KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 3, Page: 130-142, September 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Bandwidth Usage Pattern Clustering in Local Government Agencies Using a SOM and K-Means Hybrid Approach

Hikko Razaki[1], Sarjon Defit[2], Rini Sovia[3]
[1] Communication and Informatics Department, Solok, 27314, Indonesia
[2,3] Information Technology Study Program, Faculty of Computer Science, Universitas Putra Indonesia YPTK, Padang, 25221, Indonesia

Abstract

Differences in the intensity of digital services across Regional Government Organizations (OPDs) within the Solok City Government result in varying bandwidth requirements among institutions. Uniform capacity allocation does not adequately reflect these differences and may lead to network inefficiencies. This study classified bandwidth-usage patterns using a combination of the Self-Organizing Map (SOM) and K-Means Clustering to provide a basis for more proportional bandwidth allocation. The dataset comprised 1,326 daily records from 26 OPDs collected between March and May 2026, including upload volume, download volume, number of employees, and number of active applications. Data processing involved outlier detection, aggregation, Min-Max normalization, SOM training on a 5 × 5 grid, determination of the optimal number of clusters using the Elbow Method, selection of initial centroids from the extreme points of the SOM map, and final clustering using K-Means. The SOM mapping produced a Quantization Error of 0.040207 and a Topological Error of 0.076923. Three clusters were identified as the optimal conFigureation. The SOM–K-Means model achieved a Silhouette Score of 0.448335, a Davies–Bouldin Index of 0.817415, and a WCSS value of 2.702929, outperforming conventional K-Means. The final results classified 11 OPDs into the high-usage category, 4 into the moderate-usage category, and 11 into the low-usage category. These findings provide an empirical basis for allocating bandwidth according to the usage characteristics of each OPD.

Kata kunci: bandwidth, SOM, K-Means, OPD, Kota Solok

KomtekInfo Journal is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.

1. Introduction

The expansion of digital government has made internet connectivity a core component of routine public administration rather than merely a supporting facility. Activities such as exchanging documents, maintaining databases, coordinating across agencies, delivering public services, and operating governmental applications require a reliable network [1], [2]. Bandwidth, however, is a shared resource that must serve organizational units whose workloads and patterns of use are not identical. Network demand at each Regional Government Organization (Organisasi Perangkat Daerah, OPD) is affected by factors such as staff size, the number of applications in operation, the intensity of data transfer, and the nature of the services provided. Consequently, assigning the same bandwidth capacity to every OPD may be neither fair nor efficient. Units with heavier traffic can experience reduced performance at busy times, while capacity provided to lightly used units may remain idle [3], [4]. This consideration is particularly important for the Solok City Government, whose main bandwidth was reduced from 1.5 Gbps to 700 Mbps as part of budget-efficiency measures. With a more constrained capacity, allocation should be guided by recorded usage rather than an equal administrative distribution.

Evidence of unequal bandwidth demand has previously been reported in a K-Means study of OPD internet usage in the Padang Pariaman Regency Government. That study formed three groups: six OPDs with high use, fourteen with moderate use, and twelve with low use [5]. These findings indicate that agencies operating within the same governmental system may nevertheless exhibit substantially different network requirements, making data-based grouping more informative than treating all units as equivalent. However, using K-Means alone introduces a methodological concern because the resulting partition can depend strongly on the initial centroids. Starting points that poorly represent the data distribution may direct the iterative procedure toward different or less sTable solutions. A preliminary technique capable of learning the proximity structure of the data is therefore needed to provide more representative starting information for K-Means.

To provide this preliminary representation, the study employed a Self-Organizing Map (SOM). SOM is an unsupervised technique that maps high-dimensional observations to a lower-dimensional space while retaining their main neighborhood relationships [6].

Submitted: 09-06-2026 | Revised: 10-07-2026 | Accepted: 17-08-2026 | Published: 30-09-2026

During training, neurons compete to represent each input vector, and the weights of the Best Matching Unit and nearby neurons are adjusted [7]. As a result, observations with comparable characteristics tend to occupy the same or neighboring locations. This property is useful for bandwidth analysis because upload volume, download volume, employee count, and active-application count must be considered jointly. Although two OPDs may differ on an individual variable, they may still share a similar overall profile when all four indicators are evaluated together. The SOM representation reveals these relationships before final cluster boundaries are defined, allowing centroid initialization to be based on the learned data structure rather than on arbitrary random values.

K-Means Clustering was then used to produce the final groups by assigning each observation to its nearest centroid and recalculating cluster centers until membership no longer changed [8]. The algorithm was selected because its procedure is transparent, relatively simple to trace, and produces groups that network administrators can interpret directly. In the proposed combination, neuron weights at the outer regions of the SOM grid served as the initial centroids, providing well-separated starting points that reflected variation among OPD profiles. The candidate number of clusters was assessed with the Elbow Method through changes in Within-Cluster Sum of Squares (WCSS). The resulting partition was evaluated with the Silhouette Index, Davies-Bouldin Index, and WCSS. Together, these indicators describe member suitability and separation, the relationship between compactness and intercluster distance, and the amount of variation within clusters. Their use provides quantitative support for the selected grouping rather than relying only on visual inspection.

A previous study combined SOM and K-Means to examine network usage in a university setting in Bandung. After optimization, the Silhouette Index rose from 0.321 to 0.773, whereas the Davies-Bouldin Index fell from 0.896 to 0.623 [9]. These results suggest that an initial mapping stage can help K-Means generate more cohesive and better-separated clusters. Nevertheless, traffic behavior in a university differs from that of local government agencies. Campus usage is shaped largely by academic schedules and user-access periods, while OPD traffic arises from public-service delivery, administrative work, sectoral data management, and institution-specific government applications. Staff numbers also do not always correspond directly to data-transfer intensity because some services rely heavily on particular applications or process large volumes of documents. Therefore, results obtained in a campus context cannot be assumed to represent the conditions of the Solok City Government. Testing SOM-based centroid initialization on OPD data is necessary to determine whether the same improvement appears in a local-government network environment.

Accordingly, this research processes bandwidth records from Solok City OPDs through outlier screening, aggregation, normalization, SOM training, determination of the cluster count with the Elbow Method, selection of initial centroids from extreme SOM nodes, K-Means clustering, and internal performance evaluation. The analysis covers 26 OPDs represented by four variables associated with network activity and digital operational capacity. To support policy interpretation, the resulting clusters are described as low-, moderate-, and high-usage groups. The study does not prescribe an exact bandwidth value for each institution; instead, it supplies empirical evidence of differences in demand before technical allocation decisions are made by the responsible office. The combined model is also compared with conventional K-Means to assess the effect of SOM-derived initialization. The principal objective is to identify usage patterns and group OPDs according to their observed characteristics. Methodologically, the study evaluates extreme SOM nodes as starting centroids for K-Means. Practically, it provides measurable information that can assist the Solok City Communication and Informatics Office in designing a more proportional, efficient, and needs-oriented bandwidth-allocation policy.

