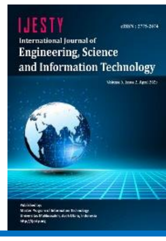
International Journal of Engineering, Science and Information Technology
Volume 5 No. 2 (2025) pp. 455-461
ISSN 2775-2674 (online)
Website: http://ijesty.org/index.php/ijesty
DOI: https://doi.org/10.52088/ijesty.v5i2.863
Research Paper, Short Communication, Review, Technical Paper



Evaluating the Quality of Agglomerative Hierarchical Clustering on Crime Data in Indonesia

Dini Dara Rizkya*, Sujacka Retno, Zara Yunizar

Department of Informatics, Faculty of Engineering, Universitas Malikussaleh, Aceh, Indonesia

**Corresponding author Email: dini.200170050@mhs.unimal.ac.id*

The manuscript was received on 18 November 2024, revised on 28 December 2024, and accepted on 28 March 2025, date of publication 2 May 2025

Abstract

This study evaluates the quality of Agglomerative Hierarchical Clustering with single Linkage, complete Linkage, average Linkage, and ward linkage on the dataset of the number of criminal cases in Indonesia (2000-2023). The analysis compares clustering performance on the original and normalized datasets using the Davies-Bouldin Index (DBI), Silhouette Score (SS), Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Callinski-Harabasz Index (CH). The results showed that Ward Linkage provided the best clustering results, with the highest CH increasing from 65.826 to 66.873, clear cluster separation, and a stable structure (NMI = 0.5855, ARI = 0.6298). Single Linkage experienced a chaining effect, although it showed improvement in DBI from 0l.1793 to 0.1765 and SS from 0.6271 to 0.6400, with NMI and ARI stable at 0.4537 and 0.5865, but CH decreased from 21.731 to 21.072. Complete Linkage was too aggressive in separating the data, shown by an increase in DBI from 0.5327 to 0.7116 and a decrease in SS from 0.6336 to 0.5830, although CH increased from 64.244 to 66.873. Average Linkage showed stable results, with NMI = 0.6481 and ARI = 0.7993 remaining, but a slight decrease in DBI from 0.3874 to 01.4091, SS from 0.6839 to 0.6825, and CH from 42.358 to 40.251. Data normalization generally helps to improve clustering quality by reducing the influence of feature scale differences. Several metrics showed improved cluster separation on normalized data, although the impact varied depending on the linkage method. Overall, Ward Linkage with normalization is recommended as the best method to produce more accurate clustering in Indonesia's crime data analysis.

Keywords: *Agglomerative Hieral Clustering, Crime Analysis, Evaluation Metrics, Data Normaization.*

1. Introduction

The number of crimes by regional police is data or statistics on cases of regional violations recorded and reported by the police in a particular area or region. This data includes various types of crimes that occur in the area and are recorded by police officers as part of their duties in maintaining public security and order [1].

According to data from (BPS, 2012), the total number of criminal acts in Indonesia in 2012 was 584,991,0101. Several ways can be used to categorize the level of vulnerability to criminal acts, according to the regional police. To better understand the level of vulnerability to criminal acts in several regions, a method is needed to group areas based on their level of vulnerability. One method that can be used is clustering, a technique in data mining that groups data based on the same characteristics. Clustering is divided into hierarchical and non-hierarchical—hierarchical clustering groups data based on the same characteristics, including algebraic and divisive clusters.

Agglomerative has several models: single Linkage, complete Linkage, average Linkage, and ward linkage. These models can be used to analyze the level of vulnerability to crime in regional police, which has significant value in Indonesia. This study evaluates the Agglomerative Hierarchical Clustering method with four linkage models (single, complete, average, and ward) to classify crimes globally based on regional police in Indonesia. Model performance is measured using the Davies-Bouldin Index (DBI), Normalized Mutual Information (NMI), Callinski-Hallarbalsz, Silhouette Score, and Adjusted Rand Index (ARI) metrics to determine which linkage model is the most effective and efficient [2].

2. Literature Review

2.1. Previous Research

Yanuwar Reinaldi, Nurissalidah Ulinnuhal, Tony Hartono, and Moh conducted research. Hafiyusholeh (2021), entitled “Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare in East Java”, In this study, the results of the calculation found that the average linkage method with three clusters is the best calculation with a silhouette index value of 0.6054, with the 1st cluster there are 23 regions, namely cities/districts with the highest community welfare, the 2nd cluster there are 11 regions namely cities/districts with moderate community welfare, and the 3rd cluster there are 4 regions, namely cities/districts with the lowest community welfare[3].

