Regional Mapping Based on Tourism Destinations in West Java: K-Medoid Clustering Analysis

Nafis Almajid [1], Prima Dina Atika [2], Khairunnisa Fadhilla Ramdhania [3*]

[1,2,3]*Department of Informatics, Universitas Bhayangkara Jakarta*
Marga Mulya, Bekasi, West Java, Indonesia

[*] khairunnisa.fadhilla@dsn.ubharajaya.ac.id

Abstract

The growth of the tourism sector in West Java demands an optimal development strategy. This study aims to cluster regions in West Java based on the characteristics of their tourist destinations using the K-Medoid algorithm. This algorithm was chosen because of its superiority in producing optimal clusters and robustness to outliers. Data on tourist destination characteristics in West Java were analyzed using the K-Medoid algorithm and the Elbow method to determine the optimal number of clusters. The evaluation was conducted using the Davies-Bouldin Index. As a result, three clusters with different characteristics were formed. The first cluster, "Medium potential and medium achievement", consists of 1 region with unoptimized potential for campsite tourism. The second cluster, "High potential and moderate achievement", consists of 25 regions with a diversity of attractions and a high number of visits. Finally, the third cluster, "Medium potential and high achievement", consists of 1 region with popular historical and cultural attractions and high visitation. The model evaluation showed a Davies Bouldin Index score of 0.08, indicating good clustering quality. This research is expected to provide insights for the government and related stakeholders to formulate targeted tourism development policies in West Java. The K-Medoid algorithm helps identify certain patterns, providing deeper insights into regional differences in terms of tourism.

Keywords: Tourism Destinations, Tourism, West Java, regional analysis, K-Medoid

I. INTRODUCTION

The tourism sector has experienced significant growth in recent times due to the emergence of more tourist destinations in various locations. Information about these tourist spots, widely spread across various media, has made it easier for visitors to access. Every year, the Indonesian Central Statistics Agency (BPS) periodically releases data on the development of tourism in Indonesia. This data is presented in both raw and infographic formats that are easy for the public to understand. The purpose of presenting data in infographic form is to help the public absorb information related to the state of tourism in the country. The Central Statistics Agency regularly publishes data on Tourism Development in Indonesia every quarter. In the first quarter of 2024, the number of foreign tourist arrivals increased by 24.85% compared to the first quarter of 2023, and the number of domestic tourist visits rose by 19.78% from the first quarter of 2023[1]. However, there has not yet been a comprehensive regional grouping based on the similarities in the characteristics of tourist destinations [2][3]. The arrival of foreign tourists to Indonesia can increase foreign exchange earnings and stimulate economic growth in tourist areas. For foreigners wishing to visit Indonesia, West Java is one of the best places to visit.

The K-Medoid algorithm can be used to cluster tourism data based on its characteristics. The results of this clustering allow the government and relevant stakeholders to formulate more accurate and effective policies tailored to the specific characteristics and needs of each group [4].

II. LITERATURE REVIEW

According to the Tourism Law No. 10 of 2009, nature, flora, fauna, ancient relics, history, art, and culture are blessings from God Almighty. Therefore, they can be utilized as assets in tourism development and serve as resources that are systematically and sustainably managed while preserving religious, cultural, and environmental values. This is aimed at achieving national development goals with a community-oriented approach [5].

Clustering is a reliable data mining method and valid tool for addressing complex problems in computer science and statistics [6] Clustering works by partitioning objects within a dataset into homogeneous clusters [2]. The clustering analysis commonly used in research includes the K-Means and K-Medoid algorithms. The advantage of K-Medoid is its abiity to produce optimal clusters that are robust to outliers, particularly when compared to the K-Means algorithm [3]. Given these considerations, the researcher has conducted a study entitled "Regional Mapping Based on Tourist Destinations in West Java: A K-Medoid Clustering Analysis."

K-Medoid is a clustering technique that uses representative objects (medoids) as representatives for the cluster center point. This algorithm uses an iterative and greedy method. Like the K-Means algorithm, the initial representative objects are selected randomly. The iterative process of replacing representative objects with non-representative objects continues until the quality of the resulting grouping cannot be improved by any replacement. This quality is measured by the deviation value between the new total distance value and the previous total distance value [7].

