JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 1, June 2025, pp. 278~287
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: 10.36378/jtos.v8i1.4400

Performance Comparison of K-Means Algorithm and BIRCH Algorithm in Clustering Earthquake Data in Indonesia with Web-Based Map Visualization

Baromim Triwijaya[1], Setyoningsih Wibowo[2], Nur Latifah Dwi Mutiara Sari[3]
[1,2,3]Program Studi Informatika, Universitas PGRI Semarang, Indonesia

Article Info

Article history:

Received 05 22, 2025
Revised 06 01, 2025
Accepted 06 18, 2025

Keywords:

Clustering
Earthquake
K-Means
BIRCH
Website

ABSTRACT

This study applies the K-Means and BIRCH algorithms to cluster earthquake data in Indonesia based on geographic coordinates (latitude and longitude), depth, and magnitude from 2008 to 2023. Due to its position at the intersection of three major tectonic plates, Indonesia is highly prone to earthquakes, making the mapping of vulnerable regions essential for disaster risk reduction. K-Means is selected for its simplicity and clustering effectiveness, while BIRCH is known for its scalability and efficiency in processing large datasets. The clustering process involves data preprocessing and normalization, followed by determining the optimal number of clusters using the Elbow method. Initial findings indicate that K-Means produces more distinct and well-separated clusters than BIRCH, with Silhouette Scores of 0.3501 and 0.2247, respectively. However, after expanding the dataset to 121,123 records and incorporating additional attributes such as mag_type, phasecount, and azimuth_gap, BIRCH demonstrated a significant improvement in performance, achieving a Silhouette Score of 0.3489—surpassing K-Means, which dropped to 0.1293. These results suggest that BIRCH is more effective for clustering large and complex datasets. The final clustering results are visualized on a web-based map to support spatial analysis and the identification of earthquake-prone zones.

Corresponding Author:

Baromim Triwijaya
Program Studi Informatika
Universitas PGRI Semarang
Jawa Tengah, Indonesia
Email: baromim08@gmail.com

1. Introduction

Indonesia is situated at the intersection of three major tectonic plates: the Indo-Australian, Eurasian, and Pacific plates. This geological positioning makes the country highly susceptible to seismic events, particularly earthquakes [1]. The frequent occurrence of such events each year can lead to considerable physical, economic, and social consequences [2]. Hence, identifying and mapping regions vulnerable to earthquakes is a critical early measure in disaster risk mitigation efforts.

One approach that can be done in grouping earthquake-prone areas is through clustering techniques in data mining. This technique is used to group data based on similar characteristics so that areas with similar levels of earthquake vulnerability will be included in one group [3]. In this research, the K-Means algorithm

is used, which is one of the non-hierarchical clustering methods that is simple, efficient, and widely used in various scientific studies [4].

Several previous studies have applied K-Means to earthquake data and showed quite good results in clustering earthquake-prone areas [5]. However, this research presents a novelty by integrating the clustering results into a web-based interactive map visualization, and comparing the performance of the K-Means and BIRCH algorithms in clustering regions based on earthquake characteristics in Indonesia. The selection of BIRCH is based on its ability to handle large datasets efficiently, because BIRCH summarizes the data into subclusters first before obtaining the final cluster, making it suitable for processing earthquake data that has a large amount of data [6].

Several previous studies have applied the K-Means and BIRCH algorithms for data clustering analysis. Malik et al. (2023), in their study titled "Earthquake Distribution Mapping in Indonesia Using K-Means Clustering Algorithm," utilized K-Means to map the distribution of earthquake points in Indonesia. The research, which was based on data from the Meteorology, Climatology, and Geophysics Agency (BMKG) from 2008 to 2023, focused on clustering earthquake events by depth, magnitude, and a combination of both. The optimal number of clusters was determined using the Silhouette Score, with the two-cluster scenario achieving the best result. This indicates that K-Means was effective in separating earthquake data and provided valuable insights into seismic activity patterns across Indonesia, contributing to disaster mitigation efforts [7].

In contrast, a study by Rizalde et al. (2023) titled "Comparison of K-Means, BIRCH, and Hierarchical Clustering Algorithms in Clustering OCD Symptom Data" evaluated the performance of these three algorithms on a dataset containing 1,500 records of OCD (Obsessive-Compulsive Disorder) symptoms. The analysis showed that BIRCH achieved the best Davies-Bouldin Index (DBI) score of 1.3 under the K10 scenario, outperforming K-Means (1.36) and Hierarchical Clustering (2.03). These results highlight BIRCH's superior accuracy and efficiency in managing large and complex datasets [8].

