**JOIN (Jurnal Online Informatika)**
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 1 | June 2025: 66-76
DOI: 10.15575/join.v10i1.1438

# Application of Self-Organizing Map and K-Means to Cluster Bandwidth Usage Patterns in Campus Environment

**Yusup Miftahuddin[1], Abdur Rafi Syach Ridwan[2]**
[1,2]Department of Informatics, Faculty of Industrial Technology, Institut Teknologi Nasional Bandung, Indonesia

**Article Info**

***Article history:***

Received August 12, 2024
Revised January 05, 2025
Accepted February 06, 2025
Published April 01, 2025

***Keywords:***

Bandwidth
Clustering
Davies Bouldin Index
Self-Organizing Map
Silhouette Index

**ABSTRACT**

Unequal bandwidth distribution in campus environments often stems from a lack of understanding of WiFi usage patterns, as seen at Itenas Bandung. Here, bandwidth is allocated equally across all buildings, ignoring differences in demand, leading to inefficiencies in high-usage areas and poor money management due to unnecessary allocation of resources to low-demand buildings. This study aims to optimize bandwidth allocation by analyzing usage patterns using a combination of Self-Organizing Map (SOM) and K-Means clustering methods. SOM is used to group buildings into low, medium, and high bandwidth usage categories, while K-Means refines these clusters to enhance accuracy. The proposed approach demonstrated significant improvements in clustering quality, with the Silhouette Index increasing from 0.321 to 0.773 and the Davies-Bouldin Index dropping from 0.896 to 0.623 in the first test. Similar enhancements were observed in subsequent tests, highlighting the effectiveness of this method in addressing unequal bandwidth distribution. This research offers a practical solution for more efficient network and financial management in educational institutions.

***Corresponding Author:***

Yusup Miftahuddin,
Informatics Study Program, Faculty of Industry Technology, Institut Teknologi Nasional Bandung
Jl. PH. H. Mustofa No.23, Bandung 40124, West Java, Indonesia
Email: yusufm@itenas.ac.id

## 1. INTRODUCTION

The internet has become a basic need for private institutions, government agencies, companies, and especially in the world of education, especially in universities [1]. The implementation of WiFi (Wireless Fidelity) on campus as a supporting facility for internet users, such as students, lecturers, and teaching staff, is primarily aimed at providing optimal services to them. The internet makes it easier for students to complete assignments and find references for study [2].

Although the internet can be accessed from various places, there is often a disparity in access speed for each user on a network. This indicates the need for efficient network management, often referred to as bandwidth management [3]. The problem of unequal bandwidth distribution is also experienced by Institut Teknologi Nasional (Itenas) Bandung. Lack of understanding of how the academic community uses WiFi causes problems such as uneven bandwidth distribution at each hotspot point on 24 buildings. Currently, bandwidth is shared equally across all buildings in Itenas, without regard to buildings that should be prioritized. As a result, areas with higher internet needs are not prioritized, potentially hampering cost efficiency in network management.

A previous study aims to determine the level of bandwidth needs of regional organizations in the Purwakarta Regency Government. The study used the DBSCAN method to classify these needs. From

43 bandwidth distribution data, two clusters and one noise were formed. Cluster 1 consists of 15 data showing a low level of bandwidth requirements, cluster 2 consists of 21 data showing a medium level of bandwidth requirements, and noise consists of 15 data showing a high level of bandwidth requirements. The results of this study can contribute to the Purwakarta Regency Government in an effort to equalize the distribution of bandwidth in each regional apparatus organization [4].

Additionally, other studies have demonstrated that the K-Means clustering method can aid companies in decision-making to increase bandwidth for their customers. By applying the K-Means algorithm in data mining, companies can identify the bandwidth improvement potential of FTTH Broadband customers. This algorithm clusters data by identifying similarities, making it easier to determine potential clusters. A study analyzing 263 FTTH Broadband customers identified five groups: 34 customers (12.92%) were categorized as highly potential, 29 (11.02%) as potential, 56 (21.30%) as moderately potential, 54 (20.53%) as less potential, and 90 (34.22%) as not potential at all [5].