2. Methods

The research procedure followed a sequential data-analysis workflow, beginning with problem identification and ending with the development of recommendations. Bandwidth-use records were obtained from Regional Government Organizations (Organisasi Perangkat Daerah, OPDs) of the Solok City Government and checked for analytical suitability. Preprocessing included handling potential outliers, summarizing observations by OPD, normalizing the variables, and examining relationships among features. The processed profiles were subsequently mapped with SOM, grouped through K-Means, and evaluated using internal clustering measures. Figuree 1 summarizes the complete research workflow.

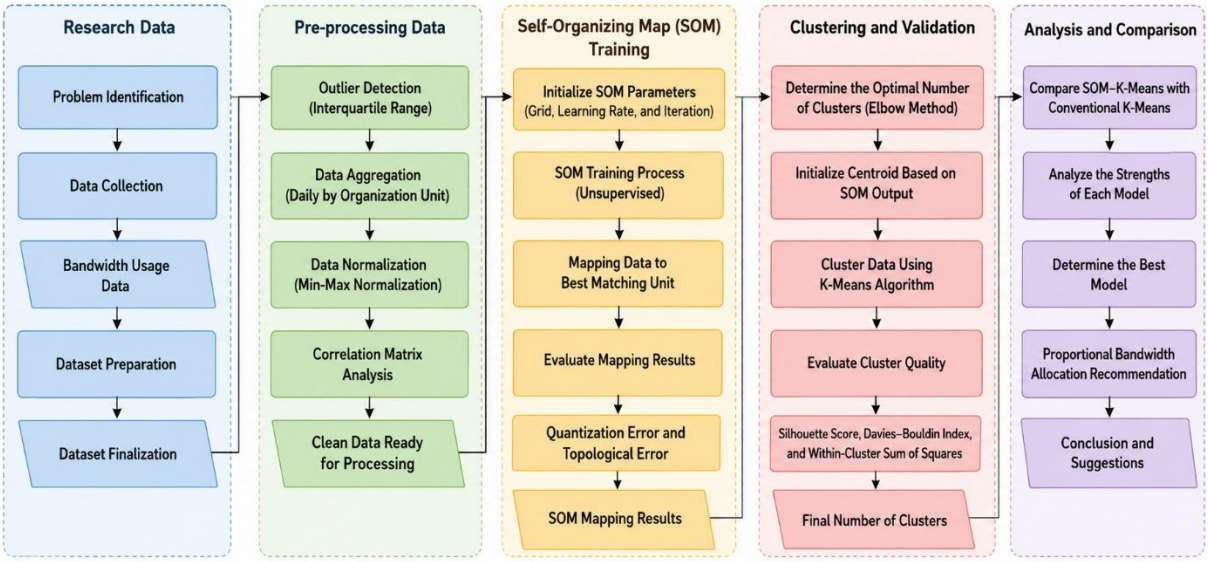


Figure 1 Research Framework

As shown in Figure 1, the preprocessed profiles were first entered into SOM to obtain an initial map of similarities among OPDs. Information from the trained map was then used to define the starting centroids for K-Means. The Elbow Method was applied to the pattern of WCSS reduction to select the number of clusters, while the Silhouette Score, Davies-Bouldin Index, and WCSS were used to evaluate the resulting structure. After a satisfactory partition was obtained, cluster characteristics were interpreted on the original measurement scale and assigned to low-, moderate-, or high-usage categories. The final stage converted these analytical results into bandwidth-allocation recommendations.

2.1 Preprocessing Data

Preprocessing was performed to convert the raw records into numerically valid and comparably scaled inputs for SOM and K-Means. It covered the cleaning, transformation, and organization required for consistent computational analysis [10], [11]. The preprocessing stage comprised numerical data cleaning, IQR-based outlier detection, OPD-level aggregation, and Min–Max normalization, this stage included cleaning and converting numerical features, detecting outliers with the Interquartile Range (IQR), calculating mean values for each OPD, and applying Min-Max normalization. Cleaning involved standardizing column labels, changing decimal commas to decimal points, converting the four features to numeric types, and excluding rows with missing feature values or absent OPD identifiers.

Outliers are data points that differ markedly from the main distribution and can affect averages as well as distance-based clustering results [12]. For their identification, the positions of the first and third quartiles (Q1 and Q3) were derived from the number of observations, n, using Formulas (1) and (2), following the formulation applied in the thesis.

$$Q1 = 25\%\ x\ n \quad (1)$$

$$Q3 = 75\%\ x\ n \quad (2)$$

After Q1 and Q3 had been located, the interquartile range was obtained by subtracting Q1 from Q3, as shown in Formula (3). Formulas (4) and (5) were then used to establish the lower and upper boundaries. These IQR-based limits define the central range of the dataset and identify observations beyond the accepted interval [12].

$$IQR = Q3 - Q1 \quad (3)$$

$$Lower\ Bound = Q1 - (1.5\ x\ IQR) \quad (4)$$

$$Upper\ Bound = Q3 + (1.5\ x\ IQR) \quad (5)$$

A record was flagged as outside the principal range when at least one feature was below its lower boundary or above its upper boundary. The procedure was conducted independently for upload volume, download volume, employee count, and the number of active applications.

Data aggregation. The daily observations were grouped by OPD, and the mean of every feature was calculated for each group. Mean aggregation was chosen because network traffic can fluctuate from day to day and from week to week [13]. Summarizing the full observation period reduces the effect of temporary variations and yields a more representative usage profile before clustering [14][15]. Formula (6) presents the calculation of the OPD-level mean.

$$\bar{x}_{g,j} = \frac{1}{n_g} \sum_{i=1}^{n_g} x_{i,g,j} \quad (6)$$

In Formula (6), $\bar{x}_{g,j}$ denotes the mean of feature j for OPD g; n_g is the number of records associated with that OPD; and $x_{i,g,j}$ denotes the value recorded in observation i for feature j. The aggregation yields one profile for each OPD comprising the average upload volume, download volume, employee count, and number of active applications.

Submitted: 09-06-2026 | Revised: 10-07-2026 | Accepted: 17-08-2026 | Published: 30-09-2026

Data normalization. The aggregated features were expressed in different units and ranges. Because both SOM and K-Means calculate Euclidean distances, a variable with a wider scale could dominate the result. Each feature was therefore rescaled to the interval 0-1 with Min-Max Scaling [16]. Min–Max normalization was performed using Formula (7).

$$x_{Normalization} = \frac{(x - x_{minimal})}{(x_{maksimal} - x_{minimal})} \quad (7)$$

In Formula (7), $x_{\text{normalized}}$ is the normalized value, x is the original feature value, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of that feature, respectively. The resulting normalized matrix was saved as data_scaled and supplied to the SOM training process.