From these findings, it can be concluded that while K-Means is effective in identifying spatial patterns in earthquake data, BIRCH demonstrates advantages in processing large-scale datasets with high computational efficiency. These considerations support the use of both algorithms in the present study to cluster earthquake-prone areas in Indonesia based on earthquake characteristics.

2. Research Method

2.1. Data Collection

Data collection is a systematic process of obtaining relevant information or facts to support research objectives [9]. The first stage in this research is the collection of data on earthquakes that occurred in Indonesia during the period 2008 to 2023 and the period 2008 to 2025. The data was obtained through the Kaggle platform, which based on its description is sourced from the Meteorology, Climatology and Geophysics Agency (BMKG) [10].

2.2. Data Preprocessing

Data preprocessing is a crucial step in data analysis that aims to prepare data into a cleaner format and ready to be used in further modeling or analysis [11]. The main purpose of preprocessing is to address issues such as missing values, remove irrelevant data, and transform the data to make it more consistent and ready to be used in machine learning algorithms. This process includes several stages, including addressing missing values, data cleaning, data normalization or standardization, and removal of unnecessary features [12].

In this research, the first step is to check for missing values in the dataset. Missing values can arise for various reasons, such as measurement error or data absence. To overcome this, features that have many missing values will be removed. The features selected for deletion are strike1, dip1, rake1, strike2, dip2, and rake2, as they are considered irrelevant and have many missing values, which may affect the quality of further analysis. The analyzed features only include latitude (lat), longitude (lon), magnitude (mag), and depth.

Table 1. The Dataset

No	tgl	ot	lat	lon	depth	mag	remark
1	2008/11/01	21:02:43	-9.18	119.06	10	4.9	Sumba Region -Indonesia
2	2008/11/01	20:58:50	-6.55	129.64	10	4.6	Banda Sea
3	2008/11/01	17:43:12	-7.01	106.63	121	3.7	Java - Indonesia
4	2008/11/01	16:24:14	-3.30	127.85	10	3.2	Seram - Indonesia
5	2008/11/01	16:20:37	-6.41	129.54	70	4.3	Banda Sea

… … … … … … … …

After removing unnecessary features, the next step is to normalize the remaining data. Normalization aims to change the scale of the data to make it more uniform, avoiding problems that arise due to differences in scale between features [13]. For this reason, the Standard Scaler method is used in this study, where each feature will be processed so that it has a distribution with a mean of 0 and a standard deviation of 1.

Table 2. Data Preprocessing Results

No	lat	lon	depth	mag
1	-1.326	-0.009	-0.508	1.567
2	-0.722	0.967	-0.508	1.207
3	-0.827	-1.156	0.937	0.128
4	-0.162	0.767	-0.508	-0.491
5	-0.697	0.947	0.382	0.847
…	…	…	…	…

2.3. K-Means Clustering

The K-Means algorithm functions by grouping data into several clusters based on the similarity between data, where each cluster is represented by a centroid [14]. The K-Means process aims to reduce the total squared distance between each data point and its nearest centroid [15]. A commonly used method to measure the distance between two points in the K-Means algorithm is the Euclidean distance, formulated as follows [16]:

$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

Description:

- x and y are two points in n-dimensional space.
- $x_1, x_2, \ldots, x_n$ are the x point components.
- $y_1, y_2, \ldots, y_n$ are the y point components.

After the Euclidean distance between each data and all centroids is calculated, each data is then classified into the cluster that has the closest distance to the centroid [17]. The next step is to update the centroid position for each cluster. The new centroid is obtained by calculating the average position of all data points in the cluster, using the equation [18]:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

Description:

- μ_k = centroid point of the k-cluster
- N_k = number of data in the k-cluster
- x_q = q-data in the k-cluster

Using this method, data can be grouped into clusters that have similar characteristics based on the minimum distance to the centroid of each cluster [19].

2.4. Elbow Method

The Elbow method is used to determine the optimal number of clusters by looking at changes in the Within-Cluster Sum of Squares (WCSS) value against the number of clusters (K). The WCSS value will decrease as K increases, but at a certain point the decrease will slow down and form an elbow. This point is considered the optimal number of clusters, as adding clusters after this point does not provide a significant improvement in clustering performance [20].