Clustering methods, such as the Self-Organizing Map (SOM) algorithm, can cluster student's bandwidth usage patterns. SOM maps datasets objectively and reduces complex data into an easily understandable two-dimensional representation [6]. A previous study used SOM to optimize clustering for family welfare in Siak Regency's Social Service. Clustering experiments with various combinations of clusters (3, 4, 5), learning rates (0.05, 0.10, 0.15, 0.20), and iterations (500, 750, 1000) showed the optimal cluster, based on Davies Bouldin Index (DBI) validation, was with 3 clusters, a learning rate of 0.20, and 500 iterations, resulting in a DBI of 0.940. The highest average DBI was at 1000 iterations with a value of 0.986, while the best iteration was 500 with the lowest DBI of 0.940 [7].

Based on the results of previous research, this research cluster bandwidth usage patterns. The aim is to provide information regarding bandwidth usage patterns at various campus locations, so as to identify areas with high and low usage levels. With this data, Itenas can optimize bandwidth distribution and improve user experience across campus. It is hoped that the results of this study will help make better decisions regarding network infrastructure and WiFi usage policies on campus.

## 2. METHOD

This research focuses on clustering bandwidth usage in a campus environment. Initially, bandwidth usage data is collected and clustered using the SOM method to objectively and unbiasedly identify hidden usage patterns. The SOM clustering results are then optimized using the K-Means method to produce more structured and interpretable bandwidth usage groups. The validity of each group is evaluated using the Silhouette Index (SI) and Davies Bouldin Index (DBI). SI measures how similar a data point is to its own cluster compared to other clusters, while DBI measures the average ratio of within-cluster distance to between-cluster distance. These indices ensure the clustering accurately reflects existing bandwidth usage patterns, not random or biased results. The research steps are illustrated in the block diagram in Figure 1.

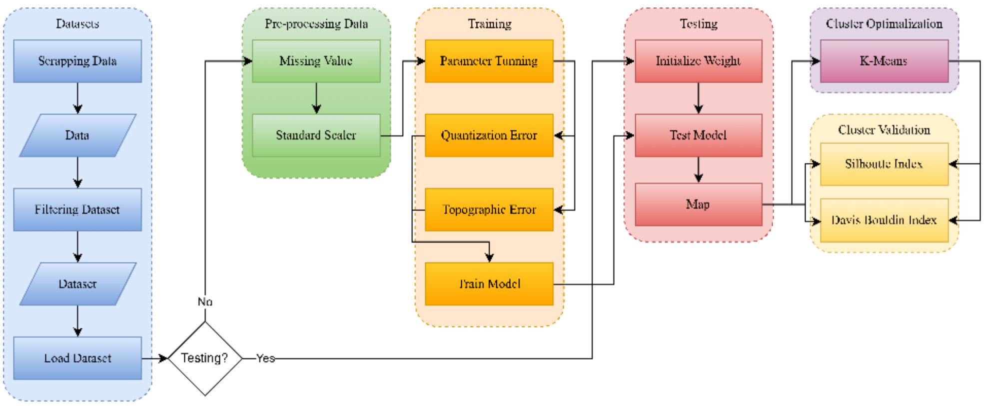


Figure 1 Research Method

### 2.1. Dataset Description

Two characteristics of bandwidth usage—the quantity of data downloaded and uploaded by the academic community on campus—are included in the tabular data used in this study. The source of this data is UPT - TIK Itenas. Structured data that is arranged in rows and columns, much to a spreadsheet, is called tabular data. A sample or observation is shown by each row, while a feature or variable is shown by each column. [8]. Table 1 shows an example of the dataset that will be used.

Table 1 Bandwidth Sample Dataset

| No | Location | Timestamp | Download | Upload |
|---|---|---|---|---|
| 141 | GD 1 BKA | 27/05/2024 07:00 | 156288 | 389328 |
| 4633 | GD 1 DI | 27/05/2024 07:00 | 360984 | 2845256 |
| ... | ... | ... | ... | ... |
| 90752 | YAYASAN | 07/06/2024 16:57 | 2437768 | 21032504 |
| 90753 | YAYASAN | 07/06/2024 17:00 | 1106680 | 16089088 |

## 2.2. Normalization Data

Data preprocessing is essential in machine learning to prepare data for analysis. It starts with cleaning missing values to ensure accuracy, followed by normalization to put everything on the same scale for better comparisons. Since raw data is rarely ready to use, these steps help make it reliable and suitable for different algorithm [9]. Table 2 displays the dataset after normalization process.