2.2 Self-Organizing Maps

SOM is an unsupervised neural-network model that represents high-dimensional observations on a lower-dimensional grid while retaining spatial similarity and topological relationships [6], [17]. Its competitive and cooperative learning mechanism places comparable inputs on the same neuron or on neighboring neurons [7], [18]. In this research, SOM was used to learn the similarity structure of the OPD profiles and to indicate their relative positions before K-Means clustering. It did not produce the final usage labels; instead, the trained map supplied a structured basis for selecting the initial K-Means centroids, consistent with the combined SOM-K-Means approach reported in [10].

The SOM was trained with 26 OPD profiles described by four normalized features. A 5 x 5 lattice, consisting of 25 neurons, was selected. The neuron weights were initialized through PCA so that their starting arrangement followed the dominant directions of variation in the data and supported faster convergence. Every neuron contained a weight vector with the same dimensionality as an input profile.

For each input vector x_i, its Euclidean distance to every neuron weight vector was computed. The neuron producing the smallest distance was designated as the Best Matching Unit (BMU), as stated in Formula (8).

$$d(x_i, w_j) = \sqrt{\sum_{p=1^m} (x_{ip} - w_{jp})^2} \quad (8)$$

Once the BMU was found, a Gaussian neighborhood function was used to determine its influence on the surrounding neurons. Neurons closer to the winning unit received larger weight changes than more distant neurons. Formula (9) expresses the strength of this neighborhood effect.

$$h_{cj}(t) = \exp\{-\frac{|r_c - r_j|^2}{[2\sigma^2(t)]}\} \quad (9)$$

In Formula (9), r_c identifies the BMU coordinates, r_j identifies the coordinates of neuron j, and $\sigma(t)$ denotes the neighborhood radius at iteration t. The resulting neighborhood coefficient was then used in the neuron-weight update defined by Formula (10).

$$w_j(t+1) = w_j(t) + \alpha(t) h_{cj}(t)[x_i - w_j(t)] \quad (10)$$

The learning-rate parameter $\alpha(t)$ controls the size of each weight adjustment. BMU selection and weight updating were repeated across the dataset until changes in the map gradually diminished. After training, every OPD profile was assigned to the neuron with the closest weight vector.

Map quality was examined through Quantization Error (QE) and Topological Error (TE). QE is the mean distance between observations and their respective BMUs, whereas TE is the proportion of observations whose first and second BMUs are not adjacent. Their calculations are provided in Formulas (11) and (12).

$$QE = (\frac{1}{N}) \sum_{i=1^n} ||x_i - w_{c(i)}|| \quad (11)$$

$$TE = (\frac{1}{N}) \sum_{i=1^n} u(x_i)|| \quad (12)$$

The indicator function $u(x_i)$ equals 0 when the first and second BMUs are adjacent and 1 when their neighborhood relation is not maintained. Smaller QE values indicate closer representation of the input profiles by neuron weights, while smaller TE values indicate better preservation of the original neighborhood structure.

2.3. K-Means Clustering

K-Means divides the observations into K groups by repeatedly assigning each item to the nearest centroid. Its objective is to minimize variation within clusters so that similar profiles are placed together and dissimilar profiles are separated [8], [19]. Because the final partition may depend on the starting centers, centroid initialization is a critical step [19]. In this study, K-Means processed the 26 normalized OPD profiles using selected SOM neuron weights as its initial centroids.

Candidate solutions for K=1 through K=10 were assessed with the Elbow Method. For each candidate, WCSS was computed as the total squared distance between every observation and the centroid of its assigned cluster. The optimal value of K was selected at the elbow point, where increasing the number of clusters produced only marginal reductions in WCSS, meaning that additional clusters produced only a relatively small decrease in WCSS. Formula (13) defines this calculation.

$$WCSS = \sum_{k=1^k} \sum_{i \in C_k} (x_i - c_k)^2 \qquad (13)$$

After K had been selected, neuron-weight vectors at coordinates $(0, 0)$, $(0, 4)$, and $(4, 0)$ were adopted as the three starting centroids. These locations lie in widely separated parts of the SOM grid and represent different profile patterns. Thus, the initialization reflected the structure learned by SOM instead of depending solely on random starting points.

In each K-Means iteration, the Euclidean distance from every OPD profile to each centroid was calculated according to Formula (14).

$$d(x_j, \mu_k = \sqrt{\sum_{j=1^m} \left(X_{ij} - \mu_{kj}\right)^2} \qquad (14)$$

Each OPD was allocated to the cluster associated with its smallest centroid distance, following the membership rule in Formula (15).

$$Z_i = \arg\min \text{ k } d(X_j, \mu_k) \qquad (15)$$

After cluster assignments were completed, each centroid was recalculated as the mean profile of the members in that cluster, as specified in Formula (16).

$$\mu_k = \frac{1}{n_k} \sum_j Z_{(j=k)} X_i \qquad (16)$$

The distance, assignment, and centroid-update steps were repeated until cluster membership no longer changed or the remaining centroid movement was negligible. The procedure produced groups of OPDs with similar upload, download, employee, and active-application characteristics.

The final partition was evaluated with the Silhouette Score, Davies-Bouldin Index (DBI), and WCSS. The Silhouette Score measures how well an observation fits its assigned cluster relative to alternative clusters; values nearer to 1 indicate a stronger partition. DBI compares within-cluster dispersion with separation between centroid pairs, so smaller values are desirable. WCSS summarizes the residual variation inside the clusters. After evaluation, the mean profiles on the original scale were used to label the groups as low, moderate, or high usage.

3. Results and Discussions

This section presents the data-processing results sequentially, beginning with dataset examination and continuing through model evaluation. Each stage is discussed by relating the numerical findings to the context of network usage across OPDs. The discussion focuses not only on the formation of the clusters but also on the model's ability to represent differences in bandwidth requirements in a meaningful and contextually reasonable manne.

3.1 Dataset

The data were obtained from network traffic records maintained by the Solok City Communication and Informatics Office for the period from March to May 2026. The dataset comprised 1,326 daily observations covering 26 Regional Government Organizations (OPDs). Each record contained the date, institution name, upload volume, download volume, number of employees, and number of active applications. The latter four numerical variables were used to construct the network-usage profiles.