2.5. BIRCH Clustering

Birch Clustering (Balanced Iterative Reducing and Clustering using Hierarchies) is a clustering algorithm designed to handle large-scale datasets and is efficient in terms of memory and computation time [21]. Birch builds a tree structure called CF Tree (Clustering Feature Tree) to incrementally cluster data hierarchically by summarizing the statistical information of the data in its nodes. This process allows Birch to perform incremental and effective clustering on large data without having to load the entire data into

memory. This algorithm is suitable for high-dimensional data and performs well when the number of clusters is not too large and the data distribution is relatively balanced [22].

2.6. Silhouette Score Evaluation

Silhouette Score is used to evaluate the quality of cluster formation. The value of this score ranges from -1 to 1. The closer to 1, the better the cluster is well-defined and separated from other clusters. Values close to 0 indicate that the data is at the boundary between two clusters, while negative values indicate the possibility of misclustering [23].

3. Result and Discussion

3.1. K-Means Clustering

Determining the optimal number (K) of clusters can be done using the Elbow method. Based on the graph of the relationship between the number of clusters (K) and the Within-Cluster Sum of Squares (WCSS) value, it can be seen that WCSS decreases sharply from K = 1 to K = 4, then the decline begins to slope. The elbow point on the graph occurs at K=4, so the optimal number of clusters is chosen as 4. With this number, the model can group data effectively without causing excessive complexity.

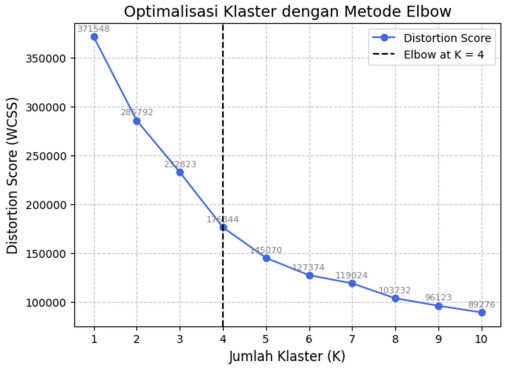

Figure 1. Elbow Method

After determining the number of clusters using the elbow method, the next process is centroid selection. Centroid selection is done with the assumption of taking directly from the available data [24]. The first centroid (Cluster 1) is taken from Data 1, the second centroid (Cluster 2) is taken from Data 2, the third centroid (Cluster 3) is taken from Data 3, and the fourth centroid (Cluster 4) is taken from Data 4. With this assumption, each initial centroid represents one unique data with different characteristics based on latitude, longitude, depth, and magnitude values. After determining the initial centroid of the data, the next step is to calculate the distance of all data to each centroid using the Euclidean Distance formula. The end result shows the division of clusters where each data is grouped according to the similarity of latitude, longitude, depth, and magnitude values after normalization, providing a more structured picture of the distribution of earthquake characteristics. So the PCA graph is obtained as follows,

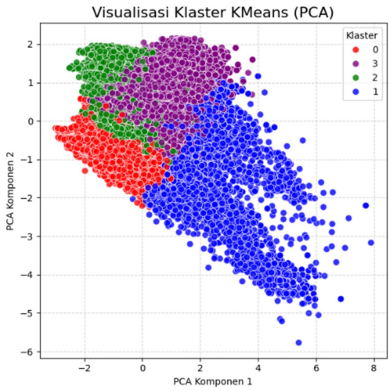

Figure 2. PCA K-Means

Figure 2 displays the clustering results using the K-Means algorithm in the form of a 2-dimensional scatter plot after processing with PCA. Each point represents earthquake data, and different colors indicate the clusters formed based on data similarity.

Table 3. K-Means Algorithm Clustering Results

No	lat	lon	depth	mag	cluster_kmeans
1	-9.18	119.06	10	4.9	0
2	-6.55	129.64	10	4.6	3
3	-7.01	106.63	121	3.7	0
4	-3.30	127.85	10	3.2	3
5	-6.41	129.54	70	4.3	3
…	…	…	…	…	…

Table 3 shows some of the results of clustering earthquakes in Indonesia using K-Means. The cluster_kmeans column shows the cluster to which each data belongs based on the similarity of parameters such as location, depth and magnitude.

