KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 1, Page: 11 - 17, March 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Analysis of Clean Water Consumption Segmentation And Classification Using K-Means Clustering and Random Forest Algorithms

Ika Melinia Sapitri[1], Sarjon Defit[2], Rini Sovia[3]
[1,2,3] Master of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia
ikameliniasafitri7@gmail.com

Abstract

The administrative grouping of PERUMDA Air Minum Kota Padang customers is not yet able to accurately represent actual customer water consumption patterns. This condition makes it difficult for the company to formulate service policies, customer management, and make appropriate data-based decisions. This study aims to analyze and map customer water consumption patterns to produce more representative customer segmentation as a basis for decision making. The research method used is a data mining approach with the application of Principal Component Analysis (PCA) for dimension reduction, K-Means Clustering for customer segmentation, and Random Forest for customer classification, using primary data from the Padang City Water Company's Customer Meter Reading Report with an initial amount of 371 data. The results of the study show that the clustering process successfully formed three customer segments, namely premium customers with high consumption bills, regular customers with moderate and stable consumption, and new customers with low consumption rates. The evaluation of the Random Forest model's performance resulted in an accuracy rate of 68.85% on the training data and 67.69% on the testing data, with an average precision value above 0.84 and an average F1-score value of around 0.68. The consistency of performance between the training data and the testing data shows that the model has fairly good generalization capabilities and does not experience overfitting.

Keywords: Customer segmentation, Water consumption patterns, Principal Component Analysis (PCA, K-Means clustering, Random Forest

1. Introduction

Human activities now generate large amounts of complex data, requiring more than just collection in terms of management and analysis. The main challenge is how to extract useful knowledge from this data [1], [2]. One rapidly developing solution is the data mining approach, which enables the discovery of hidden patterns and relationships [3]. As part of the Knowledge Discovery in Databases (KDD) process, data mining plays a crucial role in supporting effective decision making, such as trend prediction, customer segmentation, and anomaly detection in various public service sectors [4].

One popular method in data mining is K-Means Clustering. This algorithm works by grouping data based on object similarity using Euclidean distance to the group center (centroid) [5]. Computational efficiency and ease of interpretation make K-Means the top choice in customer segmentation and consumer behavior analysis [6]. However, K-Means performance often declines when faced with data with many variables that are mutually correlated. To overcome this information redundancy, the Principal Component Analysis (PCA) method is used to reduce the data dimensions without losing important information [7]. Previous research shows that applying PCA to K-Means can improve the accuracy of poverty data segmentation in Indonesia. Thus, the combination of PCA and K-Means has been proven to significantly improve the quality of clustering results [8].

Apart from clustering, the next challenge is to classify new data into existing patterns. In this case, the Random Forest algorithm is a reliable ensemble learning method. By combining multiple decision trees, Random Forest is able to provide stable, accurate predictions that are resistant to overfitting, even on high-dimensional data [9]. Previous research shows that the use of Random Forest in water quality classification and public sentiment analysis provides a level of accuracy that exceeds traditional methods such as KNN and SVM [10].

The application of data mining at the Regional Water Company (PDAM) has had a real impact on

Submitted: 10-01-2026 | Revised: 12-02-2026 | Accepted: 15-03-2026 | Published: 30-03-2026

operational efficiency [11]. Previous studies have successfully mapped customer arrears risk into three segments: low, medium, and high [12]. eanwhile, the use of clustering has also proven effective in detecting unusual consumption patterns as indicators of pipe leaks or distribution inefficiencies. The integration of this technology is crucial for public service institutions in formulating strategic policies such as water distribution planning and tariff evaluation based on customer behavior data [13].

PERUMDA Air Minum Kota Padang is currently facing a major challenge with the growth of its customer base to more than 140,000 people. Various problems have arisen, ranging from low water pressure, water loss (non-revenue water) reaching 27%, to high arrears rates. In addition, plans for gradual tariff adjustments in the 2025–2030 period require more accurate customer segmentation. Currently, customer grouping is still administrative in nature, based on tariff categories and service zones, and therefore does not reflect actual consumption patterns or heterogeneous customer payment behavior.

