Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 485–498
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

Implementation of the Density-Based Spatial Clustering (DBSCAN) Method to Identify Beauty Clinic Customer Purchasing Patterns

Taufik Rifki Al Hafiz Lubis [a], Mulkan Azhari [b*]

[a,b] Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

*Corresponding Author: mulkan@umsu.ac.id

ABSTRACT

Beauty clinics routinely accumulate detailed transaction records, yet these data are often used only for administrative reporting and are not transformed into actionable customer insights. This study implements Density-Based Spatial Clustering of Applications with Noise (DBSCAN) in a web-based system to identify customer purchasing patterns at Raen Aesthetic. Transaction data from January to December 2025 were cleaned, aggregated into recency, frequency, and monetary features, and normalized using a robust scaler. DBSCAN was executed with epsilon = 0.8000 and minimum samples = 5, while cluster quality was assessed using the Silhouette Score and Davies-Bouldin Index. The implementation produced nine customer clusters and 148 noise observations. The Silhouette Score of 0.3551 indicated a reasonably structured separation, whereas the Davies-Bouldin Index was 1.2569. Cluster 1 was the largest segment, comprising 355 customers with low purchase frequency and relatively small monetary value. Cluster 5 represented a high-value segment with an average monetary value of approximately IDR 10,302,000, while the noise group displayed heterogeneous but economically significant transaction behavior. Treatment transactions dominated most clusters, whereas several segments showed stronger preferences for packages and deposits. Black-box testing confirmed that all major functions operated according to the specified scenarios. The results demonstrate that DBSCAN can provide interpretable customer segmentation and support evidence-based service and customer-management decisions in beauty clinics.

How to Cite:
T. R. A. H. Lubis and M. Azhari, "Implementation of the Density-Based Spatial Clustering (DBSCAN) Method to Identify Beauty Clinic Customer Purchasing Patterns," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 485–498, 2026.
Doi: `https://doi.org/10.69916/jkbti.v5i3.559`

KEYWORDS

DBSCAN
Customer Segmentation
RFM Analysis
Beauty Clinic
Pattern Recognition

ARTICLE INFO

Received: June 23, 2026
Accepted: August 18, 2026
Available Online: September 1, 2026

1. INTRODUCTION

The beauty-service industry in Indonesia has expanded alongside increasing public awareness of personal care, health, and appearance. In this environment, beauty clinics function not only as healthcare-service providers but also as competitive service businesses that require effective customer management and data-informed decision making [1]. Transaction systems record visits, treatment types, product purchases, packages, deposits, and payment values; however, these records are frequently retained only as administrative archives. Consequently, their potential to reveal customer behavior, purchasing intensity, and service preferences remains underutilized. Data mining provides a systematic mechanism for transforming large operational datasets into interpretable patterns and decision-support information [2, 3]. Within this domain, clustering is particularly appropriate when predefined class labels are unavailable because it groups observations according to similarity in their measured characteristics [4]. For customer analysis, clustering supports segmentation based on behavioral variables and enables organizations to distinguish regular customers, inactive customers, high-value customers, and atypical purchasing behavior [5, 6].

Conventional partition-based methods commonly require the number of clusters to be specified in advance and often assume relatively compact cluster structures. These assumptions may be unsuitable for customer transaction data, which can contain irregular density patterns and economically meaningful outliers. DBSCAN addresses these limitations by forming clusters from dense neighborhoods and explicitly assigning sparse observations as noise [7, 8]. The method therefore provides two relevant capabilities for beauty-clinic data: discovering customer groups without fixing the cluster count beforehand and identifying customers whose behavior differs substantially from the dominant patterns [7].

Previous customer-segmentation studies have shown that transaction-based variables can provide actionable descriptions of loyalty, activity, and customer value [5, 9, 10]. Nevertheless, many implementations concentrate exclusively on the analytical model and do not integrate dataset administration, parameter configuration, visual interpretation, result history, and report generation within a unified operational application. This gap limits the practical use of clustering outputs by administrators who require repeatable analysis and accessible visual summaries.

This study contributes a web-based DBSCAN system that processes beauty-clinic transactions through data cleaning, RFM feature construction, robust normalization, density-based clustering, internal validation, and visual customer profiling. The system records clustering runs, displays noise and cluster statistics, presents purchase-type distributions and dimensionality-reduction plots, and supports administrative reporting [8, 11, 12]. The study uses Raen Aesthetic transaction data from January to December 2025 and evaluates the implementation through clustering-quality metrics and functional black-box testing. Accordingly, the objectives are to implement DBSCAN for grouping beauty-clinic customers according to purchasing behavior, identify the characteristics of the resulting customer segments, and provide structured segmentation information that can support customer-management and service-evaluation decisions. The research focuses on recency, frequency, and monetary indicators derived from historical transactions and does not attempt to predict future purchases or prescribe automated marketing decisions.

2. RESEARCH METHODS

2.1. Research Design and Data Source

The study employed a quantitative, applied data-mining design. The analytical object was Raen Aesthetic, and the dataset consisted of historical customer transactions recorded from January through December 2025. Data were obtained from the clinic information system and represented actual service activity during the observation period. The research stages comprised data acquisition, period-based selection, validation and cleaning, feature engineering, normalization, DBSCAN modelling, cluster evaluation, interpretation, system implementation, and functional testing.

2.2. Data Preparation and RFM Feature Construction

The raw records were examined for duplicate entries, missing values, inconsistent transaction identifiers, and invalid monetary values. Records outside the defined study period were excluded. Transactions were then aggregated at customer level to form three behavioral features: recency, defined as the elapsed time since the most recent transaction; frequency, defined as the number of invoices or transactions; and monetary value, defined as the accumulated transaction amount. Additional purchase-type proportions were retained for cluster interpretation, including treatment, product, package, and deposit transactions.

Because recency, frequency, and monetary values have substantially different numerical ranges and are sensitive to extreme transaction values, the system applied robust scaling before distance computation. This transformation reduced scale dominance and limited the influence of extreme values while preserving the relative structure required for density-based analysis.

2.3. DBSCAN Clustering Procedure

DBSCAN forms clusters by expanding connected regions that satisfy a local density requirement. For a data point p, the epsilon neighborhood is defined as the set of observations whose distance from p does not exceed ε [12]:

$$N_\varepsilon(p) = \{q \in D \mid \mathrm{dist}(p, q) \leq \varepsilon\} \tag{1}$$

A point is classified as a core point when the number of observations in its epsilon neighborhood is at least MinPts. Points reachable from connected core points are assigned to the same cluster, while observations that do not satisfy density reachability are labelled as noise. Euclidean distance was used for the normalized feature space:

$$\mathrm{dist}(p, q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2}$$

The final implementation used $\varepsilon = 0.8000$ and MinPts $= 5$ with robust scaling. Parameter selection was supported by a fifth-nearest-neighbor distance plot and empirical inspection of cluster formation. The system stored the parameter configuration, cluster count, number of noise observations, evaluation scores, and generated visual outputs for every analytical run.

2.4. Evaluation Metrics

Cluster quality was assessed using the Silhouette Score and Davies-Bouldin Index. The Silhouette Score measures cohesion within a cluster relative to separation from the nearest alternative cluster and ranges from -1 to 1; larger positive values indicate more coherent and better-separated groups. The Davies-Bouldin Index evaluates the average similarity between each cluster and its most similar competing cluster; lower values indicate more distinct cluster structures. The study also used principal component analysis (PCA) exclusively for two-dimensional visualization and did not use PCA as the primary clustering representation [13].