Upload and download volumes represent data-exchange activity, while the number of employees serves as a proxy for the number of potential users. The number of active applications provides information regarding the extent of digital services operated by each OPD. A sample of the dataset is presented in Table 1

Table 1 Sample Bandwidth Usage Data

Date	OPD	Upload Data Volume (GB)	Download Volume Data (GB)	Number of Employees	Number of Active Applications
2026-03-02	Badan Kepegawaian dan Pengem-bangan Sumber Daya Manusia	76,33	132,98	40	15
2026-03-02	Badan Kesatuan Bangsa dan Politik	6,70	13,79	22	15
2026-03-02	Badan Keuangan Daerah	66,22	113,57	100	19
2026-05-29	Kecamatan Lubuk Sikarah	11,80	20,86	30	16
2026-05-29	Kecamatan Tanjung Harapan	15,31	43,21	30	16

The values presented in Table 1 reveal clear differences among the OPDs. Certain institutions recorded high levels of data traffic, whereas others exhibited substantially lower usage. Variations were also observed in the number of employees and active applications. These conditions support the use of a multivariable analysis, as allocating network capacity based on a single indicator or distributing it uniformly across institutions would not adequately reflect their actual requirements

3.2 Variable Analysis

Outlier detection was performed prior to data aggregation to prevent highly deviating values from influencing the position of the data on the map and the resulting cluster structure. The four numerical variables were examined using the Interquartile Range (IQR). The lower bound was calculated as Q1 minus 1.5 times the IQR, whereas the upper bound was obtained as Q3 plus 1.5 times the IQR. A summary of the examination is presented in Table 2. All 1,326 observations remained within the lower and upper bounds of their respective variables. No records required removal or replacement. Therefore, the entire dataset was retained for the subsequent stage without any specific treatment for outliers.

Because the dataset consisted of daily records, each OPD appeared repeatedly throughout the observation period. To obtain comparable units of analysis, all records were grouped by OPD name and summarized using mean values. This procedure produced a single profile for each OPD while reducing the influence of temporary daily traffic fluctuations. The aggregation process transformed 1,326 daily records into 26 OPD profiles. Each profile represented the general pattern of network usage over the three-month observation period rather than the condition on a particular day. These values were subsequently used in the normalization process so that differences in measurement units across variables would not affect the distance calculations.

The four aggregated variables had different units and value ranges. Upload and download volumes were measured in gigabytes, whereas the numbers of employees and active applications were expressed as counts. Without scale adjustment, variables with larger ranges could dominate the Euclidean-distance calculation. Therefore, Min-Max Scaling was applied to all variables across the 26 OPDs to transform the values into a range from 0 to 1. A sample of the normalized results is presented in Table 2.

Table 2 Results of Data Normalization

OPD	Upload (X1)	Download (X2)	Number of Employees (X3)	Number of Applications (X4)
BKPSDM	0,895687	0,974882	0,111542	0
Bakesbangpol	0	0,002679	0,001856	0
BKD	0,977930	0,759592	0,628417	0,666667
Polppdam	0,826451	0,779740	0,846439	0,166667
Sekda	0,822856	0,813437	1,000000	1,000000

3.3 Training and Mapping With Self-Organizing Map

SOM was used to organize the 26 OPD profiles on a map according to the similarity of the four normalized variables. Prior to training, the grid size, learning rate, neighborhood radius, neighborhood function, and number of iterations were specified. The conFigureation employed in this study is summarized in Table 3.

Table 3 SOM Training Parameter

Evaluation Component	Result	Description
som shape	5 x 5	Dimension of SOM grid (X, Y)
input len	4	Number of the elements of the vectors in input
sigma	1,0	Spread of the neighborhood function
learning rate	0,5	Speed of learning during training
neighborhood func	gaussian	Function that defines the neighborhood influence
activation dist	euclidean	Metric used to measure the distance between the input vector and neuron
Random seed	42	Random generator number
Num iteration	1.000	Number of SOM training iterations

The parameters presented in Table 3 served as the reference conFigureation throughout the SOM training process. The 5 × 5 grid provided 25 neurons for mapping the 26 OPDs. This conFigureation was considered sufficient to capture variations among the profiles without producing an excessively sparse map. A sigma value of 1.0 limited the scope of weight adjustment to the nearest neurons, while a learning rate of 0.5 provided a sufficiently strong response during the initial stage of training. A total of 1,000 iterations was employed to allow the neuron weights adequate time to adapt to the distribution of the data.

Following the training process, the distance between each OPD profile and all neurons was recalculated. The neuron with the minimum distance was identified as the Best Matching Unit (BMU) and represented the position of each of the 26 OPDs on the grid. A sample of the

winning coordinates and their corresponding Euclidean distances is presented in Table 4

Table 4 Winners BMU Coordinates

OPD	BMU Coordinates	Euclidean Distance
BKPSDM	(0, 4)	0,019633
Bakesbangpol	(0, 0)	0,054813
BKD	(3, 4)	0,081456
Polppdam	(1, 3)	0,153648
Sekda	(4, 4)	0,035134

The distance to the BMU differed for each OPD. The Inspectorate recorded the smallest distance, at 0.002393, followed by the Population and Civil Registration Office at 0.004790 and the Communication and Informatics Office at 0.006076. Several OPDs were mapped to the same neuron, such as the National Unity and Political Agency and the Health Office at coordinate (0,0), as well as the Regional Finance Agency and the Education Office at coordinate (3,4). The Regional Disaster Management Agency, the Library and Archives Office, Lubuk Sikarah District, and Tanjung Harapan District were also positioned at coordinate (1,0). This shared placement indicates that their profiles across the four variables were relatively similar.

The coordinates of the 26 BMUs indicate only their locations on the grid. To interpret the prototype represented by each neuron, its final weight vector must also be examined. W1 to W4 correspond respectively to upload volume, download volume, number of employees, and number of active applications on the normalized scale. A sample of the winning neuron weights is presented in Table 5

Table 5 Winners BMU Weights

OPD	W1 Upload	W2 Download	W3 Employees	W4 Application
BKPSDM	0,899059	0,975064	0,119226	0,017749
Bakesbangpol	0,032583	0,018590	0,016501	0,038408
BKD	0,942804	0,783264	0,697941	0,669365
Polppdam	0,844682	0,747131	0,748743	0,279216
Sekda	0,830445	0,809850	0,975362	0,976402

The neuron weights reveal distinct profile characteristics across different regions of the map. The Regional Secretariat exhibited high values across all four dimensions, particularly for the number of employees and active applications. Bapperida showed a stronger profile in upload and download volumes, while the Communication and Informatics Office had an active-application weight approaching the maximum value. In contrast, the National Unity and Political Agency and the Health Office displayed low weights across all features. OPDs mapped to the same BMU shared an identical prototype because they were represented by the same neuron.