Research conducted by Septian Wulandari (2023) entitled “Clustering Provinces in Indonesia on the Prevalence of Toddler Stunting Using Agglomerative Hierarchical Clustering”. The results of this study are two clusters, namely Cluster 1, which has a relatively high prevalence of stunting with members in 13 provinces, and Cluster 2, which has a relatively low prevalence of stunting with members in 21 provinces. The highest chopenetic correlation value is found in World's algorithm with an a value of 0.8399978. So, it can be said that Ward's algorithm is better than the Average algorithm[4].

2.2. Data Mining

Data mining is extracting information from all data sets using algorithms and inference techniques in statistics, machine learning, and data-based marketing systems. Data mining is analyzing data from different perspectives and summarizing it into vital information that can be used to increase profits, reduce expenses, or even both. In the World of computer science, based on its function, the function of data mining is divided into six parts, namely description, estimation, prediction, classification, clustering, and association.[5]. Data mining aims to find patterns or valuable information from large, complex data sets.[6].

2.3. Clustering

Clustering is a data mining component, which is extracting interesting patterns from extensive data. Clustering is the grouping of objects into a group so that in one cluster, objects have similarities and are different from other objects in other clusters. Clustering is divided into two methods, namely hierarchical and non-hierarchical. Hierarchical is a data clustering method that begins by grouping several objects with the closest similarity, then proceeds to other objects with the closest distance so that the cluster will form a hierarchy of the most similar and the least similar [7]. There are four types of data commonly used in clustering, namely interval values, binary values, nominal values, ordinal values, and ratio values, as well as other data types. [8]. Several methods can be employed in clustering procedures, including K-Means, SOM, K-Medoids, Fuzzy C-Means, and AHC. [9].

2.4. Agglomerative Hieral Clustering

Agglomerative Hierarchical Clustering is a data clustering method that combines two clusters that have similarities[10]. In this method, the number of clusters is not determined specifically. In some cases, determining the number of clusters is also often done by combining other methods. This technique produces a hierarchical structure as a dendrogram, which describes how the clusters join at each step [11]. There are several models for measuring closeness between clusters, such as single linkage, complete linkage, average linkage and ward linkage [12].

2.5. Single Linkage

The Single Linkage method is one of the Agglomerative Hierarchical Clustering algorithm techniques used to group data based on the distance between distant points. This method defines the distance between two clusters as the minimum distance between pairs of tall points from two clusters [13].

$$d_{(UV)W} = \min\,(d_{UW}, d_{VW}) \tag{1}$$

Description:

$d_{(UV)W}$Distance between cluster (UV) and cluster W

d_{UW} and d_{VW}: distance between nearest neighbors of cluster U and W, and cluster V and W.

2.6. Complete Linkage

The Complete Linkage method uses the principle of distance between objects. The basis for the determination is the maximum distance or the farthest distance [14]. The distance between clusters is determined by the distance between two objects, one from each cluster, that is farthest away. After that the clusters U and V are merged into cluster (UV) the distance between cluster (UV) and other clusters.

$$d_{(UV)W} = \max\,(d_{UW}, d_{VW}) \tag{2}$$

Description:

$d_{(UV)W}$: distance between cluster (UV) and cluster W

d_{UW} and d_{VW}: distance between nearest neighbors of cluster U and W, and cluster V and W.

2.7. Average Linkage

Average Linkage method calculates the distance between two clusters, all the distances between clusters, namely, with d_{ik} Is the distance between object i in cluster (UV) and object k in cluster W? While NUV and Nw are the number of objects in the cluster (UV) and (W), respectively [15].

$$d_{(UV)W} = \frac{\sum i \sum k \, d_{ik}}{N_{(uv)} \, N_W} \quad (3)$$

Description:

d_{ik}: distance between object i and in cluster (UV) and object k in cluster W

$N_{(uv)}$ and $N_w$all the number of objects in the cluster (UV) and cluster W [16].

2.8. Ward Linkage

Ward's Linkage method is used in hierarchical clustering to determine the distance or similarity between clusters in the merging process. [17]. This method is designed to minimize the total variance within clusters and produce clustering that has less variance within clusters. [18].

$$\mathrm{SSE} = \Sigma(\mathrm{xj} - \mathrm{x}\,)'(\mathrm{xj} - \mathrm{x}\,)\mathrm{Nj} = 1 \quad (4)$$

Description:

Xj: is an a-column vector containing the values of object j with j=1,2...,N

3. Research Methods

The data used in this study were obtained from the Badan Pusat Statistik (BPS) through its official website, bps.go.id, one of Indonesia's primary sources of statistical data. This dataset includes data on criminal cases throughout Indonesia from 2000 to 2023, covering 34 provinces. All variable in the dataset contains information on the province, year, and total number of crimes reported. With such extensive data, this research can provide in-depth insights into crime trends in Indonesia.