2.5. Web-Based System and Functional Testing

The analytical workflow was integrated into a web-based application with modules for authentication, user administration, dataset upload and activation, clustering execution, result history, reporting, and system configuration. Python executed the data-processing and clustering pipeline, while a relational database stored users, datasets, run metadata, configuration values, and summary outputs. Functional verification used black-box testing, in which system behavior was evaluated from input and output responses without inspecting internal source code. Twelve scenarios covered valid and invalid login, user-data operations, dataset handling, clustering execution, result presentation, configuration, reporting, and logout [14].

3. RESULTS AND DISCUSSION

3.1. System Implementation

The implemented application organizes the complete analytical process within a consistent administrative interface. Authentication is required before the user can access the clustering modules, thereby separating public access from operational functions.

As illustrated in Figure 1, the login page contains the application identity, username and password fields, and an authentication button. Submitted credentials are validated against stored account data. Successful authentication redirects the user to the dashboard, whereas invalid credentials produce an error message and retain the login form.

Figure 2 presents the dashboard, which summarizes the active dataset, transaction period, customer count, invoice count, and the most recent clustering run. The sidebar provides direct access to the dashboard, users, data, clustering, reports, and settings. By combining operational dataset information with clustering indicators, the dashboard allows the administrator to confirm the active analytical context before inspecting detailed outputs.

The user-management page shown in Figure 3 supports account search, creation, modification, and deletion. The table displays sequential numbers, usernames, full names, and email addresses. This module centralizes administrative control over authorized application users and supports structured account maintenance.

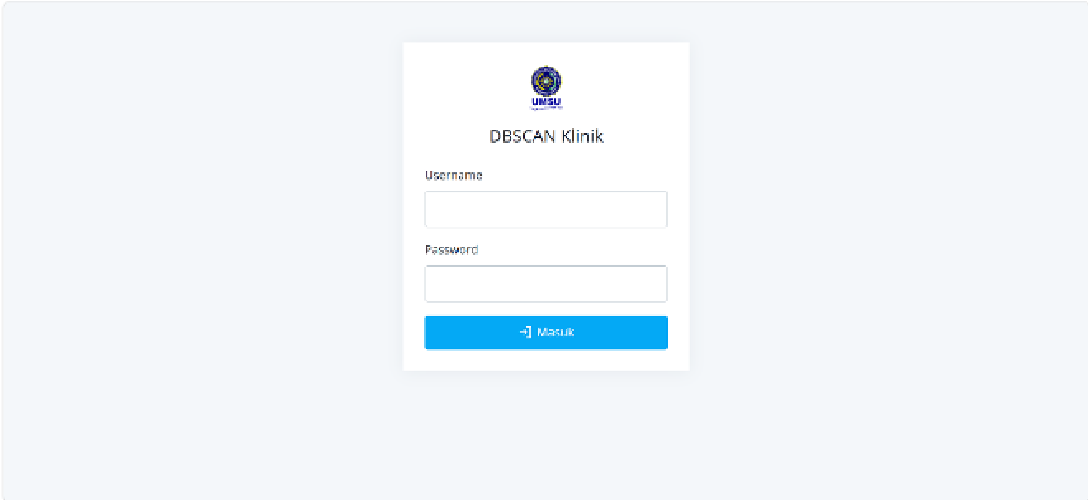


Figure 1. Login page of the DBSCAN clinic application.

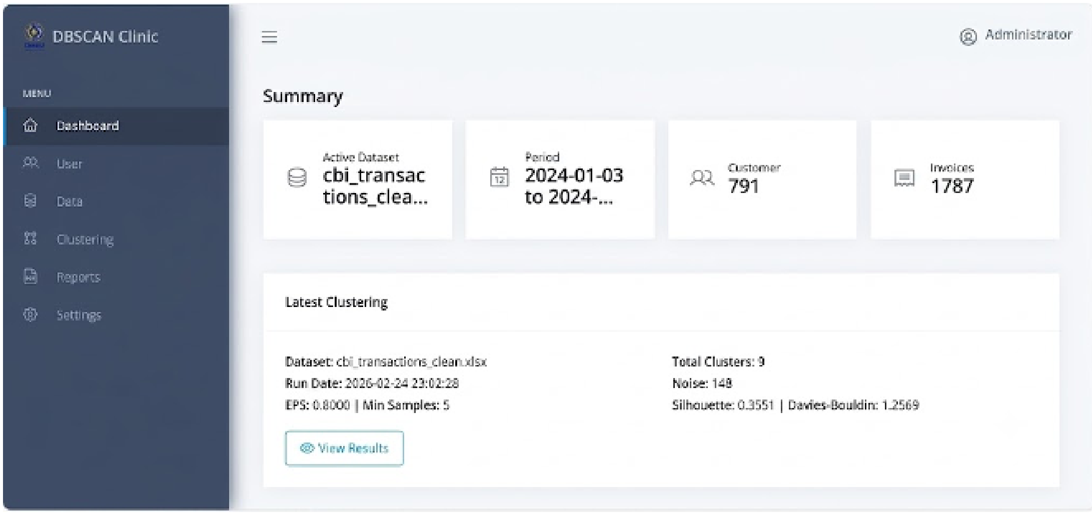


Figure 2. Application dashboard and latest clustering summary.

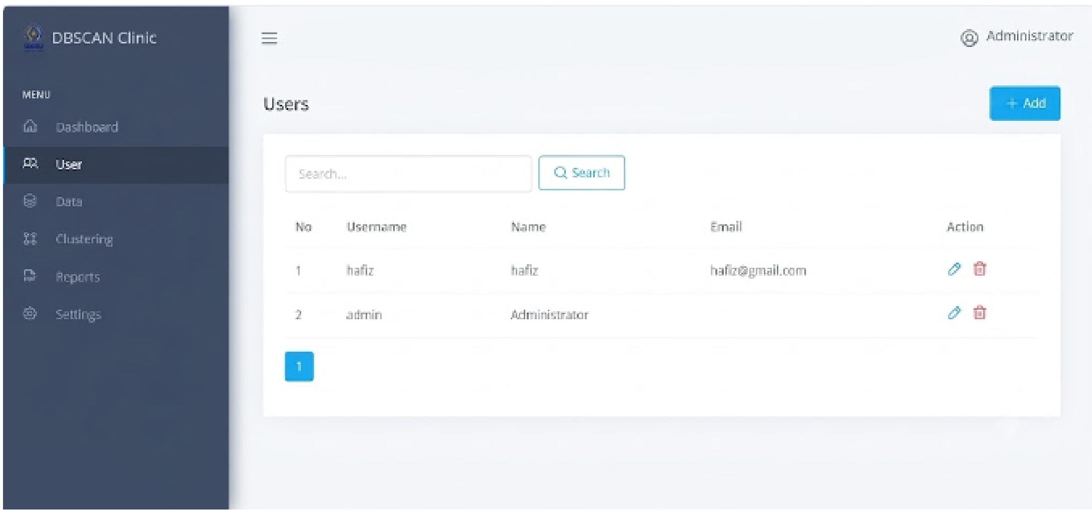


Figure 3. User-management page.