III. RESEARCH METHOD

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework. CRISP-DM was first introduced in the late 1990s by four major companies in Europe: Integral Solution Ltd. (a provider of commercial data mining solutions), NCR (a database provider), DaimlerChrysler (an automotive manufacturer), and OHRA (an insurance company). The CRISP-DM model offers several advantages, including an organized and systematic structure, comprehensive coverage of each stage in the knowledge extraction process, flexibility in application across different projects, and efficiency in developing useful models [8].

A. Business Understanding

At this step, regional clusters within West Java Province are formed based on tourist destinations using the K-Medoid algorithm. Thus, the results of this study can be used to provide recommendations to the government and local authorities regarding marketing strategies and evaluations.

B. Data Understanding

The data used in this study includes visitor counts and camping facility data, visitor counts and museum facility data, and visitor counts and homestay facility data, all of which were downloaded from the official Open Data Jabar website in .csv format. The dataset consists of six files, with a total of 1,158 records.

C. Data Preparation

Data management was conducted using Google Colab with the Python 3.10.12 programming language. The data used were cleaned by reducing or correcting inconsistencies to refine the dataset for the Modeling stage. The steps involved in Data Cleaning are as follows:

1) Dropping rows of data that are not relevant.
2) Performing data reduction through feature engineering.
3) Data integration by merging the entire dataset used.

D. Modelling

This study utilizes the K-Medoid algorithm in its modeling process. The K-Medoid algorithm uses a medoid as the center of the cluster to form groups of numerical data points that exhibit high similarity [9]. Modeling with the K-Medoid algorithm aims to identify patterns and group data into clusters based on similar characteristics.

The steps in the K-Medoid method include:
1) Determining the number of clusters (K) using the Elbow Method as an estimate for the optimal number of clusters [10] [11], with the following equation

$$SSE = \sum_{i=1}^{n} (x_i - C_k)^2 \tag{1}$$

2) The distance between each data point and its assigned cluster center is calculated using the Euclidean Distance formula [12], given by:

$$d(x_i, y_j) = \sqrt{\sum_{k=1}^{n} (x_{i,k} - y_{i,k})^2} \tag{2}$$

3) Randomly select an object from each cluster as the new medoid.
4) Calculate the Euclidean distance of each object in the cluster to the candidate new medoid.
5) Compute the total new distance and subtract the total old distance to calculate the total deviation (S). If S is less than 0, swap the object with the cluster data to form a new set of k objects as the new medoid.
6) Repeat steps 3 to 5 until no change occurs in the medoid, resulting in the final clusters and their members [6].

E. Evaluation

At this step, the modeling process is evaluated for its accuracy in achieving business objectives using the Davies-Bouldin Index, followed by an assessment based on the obtained results. The Davies-Bouldin Index measures the effectiveness of the clustering process by evaluating how well the data clusters are formed based on the attributes or features contained within the data [13] [14]. The formula for the Davies-Bouldin Index is as follows

$$DBI = \frac{1}{k} \sum_{i=1}^{k} \max_{i \neq j} R(i,j) \tag{3}$$

IV. RESULTS

This study utilizes the K-Medoid algorithm for the modeling process. To achieve the result in the form of clusters, the algorithm involves several key steps. These steps include determining the number of K (clusters) to be used. Next, a medoid is selected randomly, and the Euclidean distance of each data point to the medoid is calculated. Then, an object within each cluster is chosen as the new medoid, and the Euclidean distance between each data point and the new medoid is recalculated. The data clusters are determined based on the minimum

value of the total new distance, subtracted by the total old distance. This process is repeated iteratively until the medoid no longer changes.

A. Elbow Method

This study employs the elbow method to determine the optimal number of clusters. This method operates by running iterations from $K = 1$ to $K = n$ to identify the "elbow" point on a graph visualizing the Within-Cluster Sum of Squares (WCSS) value. The WCSS value represents the total squared distance between each data point and its respective cluster center (medoid), determined by the value of K. The "elbow" point on the graph indicates the most optimal number of clusters for the analyzed data.
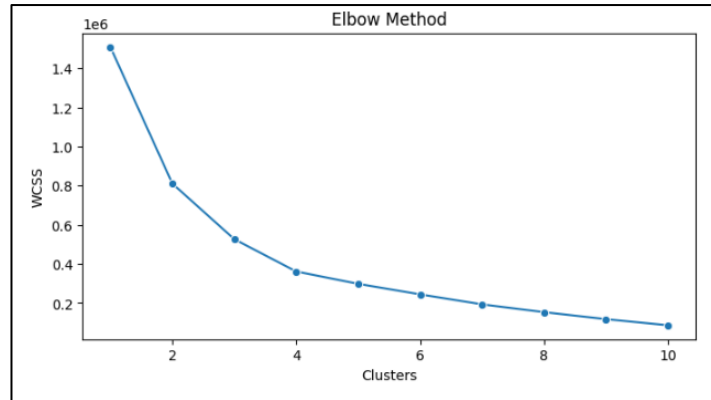

Fig. 1. Elbow Method Visualization.