To overcome these managerial and technical obstacles, this study proposes an integrated approach using a combination of PCA, K-Means Clustering, and Random Forest. PCA will simplify complex customer variables, K-Means will form segments based on actual consumption patterns, and Random Forest will be used to automatically classify new customers [14]. It is hoped that this approach will provide a strong foundation for PERUMDA Kota Padang in making fair strategic decisions, improving billing efficiency, and optimizing clean water distribution for the community.

2. Methods

The customer data analysis process in this study was conducted by integrating the Principal Component Analysis (PCA), K-Means Clustering, and Random Forest methods [15]. PCA was used to reduce data dimensions, while K-Means Clustering was applied to group customers based on similar characteristics. Furthermore, the Random Forest algorithm was used to perform classification and evaluate the clustering results [16]. The development of this research model referred to previous studies to produce accurate and objective analyses . The research process began with problem identification, data collection and preprocessing, modeling, and finally evaluation, as illustrated in the research framework in Figure 1.

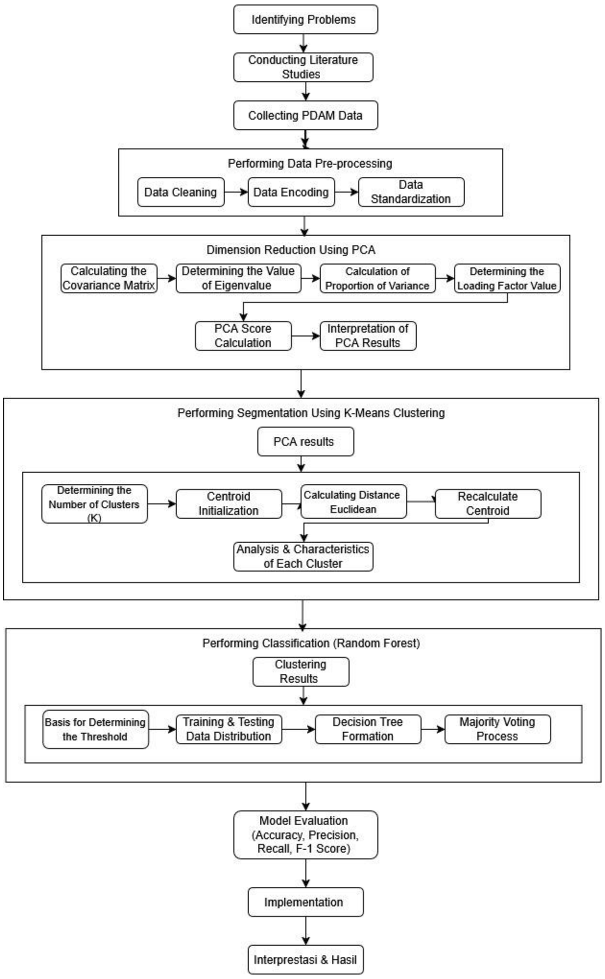

Figure 1. Research Framework

2.1. Identifying the Problem

Determine the problem to be studied, namely the analysis of segmentation and classification of clean water consumption by customers of the Padang City Water Company. Set the research objectives to reduce data dimensions, group customers based on consumption patterns, and classify customer categories.

2.2 Conducting a Literature Study

At this stage, the researcher collected references from various sources such as scientific journals, books, seminar proceedings, research reports, and official publications discussing Principal Component Analysis (PCA), K-Means Clustering, Random Forest, and their application in clean water consumption data analysis.

2.3 PDAM Data Collection

This research began with collecting secondary data from PERUMDA AIR MINUM Kota Padang as the main source. The data period used was specifically determined to be from 2024 to 2025 in order to reflect current conditions and cover consumption variations in recent years.

