JITEEHA: Journal of Information Technology Application in Education, Economy, Health and Agriculture
Vol. 02, No. 02, June 2025, pp. 48~51, e-ISSN: 3090-3939
https://journal.iteeacademy.org/index.php/jiteeha/

Clustering Wi-Fi Users Based on Activity Patterns Using K-Means Algorithm

Rini Agustina[1], Ario Fajar Dharmawan[2], Sandra Meylina Alaka Putri[3], Yank Rizky Kharisma Putra[4]

[1]Department of Information System, Universitas PGRI Kanjuruhan, Malang, Indonesia
[2,3,4] Department of Informatics Engineering, University of Widyagama Malang, Jl. Borobudur No. 35 Malang, Indonesia

Article Info

Article history:

Received March 30, 2025
Revised May 20, 2025
Accepted June 25, 2025

Keywords:

Wi-Fi
Activity pattern
Clustering
K-Means
Network management

ABSTRACT

The rapid development of Wi-Fi networks has become a key pillar in supporting the internet needs of modern society. However, the increasing number of users and their diverse activity patterns pose challenges in network management, especially with regard to bandwidth allocation and service quality. Variations in activity patterns, such as social media, streaming, and gaming, create different bandwidth requirements for each user group. This imbalance in resource utilization can result in degraded quality of service, especially during peak hours. This research aims to address these challenges by clustering Wi-Fi users based on their activity patterns using the K-Means algorithm. The data used includes access time, usage duration, connection intensity, and user activity type. After going through the analysis process, users are grouped into several clusters based on the similarity of activity patterns. The clustering results show significant differences between light, medium, and heavy users in bandwidth consumption and duration of use. The results of this study contribute to more efficient Wi-Fi network management, especially in optimizing bandwidth allocation and supporting data-based decision-making. With customized management strategies for each user group, the quality of service can be significantly improved, providing a better experience for Wi-Fi users.

Corresponding Author:

Ario Fajar Dharmawan
Department of Informatics Engineering
Faculty of Engineering, University of Widyagama Malang
Jl. Borobudur No. 35 Malang, East Java
Email info@widyagama.ac.id

1. INTRODUCTION

The rapid development of information technology has made the internet an integral part of everyday life. One of the main media for accessing the internet is Wi-Fi networks, which are widely used in various environments, such as campuses, offices, shopping centers, and residential areas. Ease of access and connection speed are the main requirements for users. However, as the number of users increases and their activity patterns diversify, the challenges in managing Wi-Fi networks are increasingly complex .[1] The activity patterns of Wi-Fi users can vary widely, from accessing social media, streaming videos, playing online games, to downloading or uploading large files. These variations create different bandwidth requirements for each user group. If not managed properly, this can cause an imbalance in network usage, which results in a decrease in service quality, especially during peak usage times .[2]

Another problem is the inability of network managers to monitor usage in real-time. This makes it difficult to know which users require more bandwidth or who tend to consume more resources than they should[3] . This imbalance can disrupt the experience of other users, which in turn affects the overall quality of service.

To overcome these challenges, analysis is needed that can provide insight into user activity patterns. One method that can be used is clustering with the K-Means algorithm, which is one of the techniques in data mining[4] . This algorithm is effective in clustering data based on certain characteristics, such as connection duration, data volume, access frequency, dominant access time, and access location. By utilizing the K-Means algorithm, service providers can identify groups of users with similar activity patterns and design more targeted network management strategies .[5] This research aims to cluster Wi-Fi users based on their activity patterns using the K-Means algorithm. The results of this research are expected to provide solutions in managing Wi-Fi networks more efficiently, improving user experience, and supporting decision-making in the development of need-based internet services. Through this research, it is expected that network managers can make better decisions in resource distribution, so that service quality is maintained and users are satisfied.

2. METHOD

This study was conducted to cluster Wi-Fi network users based on their activity patterns using the K-Means algorithm. The data used includes connection duration, data volume used, access frequency, dominant access time, and access location. The research process involved the following stages:

1. Data Collection: Data was collected from 10 Wi-Fi network users with the following attributes:
   - Connection Duration: Average connection time in hours.
   - Data Volume: Total data used in GB units.
   - Access Frequency: The number of accesses in a given period.
   - Dominant Access Time: Most access time (morning, afternoon, evening, night).
   - Access Location: The dominant location where users access the network.
2. Data Normalization: The data is normalized using the min-max normalization method to ensure that all attributes are in the range [0, 1], so that each attribute has equal weight in the distance calculation.
3. Initial Centroid Determination: Three initial centroids are randomly selected from the normalized dataset:
   - C1: [0.04, 0.07, 0.14, 0.33, 0.25] (based on Data 2).
   - C2: [1.00, 1.00, 1.00, 0.67, 0.50] (based on Data 3).
   - C3: [0.21, 0.24, 0.43, 1.00, 1.00] (based on Data 6).
4. Euclidean Distance Calculation: The distance of each data to the centroid is calculated using the formula:

   $$d = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2}$$

   Where is the i-th data attribute value and is the i-th centroid attribute value.