Table 2 Dataset after Normalization Data

| No | Location | Timestamp | Download | Upload |
|---|---|---|---|---|
| 141 | GD 1 BKA | 27/05/2024 07:00 | -0,21324055 | -0,57425269 |
| 142 | GD 1 BKA | 27/05/2024 07:03 | -0,20084279 | -0,62152673 |
| ... | ... | ... | ... | ... |
| 90752 | YAYASAN | 07/06/2024 16:57 | 0,461502722 | 2,438783926 |
| 90753 | YAYASAN | 07/06/2024 17:00 | 0,067836036 | 1,71725283 |

Based on the results, no missing data was identified, so no values were set to 0. This dataset is now ready to be used in the training process with various parameter combinations.

## 2.3. Self-Organizing Map

The Self-Organizing Map (SOM) is an unsupervised neural network that groups similar data points into clusters [10]. It has two layers: an input layer for the data and an output layer for the clusters, as shown in Figure 2. Neurons in both layers are connected by weights that adjust during training. SOM finds the best match for each data point by calculating the Euclidean distance [11]. Euclidean distance measures the distance between two points in space, based on the relationship between angles and distances. It works like the Pythagorean theorem in mathematics [12].

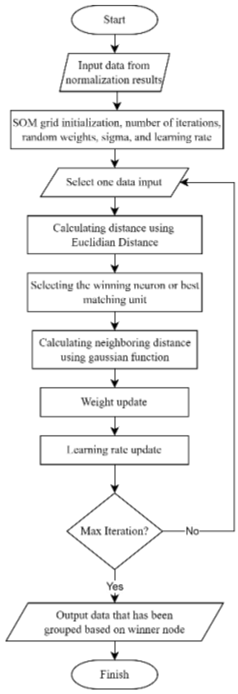


Figure 2 Flowchart of Self-Organizing Map

## 2.4. K-Means

K-Means is a clustering method that organizes data into a specific number of groups (K), initially selected at random. The "means" are the centroids, which act as the center of each cluster. For each data point, the algorithm measures its distance to all centroids using the Euclidean formula and assigns it to the nearest cluster [13][14]. The process is illustrated in Figure 3 as a flowchart.

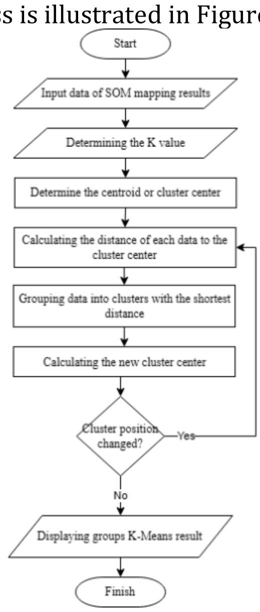


Figure 3 Flowchart of K-Means

## 2.5. Cluster Validation

Cluster validation helps us check how well a clustering algorithm has grouped the data. Two popular methods for this are the Silhouette Index and the Davies-Bouldin Index. The Silhouette Index measures how well each data point fits into its assigned cluster. It gives a score between -1 and 1, where higher scores mean better clustering [15][16][17]. The process is depicted in Figure 4 as a flowchart.

The Davies-Bouldin Index evaluates how distinct the clusters are and how close data points are to their cluster centers, where lower scores indicate better clustering [18][19]. Figure 5 presents this process in a flowchart format. [20]

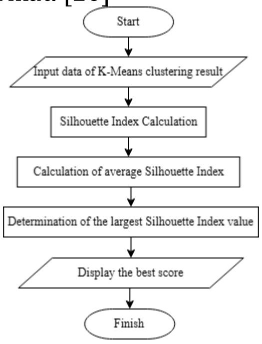


Figure 4 Flowchart of Silhouette Index

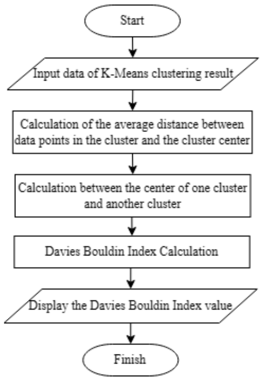


Figure 5 Flowchart of Davies Bouldin Index

## 3. RESULT AND DISCUSSION

### 3.1 Training Process

During the training process, an optimal combination of parameters is sought to create a Self-Organizing Map (SOM) model for testing. The parameters include SOM shape, input len, sigma, learning rate, neighbor function, activation distance, random seed, and epoch. Table 3 shows these parameters used during the training process.