3.2. BIRCH Clustering

The BIRCH algorithm uses an efficient hierarchical tree approach, where data is incrementally added to the Clustering Feature Tree. Each node contains summary statistics (number of data, number of vectors, and number of squares) and uses a radius threshold to determine whether to include new data in existing clusters or create new clusters [25]. This process takes place automatically, without the need for centroid initialization as in K-Means. So the PCA graph is obtained as follows,

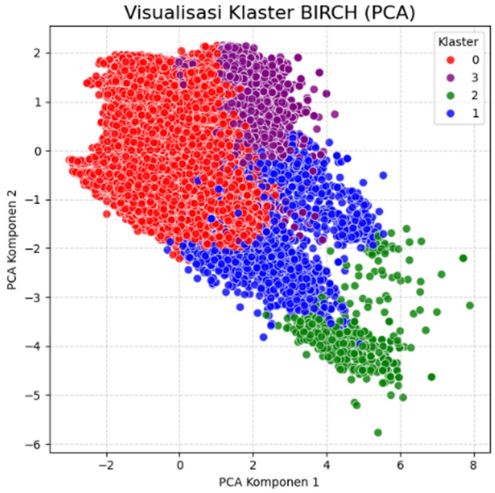

Figure 3. PCA BIRCH

In the graph above, each color indicates the cluster formed by BIRCH. It can be seen that BIRCH has successfully grouped the data based on similar characteristics, although most of the data falls into one dominant cluster,

Table 4. BIRCH Algorithm Clustering Results

No	lat	lon	depth	mag	Cluster_birch
1	-9.18	119.06	10	4.9	0
2	-6.55	129.64	10	4.6	0
3	-7.01	106.63	121	3.7	0
4	-3.30	127.85	10	3.2	0
5	-6.41	129.54	70	4.3	0
…	…	…	…	…	…

Table 4 shows a portion of the earthquake data that has been clustered using the BIRCH algorithm. All the data in this example belong to the same cluster, cluster 0. This shows that BIRCH identifies strong similarities between the data based on location, depth and magnitude parameters.

3.3. Comparison

After clustering using each algorithm, the next step was to compare the results of both. The first comparison focused on the data distribution of each cluster formed, to see how each algorithm categorized the earthquake data.

Table 5. Distribution of Clustering Results

Algorithm	Cluster 0	Cluster 1	Cluster 2	Cluster 3
K-Means	30142	7815	11884	43046
BIRCH	81267	2532	553	8535

The distribution of clustering results shows significant differences between the K-Means and BIRCH algorithms in clustering earthquake data. K-Means produces a relatively balanced distribution between clusters, with the largest amount of data in Cluster 3 (43,046 data) and the smallest in Cluster 1 (7,815 data). Meanwhile, BIRCH produces a very unequal distribution, where most of the data (81,267 data) is concentrated in Cluster 0, while the other three clusters have much less data, especially Cluster 2 which only contains 553 data. This difference indicates that K-Means tends to divide data evenly based on distance, while BIRCH is more sensitive to hierarchical structure and data density. The next step is to compare based on Silhouette Score and execution time.

Table 6. Performance Comparison of Clustering Algorithms

Algorithm	Silhouette Score	Runtime
K-Means	0.3501	0,14
BIRCH	0.2247	3,99

Based on the comparison results in the Silhouette Score and runtime tables, the K-Means algorithm performs quite better than BIRCH. K-Means produces a Silhouette Score value of 0.3501 which indicates better cluster separation and clearer cluster structure, compared to BIRCH which only achieves a score of 0.2247. In addition, the execution time of K-Means (0.14 seconds) is also faster than BIRCH (3.99 seconds), indicating that K-Means is superior in processing large amounts of earthquake data. Next is to see the comparison using map visualization.