The elbow method graph visualization in Fig. 1 shows that an increase in the number of clusters (K) correlates with a decrease in the distance between data points and their respective medoids. For instance, when $K = 2$, the WCSS value is 1,510,059.789. However, when $K = 3$, the WCSS value decreases to 811,420.329, indicating that the distance between data points and their medoids reduces as the number of clusters increases. The term "elbow method" is derived from the process of determining the optimal K value, which is done by visually identifying the "elbow" point that resembles a right angle on the resulting line chart. As shown in the Fig.1, this can be observed clearly.

B. K-Medoid Modelling

The modeling process begins by determining the number of clusters. The next stage is medoid initiation, where several data are selected randomly from the dataset used. In this study, three medoids were randomly selected from a previously processed dataset, the data selected were given in TABLE I.

TABLE I
MEDOID DETERMINATION

Regency/City	Medoid
Majalengka	0
Bogor	1
Tasikmalaya	2

The medoids randomly selected from the processed dataset will be used as the medoids for Iteration 1. The distances from each data point to these medoids will then be calculated using the Euclidean distance formula, starting from the first data point to medoids 0, 1, and 2, and continuing to the nth data point for medoids 0, 1, and 2. For example, the distance calculation for the first data point, Kabupaten Bogor, and the second data point, Kabupaten Sukabumi, for medoids 0, 1, and 2 can be performed. Specifically, the calculation for the first data

point, Kabupaten Bogor, to Kabupaten Majalengka as medoid 0 (C_0) using Equation (2), serves as the following illustrative example.

$$d(x_1, y_0) = \sqrt{(x_{1,1} - y_{0,1})^2 + (x_{1,2} - y_{0,2})^2 + \cdots + (x_{1,27} - y_{0,27})^2}$$
$$= \sqrt{(0-0)^2 + (0-5)^2 + \cdots + (0-1{,}347)^2}$$
$$= 30784{,}5215.$$

Similarly, it is calculated using the Euclidean distance formula. For the first data point, Kabupaten Bogor, the calculation shows 30.784,5215 to medoid C_0, 0 to medoid C_1, and 2.482,154105 to medoid C_2. For the second data point, the distances are 286.043,7964 to C_0, 284.523,702 to C_1, and 283.654,9796 to C_2. The Euclidean distance calculation is then performed for data points 4 through 27. The next step involves selecting the minimum value among the distances to C_0, C_1, and C_2 for each data point.

In the second iteration, the first step involves randomly selecting a non-medoid data point to undergo the Euclidean distance calculation process, similar to the procedure in the first iteration. The selected non-medoid data points are given in TABLE II.

TABLE II
DETERMINING THE MEDOID FOR THE SECOND ITERATION.

Regency/City	Medoid
Majalengka	0
Bogor	1
Bogor Baru	2

The newly selected medoids are subsequently used to perform calculations identical to those conducted in the first iteration. The next step involves calculating the total cost or deviation, obtained by subtracting the total Euclidean distance of the previous iteration from the total Euclidean distance of the current iteration. This value is then used to guide the decision-making process for the subsequent iteration. The goal is to identify the iteration with the total cost value closest to zero, based on the calculations and results of total deviation $-297.359{,}3712$.

The total cost value obtained by subtracting the results of the second iteration from the first iteration indicates that the total cost remains too low, below 0. Therefore, it can be concluded that further iterations, starting from the third, are necessary to achieve a total deviation value closer to 0 or convergence. In this study, the calculation converged in the fourth iteration, with the results detailed in TABLE III.

TABLE III
LAST ITERATION RESULT.