Table 3 Parameters Used for Training

| Parameter | Value | Description |
|---|---|---|
| som shape | 10x10, 15x15, 20x20, 25x25, 30x30, 20x10, 25x15, 35x35, 40x40, 90x90 | Dimension of SOM grid (X, Y) |
| input len | 2 | Number of the elements of the vectors in input |
| sigma | 0.5, 1.0, 1.5, 2.0, 0.75, 0.3, 0.1, 0.01 | Spread of the neighborhood function |
| learning rate | 0.1, 0.5, 1.0, 2.0, 0.01 | 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 neurons |
| random seed | 42 | Random generator number |

The goal of the training process is to find the smallest quantization error and topography error values, ideally close to 0. As a result, 130 different parameter combinations were evaluated to build the Self-Organizing Map model. From these 130 combinations, the 3 best parameter combinations were taken based on the smallest quantization error, the smallest topographic error, and the combination of the smallest quantization and topographic errors. Table 4 shows the 3 best parameter combinations.

Table 4 The Best Parameter Combinations

| som shape | input len | sigma | learning rate | neighborhood func | activation dist | random seed | epoch | q error | t error |
|---|---|---|---|---|---|---|---|---|---|
| 90x90 | 2 | 0,01 | 0,1 | Gaussian | Euclidean | 42 | 1000 | 0,012639 | 0,999316 |
| 10x10 | 2 | 2,0 | 0,1 | Gaussian | Euclidean | 42 | 1000 | 0,225549 | 0,065462 |
| 30x30 | 2 | 2,0 | 1,0 | Gaussian | Euclidean | 42 | 1000 | 0,143316 | 0,146389 |

During training to find the best parameter combination, graphs showing the quantization error and topographic error across epochs are generated for each combination. Figure 6 shows the error graph for the first combination, Figure 7 for the second, and Figure 8 for the third.

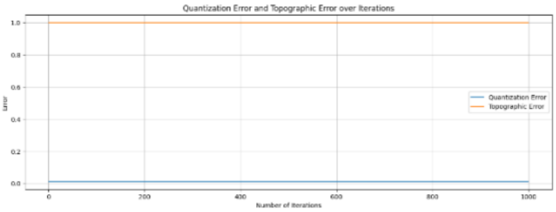


Figure 6 Error Graph of The First Parameter Combination

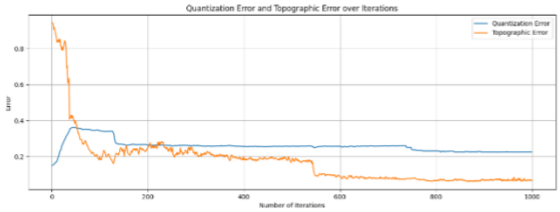


Figure 7 Error Graph of The Second Parameter Combination

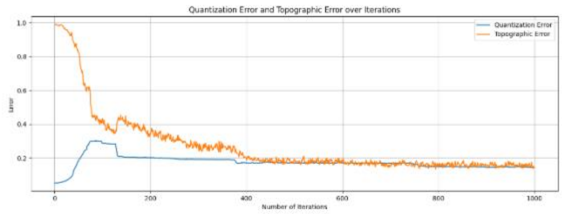


Figure 8 Error Graph of The Third Parameter Combination

The quantization and topographic error values show that the SOM shape, sigma, and learning rate have a big impact on the results. A larger SOM shape with a very small sigma and learning rate tends to reduce quantization error. On the other hand, a smaller SOM shape with a larger sigma and a small learning rate generally leads to a lower topographic error.

### 3.2 Testing Process

In the testing phase, the three parameter combinations are applied to the Self-Organizing Map (SOM) model, followed by clustering optimization using the K-Means method. This process categorizes the usage of locations or buildings at Itenas Bandung into three levels: low, medium, or high. The combination with the smallest quantization error is tested first, and with these parameters, the SOM model successfully maps the dataset into 8,100 nodes, as shown in Figure 9.