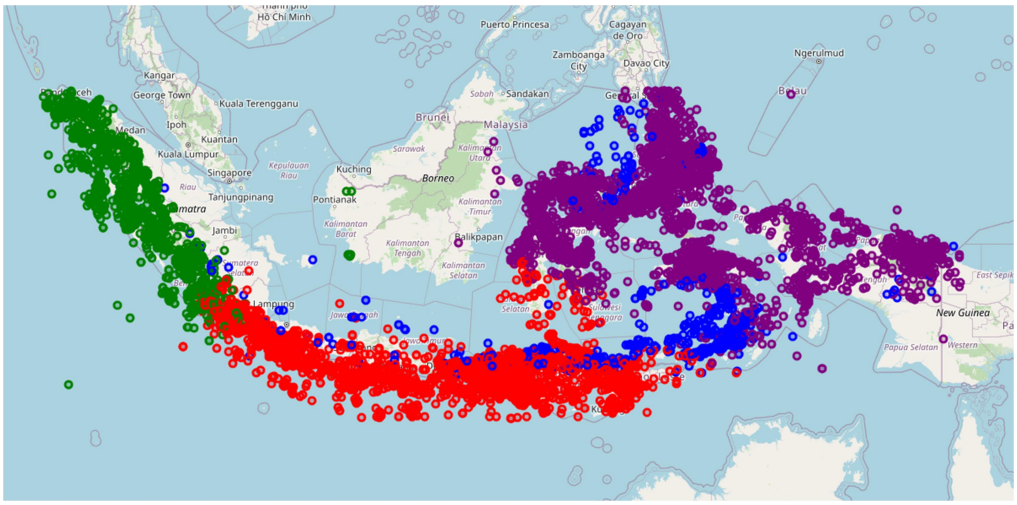

Figure 4. K-Means Clustering Result Map

Figure 4 displays the clustering results using the K-Means algorithm, which shows a more balanced distribution of clusters. Each color on the map represents a cluster that successfully groups the earthquake data based on clear geographical patterns, such as the western, central and eastern regions of Indonesia. K-Means is able to better distinguish earthquake-prone areas, creating a sharp and spatially meaningful cluster separation. This suggests that K-Means has superior performance in identifying earthquake distribution patterns based on location, producing more informative and representative visual results.

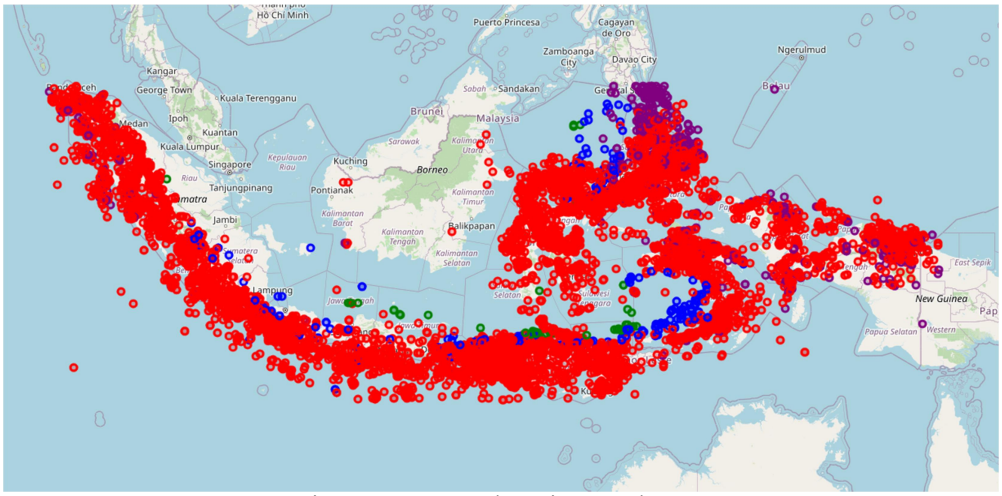

Figure 5. BIRCH Clustering Result Map

Figure 5 shows the results of clustering visualization using the BIRCH algorithm on earthquake data in Indonesia. It can be seen that the distribution of data on the map is dominated by one color (red), indicating that most of the earthquake points belong to one large cluster. This reflects that BIRCH is less than optimal in distinguishing earthquake characteristics based on the attributes used, as only a few clusters are clearly formed. This unbalanced distribution of clusters indicates that the BIRCH model tends to group the data in a general way without sharp separation, in contrast to the more structured K-Means results.

Table 7. Clustering Result

No	lat	lon	depth	mag	cluster_kmeans	cluster_birch
1	-9.18	119.06	10	4.9	0	0
2	-6.55	129.64	10	4.6	3	0
3	-7.01	106.63	121	3.7	0	0
4	-3.30	127.85	10	3.2	3	0
5	-6.41	129.54	70	4.3	3	0
…	…	…	…	…	…	…
92882	3.24	127.18	10	4.0	3	0
92883	2.70	127.10	10	3.9	3	0
92884	-7.83	121.07	10	3.8	0	0
92885	3.00	127.16	10	4.1	3	0
92886	-8.87	118.95	10	2.4	0	0

Table 7 shows the results of clustering earthquake data based on latitude, longitude, depth, and magnitude (mag) using the K-Means and BIRCH algorithms. It can be seen that BIRCH tends to group most of the data into one cluster (cluster 0), while K-Means is able to divide the data more evenly into several different clusters. This is consistent with the previous visualization results and silhouette score values, which show that K-Means has better clustering quality.