The relative positions of the BMUs were visualized on a 5×5 grid to examine the proximity relationships among the data points. Labels assigned to the same coordinates were slightly displaced to maintain readability; this adjustment was made solely for visualization purposes and did not alter the actual BMU locations. The resulting trained map is presented in Figuree 4

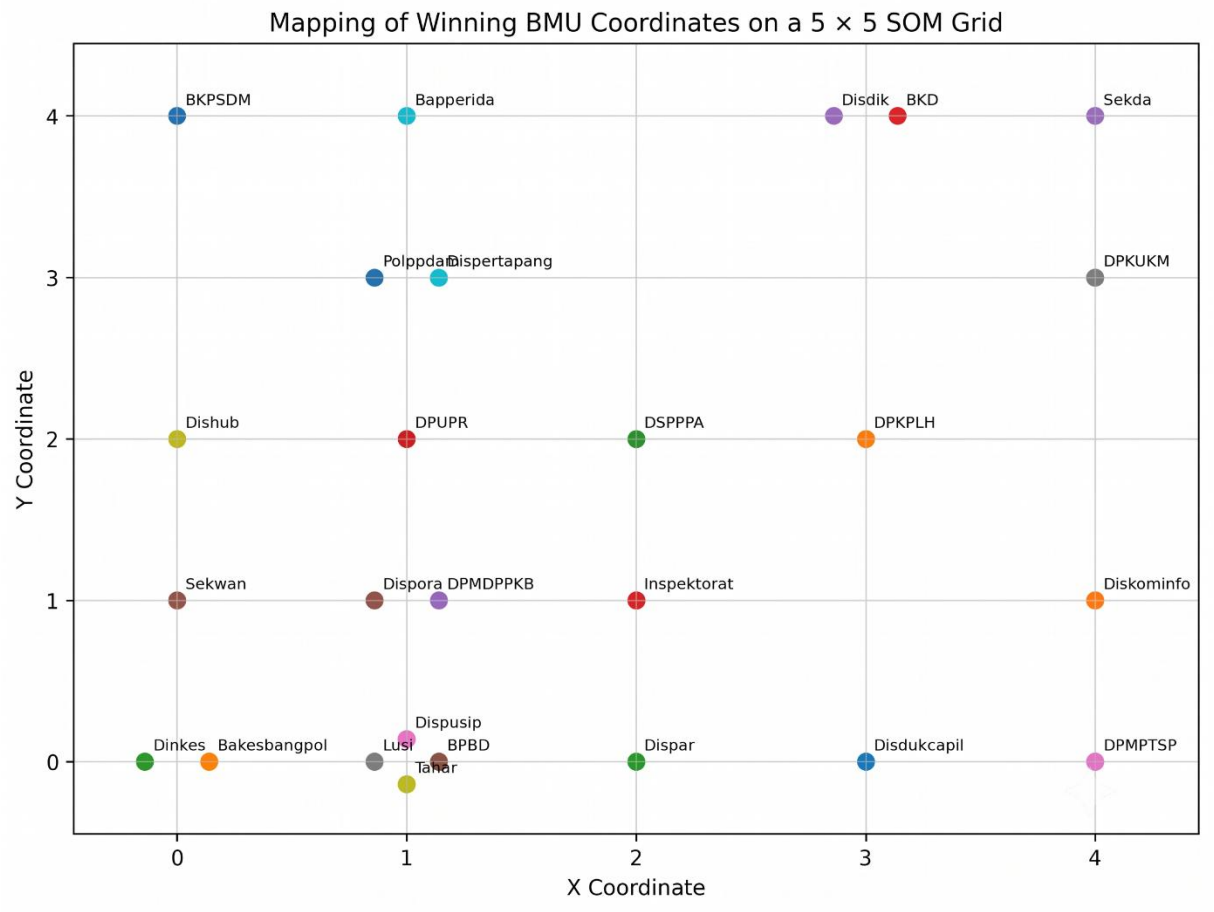


Figure 1 Winners BMU Mapping

The map presented in Figure 2 shows that OPDs with similar profiles tended to be located on the same neuron or on neighboring neurons. Several OPDs were mapped to the same coordinate, whereas more distinctive profiles occupied separate positions, such as the Communication and Informatics Office at (4,1), the Regional Secretariat at (4,4), and Bapperida at (1,4). The X and Y coordinates do not directly represent the magnitude of any particular variable; rather, the main information lies in the proximity among the mapped points. This spatial structure was subsequently used to identify the initial points for K-Means

3.4 Bandwidth Clustering

The determination of the optimal number of clusters began by calculating WCSS for K=1 to K=10. WCSS represents the sum of the squared distances between all data points and their respective cluster centroids. Because this value consistently decreases as the number of clusters increases, the optimal K was not selected solely on the basis of the smallest WCSS value. Instead, K was determined at the point where the curve began to change direction and subsequent reductions became less substantial than those observed previously. A sample of the WCSS calculation results is presented in Table 6.

The largest reduction occurred between K=1 and K=2, with WCSS decreasing from 9.505554 to 3.604157. At K=3, the WCSS value declined further to 2.702285. Beyond this point, the value continued to decrease, although the reduction associated with each additional cluster became progressively smaller. Considering this pattern and the need to establish low-, moderate-, and high-usage categories, three clusters were selected as the final conFigureation.

Table 6 WCSS Values for Determining the Optimal Number of Clusters

Number of Clusters (K)	WCSS Value
1	9,505554
2	3,604157
3	2,702285
….	……
8	0,593309
9	0,472129
10	0,362420

The pattern of WCSS reduction is visualized in the Elbow plot presented in Figuree 3. The horizontal axis represents the number of clusters, while the vertical axis shows the corresponding WCSS values. The point at K=3 is highlighted as the selected turning point of the curve.

The graph shows a sharp decline across the first two values of K, followed by a gradual flattening of the curve around K=3. Although adding more clusters beyond this point further reduced WCSS, the resulting improvement was not proportional to the increased complexity of the clustering structure. Therefore, three clusters were considered sufficient to preserve data compactness while maintaining an interpreTable solution.

Submitted: 09-06-2026 | Revised: 10-07-2026 | Accepted: 17-08-2026 | Published: 30-09-2026

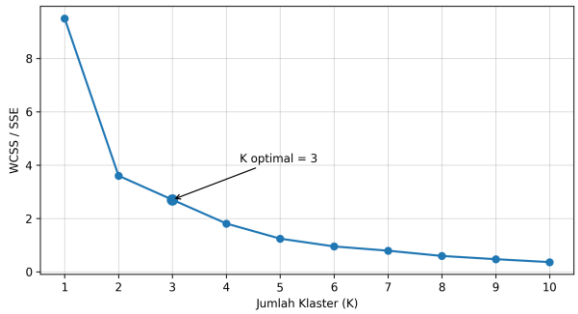


Figure 2 Elbow Graphic

After the number of clusters had been determined, the initial centroids were selected from three extreme points on the SOM map, namely (0,0), (0,4), and (4,0). These points were located in relatively distant regions of the map and therefore represented distinct initial profiles. The weight vector of each selected neuron was used as the value of centroids C1, C2, and C3. The corresponding details are presented in Table 7.