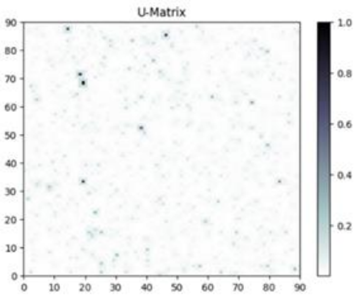


Figure 9 SOM U-Matrix Based on Smallest Quantization Error

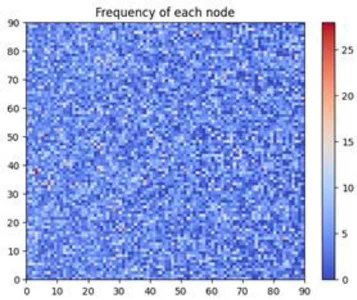


Figure 10 Data Distribution on SOM Model Based on Smallest Quantization Error

As shown in Figure 9, the SOM model effectively mapped the data, with light colors dominating the U-Matrix, indicating strong similarities among data points. Figure 10 supports this, showing evenly distributed data without concentration in specific nodes. This is due to the large grid size (90 x 90), which allows for precise mapping of data characteristics. Clustering was then applied to categorize each building's bandwidth usage by averaging download and upload values per node. The clustering results and detailed bandwidth usage for each building are shown in Figure 8 and Table 5.

p-ISSN: 2528-1682
e-ISSN: 2527-9165


Figure 11 SOM Clustering Results Based on the Smallest Quantization Error Value

Table 5 Description of Bandwidth Usage by SOM Based on the Smallest Quantization Error Value

| Category | Location |
|---|---|
| Low | GD 14 FTSP, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL |
| Medium | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 1 DP, GD 10 TI, GD 12 SIPIL, GD 14 FTI, GD 18 GEODESI, GD 9 PERPUS |
| High | GD 11 MESIN, GD 14 FAD, GD 15 BKU, YAYASAN |

As shown in Figure 9, the SOM model effectively mapped the data, with light colors dominating the U-Matrix, indicating strong similarities among data points. Figure 10 supports this, showing evenly distributed data without concentration in specific nodes. This is due to the large grid size (90 x 90), which allows for precise mapping of data characteristics. Clustering was then applied to categorize each building's bandwidth usage by averaging download and upload values per node. The clustering results and detailed bandwidth usage for each building are shown in Figure 8 and Table 5.

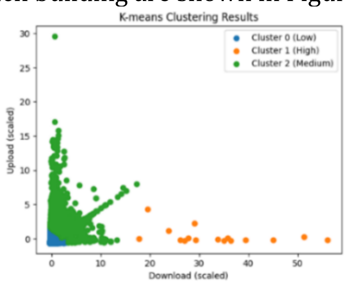


Figure 12 K - Means Clustering Results Based on the Smallest Quantization Error Value

After applying the K-Means method, the evaluation improved significantly, with the Silhouette Index increasing to 0.773 and the Davies Bouldin Index decreasing to 0.623. This improvement is due to K-Means grouping the SOM results into 3 focused clusters, aligning data more accurately with node characteristics. Unlike SOM, which categorizes buildings based on average node values, K-Means determines usage categories by averaging download and upload values within each cluster, as shown in Table 6. This approach results in more precise and meaningful groupings.

Table 6 Description of Bandwidth Usage by K - Means Based on the Smallest Quantization Error Value

| Category | Location |
|---|---|
| Low | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 1 DP, GD 10 TI, GD 11 MESIN, GD 12 SIPIL, GD 14 FAD, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL, GD 9 PERPUS |
| Medium | GD 15 BKU, YAYASAN |
| High | - |

Furthermore, the second parameter combination, which is based on the smallest topography value, is also applied. Figures 13, 12, and 15 show the U-Matrix, data distribution, and clustering results of the Self-Organizing Map test with the second parameter combination.

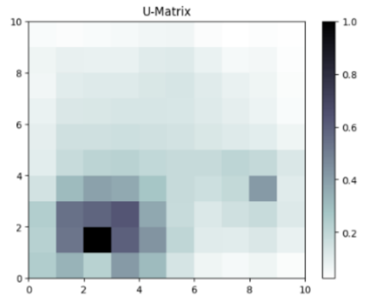

Figure 13 SOM U-Matrix Based on Smallest Topographic Error

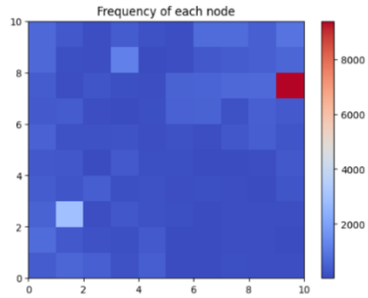

Figure 14 Data Distribution on SOM Model Based on Smallest Topographic Error

Figure 15 SOM Clustering Results Based on the Smallest Topographic Error Value

With the second parameter combination, as shown in Figures 13 and 14, the SOM mapping results are less effective. The U-Matrix is dominated by dark colors, with few light areas, due to the small SOM shape of 10 x 10, resulting in only 100 nodes. This limited size prevents the dataset from being properly distributed across the nodes, leading to inaccurate data representation. Many nodes contain thousands of data points with different characteristics, causing dissimilar data to be grouped in the same node.