However, this was the case for the initial dataset, which was only from 2008-2023 and used key attributes such as latitude, longitude, depth, and magnitude. To test the performance of the algorithm with more and more recent data from 2008-2025, an experiment was conducted using a dataset of 121,123 data and the addition of attributes such as mag_type, phasecount, azimuth_gap, and several other parameters. The BIRCH clustering results are more evenly distributed than the previous experiments, with a clearer and more balanced cluster distribution.

Table 8. Comparison of Clustering Algorithm Performance with Latest Data

Algorithm	Silhouette Score	Runtime
K-Means	0,1293	0,25
BIRCH	0,3489	0,37

This is shown by the Silhouette Score value of 0.3489, which is higher than that of K-Means which only reaches 0.1293. In terms of efficiency, the execution time of BIRCH was recorded at 0.37 seconds, slightly slower than K-Means which only required 0.25 seconds.

Table 9. Distribution of Clustering Results with Latest Data

Algorithm	Cluster 0	Cluster 1	Cluster 2	Cluster 3
K-Means	7163	39164	40525	34271
BIRCH	81267	2532	553	8535

The data distribution of the clustering results also shows significant differences. In K-Means, the data is spread across four clusters with a relatively even distribution: Cluster 2 (40,525 data), Cluster 1 (39,164 data), Cluster 3 (34,271 data), and Cluster 0 (7,163 data). Meanwhile, BIRCH produces a distribution that is more concentrated in Cluster 0 (110,973 data), followed by Cluster 1 (5,707 data), Cluster 2 (4,140 data), and Cluster 3 (303 data).

This achievement shows that BIRCH is superior in handling large datasets with complex attributes, as its hierarchical clustering process is able to absorb attribute information in more detail, resulting in a more accurate and structured separation even though it requires a slightly longer execution time.

3.4. Website Implementation

After the clustering process is complete, the clustering results are visualized in the form of a web-based interactive map. This map displays the distribution of earthquakes based on the clusters formed, making it easier to identify earthquake-prone areas more clearly. Flask technology is used as the backend to connect the clustering results with the web interface.

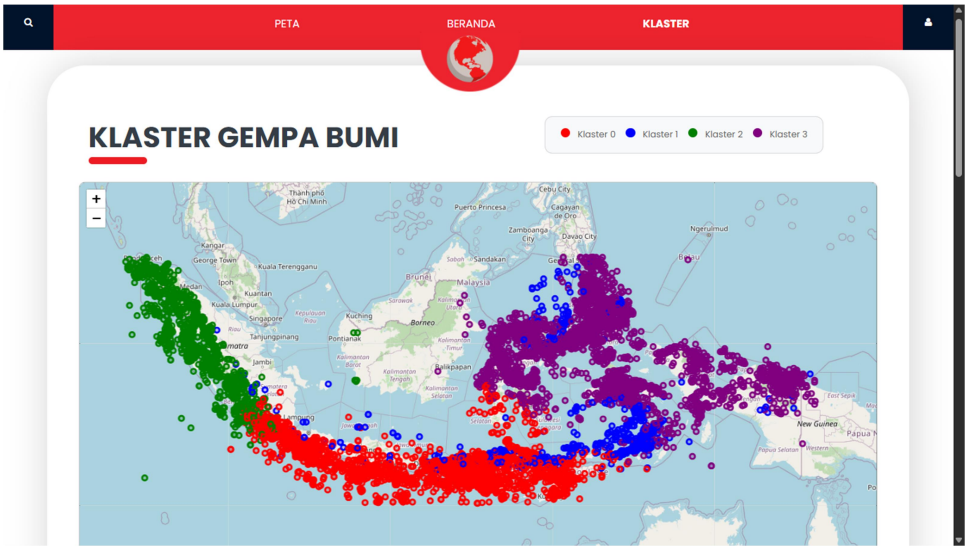

Figure 6. Earthquake Cluster Website

This map shows that each cluster groups earthquake data that are similar in terms of both geographic location and seismic characteristics. The results of this clustering help identify areas with similar earthquake characteristics, which can later be visualized through an interactive map to understand the distribution of earthquake-prone areas in Indonesia more systematically.

