 4.0," *Sci-Tech J.*, vol. 2, no. 2, pp. 148–156, 2023.

[3] R. Ardianto *et al.*, "Transformasi digital dan antisipasi perubahan ekonomi global dalam dunia perbankan," *MARAS J. Penelit. Multidisiplin*, vol. 2, no. 1, pp. 80–88, 2024.

[4] U. Rahardja, T. Hongsuchon, T. Hariguna, and A. Ruangkanjanases, "Understanding impact sustainable intention of s-commerce activities: The role of customer experiences, perceived value, and mediation of relationship quality," *Sustainability*, vol. 13, no. 20, p. 11492, 2021.

[5] K. Tabianan, S. Velu, and V. Ravi, "K-means clustering approach for intelligent customer segmentation using customer purchase behavior data," *Sustainability*, vol. 14, no. 12, p. 7243, 2022.

[6] F. Sudirjo, "Marketing strategy in improving product competitiveness in the global market," *J. Contemp. Adm. Manag.*, vol. 1, no. 2, pp. 63–69, 2023.

[7] G. Nwachukwu, "Enhancing credit risk management through revalidation and accuracy in financial data: The impact of credit history assessment on procedural financing," *Int. J. Res. Publ. Rev.*, vol. 5, no. 11, pp. 631–644, 2024.

[8] U. Turksen, V. Benson, and B. Adamyk, "Legal implications of automated suspicious transaction monitoring: enhancing integrity of AI," *J. Bank. Regul.*, vol. 25, no. 4, pp.

359–377, 2024.

[9] B. Srivastava, S. Singh, and S. Jain, "Bank competition, risk-taking and financial stability: insights from an emerging economy," *Compet. Rev. An Int. Bus. J.*, vol. 33, no. 5, pp. 959–992, 2023.

[10] N. Nassar, "Sustainable management for green competitiveness in the banking sector." Vilniaus Gedimino technikos universitetas., 2025.

[11] M. Karmenova *et al.*, "An approach for clustering of seismic events using unsupervised machine learning," *Acta Polytech. Hungarica*, vol. 19, no. 5, pp. 7–22, 2022.

[12] N. Hafizah, A. L. Hananto, F. Nurapriani, and E. Novalia, "SEGMENTASI NASABAH UMKM BERDASARKAN KINERJA DAN KEUNTUNGAN MENGGUNAKAN K-MEANS CLUSTERING," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 5, pp. 8661–8665, 2025.

[13] A. S. Paramita and T. Hariguna, "Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce," *J. Digit. Mark. Digit. Curr.*, vol. 1, no. 1, pp. 43–62, 2024, doi: 10.47738/jdmdc.v1i1.3.

[14] A. T. Thuse, "Customer Segmentation for Targeted Marketing: Exploring Dbscan & K-Means," *Preprints.org*, 2025, [Online]. Available: https://www.preprints.org/manuscript/202504.1434/v1

[15] I. Shafi *et al.*, "A review of approaches for rapid data clustering: Challenges, opportunities, and future directions," *IEEE Access*, vol. 12, pp. 138086–138120, 2024.

[16] Y. Kim and M. Vasarhelyi, "Anomaly detection with the density based spatial clustering of applications with noise (DBSCAN) to detect potentially fraudulent wire transfers.," *Int. J. Digit. Account. Res.*, vol. 24, 2024.

[17] M. Jahanian, A. Karimi, N. O. Eraghi, and F. Zarafshan, "Adaptive clustering for medical image analysis using the improved separation index," *Sci. Rep.*, vol. 15, no. 1, p. 28191, 2025.

[18] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, "Binning-based silhouette approach to find the optimal cluster using K-means," *IEEE Access*, vol. 10, pp. 115025–115032, 2022.

[19] M. Polatgil, "Investigation of The Effect of Data Normalization on Classification and Feature Selection in Intrusion Detection System," *Indones. J. Comput. Sci.*, vol. 11, no. 1, 2022.

[20] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-means algorithm for clustering corn planting feasibility area in south lampung regency," in *Journal of Physics: Conference Series*, 2021, vol. 1751, no. 1, p. 12038.

[21] K. Giri and T. K. Biswas, "Determining optimal epsilon (eps) on dbscan using empty circles," in *International Conference on Artificial Intelligence and Sustainable Engineering: Select Proceedings of AISE 2020, Volume 1*, 2022, pp. 265–275.

[22] T. T. Cai and R. Ma, "Theoretical foundations of t-sne for visualizing high-dimensional clustered data," *J. Mach. Learn. Res.*, vol. 23, no. 301, pp. 1–54, 2022.

[23] E. S. Dalmaijer, C. L. Nord, and D. E. Astle, "Statistical power for cluster analysis," *BMC Bioinformatics*, vol. 23, no. 1, p. 205, 2022.