Cluster evaluation for this parameter combination yields a Silhouette Index of 0.303 and a Davies-Bouldin Index (DBI) of 0.996. Since DBI values close to 1 indicate poor clustering quality, the results are suboptimal. However, despite these limitations, the SOM model still provides valuable usage information for each building in the dataset, as detailed in Table 7.

Table 7 Description of Bandwidth Usage by SOM Based on the Smallest Topographic Error Value

| **Category** | **Location** |
|---|---|
| Low | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 10 TI, GD 12 SIPIL, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL |
| Medium | - |
| High | GD 1 DP, GD 11 MESIN, GD 14 FAD, GD 9 PERPUS, GD 15 BKU, YAYASAN |

The SOM model results from this test will be clustered using the K-Means method, which will divide the data into 3 groups based on the winning nodes. Figures 16 and Table 8 show the clustering graph and usage description after the application of K-Means.

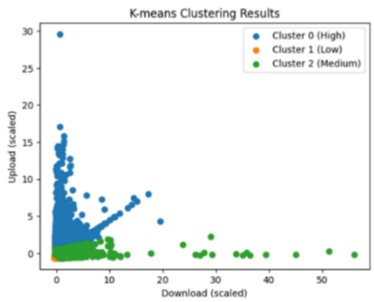

Figure 16 K-Means Clustering Results Based on the Smallest Topographic Error Value

Table 8 Description of Bandwidth Usage by K - Means Based on the Smallest Topographic Error Value

| Category | Location |
|---|---|
| Low | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 10 TI, GD 12 SIPIL, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL, GD 9 PERPUS |
| Medium | GD 1 DP, GD 14 FAD, |
| High | GD 11 MESIN, GD 15 BKU, YAYASAN |

After applying K-Means, the Silhouette Index value increases to 0.684 and the Davies-Bouldin Index to 0.659. Although the bandwidth usage information changes, the SOM model integrated with K-Means can still provide clustering results.

In the last test, the third parameter is applied, which is the combination with the smallest quantization error and topography error values. The test results with these parameters are shown in Figures 17, 18, and 19, which show the U-Matrix, data distribution, and clustering results.

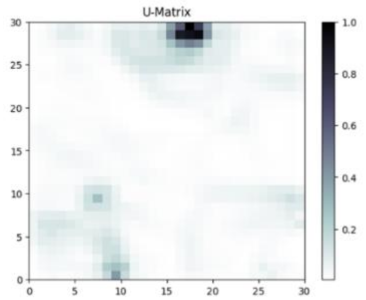


Figure 17 U-Matrix SOM Based on the Smallest Quantization and Topographic Error

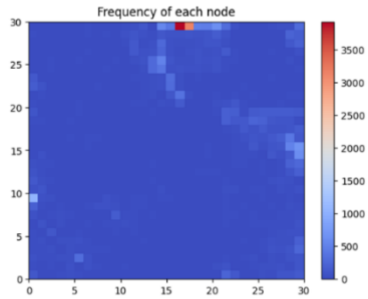


Figure 18 Data Distribution on SOM Model Based on Smallest Quantization and Topographic Error


Figure 19 SOM Clustering Results Based on the Smallest Quantization and Topographic Error Value

The results improved SOM mapping, with a more evenly distributed U-Matrix. The 30 x 30 SOM size, resulting in 900 nodes, enhanced clustering by similar characteristics. Cluster evaluation showed a Silhouette Index of 0.302 and a Davies-Bouldin Index of 0.998, indicating that the model was still able to determine the category of use, as shown in Table 9.

Table 9 Description of Bandwidth Usage by SOM Based on the Smallest Quantization and Topographic Error Value

| Category | Location |
|---|---|
| Low | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 10 TI, GD 12 SIPIL, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL |
| Medium | - |
| High | GD 1 DP, GD 11 MESIN, GD 14 FAD, GD 9 PERPUS, GD 15 BKU, YAYASAN |

To optimize the clustering results of the SOM model, the K-Means method is applied by dividing the data into 3 groups. Figures 20 and Table 10 show the clustering graph and usage description after the application of K-Means.