4. Conclusion

The implementation of K-Means and BIRCH clustering algorithms on earthquake data in Indonesia (based on location, depth, magnitude, as well as additional attributes such as mag_type, phasecount, and azimuth_gap) shows that both are able to form earthquake clusters with certain patterns. In the initial experiment, K-Means recorded a Silhouette Score of 0.3501, higher than BIRCH (0.2247). However, after adding attributes and increasing the amount of data to 121,123, BIRCH's performance improved significantly (0.3489), surpassing K-Means (0.1293). This shows that BIRCH is more optimal for large and complex data, thanks to its hierarchical clustering approach. In terms of efficiency, BIRCH takes 0.37 seconds, slightly longer than K-Means (0.25 seconds), but produces more representative clusters. Visualization in a web-based interactive map facilitates spatial analysis of earthquake patterns in Indonesia.

This result opens up opportunities for further development of spatial-temporal analysis. For future research, it is recommended to consider additional spatial/temporal attributes and alternative algorithms such as DBSCAN or HDBSCAN that are more adaptive to noise and irregular distribution.

References

[1] B. T. Ujianto and Amar Rizqi Afdholy, "Kearifan Lokal Dalam Desain Tahan Gempa: Studi Komparatif Rumah Tradisional Di Wilayah Indonesia Barat," *Pawon J. Arsit.*, vol. 8, no. 02, pp. 255–272, Jul. 2024, doi: 10.36040/pawon.v8i02.10768.

[2] Program Studi PWK Universitas Teknologi Yogyakarta and B. V. T. Dewi, "Pemetaan Perubahan Kondisi Sosial Ekonomi Masyarakat Pasca Gempa Bumi di Kecamatan Tanjung, Kabupaten Lombok Utara," *Tata Kota Dan Drh.*, vol. 12, no. 2, pp. 83–93, Dec. 2020, doi: 10.21776/ub.takoda.2020.012.02.5.

[3] T. S. Chairani, H. Listia, S. Wardaniah, S. Wulandari, P. T. Agustina, and A. Piliang, "Klasterisasi Daerah Kriminalitas Di Indonesia Dengan Metode K-Means Clustering," 2024.

[4] I. Saleh, G. Mandar, and J. Noh, "Analisis Data Gempa Di Maluku Utara Menggunakan Algoritma K-Means Dan Liner Regression," . *September*, vol. 16, no. 2, 2023.

[5] N. Dwitiyanti, S. A. Kumala, and S. D. Handayani, "Penerapan Metode K-Means Pada Klasterisasi Wilayah Rawan Gempa Di Indonesia".

[6] D. P. Az-Zahra, Y. S. Sidabutar, and A. J. M. Khan, "Implementasi Algoritma Birch Dalam Klasterisasi Kasus Biaya Hidup Di Kota Pada Beberapa Negara".

[7] D. Malik, I. M. Artha Agastya, A. Y. Anjaya, Kusrini, and S. Hartantyo, "Earthquake Distribution Mapping in Indonesia Using K-Means Clustering Algorithm," in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, Yogyakarta, Indonesia: IEEE, Aug. 2024, pp. 499–504. doi: 10.1109/ICITISEE63424.2024.10730706.

[8] A. R. Rizalde, H. A. Mubarak, G. Ramadhan, and Mohd. A. Fatan, "Comparison of K-Means, BIRCH and Hierarchical Clustering Algorithms in Clustering OCD Symptom Data," *Public Res. J. Eng. Data Technol. Comput. Sci.*, vol. 1, no. 2, pp. 102–108, Feb. 2024, doi: 10.57152/predatecs.v1i2.1106.

[9] Y. Rifa'i, "Analisis Metodologi Penelitian Kulitatif dalam Pengumpulan Data di Penelitian Ilmiah pada Penyusunan Mini Riset," *Cendekia Inov. Dan Berbudaya*, vol. 1, no. 1, pp. 31–37, Jun. 2023, doi: 10.59996/cendib.v1i1.155.

[10] BMKG and USGS, "Earthquakes in Indonesia." Kaggle. doi: 10.34740/KAGGLE/DSV/11641367.

[11] A. E. Satriatama *et al.*, "Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 3, pp. 172–182, Jul. 2023, doi: 10.47233/jteksis.v5i3.828.

