IJMR E-ISSN: 3109-9777
INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEACH

APPLICATION OF THE K-MEANS ALGORITHM FOR ANALYZING THE EFFECTIVENESS OF LOCAL GOVERNMENT WEBSITES: A CASE STUDY OF SOUTH SUMATRA PROVINCE

Naufal[1*], Bintang Adji Pratama[2], M Zaki Zain[3], Fenny Purwani[4]
[1] Universitas Islam Negeri Raden Fatah Palembang, Indonesia
[2] Universitas Islam Negeri Raden Fatah Palembang, Indonesia
[3] Universitas Islam Negeri Raden Fatah Palembang, Indonesia
[4] Universitas Islam Negeri Raden Fatah Palembang, Indonesia
Email : 23041450154@radenfatah.ac.id[1], bintangadjipratama_23041450157@radenfatah.ac.id[2], muhamadzakizain@gmail.com[3], fennypurwani_uin@radenfatah.ac.id[4]

E-ISSN : 3109-9777

Received: November 2025	Accepted: November 2025	Published: Desember 2025

Abstract :

The development of information technology has encouraged local governments to improve the quality of public services through digital media, one of which is the official government website. However, the effectiveness of local government websites still varies and not all of them provide an optimal user experience. This study aims to analyze the effectiveness of local government websites in South Sumatra Province using the K-Means clustering algorithm with the help of the RapidMiner Studio application. The data used is secondary data from SimilarWeb, covering four main indicators, namely total visits, average visit duration, pages per visit, and bounce rate. The results of the analysis show that local government websites are divided into three main clusters, namely effective, less effective, and ineffective. Effective websites have high visit rates, long visit durations, and low bounce rates, while ineffective websites have the lowest performance on all indicators. The variables of average visit duration and bounce rate proved to be the most dominant factors in determining website effectiveness. These findings are expected to serve as a basis for local governments in evaluating and optimizing their websites so that digital public services are more efficient, interactive, and responsive to the needs of the community.

Keyword: *K-Means Clustering, Website Effectiveness Analysis, Local Government Websites*

INTRODUCTION

The development of information technology in the digital era has brought major changes to government systems in various regions in Indonesia. The government is required to adapt to technological advances in an effort to improve efficiency, transparency, and the quality of public services. One form of technology application is through the official government website, which serves as a medium of communication and public service that can be accessed by the public anytime and anywhere (Khomarudin, 2003).

Local government websites are one of the main components of e-government implementation, which is expected to bring the government and the public closer together. Websites are not only a means of disseminating information, but also a medium for online public services, ranging from permit applications and tax information to the publication of government activities. However, research by Rahman et al. shows that although websites are available in various government agencies, their effectiveness is still low due to a lack of

content updates, minimal user interaction, and an unattractive interface(Gartika & Widiyanto, 2024).

South Sumatra Province is one of the regions that has implemented the Electronic-Based Government System (SPBE) in various sectors. A number of agencies under the South Sumatra Provincial Government have official websites such as sumselprov.go.id, bapenda.sumselprov.go.id, and diskominfo.sumselprov.go.id. However, the extent to which these websites are effective in attracting users and providing optimal information services is still unknown. Therefore, data-based analysis is needed to evaluate the effectiveness of government websites in the region

In this study, secondary traffic data from SimilarWeb was used, which includes several indicators such as total visits, bounce rate, and average visit duration. This data was used to describe the level of visits and user interaction with government websites. The analysis method applied is the K-Means Clustering algorithm, which is an unsupervised learning method that functions to group data based on similarity of characteristics. The K-Means algorithm works by finding the center point (centroid) of each data group and updating it iteratively until the most optimal grouping results are obtained.

Through the application of the K-Means algorithm, it is hoped that the effectiveness of the South Sumatra Provincial government website can be grouped into three categories, namely effective, quite effective, and less effective. The results of this study are expected to contribute to the South Sumatra Provincial Government in evaluating and improving the quality of regional agency websites to be more informative, responsive, and in line with the needs of the community.