Regency/City	C_0	C_1	C_2	min	cluster
Bogor	284.523,702	3	305.959,8598	3	C1
Sukabumi	0	284.523,7017	415.628,3297	0	C0
Cianjur	284.523,6782	55,83009941	305.959,8622	55,83009941	C1
Bandung	290.507,7303	71.534,72675	314.209,5025	71.534,72675	C1
Garut	284.222,3374	9.079,973733	305.492,3473	9.079,973733	C1
Tasikmalaya	256.273,0301	51.063,17697	310.094,4932	51.063,17697	C1
Ciamis	286.267,2562	32.374,35948	307.535,0169	32.374,35948	C1
Kuningan	300.139,9745	98.903,88978	304.143,1087	98.903,88978	C1
Cirebon	284.511,888	1.212,520515	305.411,7647	1.212,520515	C1
Majalengka	286.043,7964	30.784,52204	306.275,9625	30.784,52204	C1
Sumedang	287.573,0165	42.602,56369	308.654,4903	42.602,56369	C1
Indramayu	284.554,8326	7.660,173954	302.338,4313	7.660,173954	C1
Subang	284.070,9788	14.492,08498	297.855,177	14.492,08498	C1
Purwakarta	283.983,6611	20.846,5825	297.361,6424	20.846,5825	C1
Karawang	297.720,0566	90.846,19488	293.311,2844	90.846,19488	C1
Bekasi	284.546,0503	5.060,515784	305.993,6236	5.060,515784	C1

Regency/City	C_0	C_1	C_2	min	cluster
Bandung Barat	282.601,424	5.853,379195	306.015,1428	5.853,379195	C1
Pangandaran	284.522,1532	556.6264457	305.960,3407	556,6264457	C1
Bogor City	415.628,3297	305.959,8598	0	0	C2
Sukabumi City	284.400,9778	3.931,438159	303.382,2054	3.931,438159	C1
Bandung City	284.523,5889	200,706253	305.959,9058	200,706253	C1
Cirebon City	286.285,6093	35.729,96492	278.286,6256	35.729,96492	C1
Bekasi City	284.521,362	400,005	305.803,1979	400,005	C1
Depok City	284.523,7017	0	305.959,8598	0	C1
Cimahi City	284.073,0144	1.220,823902	305.961,8515	1.220,823902	C1
Tasikmalaya City	283.654,9796	2.482,153903	305.969,5112	2.482,153903	C1
Banjar City	284.523,7028	1,414213562	305.959,8599	1,414213562	C1

C. Clustering Result

The results of the modeling process conducted earlier can be further analyzed to determine the categories of the clusters obtained through Exploratory Data Analysis (EDA). These analytical findings can then be used to define the categories of the clusters as follows:

1) Cluster 0: Moderate Potential and Moderate Performance

This cluster is characterized by underutilized potential in camping tourism and museum tourism, as indicated by relatively low visitor numbers for these attractions. Significant efforts are needed in infrastructure development, promotion, and diversification of tourist attractions. However, there is growth potential, as evidenced by a relatively high number of domestic visitors, particularly for homestay accommodations.

2) Cluster 1: High Potential and Moderate Performance

This cluster exhibits a diversity of tourist attractions and moderate visitor numbers. Efforts should focus on improving service quality, implementing innovative marketing strategies, and ensuring sustainable environmental management.

3) Cluster 2: Moderate Potential and High Performance

This cluster features popular historical and cultural attractions with high visitor numbers and relatively significant economic impacts, particularly in the Meeting, Incentive, Convention, and Exhibition (MICE) sector. To maximize its potential, efforts should include expanding market segmentation and fostering innovation and creativity.

Next, a function will be created to map cluster values to their respective categories: Cluster 0 will be categorized as "Moderate Potential and Moderate Performance," Cluster 1 as "High Potential and Moderate Performance," and Cluster 2 as "Moderate Potential and High Performance." This function will then be applied to the cluster column, and the results will be stored in a new column named 'Cluster Description.' TABLE IV

TABLE IV
DATASET RESULTS WITH LABELS.