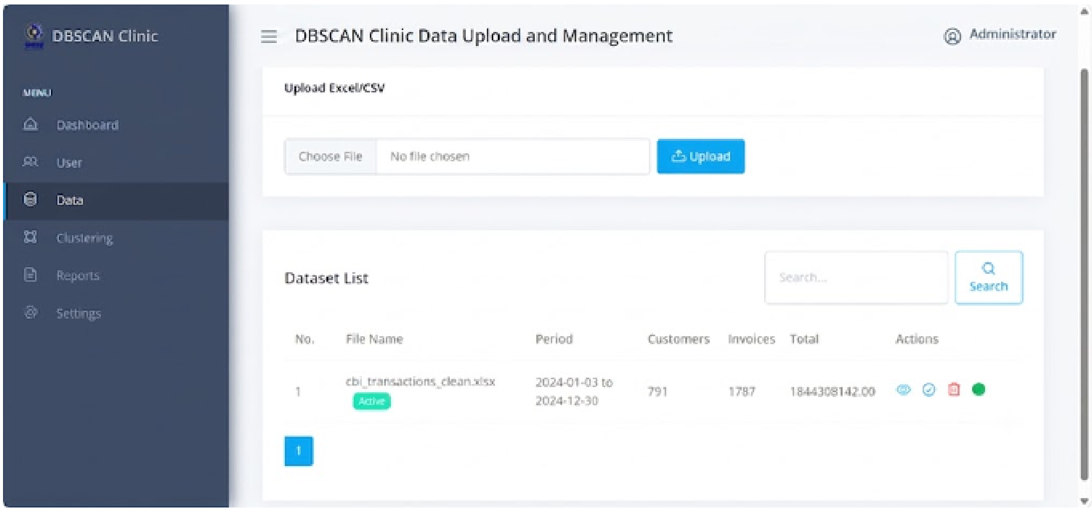


Figure 4. Dataset upload and activation page.

Figure 4 shows the dataset module. Excel or CSV files can be uploaded and recorded with their transaction period, number of customers, number of invoices, and total transaction value. A dataset can be designated as active for subsequent analysis, while additional actions provide detail inspection and deletion. This design prevents clustering from being executed against an unintended data source.

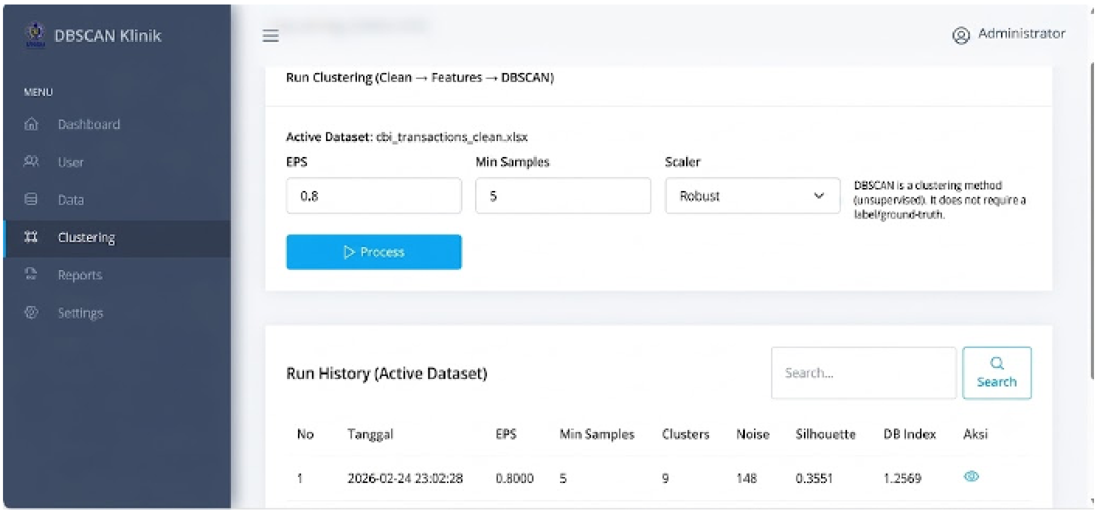


Figure 5. DBSCAN execution page and run history.

The clustering interface in Figure 5 integrates data cleaning, feature construction, scaling, and DBSCAN execution. The displayed run used the active file `cbi_transactions_clean.xlsx`, $\varepsilon = 0.8$, MinPts $= 5$, and robust scaling. The resulting history record reports nine clusters, 148 noise observations, a Silhouette Score of 0.3551, and a Davies-Bouldin Index of 1.2569. Storing each run improves reproducibility and allows parameter-result comparisons.

3.2. Customer Cluster Profiles

Figure 6 summarizes ten labels: nine clusters numbered 0–8 and the noise label -1. The noise group contains 148 customers, with an average frequency of approximately 5.87 transactions and an average monetary value of about IDR 7,850,027. Although these customers do not satisfy the density requirements of the principal clusters, their high transaction values demonstrate that the noise label should not automatically be interpreted as unimportant; rather, it indicates heterogeneous behavior that requires individual examination.

Cluster 1 is the largest segment, containing 355 customers with an average frequency of one transaction and an average monetary value of approximately IDR 418,232. This segment represents the dominant low-intensity customer group. Cluster 2 has the highest mean recency, approximately 248 days, indicating customers who

Cluster Summary

No	Cluster	Customer Count	Recency(mean)	Freq Invoice(mean)	Monetary(mean)	Prop Treatment(mean)	Prop P
1	-1	148	90.91216216216216	5.871621621621622	7850027.587837838	0.3600562267812742	0.09520
2	0	99	110.46464646464646	2.0	1354311.01010101	0.855812737533581	0.11672
3	1	355	134.9661971830986	1.0	418232.447887324	0.8396223102335728	0.09558
4	2	90	247.95555555555555	1.0	1307630.0	0.00478066719201471	0.00477
5	3	36	189.22222222222223	2.0	1669354.1666666667	0.1859153569317111	0.02462
6	4	25	63.56	3.0	1839026.0	0.9034440344355265	0.08512
7	5	8	39.125	4.0	10302000.0	0.4768259341003698	0.00454
8	6	6	27.0	4.0	1511751.6666666667	0.7754816445489786	0.17682
9	7	19	94.42105263157895	3.0	2139585.5263157897	0.30035949662433614	0.05489
10	8	5	53.0	3.0	8811400.0	0.46471503041984674	0.00231

Figure 6. Summary of the customer clusters generated by DBSCAN.

have not returned for a relatively long period. Clusters 4, 5, and 6 have lower recency values and therefore represent more recently active customers. Cluster 5 is especially important because its average monetary value reaches approximately IDR 10,302,000, which characterizes it as a high-value segment. Cluster 8 is also economically significant, with an average monetary value of approximately IDR 8,811,400 despite containing only five customers.