This study aims to analyze the effectiveness level of local government websites in South Sumatra Province based on performance data obtained from SimilarWeb. Through a data mining approach, the K-Means clustering algorithm was applied to group these websites based on effectiveness indicators such as the number of visits (total visits), average visit duration, bounce rate, and number of pages visited (pages per visit). From the results of this grouping, this study also seeks to identify the characteristics of each group of websites and determine the factors that distinguish between websites with high, medium, and low effectiveness. Thus, this study is expected to provide an empirical picture of the effectiveness of local government websites and serve as a basis for efforts to improve the quality of public digital services in South Sumatra(Mulyani & Setiawan, 2024).

RESEACH METHOD

Data Collection Stage

The data used in this study is secondary data obtained from the SimilarWeb platform, which is one of the websites that provides global website performance statistics. SimilarWeb was chosen because it provides various quantitative indicators such as the number of visits, access time, and user interaction rates, which are relevant for assessing the effectiveness of a

website(Al-Kautsar Maktub et al., 2025).
Data Processing Stage The variables used in this study consist of four main indicators, namely:

- Total Visits – the total number of user visits during a certain period, describing the popularity and visibility of the website.
- Average Visit Duration – the average time spent by users when accessing the website, indicating the level of user interest and comfort.
- Pages per Visit – the average number of pages opened by users in a single visit session, reflecting the depth of information exploration.
- Bounce Rate – the percentage of users who leave the website after opening one page, which is an indicator of satisfaction and content relevance.

Clustering Stage

The clustering stage is the core of the analysis process in this study. After the website performance indicator data was normalized, a clustering process was carried out using the K-Means algorithm to identify patterns of effectiveness of local government websites in South Sumatra Province. This algorithm falls under the category of unsupervised learning, which aims to group data based on the level of similarity between attributes.The optimal number of clusters was determined using the Elbow method, which is a graphical approach to determine the point at which the decline in the Sum of Squared Error (SSE) value begins to slow down significantly. Based on the calculations using the RapidMiner Studio application, it was found that the optimal value was K = 3, so the data was grouped into three categories, namely effective, quite effective, and less effective (Emigawaty et al., 2023).

Next, the system randomly initializes the initial centroids from the data set. Each centroid represents the center of a data group (cluster). The distance between each data point and the centroid is calculated using the Euclidean Distance formula as shown in Equation (1):

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n}(x_{ik} - c_{jk})^2}$$

Explanation:
$d(x_i, c_j)$= distance between data point i and centroid j
x_{ik}= the value of the kth attribute of the ith data point
c_{jk}= the value of the kth attribute of the jth centroid
n= number of attributes used.

The clustering process was performed using the *K-Means* algorithm, which belongs to the *unsupervised learning* category. The optimal number of clusters was determined using the Elbow method, where the best value was obtained at K = 3, representing three levels of website effectiveness: high, medium, and low (Nurdiansyah et al., 2024).

Each initial centroid is determined randomly from the data set, then the

distance between each data point and the centroid is calculated using the following Euclidean Distance formula:

$$D = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 + (x_4 - c_4)^2}$$

The smallest distance value determines the membership of each data point to a particular cluster. The iteration process is repeated until the centroid position does not change (converges) (Ambarsari et al., 2024).

Analysis Stage

The analysis stage is a follow-up process after the K-Means algorithm produces the optimal cluster division[8]. This stage is carried out to interpret the grouping results, understand the characteristics of each cluster, and identify the factors that most influence the effectiveness of local government websites in South Sumatra Province. The clustering results using the RapidMiner Studio application show that government websites are divided into three main clusters, namely effective, fairly effective, and ineffective. The four main indicators used total visits, average visit duration, pages per visit, and bounce rate show significant variations in each cluster, reflecting differences in the level of user engagement with each website(Ikotun et al., 2023).