5. Cluster Determination: Each data is grouped into the cluster with the shortest distance to the centroid.
6. Calculation of New Centroid: The new centroid is calculated as the average of all the data in each cluster.
7. Iteration: Steps 4 to 6 are repeated until the centroid position does not change significantly.

3. RESULTS AND DISCUSSION

The following is a detailed K-Means clustering calculation, explained in a simple way.

3.1. Initial Data

The data to be used as input for clustering is as follows:

ID	Connection Duration	Data Volume	Access Frequency	Dominant Access Time	Access Location
1	8.5	3.2	5	3 (Night)	1 (Campus Area)
2	1.0	0.8	2	2 (Afternoon)	2 (Cafe)
3	12.0	7.5	8	3 (Night)	3 (House)

4	0.5	0.3	1	1 (Morning)	4 (Office)
5	5.0	1.2	3	2 (Afternoon)	1 (Campus Area)
6	3.0	2.0	4	4 (Afternoon)	5 (Public Park)
7	9.0	5.0	6	3 (Night)	3 (House)
8	2.5	1.0	2	2 (Afternoon)	2 (Cafe)
9	6.0	2.8	5	4 (Afternoon)	1 (Campus Area)
10	4.0	1.5	3	1 (Morning)	4 (Office)

3.2. Data Normalization

To equalize the scale between features, the data is normalized using the Min-Max Scaling method:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Normalization result (example **Connection Duration** feature):

$$x'_{\text{Durasi}} = \frac{x - \min(x_{\text{Durasi}})}{\max(x_{\text{Durasi}}) - \min(x_{\text{Durasi}})}$$

After all the features are normalized, the data becomes like this:

ID	Connection Duration	Data Volume	Access Frequency	Dominant Access Time	Access Location
1	0.68	0.40	0.57	0.67	0.00
2	0.04	0.07	0.14	0.33	0.25
3	1.00	1.00	1.00	0.67	0.50
4	0.00	0.00	0.00	0.00	0.75
5	0.39	0.12	0.29	0.33	0.00
6	0.21	0.24	0.43	1.00	1.00
7	0.73	0.64	0.71	0.67	0.50
8	0.18	0.10	0.14	0.33	0.25
9	0.50	0.35	0.57	1.00	0.00
10	0.32	0.17	0.29	0.00	0.75

3.3. Centroid Initialization

Randomly, three data are selected as the initial centroids. For example:

- **Centroid 1 (C1):** Data ID 2 [0.04, 0.07, 0.14, 0.33, 0.25]
- **Centroid 2 (C2):** Data ID 3 [1.00, 1.00, 1.00, 0.67, 0.50]
- **Centroid 3 (C3):** Data ID 6 [0.21, 0.24, 0.43, 1.00, 1.00]

3.4. Data Grouping

Calculate the Euclidean distance between each data to the centroid, with the formula:

- Data 1 to C1: 0.9389, to C2: 0.9473, to C3: 1.1726
- Data 2 to C1: 0.0000, to C2: 1.6444, to C3: 1.0739
- Data 3 to C1: 1.7262, to C2: 0.0000, to C3: 1.2675

3.5. New Centroid Calculation

After all the data is clustered to the nearest centroid, calculate the average of each feature for the data in the cluster as the new centroid:

$$c' = \frac{\sum x}{N}$$

- C1: [0.196, 0.094, 0.172, 0.198, 0.370]
- C2: [0.865, 0.820, 0.855, 0.670, 0.500]
- C3: [0.465, 0.333, 0.570, 0.890, 0.333]

3.6. Clustering Result

After the iteration is complete, here are the results of grouping the data into three clusters:

- Low Activity Cluster: Data 2, Data 4, Data 5, Data 8, Data 10
- Medium Activity Cluster: Data 1, Data 9
- High Activity Cluster: Data 3, Data 6, Data 7

ID	Cluster
1	2
2	1
3	3
4	1
5	1
6	3
7	3
8	1
9	2
10	1

Clustering Wi-Fi users based on their activity patterns using the K-Means algorithm. From the results of the clustering process, users were successfully grouped into three main **clusters** based on different activity patterns**,** namely **light cluster, medium cluster,** and **heavy cluster**. Each cluster reflects the unique characteristics of users who have different levels of network usage.