The following is Dalltall's presentation on criminal cases based on regional police forces in all provinces in Indonesia.

Table 1 Original Dataset

Province	2000	2001	2002	2003	2004	2005	2006	…	2023
ACEH	4.286	3.420	1.668	2.724	1.873	2.181	986		8.159
NORTH SUMATRA	15.887	15.395	15.063	17.530	20.924	25.111	27.785		35.366
WEST SUMATRA	4.464	4.879	4.845	5.842	5.387	7.203	9.953		9.073
RIAU	4.542	5.341	5.571	7.020	7.151	6.855	6.277		8.382
JAMBI	1.667	1.493	1.554	1.793	1.984	2.202	1.969		5.386
SOUTH SUMATRA	10.754	10.152	10.502	7.534	7.328	8.579	8.294		12.617
BENGKULU	941	676	1.170	1.159	1.086	1.100	1.654		3.456
LAMPUNG	5.473	5.265	3.290	3.697	4.624	4.253	6.052		9.175
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PAPUA	2.678	2.522	3.555	3.694	4.749	5.387	5.549		7.017

In this study, we tested clustering using the original dataset and the standardized normalized with Stallard's Scale to see which works better for clustering with the ALL hierarchical clustering algorithm. The default normalized with StandardScaler is shown in Table 2 below.

Table 2. Normalized Dataset

Province	2000	2001	2002	2003	2004	2005	…	2023
ACEH	-0,12828	-0,28485	-0,50057	-0,37909	-0,45692	-0,48246		-0,02554
NORTH SUMATRA	1,759196	1,349846	1,285407	1,450343	1,425751	1,579815		3,239503
WEST SUMATRA	-0,09932	-0,08568	-0,07698	0,006167	-0,10966	-0,03079		0,084143
RIAU	-0,08663	-0,02261	0,019823	0,151721	0,064665	-0,06209		0,001218
JAMBI	-0,55439	-0,5479	-0,51577	-0,49413	-0,44595	-0,48057		-0,35833
SOUTH SUMATRA	0,924058	0,634131	0,677281	0,215232	0,082157	0,092961		0,50945
BENGKULU	-0,67251	-0,65943	-0,56697	-0,57247	-0,5347	-0,57968		-0,58994
LAMPUNG	0,064841	-0,03299	-0,28431	-0,25887	-0,18506	-0,29611		0,096384
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
PAPUA	-0,38991	-0,40743	-0,24897	-0,25924	-0,17271	-0,19412		-0,16259

In this study, we used the Agglomerative Hieral Clustering method on the original crime dataset and the normalized dataset. The idea is to use these datasets for clustering and to check the results with various metrics to ensure all analyses. Application of Agglomerative Hieral Clustering Method. This process includes calculating each model, namely Single Linkage, Complete Linkage, Average Linkage, and Ward Linkage, to group data on the number of criminal calls reported to regional police in all provinces in Indonesia. And then we perform method validation. This process includes measuring the level of accuracy for each method determined using five evaluation metrics, namely the Davies-Bouldin Index (DBI), Silhouette Score, Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Callinski-Harabasz. [19].

The research method involves all activities to obtain detailed principles that will be used as guidelines in conducting research, including collecting, writing, and studying systematically. The stages carried out in this study are as follows:

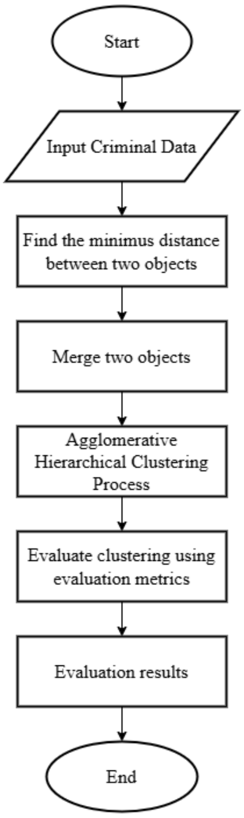


Fig 1. Flowchart system

Based on the system framework designed above, the system that will be run will go through stages from start to end, where each stage will be explained.