RESULTS AND DISCUSSION

1) Data Collection

In this study, the data used is secondary data obtained from the SimilarWeb platform, which provides statistical information on the performance of local government websites in South Sumatra Province. This data includes four main parameters used as grouping variables, namely *Total Visits*, *Bounce Rate, Average Visit Duration, and Pages per Visit[3].* The data was taken from ten local government websites, including: the Communication and Information Agency (Kominfo Sumsel), Regional Revenue Agency (Bapenda Sumsel), Trade Agency (Disdag Sumsel), Transportation Agency (Dishub Sumsel), Regional Financial and Asset Management Agency (BPKAD Sumsel), Central Statistics Agency (BPS Sumsel), Energy and Mineral Resources Agency (ESDM Sumsel), and the Ministry of Law and Human Rights of South Sumatra.

Table 1. Data Collection Title

Department Name	Total Visits	Bounce Rate	Visit Duration	Pages per Visit
South Sumatra Ministry of Communication and Information Technology	14,128	23.71	1:03	2.16
South Sumatra Regional Revenue Agency	23,488	65.83%	0:25	5.21
Palembang Fire Department	21,396	70.56%	1:14	1.98
South Sumatra Trade and Industry Office	323	28.92%	0:25	1
Transportation Agency of South Sumatra Province	398	44.41	0:19	1.69
South Sumatra Regional Financial and Asset Management Agency	1,033	25.39	0:23	1.85
BPS South Sumatra	74,424	56.31%	1:31	3.03
ESDM South Sumatra	4,165	14.03%	0:30	2

Ministry of Law and Human Rights of South Sumatra	23,999	66.00%	1:20	5.38
South Sumatra Legal Information Center	6,829	34.88%	0:35	1.04

(Source: SimilarWeb)

2) Determining the Number of Clusters (k)
   Data grouping was performed using the K-Means Clustering algorithm with k = 3, namely:
   - Cluster 1 (C1) Websites with low effectiveness
   - Cluster 2 (C2) Websites with moderate effectiveness
   - Cluster 3 (C3) Websites with high effectiveness
3) Determining the Initial Centroid
   Initial centroids were determined randomly based on the data in Table 1, with the following results:

Table 2. Initial cluster centers

Variable	C1 (Less Effective)	C2 (Moderately Effective)	C3 (Effective)
Total Visits	323	14,128	74,424
Bounce Rate (%)	28.92	23.71	56.31
Visit Duration	0:25	1:03	1:31
Pages per Visit	1.00	2.16	3.03

C1 is represented by the South Sumatra Trade and Industry Agency, C2 by the South Sumatra Communication and Information Agency, and C3 by the South Sumatra Statistics Agency.

4) Calculating the Distance of Each Object to the Centroid
   The formula for calculating the distance from each website to each centroid is:

$$d(x_i, c_j) = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2 + (x_4 - c_4)^2}$$

   The results of the distance calculations between the data and each centroid are presented in Table 3 below.

Table 3. Centroid distance calculation for the first iteration

Website	DC1	DC2	DC3
South Sumatra Ministry of Communication and Information Technology	12.12	0.00	53.21
South Sumatra Regional Revenue Agency	11.93	22.41	37.85
Palembang Fire Department	18.60	25.09	51.34
South Sumatra Trade Agency	0	13.57	60.42
South Sumatra Transportation Agency	2.32	11.12	57.03
South Sumatra Regional Treasury Office	1.01	12.43	55.68
South Sumatra Statistics Agency	60.42	53.21	0.00
South Sumatra ESDM	8.94	9.13	47.10
Ministry of Law and Human Rights of South Sumatra	21.33	19.54	32.66
South Sumatra Judicial Information System	4.35	10.44	50.31

5) Each website is allocated to the cluster with the closest distance based on the minimum value in the table above.