Regency/City	Cluster	Label
Bogor	1	High Potential and Moderate Performance
Sukabumi	0	Moderate Potential and Moderate Performance
Cianjur	1	High Potential and Moderate Performance
Bandung	1	High Potential and Moderate Performance
Garut	1	High Potential and Moderate Performance
Tasikmalaya	1	High Potential and Moderate Performance
Ciamis	1	High Potential and Moderate Performance
Kuningan	1	High Potential and Moderate Performance
Cirebon	1	High Potential and Moderate Performance
Majalengka	1	High Potential and Moderate Performance

Regency/City	Cluster	Label
Sumedang	1	High Potential and Moderate Performance
Indramayu	1	High Potential and Moderate Performance
Subang	1	High Potential and Moderate Performance
Purwakarta	1	High Potential and Moderate Performance
Karawang	1	High Potential and Moderate Performance
Bekasi	1	High Potential and Moderate Performance
Bandung Barat	1	High Potential and Moderate Performance
Pangandaran	1	High Potential and Moderate Performance
Bogor City	2	Moderate Potential and High Performance
Sukabumi City	1	High Potential and Moderate Performance
Bandung City	1	High Potential and Moderate Performance
Cirebon City	1	High Potential and Moderate Performance
Bekasi City	1	High Potential and Moderate Performance
Depok City	1	High Potential and Moderate Performance
Cimahi City	1	High Potential and Moderate Performance
Tasikmalaya City	1	High Potential and Moderate Performance
Banjar City	1	High Potential and Moderate Performance

D. Evaluation

This evaluation is conducted by measuring the proximity between data points, the distance of data points to respective cluster centers, and the distances between clusters. To measure the quality of the resulting model, this study employs the Davies-Bouldin Index (DBI) as the evaluation metric. The DBI evaluates the performance of the clustering model by measuring the degree of separation between clusters. A lower DBI score, closer to 0, indicates better-separated clusters, reflecting a more optimal model. The steps for calculating the DBI are as follows:

1) One method for measuring the similarity or closeness of data points within a cluster to its center is by calculating the Sum of Squares Within Cluster (SSW), also known as the cohesion value. This calculation is performed using the following equation.

$$SSW = \frac{1}{m} \sum_{i=1}^{k} d(x_i, y_j) \tag{4}$$

For example, calculating the distance of the 1st data to the medoid cluster

$$d(x_i, y_j) = \sqrt{\sum_{k=1}^{3} (x_{i,k} - y_{i,k})^2}$$
$$= \sqrt{(x_{1,1} - y_{0,1})^2 + (x_{1,2} - y_{0,2})^2 + \cdots + (x_{1,27} - y_{0,27})^2}$$
$$= \sqrt{(0 - 0)^2 + (0 - 153)^2 + \cdots + (0 - 0)^2}$$
$$= 284523.702$$

The calculation above is then applied to all data points in the dataset. The next step is to determine the SSW value for each cluster by calculating the average distance of all cluster members from their respective cluster centers.

$$SSW_0 = \frac{0}{1} = 0$$
$$SSW_1 = \frac{3 + 55.83009941 + \cdots + 1.414213562}{25} = 21075.86509$$
$$SSW_2 = \frac{0}{1} = 0$$

2) The next step is to calculate the Sum of Squares Between Clusters (SSB)[15]. The SSB value reflects the degree of separation between clusters, using distance formula (5).

$$SSB_{i,j} = d(y_i, y_j) \tag{5}$$

As follows,

$$SSB_{0,1} = \sqrt{(0-0)^2 + (153-1)^2 + \cdots + (0-0)^2}$$
$$SSB_{0,1} = 284523.7017$$

the calculation above is performed for each cluster. In this study, which uses three clusters, the calculation is repeated for all clusters, yielding the results showing in TABLE 5.

TABLE 5
SBB CALCULATION RESULTS

SSB	C0	C1	C2
C0	0	284523.7017	415628.3297
C1	284523.7017	0	305959.8598
C2	415628.3297	305959.8598	0

3) Once the SSW and SSB values are obtained, the next step is to calculate the ratio to compare the variability within clusters with the variability between clusters. A smaller ratio value indicates better clustering, as it reflects more homogeneous clusters that are clearly separated. The ratio is calculated using the formula (6).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{6}$$

Thus, the following results are obtained:

$$R_{1,2} = \frac{0 + 21075.86509}{284523.7017} = 0.074074198$$
$$R_{1,3} = \frac{0 + 0}{415628.3297} = 0$$
$$R_{2,3} = \frac{21075.86509 + 0}{305959.8598} = 0.068884412$$

4) The calculated ratio values are then used to compute the DBI value using formula (3):

$$DBI = \frac{1}{3}(0.074074198 + 0.074074198 + 0.068884412) = 0.07234427$$

The obtained DBI score can be considered a favorable result, as it is close to 0. The evaluation process is conducted again to confirm that $k = 3$ is the optimal number of clusters, using the Davies-Bouldin Index for several potential cluster configurations.