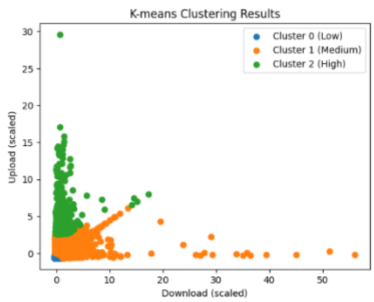


Figure 20 K - Means Clustering Results Based on the Smallest Quantization and Topographic Error Value

Table 10 Description of Bandwidth Usage by K - Means Based on the Smallest Quantization and Topographic Error Value

| **Category** | **Location** |
|---|---|
| Low | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 1 DP, GD 10 TI, GD 12 SIPIL, GD 14 FAD, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL, GD 9 PERPUS |
| Medium | GD 11 MESIN, GD 14 FAD, GD 15 BKU, YAYASAN |
| High | - |

After applying K-Means to the SOM model in the third test, the Silhouette Index increased to 0.598 and the Davies Bouldin Index to 0.848, proving that K-Means improves SOM clustering results. Table 11 presents information on buildings that may require increased bandwidth capacity, aiding campus network infrastructure decisions.

Table 11 Bandwidth Upgrade Potential Information

| **Category** | **Location** |
|---|---|
| Not Considered | GD 1 BKA, GD 1 DI, GD 1 DKV, GD 10 TI, GD 12 SIPIL, GD 14 FAD, GD 14 FTI GD 14 FTSP, GD 18 GEODESI, GD 19 KIMIA, GD 20 ELEKTRO, GD 3 PWK, GD 4 SI, GD 8 TL, GD 9 PERPUS |
| Considered | GD 1 DP |
| Highly Considered | GD 11 MESIN, GD 14 FAD, GD 15 BKU, YAYASAN |

The consideration of increased bandwidth capacity is based on K-Means test results, validated by increased Silhouette and Davies-Bouldin Index values. The assessment uses the mode of the three tests to categorize each building's bandwidth utilization: low (not considered), medium (considered), and high (highly considered). The results of this study aim to help network administrators at Itenas make better decisions about network infrastructure, especially WiFi, by considering actual bandwidth usage patterns. Buildings with high bandwidth usage can be prioritized for capacity upgrades to meet demand, while those with low usage may have their allocation adjusted to reduce unnecessary costs and optimize overall spending on internet services.

## 4. CONCLUSION

This research demonstrates that the combination of Self-Organizing Map (SOM) and K-Means effectively resolves the issue of unequal bandwidth distribution at Itenas Bandung. By grouping buildings into low, medium, and high usage categories, the method enables a more efficient allocation of resources. The integration of K-Means significantly enhances SOM's clustering accuracy, with improvements in the Silhouette Index of up to 264.1% (from 0.212 to 0.773) and reductions in the Davies-Bouldin Index by 30.4% (from 0.897 to 0.624) in the first test. These findings highlight the efficiency and reliability of the proposed approach in analyzing bandwidth usage patterns. Future studies could refine this method further by incorporating additional features, such as time-based usage trends or real-time clustering, to improve its scalability and adaptability in dynamic environments.

## ACKNOWLEDGEMENTS

This study is funded by the Informatics Study Program at Institut Teknologi Nasional Bandung, which has provided us the opportunity to carry out this research.

## REFERENCES


[1] Irwansyah, H. Yudiastuti, and F. Panjaitan, "Analisis dan Monitoring WIFI pada Universitas Islam Negeri Palembang," *Journal of Information Technology Ampera*, vol. 3, no. 2, pp. 277–290, 2022, [Online]. Available: https://journal-computing.org/index.php/journal-ita/index

[2] M. Saski, I. Iskandar, Novriyanto, and Pizaini, "Optimasi Kualitas Jaringan WIFI Fakultas Melalui Redesain Topologi Dengan Menggunakan Network Simulator 2," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 2, pp. 1032–1041, 2023, doi: 10.30865/klik.v4i2.1272.

[3] M. G. Nasrullah, N. Heryana, and A. Solehudin, "Manajemen Bandwith Menggunakan Metode Hirarchical Token Bucket Pada Pembatasan Kecepatan Internet (Studi Kasus : Kantor Kelurahan Ciparage Jaya)," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2291–2296, 2024.