Table 4. Results of the first iteration of cluster grouping

Website	DC1	DC2	DC3	C1	C2	C3
South Sumatra Ministry of Communication and Information Technology	12.1	0	53.2	0	1	0
South Sumatra Regional Revenue Agency	11.9	22.4	37.9	0	0	1
Palembang Fire Department	18.6	25.1	51.3	0	1	0
South Sumatra Trade Agency	0	13.6	60.4	1	0	0
South Sumatra Transportation Agency	2.32	11.1	57	1	0	0
South Sumatra Regional Financial and Asset Management Agency	1.01	12.4	55.7	1	0	0
South Sumatra Statistics Agency	60.4	53.2	0	0	0	1
ESDM South Sumatra	8.94	9.13	47.1	0	1	0
Ministry of Law and Human Rights of South Sumatra	21.3	19.5	32.7	0	0	1
(Anonymous)	4.35	10.4	50.3	1	0	0

From the table above, we obtain:

- The closest distance to C1 is 4 websites
- The closest distance to C2 is 3 websites
- The closest distance to C3 is 3 websites

6) Determining the New Centroid (Iteration 1)
The new centroid is calculated using the average formula for each variable from the websites included in one cluster

$$c_{ij} = \frac{1}{N_i} \sum_{k=1}^{N_i} x_{kj}$$

Table 5. New centroid iteration 1

Variable	C1	C2	C3
Total Visits	2,646	13,230	40,637
Bounce Rate (%)	33.40	37.45	62.71
Visit Duration	0:25	0:56	1:27
Pages per Visit	1.39	2.05	4.54

7) iteration 2 to Convergence
With the new centroid, the distance calculation is performed again until there is no change in cluster members. After three iterations, the position of the cluster members does not change anymore, meaning that the results have converged.

Table 6. Final (converged) centroids

Variable	C1 (Low)	C2 (Medium)	C3 (High)
Total Visits	2,646	13,230	40,637
Bounce Rate (%)	33.40	37.45	62.71
Visit Duration	0:25	0:56	1:27
Pages per Visit	1.39	2.05	4.54

8) Final Clustering Results
Based on the final results of the iteration process, the website clustering results are obtained as shown in Table 6.

Table 7. Final Clustering Results

Cluster	Website	Description
C1 (Low Effectiveness)	Disdag Sumsel, Dishub Sumsel, BPKAD Sumsel, (Anonymous)	Website with low total visits, short duration, and few pages.
C2 (Moderate Effectiveness)	South Sumatra Communication and Information Agency, Palembang Fire Department, South Sumatra Energy and Mineral Resources Agency	Website with moderate visits, medium access time, and fairly good user interaction.
C3 (High Effectiveness)	BPS South Sumatra, Bapenda South Sumatra, Kemenkumham South Sumatra	Website with high traffic, long duration, and a large number of pages explored.

9) Interpretation
The clustering results show that:
- C3 (high effectiveness) is dominated by websites with high traffic and deep interaction, such as BPS Sumsel and Kemenkumham Sumsel, which have a lot of informative content and online services.
- C2 (moderate) indicates sites with good user engagement, but still require interface and navigation optimization, such as Kominfo South Sumatra and ESDM South Sumatra.
- C1 (low) includes sites with low visits and short durations, tending to contain static information and less interactivity, such as Disdag Sumsel and Dishub Sumsel.

10) The clustering process uses the RapidMiner application
- The first step is to import or add the data to be clustered by clicking "Add Data" in the top-right corner of the navigation bar

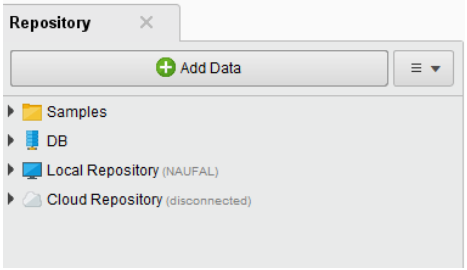

Figure 1. Data *import* process

- Next, select the file to be processed and click "Next" when you reach the desired file