2.4 Data Preprocessing

This stage consists of three stages of data preprocessing: Data Cleaning to remove duplicate data, address missing values, and correct input errors. Encoding categorical variables to convert categorical data into numerical form. Data Standardization to equalize variable scales for more optimal distance calculations and PCA.

2.5 Dimension Reduction Using PCA

At this stage, dimension reduction is performed to simplify the data without losing important information. PDAM customer data generally has many variables that are correlated with each other, so the Principal Component Analysis (PCA) technique is needed to make the analysis more efficient and focused on the most influential variables [17]. PCA helps overcome multicollinearity issues while highlighting the main patterns in the data.

2.6 Performing K-Means Clustering

At this stage, the data used as input is the result of dimension reduction using Principal Component Analysis (PCA) so that the information entered into the algorithm is more concise but still represents the main variance in the data [18]. The K-Means stage begins with determining the optimal number of clusters to produce an output in the form of data grouping.

2.7 Performing Classification (Random Forest)

In this study, Random Forest is used after the clustering process, where the cluster results are used as the target class. Thus, the model is not only able to understand existing patterns, but also predict new customer segments based on the input data [19], [20].

2.8 Model Evaluation

Model performance is tested using evaluation metrics such as accuracy, precision, recall, and F1-score to ensure the reliability of Random Forest in performing classification.

2.9 Implementation

The algorithm was implemented in the program by building a series of codes containing all stages of data processing, starting from customer dataset input, pre-processing (cleaning, Z-Score standardization, and encoding), to PCA calculation, K-Means cluster formation, and classification using Random Forest.

2.10 Interpretation and Results

The analysis results were used to interpret the characteristics of each customer group and generate strategic recommendations for PERUMDA Air Minum Kota Padang regarding tariff policies, services, and water resource management.

3. Results and Discussions

This stage presents the results of processing and analyzing PERUMDA Air Minum Kota Padang customer data based on the established research methodology stages, starting from data preprocessing, dimension reduction, customer segmentation, classification, to model evaluation. The discussion focuses on interpreting the results obtained to describe water consumption patterns and customer characteristics, as well as assessing the model's performance in supporting data-driven decision making.

3.1 Data Processing

The raw data from PDAM Kota Padang still contains identity attributes, administrative data, and value variations that need to be adjusted in order to be ready for processing. Therefore, data cleaning, variable coding, and standardization were carried out so that the dataset would be relevant, consistent, and suitable for analysis. This study uses the Z-Score Standardization method, which transforms each value into a distribution with a mean of 0 and a standard deviation of 1. This method ensures that all variables contribute equally to the distance calculation and cluster formation processes. The results of the overall data standardization can be seen in Table 1.

Table 1. Overall Data Standardization Results

Class Rate	Region	Subscription Duration	Usage (m^3)	Total Bill (Rp.)
-0.63614	-1.77666	-1.628	-0.22173	-0.16843
-0.16086	-1.68446	-1.41525	-0.33335	-0.03831
-0.16086	-1.68446	-1.27342	-0.37055	-0.14037
-0.16086	-1.68446	-1.27342	-0.24405	-0.16843
-0.63614	-1.68446	0.286716	-0.16964	-0.16843
-0.16086	-1.68446	0.570377	-0.31102	-0.14037
-0.16086	-1.68446	0.570377	-0.1622	-0.09585
…	…	…	…	…
1.740242	1.634536	0.286716	-0.36311	-0.16843
1.740242	1.634536	1.350446	-0.34079	0.057672
1.740242	1.634536	0.570377	0.128013	-0.16843
1.740242	1.634536	0.712208	0.425663	-0.16843

This standardization process is carried out so that each variable has a comparable scale, with an average of 0 and a standard deviation of 1. This step is important to avoid the domination of variables with large units, such as the number of bills, over other variables when performing Principal Component Analysis (PCA), K-Means Clustering, and Random Forest.