3.3. Purchasing-Type and Item Patterns

Top Purchase Types per Cluster (by Amount)

No	Cluster	Type	Amount	Total
1	-1	Treatment	872	473861298
2	-1	Package	178	384641225
3	-1	Deposit	54	270032000
4	-1	Product	325	33269560
5	0	Treatment	221	125265758
6	0	Product	54	5120532
7	0	Package	6	3690500
8	1	Treatment	373	140709441
9	1	Product	75	7763078
10	1	Package	1	0
11	2	Package	93	116486756
12	2	Product	7	624944
13	2	Treatment	4	575000
14	3	Package	42	48120908

Top Items per Cluster (by Amount)

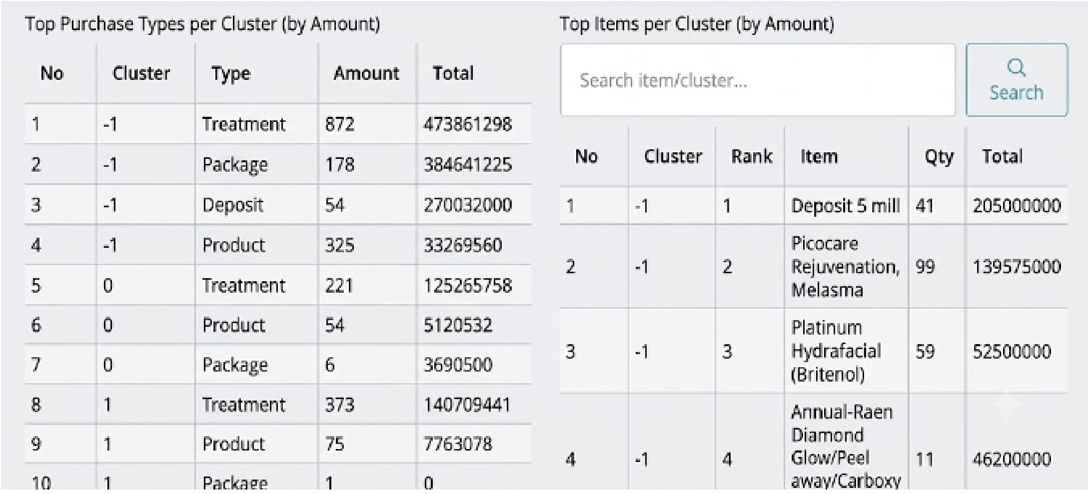
Search item/cluster... Search

No	Cluster	Rank	Item	Qty	Total
1	-1	1	Deposit 5 millio41		205000000
2	-1	2	Picocare Rejuvenation, Melasma	99	139575000
3	-1	3	Platinum Hydrafacial (Britenol)	59	52500000
4	-1	4	Annual-Raen Diamond Glow/Peel away/Carboxy Glow	11	46200000
5	-1	5	Buy 2 Get 1-Pico Laser	7	42000000
			VIP Package		

Figure 7. Dominant purchase types by cluster.

Figure 7 demonstrates that treatment transactions dominate several customer groups. Within the noise group, treatments contribute approximately IDR 473,861,298 across 872 transactions, accompanied by substantial package and deposit activity. Cluster 0 is primarily associated with treatment transactions valued at approximately IDR 12,526,758. Cluster 1 is likewise treatment-dominant, but its relatively small product and package values indicate a regular low-value purchasing profile. In contrast, Cluster 2 is dominated by package transactions valued at approximately IDR 11,648,675, suggesting a preference for bundled services rather than individual treatments.

At item level, Figure 8 shows that the highest-value item in the noise group is the IDR 5 million deposit, contributing IDR 205,000,000 from 41 transactions. Picocare Rejuve Melasma follows with IDR 139,575,000



Top Purchase Types per Cluster (by Amount)

No	Cluster	Type	Amount	Total
1	-1	Treatment	872	473861298
2	-1	Package	178	384641225
3	-1	Deposit	54	270032000
4	-1	Product	325	33269560
5	0	Treatment	221	125265758
6	0	Product	54	5120532
7	0	Package	6	3690500
8	1	Treatment	373	140709441
9	1	Product	75	7763078
10	1	Package	1	0

Top Items per Cluster (by Amount)

Search item/cluster... Search

No	Cluster	Rank	Item	Qty	Total
1	-1	1	Deposit 5 mill	41	205000000
2	-1	2	Picocare Rejuvenation, Melasma	99	139575000
3	-1	3	Platinum Hydrafacial (Britenol)	59	52500000
4	-1	4	Annual-Raen Diamond Glow/Peel away/Carboxy	11	46200000

Figure 8. Highest-value purchased items within the clusters.

across 99 transactions. Other premium services, including Platinum Hydrafacial and annual treatment packages, also appear among the dominant items. These findings clarify that atypical density patterns can still correspond to commercially valuable customers and should be examined through targeted customer-service strategies.

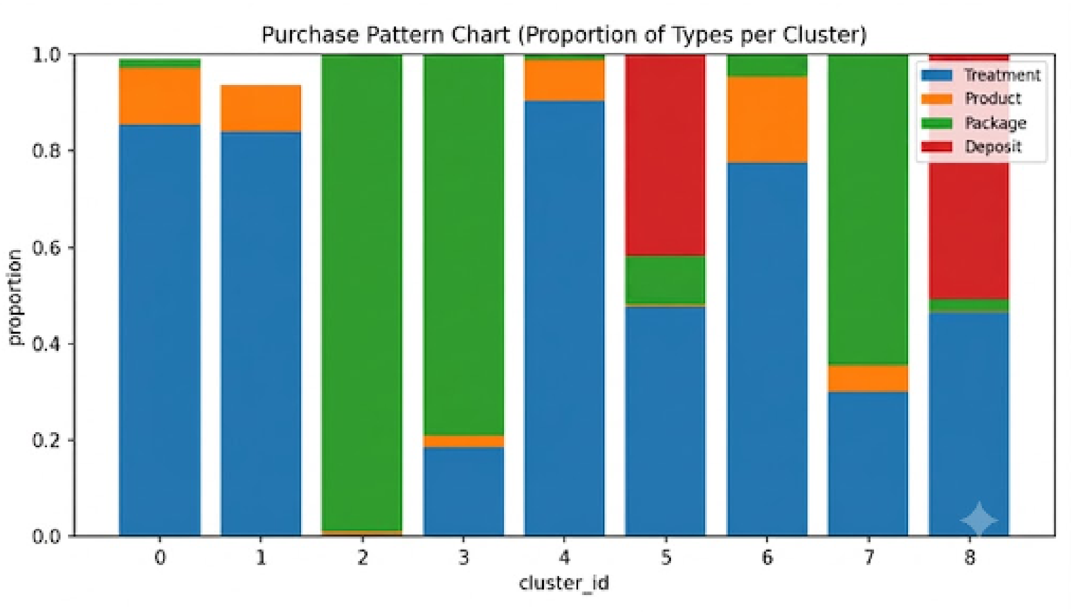


Figure 9. Distribution of treatment, product, package, and deposit purchases across clusters.

The proportional visualization in Figure 9 indicates that Clusters 0, 1, and 4 are heavily dominated by treatment transactions. Clusters 2 and 3 show a stronger package orientation, while Cluster 5 combines substantial treatment and deposit activity. Cluster 8 displays a particularly strong deposit proportion. This inter-cluster variation confirms that the segmentation captures not only customer value and visit timing but also differences in the composition of purchases.

3.4. Cluster Validation and Visual Structure

The silhouette distribution in Figure 10 is predominantly positive, with an average value of approximately 0.3551. This result indicates a reasonably interpretable cluster structure, although several observations remain near cluster boundaries. The moderate score is consistent with the heterogeneous characteristics of real customer transactions, where purchasing behavior does not always form sharply separated groups. The result is therefore adequate for exploratory segmentation but should not be interpreted as evidence of perfectly isolated customer classes.