Table 7 Initial Centroids Based on SOM Extreme Points

Centroid	Coordinates	OPD	W1	W2	W3	W4
C1	(0, 0)	Bakesbangpol	0,032583	0,018590	0,016501	0,038408
C2	(0, 4)	BKPSDM	0,899059	0,975064	0,119226	0,017749
C3	(4, 0)	DPMPTSP	0,082021	0,111629	0,095506	0,831669

Centroid C1 was derived from the neuron representing Bakesbangpol at coordinate (0,0), C2 from the neuron representing BKPSDM at (0,4), and C3 from the neuron representing DPMPTSP at (4,0). These three prototypes exhibited distinct weight conFigureations, enabling them to serve as spatially dispersed initial points for the K-Means clustering process. The positions of the three extreme points are visualized in Figuree 4 to provide a clearer representation of the spatial relationships among the initial centroids.

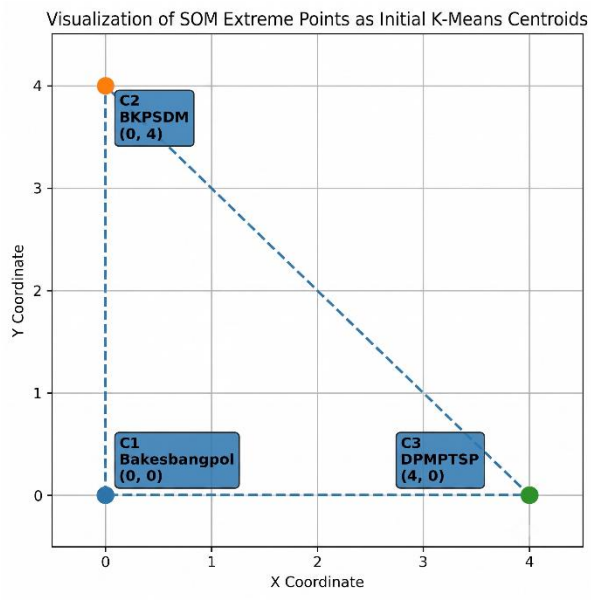


Figure 3 SOM Extreme Node Visualization

The three points formed a triangular conFigureation along the outer boundaries of the grid. C1 represented a profile with relatively low weights across most features, C2 was characterized by high upload and download traffic, whereas C3 was distinguished by a high number of active applications. This spatial distribution provided K-Means with diverse initial positions before the cluster centers were recalculated based on the members assigned to each group.

The distance from each OPD to C1, C2, and C3 was then calculated. Each OPD was assigned to its nearest centroid, after which the cluster centers were updated using the mean values of their respective members. Cluster membership became sTable at the second iteration. C1 contained 11 OPDs, C2 contained 11 OPDs, and C3 contained 4 OPDs. The complete cluster membership is presented in Table 8.

Table 8 OPD Membership in Each Cluster

Cluster	OPD
C1	Badan Kesatuan Bangsa dan Politik Badan Penanggulangan Bencana Daerah Dinas Kesehatan Dinas Pariwisata Dinas Pemberdayaan Masyarakat dan Desa, Pengendalian Penduduk Dinas Pemuda dan Olahraga Dinas Perpustakaan dan Kearsipan Inspektorat Daerah Kecamatan Lubuk Sikarah Kecamatan Tanjung Harapan Sekretariat Dewan Perwakilan Rakyat Daerah
C2	Badan Kepegawaian dan Pengembangan Sumber Daya Manusia Badan Keuangan Daerah Badan Perencanaan Pembangunan, Riset dan Inovasi Daerah Dinas Pekerjaan Umum dan Penataan Ruang Dinas Pendidikan Dinas Perdagangan dan Koperasi Usaha Kecil dan Menengah Dinas Perhubungan Dinas Pertanian dan Pangan Dinas Perumahan, Kawasan Permukiman dan Lingkungan Hidup Satuan Polisi Pamong Praja dan Pemadam Kebakaran Sekretariat Daerah
C3	Dinas Kependudukan dan Pencatatan Sipil Dinas Komunikasi dan Informatika Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu Dinas Sosial, Pemberdayaan Perempuan dan Perlindungan Anak

C1 and C2 each accounted for 42.31% of the study objects, whereas C3 represented 15.38%. Cluster membership remained unchanged after the centroids were updated in the second iteration, indicating that the clustering process had converged. At this stage, the

labels C1, C2, and C3 served only as cluster identifiers; their corresponding usage levels were determined through an analysis of the mean cluster profiles.

Cluster labeling was performed using the original aggregated values to ensure that the results remained readily interpreTable in an operational context. Mean upload and download volumes were expressed in gigabytes, while the numbers of employees and active applications remained in count units. A comparison of the clusters and their final labels is presented in Table 9

Table 9 Mean Variable Values and Cluster Labels

Variabel	C1	C2	C3
Upload	20,146	45,751	22,953
Download	39,145	106,872	46,258
Pegawai	39,127	92,752	44,745
Aplikasi	16,000	17,818	19,250
Total	114,417	263,194	133,206
Label	**Rendah**	**Tinggi**	**Sedang**

C2 recorded a mean upload volume of 45.751, a mean download volume of 106.872, an average of 92.752 employees, and 17.818 active applications. This combination represented the highest overall profile; therefore, C2 was classified as the high-usage cluster. C3 occupied an intermediate position, with a total characteristic value of 133.206, and was consequently labeled as the moderate-usage cluster. C1 showed the lowest values across most variables, with a total characteristic value of 114.417, and was therefore categorized as the low-usage cluster.