1. **Lightweight Cluster**
   This cluster consists of users with short connection duration, low data volume, and infrequent access frequency. Most users in this cluster access Wi-Fi networks in locations such as cafes or offices with access times predominantly in the morning to afternoon. These characteristics indicate that users in the

light cluster tend to use Wi-Fi networks for simple activities, such as browsing or accessing social media for a short period of time.

2. **Medium Cluster**
   Users in this cluster have moderate activity patterns with moderate connection duration and data volume, and relatively regular access frequency. The dominant location of users in this cluster is the campus area, indicating that they use the Wi-Fi network for academic or work needs with moderate intensity. The dominant access time varies, ranging from noon to evening.
3. **Heavy Cluster**
   The heavy cluster consists of users with long connection duration, high data volume, and frequent access. The majority of users in this cluster access the Wi-Fi network from home, which shows intensive usage patterns such as streaming videos, playing online games, or downloading large files. Access times are predominantly at night, which is the peak time of network usage.

The clustering process starts with data normalization to ensure all attributes are in the same range, so that the weight of each attribute in the Euclidean distance calculation is balanced. After that, an initial centroid is randomly selected, and a distance calculation is performed to determine the initial cluster. This process is repeated several times until the centroid is stable and does not undergo significant changes .[6]

4. CONCLUSION

The conclusion of this study shows that clustering Wi-Fi users based on activity patterns using the K-Means algorithm can provide significant insights in network management. By grouping users into different clusters-light, medium, and heavy-service providers can more effectively allocate bandwidth and improve service quality. This research also highlights the importance of understanding variations in user activity patterns to avoid imbalances in resource usage, which can affect the overall user experience. The clustering results are expected to support data-driven decision-making in the development of better internet services.

ACKNOWLEDGEMENTS

The authors would like to express their deepest gratitude to Mrs. Fitri Marisa, S.Kom., M.Pd., Ph.D., for her invaluable guidance, encouragement, and support during the preparation of this research paper. Her critical feedback and insightful suggestions greatly enhanced the quality of this study. We also extend our heartfelt appreciation to the entire research team, namely Ario Fajar Dharmawan, Sandra Meylina Alaka Putri, and Yank Rizky Kharisma Putra, for their collaborative spirit, hard work, and commitment in conducting this research. Their dedication and teamwork were instrumental in achieving the research objectives.

REFERENCES

[1] J. Widianto, P. Putra, E. A. Suganda, and I. Purnamasari, "Journal of Informatics and Software Engineering Application of RapidMiner with the K-Means Method in Determining the Cluster of WIFI Network Interference Provider PT.XYZ in Karawang Area," *J. Inform. and Software Engineering*, vol. 4, no. `, pp. 31-35, 2022.

[2] A. P. Kuncoro, D. P. Hutomo, and M. Zulfadhilah, "Grouping Internet Users Using K-Means Method on Server Access Log Data," *Citisee 2017*, pp. 149-153, 2017.

[3] R. O. Felani, "Analysis of e-learning User Behavior using K-Means Clustering Algorithm," *Jusikom J. Sist. Comput. Musirawas*, vol. 7, no. 1, pp. 61-73, 2022, doi: 10.32767/jusikom.v7i1.1538.

[4] T. I. Saputra and R. Arianty, "Implementation of K-Means Clustering Algorithm on Sentiment Analysis of Indosat User Complaints," *J. Ilm. Inform. Comput.*, vol. 24, no. 3, pp. 191-198, 2019, doi: 10.35760/ik.2019.v24i3.2361.

[5] N. Syamsu and W. Al, "Application of k-means clustering algorithm on independent teaching user reviews in play store," 2024.

[6] E. Affandi, S. Syam, Y. Syahra, and A. Amrullah, "Data Mining in Grouping Indihome Customer Data Using the K-Means Clustering Method at PT.Telkom Akses," *J. Mantik*, vol. 4, no. 4, pp. 2604-2612, 2021, [Online]. Available: https://iocscience.org/ejournal/index.php/mantik