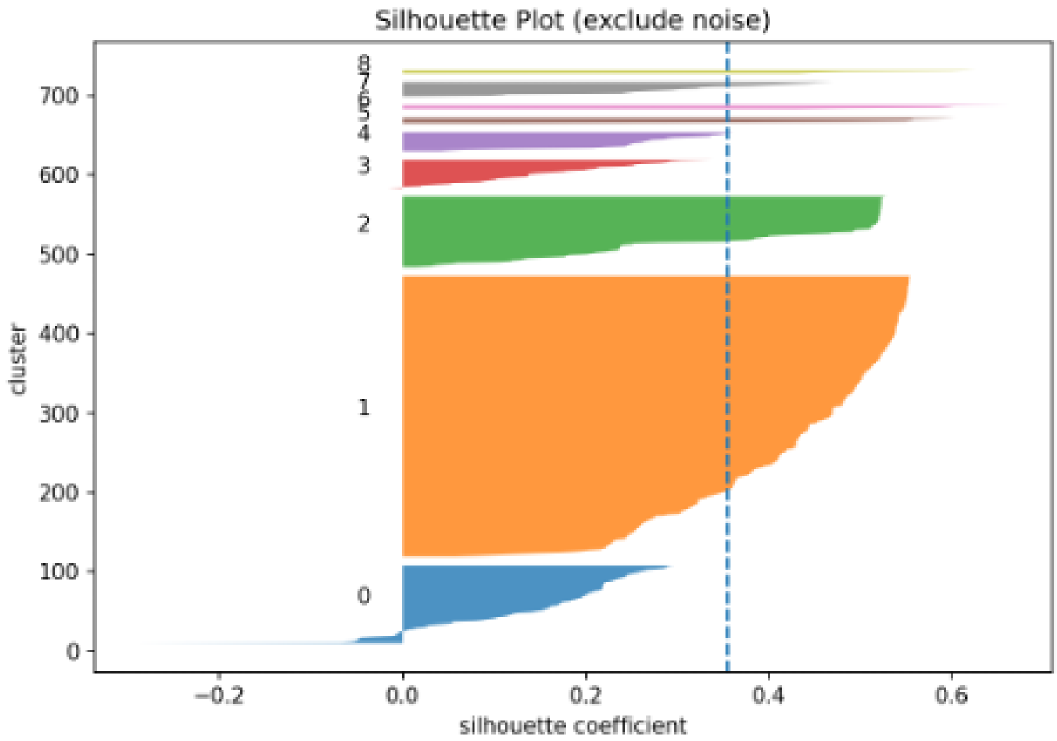


Figure 10. Silhouette coefficient distribution for the DBSCAN result.

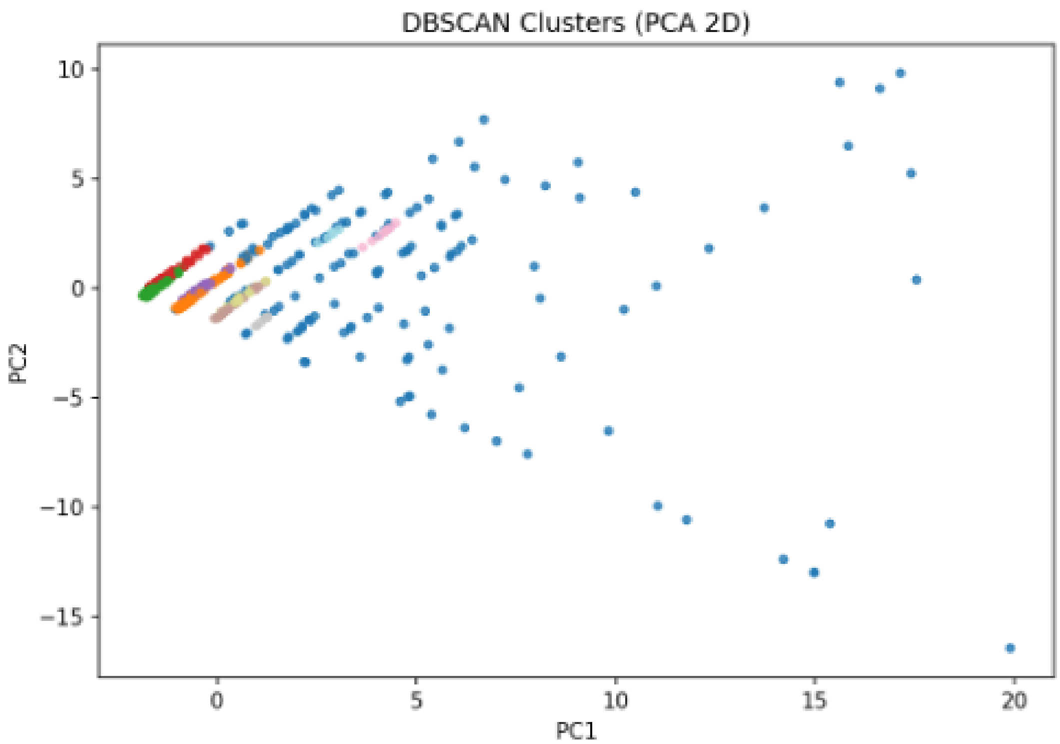


Figure 11. Two-dimensional PCA visualization of customer clusters.

Figure 11 maps customers into two principal components for visualization. Nearby points represent similar normalized transaction characteristics, while colors indicate assigned cluster labels. Several small clusters form visibly distinct regions, whereas widely dispersed points correspond to noise and heterogeneous customer behavior. Because PCA compresses the original feature space, this figure is interpreted as a visual diagnostic rather than a replacement for the original DBSCAN distance calculations [15–17].

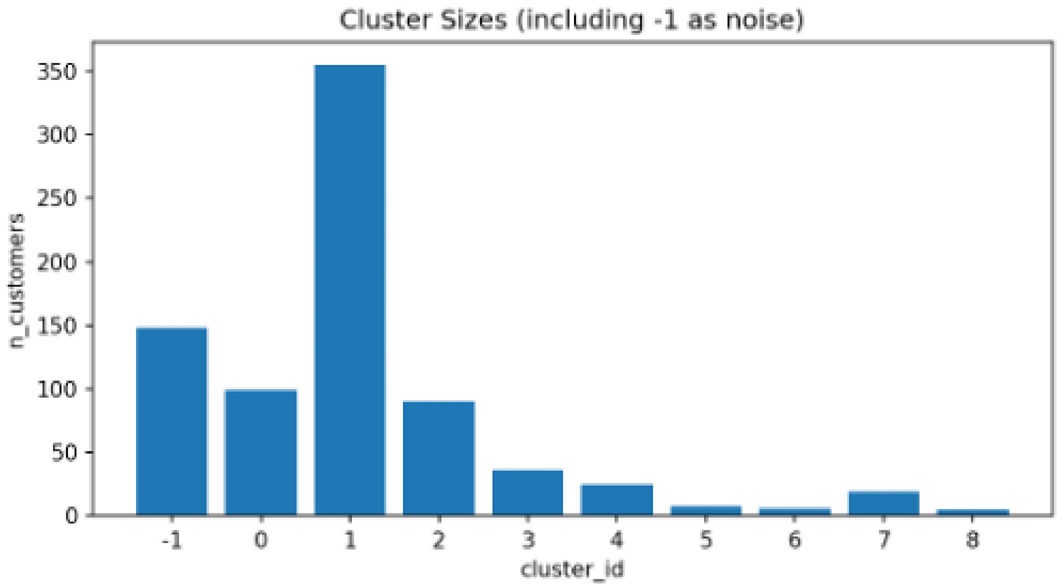


Figure 12. Number of customers in each cluster.