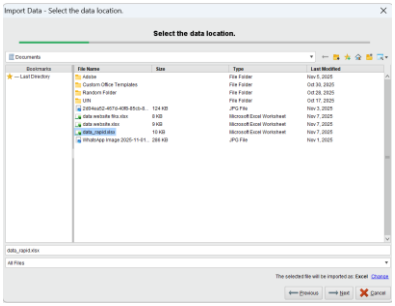

Figure 2. *Selecting* data to import

- After the data has been successfully imported, a preview of the data will appear as shown in Figure 3

ExampleSet (10 examples, 0 special attributes, 5 regular attributes)

Row No.	Nama Dinas	Total Visit	Bounce Rate	Visit Duratio...	Page per Visit
1	Kominfo Sum...	14128	23.710	63	2.160
2	Bapenda Su...	23488	65.830	25	5.210
3	Damkar Pale...	21396	70.560	74	1.980
4	Disdag Sum...	323	28.920	25	1
5	Dishub Sum...	398	44.410	19	1.690
6	BPKAD Sum...	1033	25.390	23	1.850
7	BPS Sumsel	74424	56.310	91	3.030
8	ESDM Sumsel	4165	14.030	30	2
9	Kemenkumh...	23999	66	80	5.380
10	JDIH Sumsel	6829	34.880	35	1.040

Figure 3. *Preview* of imported data

- Next, click *Design* on the top main menu, then drag and drop the imported data. After that, find and drag k-means, and the process will appear and be connected. After that, it will appear as shown in Figure 4.

Figure 4. The *clustering* process design

- Next, set the number *of clusters* in the "*Parameters*" menu

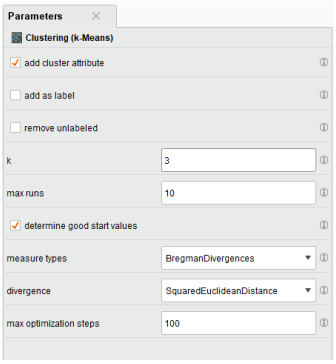

Figure 5. Process of determining the number of clusters

- Then click the triangle above the navigation menu to display the results of the *clustering* process in Figure 6.

Row No.	id	cluster	Total Visit	Bounce Rate	Visit Duratio...	Page per Visit
4	4	cluster_0	323	28.920	25	1
5	5	cluster_0	398	44.410	19	1.690
6	6	cluster_0	1033	25.390	23	1.850
8	8	cluster_0	4165	14.030	30	2
10	10	cluster_0	6829	34.880	35	1.040
7	7	cluster_1	74424	56.310	91	3.030
1	1	cluster_2	14128	23.710	63	2.160
2	2	cluster_2	23488	65.830	25	5.210
3	3	cluster_2	21396	70.560	74	1.980
9	9	cluster_2	23999	66	80	5.380

Figure 6. Results of *the clustering process*

- Next, Figure 7 will display a chart plot

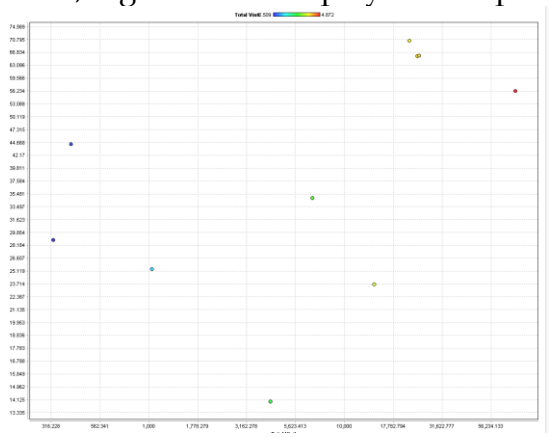

Figure 7. *Chart* graph of the *clustering* results

Based on the results of testing the data using the RapidMiner application, the *clustering* process with the K-Means algorithm produced three main groups as shown in the *Centroid Cluster Model* display.

a) Cluster 0 (C1) has 5 data points, indicating that the first group contains websites with low effectiveness. Websites in this group have low visit numbers, short access durations, and few pages opened by users.
b) Cluster 1 (C2) has 1 data point, which describes the group with a high level of effectiveness. Websites in this cluster have a high number of visits and long access duration, thus showing the best performance compared to other groups.
c) Cluster 2 (C3) consists of 4 data points, representing a group with moderate effectiveness. Websites in this group have a fairly good number of visits and user interactions, but still need improvement in terms of retention and user experience.