3.2 Dimension Reduction Using PCA

Principal Component Analysis (PCA) in this study was conducted to reduce nine variables into a number of principal components that could represent most of the data information. The process began with data standardization and covariance matrix calculation to identify linear relationships between variables, which were then used as the basis for determining eigenvalues and eigenvectors.

The calculation of the first Principal Component Score for observation 1 was performed by multiplying the z-score of each variable by the corresponding PCA1 Loading Factor, then summing the results. This process used Equation :

Variable 1 =
$$0{,}245872 \times -0{,}636138976 = -0{,}15609$$

Variable 2 =
$$0{,}087875 \times -1{,}77666 = -0{,}15609$$
$$\vdots$$

Variable 9 =
$$0{,}218794 \times -0{,}16843 = -0{,}03683$$
$$F_1, Data\ 1 = (-0{,}15609) + (-0{,}15610) + \cdots + (-0{,}03683) = -1{,}16562$$

The same calculation process is applied to all observations and the main components are retained, resulting in a PCA score matrix that is then used as input in further analysis, such as clustering analysis, to facilitate data grouping based on the main patterns formed. The PCA score results can be seen in Table 2.

Table 2. PCA Score Calculation Results

PCA1	PCA2	PCA3	PCA4
0,245872	0,589139	0,151168	0,266936
0,087875	0,109052	0,380862	-0,60961
0,164894	0,149746	0,16113	-0,6438
0,516147	-0,19188	-0,26288	-0,04461
0,517779	-0,19175	-0,26075	-0,04352
0,500081	-0,12739	-0,0213	0,05342
0,263762	0,597272	0,128423	0,204485
0,051096	-0,31742	0,592766	0,214071
0,218794	-0,26701	0,548375	0,219686

Based on Kaiser's criteria (eigenvalue ≥ 1), four main components (PCA1–PCA4) were obtained, which cumulatively explained 80.59% of the total data variation, thus considered sufficiently representative in describing the data structure. Furthermore, the eigenvectors of the four components were normalized into loading factor values to interpret the contribution of each variable to each component. The results show that PCA1 represents the dimension of water consumption, PCA2 reflects tariff characteristics and usage limits, PCA3 describes administrative and financial aspects, and PCA4 relates to regional factors and subscription duration.

3.3 Segmentation Using K-Means Clustering

The K-Means Clustering algorithm is used to segment customers based on the principal component scores generated from the dimension reduction process using Principal Component Analysis (PCA). The number of clusters is set to three (K = 3) to represent customer groups with different water usage characteristics. The clustering process begins with the random determination of initial centroids in the PCA feature space (PCA1–PCA4), followed by the calculation of the Euclidean distance between each data point and the centroid to determine cluster membership. The centroid values are updated by calculating the average principal component scores in each cluster. The distance calculation and centroid update stages were performed iteratively until convergence was achieved, i.e., when there were no significant changes in cluster membership or centroid values. The final centroid calculation results can be seen in Table 3.

Table 3. Final Centroid

Cluster	PCA1	PCA2	PCA3	PCA4
C1	1,783745	1,870745	0,509769	0,523132
C2	-0,30227	-0,27105	0,090953	-0,93688
C3	-0,64017	-0,72116	-0,37826	0,74302

Based on the final centroid analysis results, it can be concluded that cluster formation successfully identified three customer segments of PERUMDA Kota Padang. Each segment reflects a distinctive pattern based on the principal component analysis (PCA) values that form the centroid of each cluster. The cluster formation results can be seen in Table 4.

Table 4. Cluster Formation Results

Cluster	General Characteristics	Number of Members	Category
C1	High consumption, high rates, high bills	67	Premium Customers
C2	Moderate consumption, loyal, stable bills	135	Regular Customers
C3	Low consumption, low rates, new customers	123	New Customers

The clustering results show that the K-Means method is capable of forming three stable clusters with different characteristics based on consumption patterns, rates, and customer administrative aspects, so that the segmentation results can be used as a basis for further analysis and managerial decision-making.