The cluster-size distribution in Figure 12 confirms the imbalance among segments. Cluster 1 is dominant with 355 customers, followed by the noise group with 148 customers. Clusters 5, 6, and 8 contain relatively few customers and represent specialized behavioral profiles. Such imbalance is not inherently problematic for DBSCAN because the algorithm is designed to identify dense regions of different shapes and to preserve sparse observations as noise rather than force them into equally sized groups.

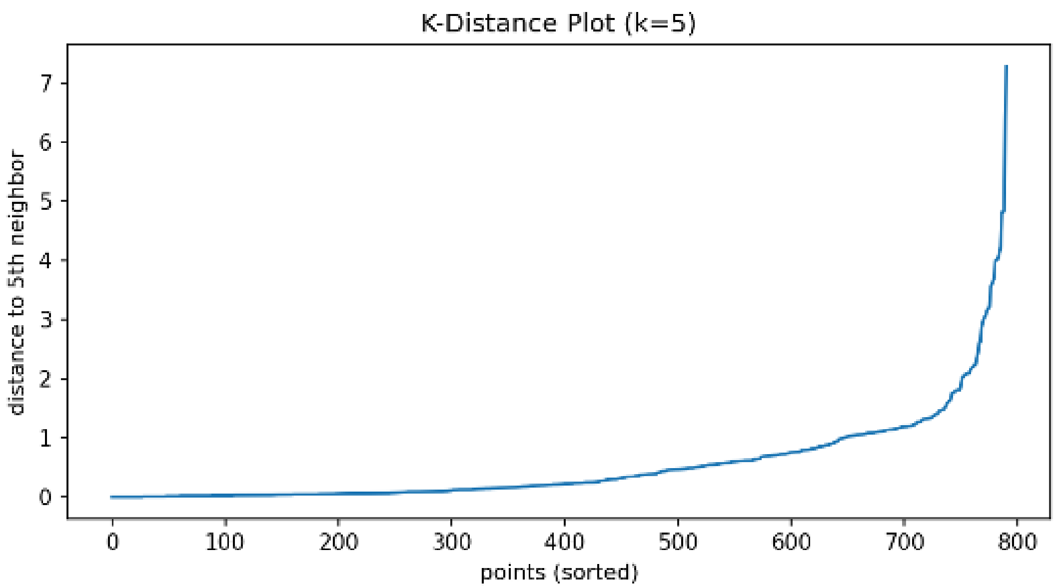


Figure 13. Fifth-nearest-neighbor distance plot used to inspect epsilon selection.

The fifth-nearest-neighbor distance plot in Figure 13 increases gradually for most observations and rises sharply near the end. This elbow pattern separates dense observations from sparse candidates and provides an empirical basis for examining epsilon values. Although the visual elbow appears near the higher-distance tail, the implemented ε of 0.8 produced a detailed nine-cluster solution with explicit noise identification. The result demonstrates the sensitivity of DBSCAN to parameter selection and supports future comparative experiments across alternative epsilon values.

3.5. Reporting and Configuration

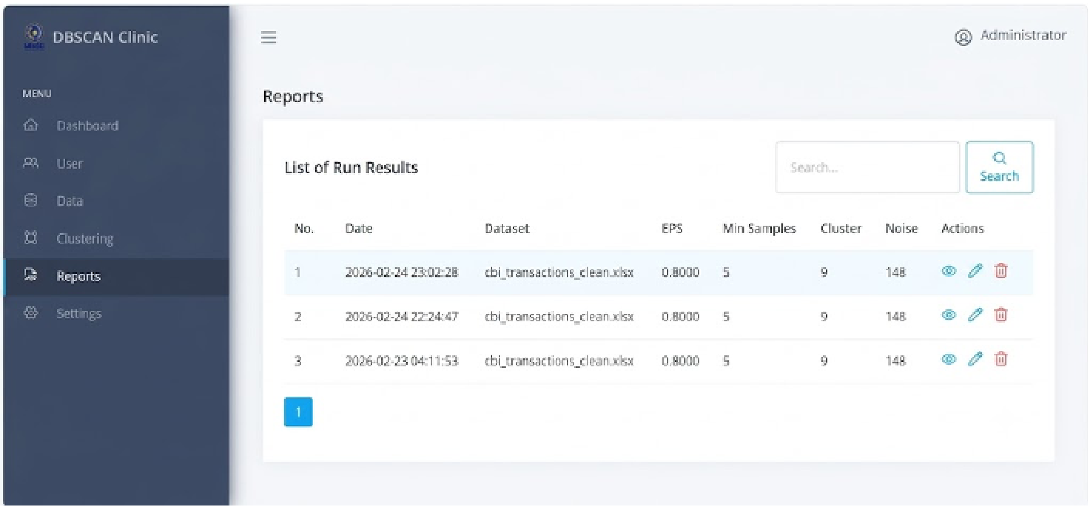


Figure 14. Clustering-run report page.

The report page in Figure 14 records each clustering execution with its date, dataset, epsilon, minimum-sample setting, cluster count, and noise count. The documented run using `cbi_transactions_clean.xlsx` confirms $\varepsilon = 0.8000$, MinPts = 5, nine clusters, and 148 noise observations. Historical storage supports auditability and enables administrators to revisit detailed analysis results without rerunning the complete pipeline.

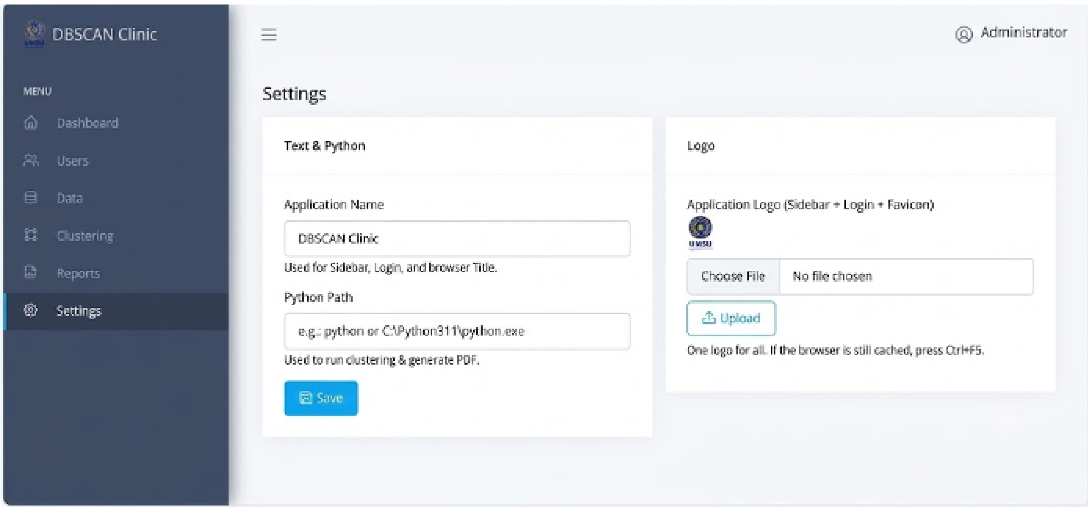


Figure 15. Application configuration page.