a. Start
   The process begins
b. Input Criminal Data
   Criminal data is input into the system. This data could include crime types, locations, dates, or other relevant features.
c. Find the Minimum Distance Between Two Objects
   The algorithm calculates the distance between all pairs of data points and identifies the two objects (data points) that are closest to each other based on a chosen distance metric (e.g., Euclidean distance).
d. Merge Two Objects
   The two closest objects all merged into one cluster.
e. Agglomerative Hieral Clustering Process
f. Each object starts in its cluster.
   The algorithm repeatedly merges the two closest clusters.
g. This continues until a stopping criterion is met (such as a desired number of clusters or all objects merged into one cluster).
   Evaluate Clustering Using Evation Metrics
h. After the clustering process is complete, the results are all evaluated using metrics such as:
   Silhouette Score, Davies–Bouldin Index, Calinski–Harabasz Index, Adjusted Rand Index, Normalized Mutual Information. These metrics help assess how well the tall has have been grouped. Evaluation Results
   The evaluation results are all presented to show the quality and effectiveness of the clustering process.
i. End
   The process ends.

4. Results and Discussion

4.1. Research Result

In this study, the application of the Agglomerative Hierarchical Clustering method aims to determine the evaluation results of the calculation process of each model, namely Single Linkage, Complete Linkage, Average Linkage, and Ward Linkage, to group data on the number of criminal calls made on regional police in all provinces in Indonesia. Then, the allclusivity level will be tested using five evaluation metrics, including the Davies-Bouldin Index (DBI), Silhouette Score, Normalized Mutual Information (NMI), adjusted Rand Index (ARI), and Callinski-Harabasz.[20]. In this study, the number of clusters formed is k=3, namely, 1 = “not vulnerable,” 2 = " vulnerable, " and 3 = “very vulnerable.”

4.2. Visualization Single Linkage Clustering Result Comparison

In this study, we used the Agglomerative Hierarchical Clustering method on the original crime dataset and the normalized dataset. The idea is to use these datasets for clustering and to check the results with various metrics to ensure all analyses.

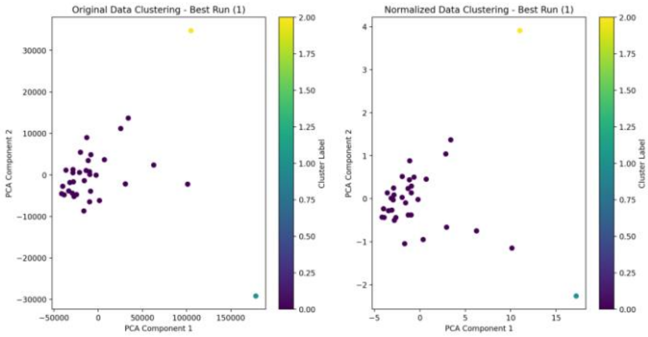


Fig 2. Single Linkage Clustering Result Comparison

The graph above compares the clustering results using PCA before and after normalization with the Single Linkage method in Agglomerative Hierarchical Clustering. The left graph shows clustering on the original data with differences in scale between features. In contrast, the right graph displays the results after normalization, resulting in a more optimal cluster distribution.

Before normalization, the Single Linkage method produces less structured clusters due to the dominance of large-value features. After normalization, the cluster separation is clearer and more stable as the feature contributions are more balanced. Normalization improves the quality of clustering with Single Linkage, ensuring all more stable clusters are separated without distortion of feature scale.

4.3. Visualization Complete Linkage Clustering Result Comparison

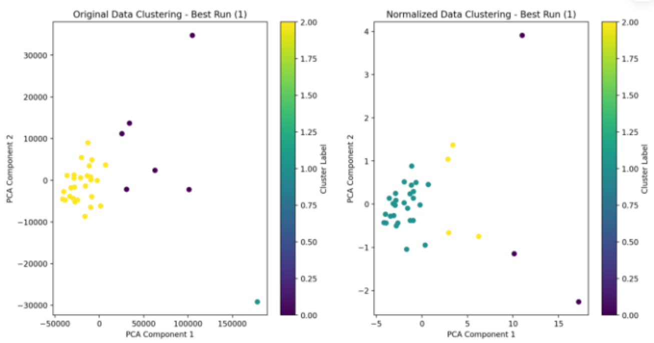


Fig 3. Complete Linkage Clustering Result Comparison

The graph above compares the clustering results using PCA before and after normalization with the Complete Linkage method in Agglomerative Hierarchical Clustering. The left graph shows clustering on the original data with differences in scale between features. In contrast, the right graph displays the results after normalization, resulting in a more even distribution of points and more optimal clustering. Complete Linkage separates clusters based on the maximum distance between points, with normalization improving cluster separation and ensuring more balanced feature contributions.