3.4 Classification Using Random Forest

The Random Forest model is created gradually and systematically through five main stages, namely using K-Means clustering results as class labels, determining decision thresholds based on PCA principal components, dividing data into training and testing data, creating a number of decision trees with different rules, and determining the final classification results through a majority voting mechanism. This series of stages ensures that the model is built in a structured, consistent manner based on the natural characteristics of the data. The results of the formation of a single decision tree can be seen in Figure 2.

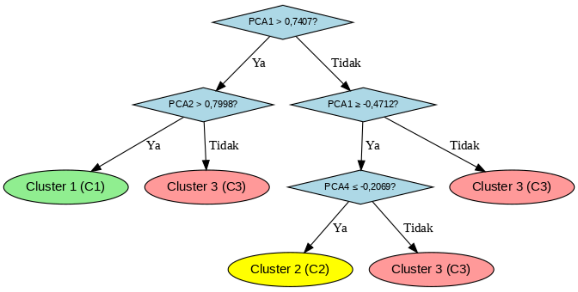

Figure 2. Decision Tree 1

If PCA1 > 0.7407 and PCA2 > 0.7998, then the data is classified into Cluster 1 (C1). If PCA1 is between -0.4712 and 0.7407 and PCA4 ≤ -0.2069, then the data is classified into Cluster 2 (C2). Other conditions are categorized into Cluster 3 (C3). Tree 1 emphasizes the contribution of PCA1 (dominance of customers with high component values) and PCA2 (variation in consumption behavior). A high combination of values on both PCAs indicates customers with premium and active characteristics (C1), while medium-low values indicate regular or new customers (C2 and C3).

3.5 Model Evaluation

Model evaluation was performed by comparing the prediction results of the Random Forest algorithm with the actual cluster labels obtained from the K-Means clustering process. Model performance measurement was carried out separately on training data and testing data. Evaluation on training data was used to determine the extent to which the model was able to learn classification patterns from the training data, while evaluation on testing data aimed to test the model's generalization ability in classifying new data that was not involved in the model formation process. The model's performance level was measured using evaluation metrics such as accuracy, precision, recall, and F1-score, thereby providing a comprehensive overview of the performance of the resulting classification model. A comparison of the test and training data evaluation results can be seen in Table 5.

Table 5. Comparison of Training Data and Testing Data

Dataset	Average Precision	Average Recall	Average F1-Score	Brief Interpretation
Training (260 data)	0,848	0,671	0,683	The model learns well, with balanced accuracy and completeness of predictions.
Testing (60 data)	0,841	0,675	0,679	Performance on new data is relatively stable, indicating good generalization ability and no overfitting.

The difference in average values between the training and testing data is very small (difference < 0.01). This shows that the Random Forest model is consistent and has the ability to recognize patterns in new data well. The high precision values in both datasets indicate that the model rarely makes mistakes in assigning cluster labels. The fairly stable recall and F1-score values illustrate the balance between the ability to detect all cluster members and classification accuracy. Overall, this model is suitable for customer clustering based on PCA dimension reduction results.

4. Conclusions

The process of performing segmentation and classification analysis took customer data from PERUMDA Air Minum Kota Padang using the PCA method for dimension reduction, K-Means Clustering for customer segmentation, and Random Forest for customer classification. The dimension reduction process using PCA shows that four main components (PCA1–PCA4) are able to explain 80.59% of the data diversity, so that most of the important information in the dataset can be represented well without losing the main characteristics of the data. The clustering results using the K-Means algorithm produced three customer groups with different characteristics, namely customers with high consumption and billing values (Cluster 1), regular customers with stable usage patterns (Cluster 2), and new customers or those with low consumption levels (Cluster 3). These differences in characteristics provide a clear and structured picture of customer segmentation. Furthermore, the Random Forest classification model built based on the clustering results showed an accuracy rate of 68.85% on the training data and 67.69% on the testing data, with relatively balanced precision, recall, and F1-score values. This indicates that the model has stable performance and fairly good generalization capabilities. Overall, the results of this study can be used as a basis to support PERUMDA Kota Padang's strategic decision-making regarding customer service management and improvement.