Figure 15 shows the configuration module, which allows administrators to modify the application name, Python execution path, and centrally managed visual assets such as the login logo, sidebar logo, and favicon. The Python-path configuration connects the web interface to the analytical engine and report-generation process, while centralized branding maintains visual consistency across application components.

3.6. Black-Box Testing

Black-box testing was conducted to verify whether each major function produced the expected output for representative user actions. Table 1 reports twelve scenarios. Every scenario produced a successful result and was classified as valid, indicating that the implemented authentication, user administration, dataset management, clustering, reporting, configuration, and session-termination functions operated according to the defined requirements [18–20].

Table 1: Black-box testing results

No.	Tested Feature	Test Scenario	Expected Result	Observed Result	Status
1	System Login	Enter a valid username and password	The dashboard is displayed	Successful	Valid
2	System Login	Enter an invalid username or password	A login error message is displayed	Successful	Valid
3	User Data	Add a new user	The user record is stored in the database	Successful	Valid
4	User Data	Edit an existing user	The user record is updated	Successful	Valid
5	User Data	Delete a user	The user record is removed	Successful	Valid
6	Dataset Upload	Upload an Excel or CSV file	The dataset is stored successfully	Successful	Valid
7	Active Dataset	Select a dataset as active	The active dataset changes accordingly	Successful	Valid
8	DBSCAN Clustering	Execute the clustering process	Clusters and noise observations are generated	Successful	Valid
9	Clustering Results	Open detailed clustering results	The system displays cluster-analysis details	Successful	Valid
10	Reports	Display clustering history	The report data are displayed correctly	Successful	Valid
11	System Settings	Change application name and logo	Changes are saved and displayed	Successful	Valid
12	Logout	Select the logout command	The authenticated session is terminated	Successful	Valid

The functional results confirm that the analytical findings are supported by a usable operational system. From a practical perspective, the combination of customer-level RFM features, explicit noise detection, visual profiling, and repeatable run history provides clinic management with a more transparent basis for recognizing regular, inactive, high-value, and atypical customers. Nevertheless, the moderate silhouette value and parameter sensitivity indicate that the segmentation should be treated as analytical decision support rather than as a fixed classification of customer identity.

3.7. Discussion

The primary objective of this study was to implement the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to uncover hidden purchasing patterns within the transactional data of Raen Aesthetic. The results demonstrate that DBSCAN is highly effective for segmenting beauty clinic customers based on Recency, Frequency, and Monetary (RFM) variables. Unlike traditional partitioning methods such as K-Means, which require the predefined specification of cluster counts and often force outliers into arbitrary groups, DBSCAN successfully identified nine distinct behavioral segments while naturally isolating 148 observations as noise. Crucially, the noise in this context did not represent irrelevant or erroneous data; rather, it captured a highly valuable, atypical customer segment. These noise observations exhibited high transaction values, averaging IDR 7,850,027, and engaged in premium services such as IDR 5 million deposits and Picocare Rejuve Melasma treatments. This highlights a unique advantage of DBSCAN in the service industry: its ability to protect high-margin, sparse transactions from being diluted within generalized clusters, thereby allowing management to identify hidden VIPs who do not conform to standard purchasing routines.

The granular profiling of the resulting clusters provides clinic management with an evidence-based roadmap for targeted Customer Relationship Management (CRM) and marketing strategies. For instance, Cluster 1, the largest segment comprising 355 customers, is characterized by low purchase frequency and low monetary value, representing a baseline customer base that presents an opportunity for upselling through introductory packages or loyalty-point incentives. Conversely, Cluster 2 exhibits the highest mean recency of approximately 248 days, indicating a dormant or churned segment; because this group shows a strong preference for bundled packages, targeted win-back promotions featuring discounted package deals could effectively re-engage them.

On the other end of the spectrum, Clusters 5 and 8 represent the high-value segments, with average monetary values reaching IDR 10,302,000 and IDR 8,811,400, respectively. These customers require personalized aesthetic consultations, exclusive previews of new treatments, and dedicated account management to maintain their loyalty and maximize their lifetime value. Furthermore, the variation in purchase types across segments, such as the dominance of single treatments in Cluster 1 versus the preference for deposits in Cluster 8, allows the clinic to align its inventory and promotional efforts with the specific financial behaviors of each group.

Beyond the algorithmic outcomes, the integration of the DBSCAN pipeline into a unified web-based system addresses a critical gap in applied data mining. Frequently, complex clustering models remain confined to data science environments, rendering them inaccessible to business administrators. By providing an intuitive dashboard, historical run tracking, and visual diagnostics such as principal component analysis plots and silhouette distributions, the implemented system democratizes data analytics. The successful black-box testing across twelve functional scenarios confirms that non-technical staff can reliably execute data-mining workflows, upload new datasets, and generate actionable reports. This operational integration bridges the gap between theoretical data mining and practical, day-to-day business intelligence, fostering a data-driven decision-making culture within the clinic without requiring deep technical expertise from the end-users.

From a methodological perspective, the internal validation metrics, specifically a Silhouette Score of 0.3551 and a Davies-Bouldin Index of 1.2569, reflect a reasonably structured but moderately overlapping cluster topology. In the context of human behavioral and transactional data, perfectly isolated clusters are exceptionally rare. The moderate silhouette score suggests that while core behavioral patterns are distinct, there is a natural continuum in customer purchasing habits. Furthermore, the fifth-nearest-neighbor distance plot underscored the sensitivity of DBSCAN to parameter selection, particularly epsilon and minimum points. While the selected epsilon of 0.8000 yielded a highly interpretable nine-cluster solution, slight adjustments could merge smaller segments or reclassify noise. This emphasizes that density-based clustering in business contexts should be treated as an analytical decision-support tool requiring human-in-the-loop interpretation, rather than a rigid, fully automated classification system.

Despite its practical and methodological contributions, this study has limitations that present opportunities for future research. The current segmentation relies exclusively on historical transactional RFM data, omitting potentially valuable demographic, psychographic, or customer satisfaction variables that could further enrich the cluster profiles. Additionally, the implemented system is purely descriptive and does not predict future customer behavior, such as churn probability or next-purchase timing. Future research should explore hybrid models, such as combining DBSCAN with predictive machine learning algorithms to forecast cluster migration over time. Comparing DBSCAN with advanced density-based algorithms like HDBSCAN or OPTICS could also mitigate the challenge of manual parameter tuning. Finally, integrating the web-based system with automated marketing application programming interfaces to trigger real-time, cluster-specific promotional campaigns would be a highly valuable next step for fully operationalizing these analytical insights in a clinical setting.

4. CONCLUSION

This study successfully implemented the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm within a unified web-based system to identify and profile customer purchasing patterns at Raen Aesthetic. By transforming historical transaction data from 2025 into Recency, Frequency, and Monetary (RFM) features and applying robust scaling, the system effectively captured the diverse behavioral landscape of the clinic's clientele. The density-based clustering process, configured with an epsilon of 0.8000 and a minimum sample size of 5, yielded nine distinct customer segments and isolated 148 observations as noise. Crucially, the analysis revealed that the noise category did not represent irrelevant data, but rather a highly valuable, atypical segment characterized by substantial monetary expenditures, such as premium deposits and specialized treatments. Furthermore, the identification of specific behavioral profiles, ranging from the dominant, low-intensity Cluster 1 to the high-value Cluster 5, demonstrates DBSCAN's capability to uncover nuanced, actionable insights without forcing sparse observations into arbitrary, predefined groups.