After the labels had been assigned, the cluster members were reorganized according to their usage categories. C2 represented the high-usage group, C3 the moderate-usage group, and C1 the low-usage group. The complete list of OPDs in each category is presented in Table 10

Table 10 Grouping of OPDs by Usage Category

Label	OPD
Tinggi	Badan Kepegawaian dan Pengembangan Sumber Daya Manusia
	Badan Keuangan Daerah
	Badan Perencanaan Pembangunan, Riset dan Inovasi Daerah
	Dinas Pekerjaan Umum dan Penataan Ruang
	Dinas Pendidikan
	Dinas Perdagangan dan Koperasi Usaha Kecil dan Menengah
	Dinas Perhubungan
	Dinas Pertanian dan Pangan
	Dinas Perumahan, Kawasan Permukiman dan Lingkungan Hidup
	Satuan Polisi Pamong Praja dan Pemadam Kebakaran
	Sekretariat Daerah
Sedang	Dinas Kependudukan dan Pencatatan Sipil
	Dinas Komunikasi dan Informatika
	Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu
	Dinas Sosial, Pemberdayaan Perempuan dan Perlindungan Anak
Rendah	Badan Kesatuan Bangsa dan Politik
	Badan Penanggulangan Bencana Daerah
	Dinas Kesehatan
	Dinas Pariwisata
	Dinas Pemberdayaan Masyarakat dan Desa, Pengendalian Penduduk
	Dinas Pemuda dan Olahraga
	Dinas Perpustakaan dan Kearsipan
	Inspektorat Daerah
	Kecamatan Lubuk Sikarah
	Kecamatan Tanjung Harapan
	Sekretariat Dewan Perwakilan Rakyat Daerah

Eleven OPDs were classified in the high-usage category, four in the moderate-usage category, and eleven in the low-usage category. This composition confirms that network requirements differ across institutions. The high-usage group exhibited a greater combination of network traffic, number of employees, and application utilization and therefore warrants higher capacity priority. The moderate-usage group requires an intermediate level of allocation, whereas the low-usage group can be assigned capacity in accordance with its actual workload. These findings fulfill the objective of the study by identifying and classifying bandwidth-usage patterns among the OPDs. From a methodological perspective, SOM successfully provided more directed initial centroids for K-Means; from a practical perspective, the resulting groups can serve as a basis for considering a more proportional allocation of bandwidth.

3.5 Evaluation of Mapping and Clustering Results

The evaluation of the SOM map focused on two aspects: representational proximity and the preservation of topological structure. Quantization Error measures the average distance between each data point and its BMU, whereas Topological Error records the proportion of data for which the two nearest BMUs are not adjacent. Low values for both metrics indicate that the map represents the data effectively while preserving its topological relationships. The calculation results are presented in Table 11

Table 11 SOM Mapping Evaluation

Evaluation	Value
Quantization Error (QE)	0,040207
Topological Error (TE)	0,076923

A Quantization Error of 0.040207 indicates that the average distance between each OPD profile and its winning neuron was low in the normalized data space. A Topological Error of 0.076923 shows that neighborhood violations occurred in only 2 of the 26 OPDs. In other words, approximately 92.31% of the topological relationships were preserved. These values support the use of the SOM map as a basis for selecting the initial centroids.

The performance of the combined method was subsequently compared with that of conventional K-Means. The principal difference between the two approaches lies in the initialization strategy: the baseline model used conventional centroids, whereas the combined model employed neuron weights obtained from the extreme points of the SOM map. Three evaluation measures were used for comparison, namely the Silhouette Index, Davies–Bouldin Index, and WCSS. A higher Silhouette Index indicates better clustering quality, whereas lower DBI and WCSS

values are preferable. The numerical comparison is presented in Figuree

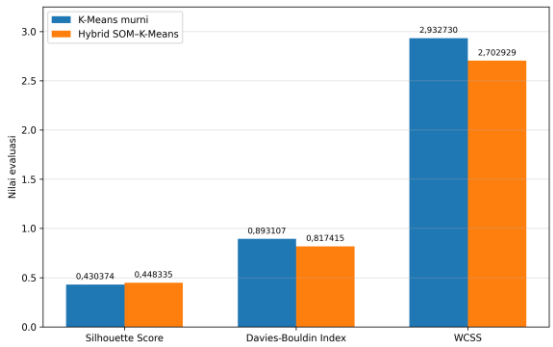


Figure 4 Comparison of Conventional K-Means and Hybrid SOM–K-Means Evaluation Results

The Silhouette Index achieved by SOM–K-Means was 0.448335, whereas conventional K-Means obtained a value of 0.430374. This difference indicates that, in the combined model, cluster members were slightly more cohesive within their assigned groups and better separated from members of other clusters. Although the improvement was modest, its direction was consistent with the intended benefit of using SOM-based initial centroids.

The Davies–Bouldin Index decreased from 0.893107 to 0.817415. WCSS also declined from 2.932730 to 2.702929. These changes indicate lower within-cluster dispersion and improved separation among clusters in the combined model. Therefore, the advantage of the proposed method was reflected across more than one evaluation measure.

All three metrics showed a consistent pattern: the Silhouette Index increased, while the Davies–Bouldin Index and WCSS decreased. These findings suggest that the structure learned by SOM helped K-Means obtain initial centroids that were better aligned with the distribution of the data. The objective of producing an objective classification of bandwidth-usage patterns was therefore achieved, although the observed improvement should be interpreted with consideration of the relatively limited dataset size. The contribution of this study lies in the application of topology-based centroid initialization and in the development of OPD groupings that may serve as an initial input for bandwidth-allocation policies within the Solok City Government.

4. Conclusions

Based on the analysis of 1,326 bandwidth-usage observations from 26 Regional Government Organizations (OPDs) within the Solok City Government, it can be concluded that the combination of a Self-Organizing Map and K-Means was effective in mapping and clustering bandwidth-usage patterns based on upload volume, download volume, number of employees, and number of active applications. The preprocessing stage, which included outlier detection, data aggregation, and normalization, produced a dataset suiTable for the modeling process. SOM mapping using a 5 × 5 grid yielded a Quantization Error of 0.040207 and a Topological Error of 0.076923, indicating that the data characteristics were represented with relatively low quantization error and limited topological distortion. The Elbow Method identified three clusters as the optimal conFigureation, while the use of neuron weights from the extreme points of the SOM map as initial centroids enabled K-Means to produce better clustering results than conventional K-Means. This improvement was demonstrated by a Silhouette Score of 0.448335, a Davies–Bouldin Index of 0.817415, and a WCSS value of 2.702929, compared with 0.430374, 0.893107, and 2.932730, respectively, for conventional K-Means. The final results classified the 26 OPDs into 11 high-usage OPDs, 4 moderate-usage OPDs, and 11 low-usage OPDs. Accordingly, the resulting clusters may contribute to the Solok City Communication and Informatics Office in formulating a more objective and proportional bandwidth-allocation policy that reflects the network-usage characteristics of each OPD.

References

[1] M. S. J. Sangaji and J. Irianto, "Transformasi Inovasi Pelayanan Publik menuju Pemerintahan Digital," *Jejaring Administrasi Publik*, vol. 17, no. 1, pp. 54–70, Jun. 2025, doi: 10.20473/jap.v17i1.72708.

[2] D. Mutiarin, Misran, N. S. Rukmana, and T. Ningtyas, "Bridging the digital divide through digital infrastructure," *Journal of Infrastructure, Policy and Development*, vol. 8, no. 8, 2024, doi: 10.24294/jipd.v8i8.6817.