References

[1] B. E. Adiana, I. Soesanti, and A. E. Permanasari, "Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik Clustering," *J. Terap. Teknol. Inf.*, vol. 2, no. 1, pp. 23–32, 2018, doi: 10.21460/jutei.2018.21.76.

[2] M. Imron and P. Korespondensi, "Implementasi Model Prediksi Churn Pelanggan Menggunakan Algoritma

Random Forest Pada Website Industri Telekomunikasi," *J. Teknol. Inf.*, vol. 1, no. 1, pp. 12–25, 2025.

[3] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-Means and Alternative Clustering Methods in Modern Power Systems," *IEEE Access*, vol. 11, no. November, pp. 119596–119633, 2023, doi: 10.1109/ACCESS.2023.3327640.

[4] K. et al 2023, "PENERAPAN DATA MINING UNTUK MENENTUKAN LOKASI PROMOSI SEKOLAH DENGAN METODE K-MEANS CLUSTERING (STUDI KASUS: SMP ISLAM AL SYUKRO UNIVERSAL)," vol. 32, no. 3, pp. 167–186, 2021.

[5] A. Dinanti and J. Purwadi, "Analisis Performa Algoritma K-Nearest Neighbor dan Reduksi Dimensi Menggunakan Principal Component Analysis," *Jambura J. Math.*, vol. 5, no. 1, pp. 155–165, 2023, doi: 10.34312/jjom.v5i1.17098.

[6] C. Auzia Nugraha, M. A. Kesuma, O. I. Cahyani, M. Wati, and Haviluddin, "Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means," *JUKI J. Komput. dan Inform.*, vol. 7, no. 1, pp. 80–88, 2025.

[7] R. Ishak, Nurmawanti, and Amiruddin, "Optimasi K-Means pada Clustering Penyakit Ibu Hamil Menggunakan Random Forest," *Jambura J. Electr. Electron. Eng.*, vol. 7, no. 1, p. 41, 2024.

[8] K. H. Izzuddin and A. W. Wijayanto, "Pemodelan Clustering Ward, K-Means, Diana, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021," *Komputika J. Sist. Komput.*, vol. 13, no. 1, pp. 41–53, 2024, doi: 10.34010/komputika.v13i1.10803.

[9] Fadil Danu Rahman, M. I. Z. Mulki, and A. Taryana, "Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest," *J. SINTA Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. 2, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15.

[10] C. Mario, R. R. Suryono, U. T. Indonesia, and B. Lampung, "Analisis Sentimen Publik pada Film Dirty Vote di Youtube Menggunakan Random Forest dan Naive Bayes," vol. 10, no. 1, pp. 111–122, 2025.

[11] E. A. Winanto, Y. Novianto, S. Sharipuddin, I. S. Wijaya, and P. A. Jusia, "Peningkatan Performa Deteksi Serangan Menggunakan Metode Pca Dan Random Forest," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 2, pp. 285–290, 2024, doi: 10.25126/jtiik.20241127678.

[12] S. Chairunnissa, D. Akbar, S. Defit, and B. Hendrik, "Segmentasi Tunggakan Pelanggan Menggunakan Algoritma K-Means Cluster pada Perusahaan Air Minum Daerah," *J. Pustaka AI*, vol. 5, pp. 349–355, 2025.

[13] K. Abdi, A. Warjaya, I. Muthmainnah, and P. H. Pahutar, "Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 2, pp. 81–88, 2024, doi: 10.54082/jiki.81.