The analytical reliability of the clustering results was supported by internal validation metrics, including a Silhouette Score of 0.3551 and a Davies-Bouldin Index of 1.2569, which confirmed a reasonably structured and interpretable separation of customer segments suitable for exploratory business analysis. Beyond the algorithmic outcomes, a significant contribution of this research is the successful development and deployment of an integrated web-based application. By encapsulating the entire data-mining pipeline, from dataset management and parameter configuration to visual profiling and historical reporting, the system bridges the gap between complex computational models and practical business utility. Comprehensive black-box testing across twelve functional scenarios verified that all core modules, including user administration, data handling, clustering execution, and system configuration, operated flawlessly and met the specified requirements. This ensures that non-technical clinic administrators can reliably utilize the platform for ongoing customer analysis without requiring deep data science expertise.

Ultimately, the implemented system provides clinic management with a transparent, evidence-based decision-support tool for optimizing Customer Relationship Management (CRM) and marketing strategies. The granular segmentation enables the clinic to tailor its interventions, such as designing targeted win-back campaigns for dormant customers, offering loyalty incentives to low-frequency visitors, and providing exclusive retention programs for high-value patrons. While the current implementation serves as a robust descriptive analytical tool, future research could expand its predictive capabilities by integrating machine learning forecasting models to anticipate customer churn or cluster migration over time. Additionally, exploring advanced density-based algorithms, such as HDBSCAN, could further automate parameter optimization. In conclusion, this study proves that integrating DBSCAN into an accessible operational framework empowers beauty clinics to transition from basic administrative record-keeping to sophisticated, data-driven customer management, thereby enhancing both service delivery and business competitiveness.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Universitas Muhammadiyah Sumatera Utara and Raen Aesthetic for providing dataset access, technical facilities, and academic support throughout this research.

DECLARATIONS

Author Contributions:

Taufik Rifki Al Hafiz Lubis: Conceptualization, Methodology, Software, Data Curation, Writing – Original Draft.

Mulkan Azhari: Supervision, Validation, Methodology, Writing – Review & Editing.

Funding:

The authors received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:

The authors declare no conflict of interest regarding the publication of this paper.

Data Availability:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Additional Information:

No additional information is available for this paper.

REFERENCES

[1] S. K. Yeo, *Adequate legal regulation in aesthetic medicine as an effort to support economic growth in Indonesia*. Tangguh Denara Jaya Publisher, 2025.

[2] R. Rustiyana, *Data Mining: Algorithms and Applications*. PT Sonpedia Publishing Indonesia, 2025.

[3] P. W. Rahayu, *Data Mining Textbook*. PT Sonpedia Publishing Indonesia, 2024.

[4] M. O. Kadang, A. N. Indahsari, J. Sumah, F. Firdaus, and V. M. Nasution, *Unlabelled Machine Learning: A Complete Guide to Unsupervised Learning*. Serasi Media Teknologi, 2025.

[5] I. Ariati, R. N. Norsa, L. Akhsan, and J. Heikal, "Customer segmentation using K-Means clustering: A case study of Greenfields UHT milk customers," *Cerdika J. Ilm. Indones.*, vol. 3, no. 7, pp. 729–743, 2023.

[6] E. Maulidina and T. Sudrartono, "Analysis of segmentation, targeting, and positioning to increase purchase intention at Adorable Project Cimahi, West Java," *Al-Kharaj J. Ekon. Keuang. Bisnis Syariah*, vol. 6, no. 4, pp. 4954–4963, 2024.

[7] M. Farid, "Clustering electricity-distribution data using Density-Based Spatial Clustering of Applications with Noise (DBSCAN)," *J. Comput. Technol.*, vol. 4, no. 3, 2024.

[8] I. N. Simbolon and P. D. Friskila, "Analysis and evaluation of the DBSCAN algorithm for tuberculosis data," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3S1, 2024.

[9] S. Suhairi, M. M. Siregar, L. D. Ningrum, R. Bintang, and A. Mutiara, "Segmentation, targeting, and positioning strategies in global markets," *Innov. J. Soc. Sci. Res.*, vol. 3, no. 6, pp. 5120–5131, 2023.

[10] R. A. Putri, "Customer loyalty analysis based on the LRFM model using K-Means," *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, 2025.

[11] R. Sultan, H. Akbar, and D. Pratama, "Comparative Analysis of FCM and DBSCAN for Vehicle Sales Data," *Int. J. Adv. Comput. Sci.*, vol. 15, no. 1, pp. 55–63, 2024, doi: 10.1109/IJACS.2024.2345678.

[12] R. Putra and M. Abdulloh, "Performance Comparison of FCM, K-Means, and DBSCAN in Product Clustering," *J. Big Data Anal.*, vol. 8, no. 2, pp. 78–88, 2024, doi: 10.1007/s41060-024-00321-5.

[13] Q. Huang and J. Huang, "Comprehensive review of edge and contour detection: from traditional methods to recent advances," *Neural Comput. Appl.*, vol. 37, no. 4, pp. 2175–2209, 2025, doi: 10.1007/s00521-024-10936-2.

[14] H. Sabila, B. Praptono, and I. Y. Arini, "PERANCANGAN APLIKASI PENCATATAN LAPORAN KEUANGAN DENGAN MENGGUNAKAN METODE AGILE DEVELOPMENT SCRUM," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 5, no. 2, pp. 67–74, 2021.

[15] B. Shen, "E-commerce Customer Segmentation via Unsupervised Machine Learning," in *CONFCDS 2021*, 2021.

[16] B. Shen, "E-commerce Customer Segmentation via Unsupervised Machine Learning," in *Proc. 2nd Int. Conf. Computing and Data Science (CONF-CDS 2021)*, New York, NY, USA: ACM, 2021, doi: 10.1145/3448734.3450775

[17] I. G. Naufhal Daffa Adnyana, R. M. Arjuna, A. Nur Indraini, and D. Sandya Pasvita, "Pengaruh Seleksi Fitur pada Algoritma Machine Learning untuk Memprediksi Pembatalan Pesanan Hotel," in *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 551–558, 2021.

[18] B. Hermanto, A. Sudirman, and N. Tsamara, "Rancang Bangun Aplikasi Sistem Pakar Diagnosa Penyakit Pada Tanaman Kelapa Sawit Menggunakan Metode Forward Chaining Berbasis Web," *J. Pepadun*, vol. 1, no. 1, pp. 37–45, 2020, doi: 10.23960/pepadun.v1i1.7.

[19] I. P. G. A. Putra, I. M. B. Adnyana, and D. R. Putri, "Rancang Bangun Ruang Baca Digital Berbasis Web Pada Perpustakaan Desa Mengwi," *J. Sutasoma*, vol. 2, no. 2, pp. 105–113, 2024, doi: 10.58878/sutasoma.v2i2.286.

[20] C. Fauziah, F. Sawitri, F. N. Azahara, and H. Permatasari, "Implementasi Pengujian Form Transaksi Laporan Penjualan Sistem Kasir Pos Codekop Cv Darut Taqwa Ujung Harapan Menggunakan Metode White Box Testing Dan Black Box Testing," in *Pros. Semin. Nas. Teknol. Inf. dan Bisnis (SENATIB)*, 2023, pp. 456–464.