[4] T. I. Hermanto and Y. Muhyidin, "Analisis Data Sebaran Bandwidth Menggunakan Algoritma Dbscan Untuk Menentukan Tingkat Kebutuhan Bandwidth Di Kabupaten Purwakarta," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 5, no. 2, pp. 130–137, Jul. 2020, doi: 10.36341/rabit.v5i2.1388.

[5] S. A. Arnomo and Y. Yulia, "Clustering the potential bandwidth upgrade of FTTH broadband subscribers," *ILKOM Jurnal Ilmiah*, vol. 13, no. 1, pp. 51–57, Apr. 2021, doi: 10.33096/ilkom.v13i1.805.51-57.

[6] R. N. Zulfahmi, M. Kristiana Daul, M. Al Ayyubi, I. Wayan, Pradnyana, and R. Dwi Bekti, "Pemetaan Kerentanan Tingkat Kriminalitas Menggunakan Metode Self Organizing Map," *Media Cetak*, vol. 2, no. 5, pp. 872–881, 2023, doi: 10.55123/insologi.v2i5.2566.

[7] S. Aziz and Mustakim, "Implementasi Algoritma Self Organizing Map untuk Identifikasi Pola Pengelompokan Tingkat Kesejahteraan Keluarga Kabupaten Siak," *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, vol. 2, no. 2, pp. 82–92, 2022.

[8] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, May 2022, doi: 10.1016/j.inffus.2021.11.011.

[9] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 1, pp. 117–122, 2020.

[10] A. Kharunnisa and B. Harianja, "Klasifikasi Kompetensi Tenaga Kerja Program Magang Ke Luar Negeri Dengan Metode Self Organizing Maps (SOM)," *Journal of Machine Learning and Data Analytics (MALDA)*, vol. 01, no. 01, 2022.

[11] A. U. Khasanah, "A Comparison Study: Clustering using Self-Organizing Map and K-means Algorithm," *Jurnal Universitas Sebelas Maret*, vol. 15, no. 1, pp. 51–58, 2016.

[12] Y. Miftahuddin, S. Umaroh, and F. R. Karim, "Perbandingan Metode Perhitungan Jarak Euclidean, Haversine, Dan Manhattan Dalam Penentuan Posisi Karyawan," *Jurnal Tekno Insentif*, vol. 14, no. 2, pp. 69–77, Aug. 2020, doi: 10.36787/jti.v14i2.270.

[13] I. Ahmad Zaki and I. A. Sembe, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," *Journal of Mathematics, Computations, and Statistics*, vol. 5, no. 2, pp. 163–176, 2022, [Online]. Available: http://www.ojs.unm.ac.id/jmathcos

[14] Zubair, M., Iqbal, M. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2024). An improved K-means clustering algorithm towards an efficient data-driven modeling. *Annals of Data Science*, *11*(5), 1525-1544.

[15] N. Ulinnuha, "Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means," *ITSMART: Jurnal Ilmiah Teknologi dan Informasi*, vol. 9, no. 1, pp. 8–12, 2020.

[16] Juanita, S., & Cahyono, R. D. (2024). K-means clustering with comparison of Elbow and silhouette methods for medicines clustering based on user reviews. *Jurnal Teknik Informatika (JUTIF)*, *5*(1), 283-289.

[17] Jatmiko, A. R., Hendrawan, N. D., Arief, R. M., Putra, F. A. I. A., & Karyudi, M. D. P. (2023). Signifikansi pengaruh akses teknologi informasi terhadap indeks pembangunan manusia di Indonesia. *JITEKH: Jurnal Ilmiah Teknologi Harapan*, *11*(2), 83-94.

[18] Dinata, R. K., & Retno, S. (2024). Optimizing the Evaluation of K-means Clustering Using the Weight Product. *Revue d'Intelligence Artificielle*, *38*(4).

[19] Muningsih, E., Kesuma, C., Sunanto, S., Suripah, S., & Widayanto, A. (2023, May). Combination of K-Means method with Davies Bouldin index and decision tree method with parameter optimization for best performance. In *AIP Conference Proceedings* (Vol. 2714, No. 1). AIP Publishing.

[20] I. T. Umagapi, B. Umaternate, S. Komputer, P. Pasca Sarjana Universitas Handayani, B. Kepegawaian Daerah Kabupaten Pulau Morotai, and B. Riset dan Inovasi, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Seminar Nasional Sistem Informasi dan Teknologi (SISFOTEK)*, vol. 7, no. 1, pp. 303–308, 2023.