[12] A. N. Haya and M. Y. Ramme, "Penerapan Algoritma Stacking Ensemble Machine Learning Berbasis Pohon untuk Prediksi Penyakit Diabetes," *Pros. Semin. Nas. SAINS DATA*, vol. 4, no. 1, pp. 954–961, Oct. 2024, doi: 10.33005/senada.v4i1.388.

[13] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *J. SISKOM-KB Sist. Komput. Dan Kecerdasan Buatan*, vol. 7, no. 1, pp. 31–40, Oct. 2023, doi: 10.47970/siskom-kb.v7i1.464.

[14] P. Suwito and H. Henny, "Clustering Penilaian Dosen Berdasarkan Indeks Kepuasan Mahasiswa," *Simtek J. Sist. Inf. Dan Tek. Komput.*, vol. 6, no. 2, pp. 122–127, Oct. 2021, doi: 10.51876/simtek.v6i2.104.

[15] D. Dona and M. Rifqi, "Penerapan Metode K-Means Clustering Untuk Menentukan Status Gizi Baik Dan Gizi Buruk Pada Balita (Studi Kasus Kabupaten Rokan Hulu)," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, vol. 7, no. 2, pp. 179–191, Jul. 2022, doi: 10.36341/rabit.v7i2.2171.

[16] A. Kusmiran, Minarti, M. F. I. Massinai, A. Zarkasi, A. A. Maharani, and R. Desiani, "Klasifikasi Kedalaman Kejadian Gempa Menggunakan Algoritma K-Means Clustering: Studi Kasus Kejadian Gempa Di Sulawesi," *JFT J. Fis. Dan Ter.*, vol. 9, no. 2, pp. 79–88, Dec. 2022, doi: 10.24252/jft.v9i2.29198.

[17] M. H. Abdurrohman, E. Haerani, F. Syafria, and L. Oktavia, "Implementasi K-Means Clustering Pada Data Pengelompokan Pendaftaran Mahasiswa Baru (Studi Kasus Universitas Abdurrab," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, vol. 9, no. 1, pp. 138–147, Jan. 2024, doi: 10.36341/rabit.v9i1.4255.

[18] N. K. Zuhal, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," vol. 1, 2022, doi: 10.29407/stains.v1i1.1495.

[19] S. Mutiah, Y. Hasnataeni, A. Fitrianto, E. Erfiani, and L. M. R. D. Jumansyah, "Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat," *Teorema Teori Dan Ris. Mat.*, vol. 9, no. 2, p. 247, Sep. 2024, doi: 10.25157/teorema.v9i2.16290.

[20] F. P. Azizah, S. S. Hilabi, and A. Hananto, "Perbandingan Algoritma K-Means dan Hierarchical Untuk Klasterisasi Data Kehadiran Karyawan," vol. 14, no. 1.

[21] N. J. Benzer, "Balanced Iterative Reducing And Clustering Using Heirarchies(Birch)," Medium. Accessed: May 06, 2025. [Online]. Available: https://medium.com/@noel.cs21/balanced-iterative-reducing-and-clustering-using-heirachies-birch-5680adffaa58

[22] J. Laurenso, D. Jiustian, F. Fernando, V. Suhandi, and T. H. Rochadiani, "Implementation of K-Means, Hierarchical, and BIRCH Clustering Algorithms to Determine Marketing Targets for Vape Sales in Indonesia," *J. Appl. Inform. Comput.*, vol. 8, no. 1, pp. 62–70, Jul. 2024, doi: 10.30871/jaic.v8i1.4871.

[23] G. M. M. Sujak, H. N. Rofiq, and F. I. Tawakal, "Implementasi K-Means Clustering untuk Optimalisasi Anggaran Penyakit Tidak Menular: Implementation of K-Means Clustering for Optimizing Non-Communicable Disease Budgets," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 1, pp. 67–74, Nov. 2024, doi: 10.57152/malcom.v5i1.1597.

[24] E. Ramadanti and M. Muslih, "Penerapan Data Mining Algoritma K-Means Clustering Pada Populasi Ayam Petelur Di Indonesia," *Rabit J. Teknol. Dan Sist. Inf. Univrab*, vol. 7, no. 1, pp. 1–7, Jan. 2022, doi: 10.36341/rabit.v7i1.2155.

[25] P. D. R. SARI, "Pengelompokan Tingkat Penjualan Smartphone Toko Offline Menggunakan Algoritma Birch Clustering," S1, Universitas Malikussaleh, 2024. Accessed: May 21, 2025. [Online]. Available: https://rama.unimal.ac.id/id/eprint/9229/