TABLE V
DBI EVALUATION RESULTS.

$k = n$	Davies Bouldin Index
$k = 2$	0.125019
$k = 3$	0.087020
$k = 4$	0.358987
$k = 5$	0.517217
$k = 6$	0.433907
$k = 7$	0.386952
$k = 8$	0.330428
$k = 9$	0.317409
$k = 10$	0.139970

the DBI can be observed from TABLE V that value for each tested number of clusters shows that $k = 3$ has the best DBI value compared to other values of K. Therefore, it can be confirmed that $k = 3$ is the optimal number of clusters in this study.

V. CONCLUSION

The results of the modeling conducted using the K-Medoids algorithm successfully identified patterns in the dataset used, producing a clustering of areas in West Java into 3 clusters. Kabupaten Sukabumi is a member of cluster 0, while Kabupaten Bogor, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, and Kota Banjar are members of cluster 1. Lastly, Kota Bogor is a member of cluster 2. The clustering results also received a good DBI evaluation score with a value of 0.08701964080920822.

Other clustering algorithms such as K-Prototype, Agglomerative Clustering, or DBSCAN can be used to test the accuracy of each model in handling data. Second, further research development on regional clustering based on tourist destinations in West Java, particularly in examining the results and influencing factors, requires synergy and collaboration across various scientific disciplines. Third, real-time system implementation can be carried out using the developed model.

REFERENCES

[1]   Badan Pusat Statistik, "Perkembangan Pariwisata April 2024," Berita Resmi Statistik. Accessed: Jun. 03, 2024. [Online]. Available: https://www.bps.go.id/id/pressrelease/2024/06/03/2351/perkembangan-pariwisata-april-2024.html

[2]   A. Neva and F. Dikananda, "Penerapan Metode K-Means Clustering dalam Mengelompokan Jumlah Wisatawan Asing di Jawa Barat," *JURNAL ALGOR*, vol. 4, Mar. 2023.

[3]   S. Y. Riska and L. Farokhah, "Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia," *INTEGER: Journal of Information Technology*, vol. 8, pp. 1–8, Mar. 2023.

[4]   I. Muthahharah and Z. Mar'ah, "Clustering Biplot on Tourist Visits in Indonesia," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 3, no. 1, pp. 43–50, Mar. 2024, doi: 10.30812/ijecsa.v3i1.3890.

[5]   D. Andriani and et al, *Perencanaan Pariwisata*, 1st ed. Bandung: Widina Bhakti Persada Bandung, 2022.

[6]   D. Masri and S. Kasnelly, "Data Mining Techniques are Being Used to Map the Number of Foreign Guests at Indonesian Star Hotels," *International Journal of Information System & Technology Akreditasi*, vol. 5, no. 2, pp. 215–221, 2021, [Online]. Available: https://www.bps.go.id/indicator/16/310/1/sum-tamu-asing-pada-hotel-star.html

[7]   J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge: Katey Birtcher, 2022.

[8]   T. A. Munandar, *Data Mining Menggunakan R: Teori dan Praktek*, 1st ed. Serang: PT. Bale Damar Publishing, 2023.

[9]   E. Tasia and M. Afdal, "Perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Daerah Rawan Banjir di Kabupaten Rokan Hilir," *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, vol. 3, no. 1, pp. 65–73, Mar. 2023.

[10]   C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm," *EURASIP J Wirel Commun Netw*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13638-021-01910-w.

[11] D. Ayu, I. C. Dewi, and K. Pramita, "Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," 2019.

[12] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean Distance Matrices: Essential theory, algorithms, and applications," *IEEE Signal Process Mag*, vol. 32, no. 6, pp. 12–30, 2015, doi: 10.1109/MSP.2015.2398954.

[13] I. Firman Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 2548–6861, 2022, [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[14] A. I. Ramadhan, P. D. Atika, and K. F. Ramdhania, "Analisis Clustering K-Means untuk Pemetaan Tingkat Pengangguran Terbuka di Provinsi-Provinsi Indonesia Tahun 2013-2023," *Journal of Students' Research in Computer Science*, vol. 5, no. 2, pp. 109–122, Nov. 2024, doi: 10.31599/WBPYDB62.

[15] N. Hidayati, "Application of Cluster Analysis to Data Distribution of COVID-19 in Bengkulu Province," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 2, pp. 743–750, Jun. 2022, doi: 10.30598/barekengvol16iss2pp743-750.