CONCLUSION

This study successfully applied the *K-Means Clustering* algorithm to analyze the effectiveness of local government websites in South Sumatra Province based on four main indicators, namely *total visits*, *average visit duration*, *pages per visit*, and *bounce rate*. Based on the results of data processing using the RapidMiner Studio application, three main clusters were obtained that describe

the effectiveness of websites, namely:

1. Cluster 1 (Low Effectiveness) – Websites with low visit numbers, short visit durations, and minimal user interaction.
2. Cluster 2 (Medium Effectiveness) – Websites with moderate performance, having fairly good traffic and user engagement but still requiring content and interface optimization.
3. Cluster 3 (High Effectiveness) – Websites with high visit numbers, long access durations, and deep interactions that reflect better user satisfaction levels.

The variables of *average visit duration* and *bounce rate* proved to be the most dominant factors in determining website effectiveness. This shows that the longer users stay on a website page and the lower the bounce rate, the higher the level of effectiveness.

Overall, the results of this study provide an empirical picture of the varying effectiveness of local government websites in South Sumatra. These findings can be used as a basis for local governments to evaluate and develop strategies to improve website quality, such as regular content updates, increased access speed, and a more interactive and user-friendly interface design. Thus, digital-based public services in South Sumatra Province can run more effectively, transparently, and responsively to the needs of the community.

REFERENCES

M. Al-Kautsar Maktub, P. W. Handayani, and F. P. Sunarso, "Citizen acceptance and use of the Jakarta Kini (JAKI) e-government: Extended unified model for electronic government adoption," *Heliyon*, vol. 11, no. 2, p. e42078, Jan. 2025, doi: 10.1016/j.heliyon.2025.e42078.

D. Gartika and A. Widiyanto, "How do local governments implement e-government: A case study from Bali Province, Indonesia," *JeDEM - eJournal of eDemocracy and Open Government*, vol. 16, no. 1, pp. 61–88, Sep. 2024, doi: 10.29379/jedem.v16i1.898.

E. Emigawaty, K. Adi, A. F. Rochim, B. Warsito, and A. Wibowo, "K-Means Clustering Algorithm for Partitioning the Openness Levels of Open Government Data Portals," *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 977–983, Sep. 2023, doi: 10.30630/joiv.7.3.1761.

A. Nur Khomarudin, "Data Mining Techniques: K-Means Clustering Algorithm," 2003. [Online]. Available: https://agusnkhom.wordpress.com

D. Nurdiansyah, M. N. P. Ma'ady, Y. Sukmawaty, M. C. C. Utomo, and T. Mutiani, "CLUSTERING ANALYSIS FOR GROUPING SUB-DISTRICTS IN BOJONEGORO DISTRICT WITH THE K-MEANS METHOD WITH A VARIETY OF APPROACHES," *BAREKENG: Journal of Mathematics and Applied Sciences*, vol. 18, no. 2, pp. 1095–1104, May 2024, doi: 10.30598/barekengvol18iss2pp1095-1104.

Erlin Windia Ambarsari, Dedin Fathudin, and Gravita Alfiani, "Utilizing K-Means Clustering to Understand Audience Interest in SEO-Optimized Media Content," *Journal of Computing and Informatics Research*, vol. 3, no. 2, pp. 208–214, Mar. 2024, doi: 10.47065/comforch.v3i2.1207.

H. Mulyani and R. A. Setiawan, "CLUSTERING THE HAPPINESS LEVEL OF PROVINCES IN INDONESIA USING K-MEANS," *Journal of Information Technology and Its Utilization*, vol. 7, no. 2, pp. 47–52, Dec. 2024, doi: 10.56873/jitu.7.2.5854.

A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf Sci (N Y)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/j.ins.2022.11.139.