[3] A. Haastrup, M. Zehri, D. Rincón, J. R. Piney, and A. Bazzi, "A Distance-Weighted Dynamic Bandwidth Allocation Algorithm for Improved Performance in Long-Reach Passive Optical Networks for Next Generation Networks," *Photonics*, vol. 10, no. 8, Aug. 2023, doi: 10.3390/photonics10080923.

[4] D. M. Amitu, R. N. Akol, and J. Serugunda, "Dynamic priority-based bandwidth allocation scheme for machine type communications," *EURASIP J. Wirel. Commun. Netw.*, vol. 2024, no. 1, Dec. 2024, doi: 10.1186/s13638-024-02400-5.

[5] R. Mubarak, S. Defit, and G. W. Nurcahyo, "Algoritma K-Means Clustering Penggunaan Bandwidth Internet (Studi Kasus di Pemerintah Daerah Kabupaten Padang Pariaman)," *Explore: Jurnal Sistem Informasi dan Telematika*, vol. 14, no. 1, p. 37, Jul. 2023, doi: 10.36448/jsit.v14i1.3037.

[6] A. Guérin, P. Chauvet, and F. Saubion, "A Survey on Recent Advances in Self-Organizing Maps," Dec. 2024, [Online]. Available: http://arxiv.org/abs/2501.08416

[7] S. Licen, A. Astel, and S. Tsakovski, "Self-organizing map algorithm for assessing spatial and temporal patterns of pollutants in environmental compartments: A review," Jun. 20, 2023, *Elsevier B.V.* doi: 10.1016/j.scitotenv.2023.163084.

[8] S. Rezvani and E. A. Jorswieck, "Optimal Water-Filling Algorithm in Downlink Multi-Cluster NOMA Systems," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.08636

[9] Y. Miftahuddin and A. R. S. Ridwan, "Application of Self-Organizing Map and K-Means to Cluster Bandwidth Usage Patterns in Campus Environment," *Jurnal Online Informatika*, vol. 10, no. 1, pp. 66–76, Apr. 2025, doi: 10.15575/join.v10i1.1438.

[10] M. Walid, N. L. N. Sahbaniya, H. Hozairi, F. Baskoro, and A. Y. Wijaya, "K-Means Clustering and Multilayer Perceptron for Categorizing Student Business Groups," *Knowledge Engineering and Data Science*, vol. 6, no. 1, p. 69, Jul. 2023, doi: 10.17977/um018v6i12023p69-78.

[11] Mei Yanda Maharani Messakh, Jusuf Ezra Rudolf Fransz, David Hermanus, Adriana Fanggidae, and Yulianto Triwahyuadi Polly, "Clustering Dataset Numerik Multivariat menggunakan Algoritma K-Means dengan Evaluasi Silhouette dan Davies–Bouldin Index," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 6, no. 1, pp. 356–368, May 2026, doi: 10.55606/jutiti.v6i1.6748.

[12] N. Aminudin, N. Hidayat, D. Feriyanto, D. Septasari, and I. Awaliyani, "Digital Landscape and Behavior in Indonesia 2024: A National Survey Analysis of Internet Penetration, Cybersecurity Risks, and User Segmentation Using K-Means Clustering and Logistic Regression," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3336–3351, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5117.

[13] Y. Seo, D. Lim, J. Lee, J. So, and H. Kim, "A Framework for Evaluating Driving Patterns of Automated Vehicle Based on On-Road Driving Records," *IEEE Access*, vol. 13, pp. 67894–67907, 2025, doi: 10.1109/ACCESS.2025.3556299.

[14] I. Panitsas and L. Tassiulas, "A Family of Open Time-Series Foundation Models for the Radio Access Network," Apr. 2026, [Online]. Available: http://arxiv.org/abs/2604.04271

[15] N. Suarna and Y. Arie Wijaya, "ANALISA PENERAPAN METODE CLUSTERING K-MEANS UNTUK PENGELOMPOKAN DATA TRANSAKSI KONSUMEN (Studi Kasus: Cv. Mitra Indexindo Pratama)," 2023.

[16] A. Brijith, "Data Preprocessing for Machine Learning."

[17] M. Ariesta, I. Setiawan, P. Sistem, T. Informasi, I. Teknologi, and D. B. Semarang, "SISTEM PENGENALAN RETINA MENGGUNAKAN SELF ORGANIZING MAP UNTUK MENDETEKSI RETINOPATI DIABETIKA," 2023.

[18] J. Homepage and S. Aziz, "IJIRSE: Indonesian Journal of Informatic Research and Software Engineering Implementation of Self Organizing Map Algorithm to Identify the Grouping Patterns of Family Welfare Level in Siak Regency Implementasi Algoritma Self Organizing Map untuk Identifikasi Pola Pengelompokan Tingkat Kesejahteraan Keluarga Kabupaten Siak," vol. 2, no. 2, 2022.

[19] A. E. Widjaja, A. Fransisko, C. A. Haryani, and Hery, "Text Mining Application with K-Means Clustering to Identify Sentiments and Popular Topics: a Case Study of the three Largest Online Marketplaces in Indonesia," *Journal of Applied Data Sciences*, vol. 4, no. 4, pp. 441–453, Dec. 2023, doi: 10.47738/jads.v4i4.134.

Biographies of Authors



Hikko Razaki was born in Solok City on November 19, 1989. He completed his bachelor's degree in Information Systems at STMIK Indonesia in 2019. His academic background has provided him with knowledge and competencies in information systems and information technology. He has a strong interest in the application and development of information technology, particularly in supporting data management and organizational needs. Hikko Razaki is known as a responsible, diligent individual with a strong commitment to continuous learning and professional development



Sarjon Defit was born in Padang Sibusuk on 7 August 1970. He is the Chancellor of Universitas Putra Indonesia YPTK Padang. He is currently active as a lecturer in Computer Science. He completed his undergraduate education at the College of Informatics and Computer Management (STMIK "YPTK" Padang), graduating in 1993. He completed his master's degree at Universiti Teknologi Malaysia, Johor Bahru, in 1998. He subsequently completed his doctoral education at Universiti Teknologi Malaysia, Johor Bahru, graduating in 2003. His fields of expertise include data mining, artificial intelligence, decision support systems, and other related areas. He can be contacted via email at sarjon_defit@upiyptk.ac.id



Rini Sovia Rini Sovia was born in Solok on April 5, 1976. She is a lecturer at Universitas Putra Indonesia YPTK Padang. She completed her undergraduate education in Informatics Management in 1999 and earned her master's degree in Informatics Engineering in 2006. She is currently pursuing a doctoral degree in Information Technology. Her areas of

	expertise include artificial intelligence (AI), expert systems (ES), data mining (DM), decision support systems (DSS), and databases. She can be reached via email at rini_sovia@upiyptk.ac.id