[14] I. A. Rosyada and D. T. Utari, "Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering," *Jambura J. Probab. Stat.*, vol. 5, no. 1, pp. 6–13, 2024, doi: 10.37905/jjps.v5i1.18733.

[15] Nugraha, "PENINGKATAN PERFORMA DBSCAN DENGAN REDUKSI DIMENSI PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM KLASTERISASI TINGKAT KEMISKINAN DI INDONESIA Muhammad," vol. 4, no. 5, pp. 595–608, 2025.

[16] S. D. Yuliani and P. Kartikasari, "Klasifikasi Pelanggan Churn Pada Kegiatan Servis Dan Penjualan Sparepart Dengan Metode Random Forest," *Pros. Semin. Nas. Sains dan Teknol. Seri IV Fak. Sains dan Teknol.*, vol. 2, no. 2, pp. 84–96, 2025.

[17] M. R. Fairuzzen, V. H. Merpaung, A. A. Putra, A. A. Malik, and Mahipal, "Interdisciplinary Explorations in Research JAGUNG DI PROVINSI-PROVINSI INDONESIA," vol. 2, pp. 1497–1516, 2024.

[18] E. Febrianty, L. Awalina, and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan

Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail Optimizing Marketing Strategies with Customer Segmentation Using K-Means Clustering on Online Retail Transactions," *J. Teknol. dan Inf.*, vol. 13, no. September, pp. 122–137, 2023, doi: 10.34010/jati.v13i2.

[19] T. S. Kasus, P. Pada, and S. Tripadvisor, "Analisis Sentimen Pelanggan Hotel di," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 8106, pp. 21–29, 2021, doi: 10.20895/INISTA.V3.

[20] D. A. Kusuma, A. R. Dewi, and A. R. Wijaya, "Perbandingan Random Forest dan Convolutional Neural Network dalam Memprediksi Peralihan Pelanggan," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 10, no. 2, pp. 186–194, 2025, doi: 10.14421/jiska.2025.10.2.186-194.

Author Biography

	**Ika Melinia Sapitri** was born in Sungai Penuh on January 17, 2000. She is known as an academic who has a high fighting spirit, is responsible, and is diligent in pursuing her education by adhering to the values she believes in. Her determination and discipline are important foundations in achieving various academic achievements. The author completed her Bachelor's degree in the Computer Science and Information Technology Education Program and obtained her (S.Pd) degree from Padang State University in 2024. As a manifestation of his commitment to the development of science, the author continued his master's education in the Informatics Engineering Study Program at the Putra Indonesia YPTK University in Padang. His areas of interest include data analysis, data mining, and the application of information technology in solving various data-based problems. Guided by the principles of humility and lifelong learning, the author is determined to continuously improve their quality and contribute to the development of science and technology. The author hopes that every step taken will bring benefits, inspire, and provide positive value to the academic community and society at large.
	**Sarjon Defit** was born Padang Sibusuk/07 August 1970. Currently active as a lecturer in Computer Science. The educational history of SI at the College of Informatics and Computer Management (STMIK "YPTK" Padang) with a graduation in 1993. An educational history of S2 at Universiti Teknologi Malaysia, Johor Bahru, graduated in 1998. Then a Doctoral Education History at Universiti Teknologi Malaysia, Johor Bahru, graduated in 2003. The field of science consists of data mining, artificial intelligence, decision support systems, and others. He can be contacted at email: sarjon_defit@upiyptk.ac.id
	**Rini Sovia** was born in Solok, April 5, 1976. Lecturer at Universitas Putra Indonesia YPTK Padang. He completed his education in informatics management in 1999 and his master's degree in informatics engineering in 2006. He is currently earning a doctorate in information technology in 2023. His areas of expertise include artificial intelligence (AI), expert systems (ES), data mining (DM), decision support systems (DSS), and databases. He can be reached at email: rini_sovia@upiyptk.ac.id