4.4. Visualization Average Linkage Clustering Result Comparison

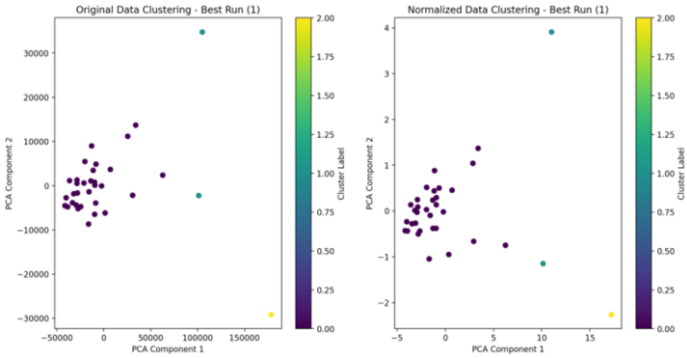


Fig 4 Average Linkage Clustering Result Comparison

The graph above compares the clustering results using PCA before and after normalization with the Complete Linkage method in Agglomerative Hierarchical Clustering. The left graph shows clustering on the original data with differences in scale between features. In contrast, the right graph displays the results after normalization, resulting in a more even distribution of points and more clustering. Complete Linkage separates clusters based on the maximum distance between points, with normalization improving cluster separation and ensuring more balanced feature contributions.

4.5. Visualization Ward Linkage Clustering Result Comparison

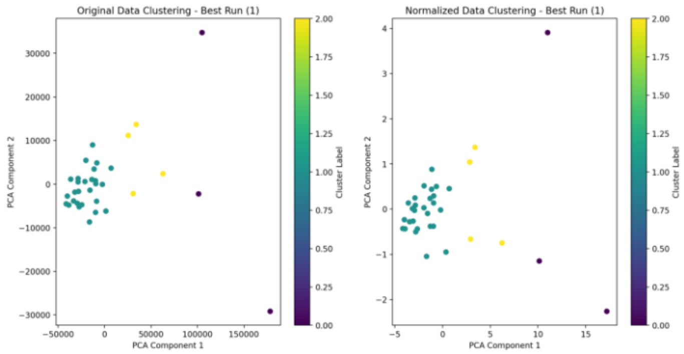


Fig 5. Ward Linkage Clustering Result Comparison

The graph compares the clustering results using PCA before and after normalization with the Ward Linkage method in Agglomerative Hierarchical Clustering. The left graph shows clustering on the original data with a large ranges of values, causing an uneven distribution of points. After normalization (right graph), the dot distribution is more regular and the cluster separation is clearer. Normalization reduces biases due to feature-scale differences, allowing Ward Linkage to form more stable and representative clusters.

4.6. Visualization of Evaluation Metrics

The evaluation metrics for single, complete, average, and ward linkage methods are presented in the following table. The results show significant differences in clustering quality between the original and normalized datasets. curacy evaluation was were performed using five metrics: Davies-Bouldin Index (DBI), Silhouette Score, Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Calinski-Harabasz.

Table 3. Original Data Evaluation Metrix Results

Method	DBI	SS	NMI	ARI	CH
Single Linkage	0.1792847	0.62711285	0.45371277	0.586494302	21.73070092
Complete Linkage	0.532748	0.63363	0.502338	0.598407	64.24351
Average Linkage	0.387364	0.683877	0.648091	0.799264	42.35754
Ward Linkage	0.719096	0.585336	0.585515	0.629804	65.82553

Table 4 Normalization Data Evaluation Metrix Results

Method	DBI	SS	NMI	ARI	CH
Single Linkage	0.1765223	0.6399625	0.45371277	0.5864943	21.071625
Complete Linkage	0.711551	0.583045	0.585515	0.629804	66.87288
Average Linkage	0.409147	0.68249	0.648091	0.799264	40.25125
Ward Linkage	0.711551	0.583045	0.585515	0.629804	66.87288

This study compared four alternative hierarchical clustering methods: Single, Complete, Average, and Ward Linkage on original and normalized data. The evaluation was are done using five main metrics: Davies-Bouldin Index (DBI), Silhouette Score, Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Callinski-Harabasz, to assess the clustering quality based on separation, density, and conformity to ground truth.

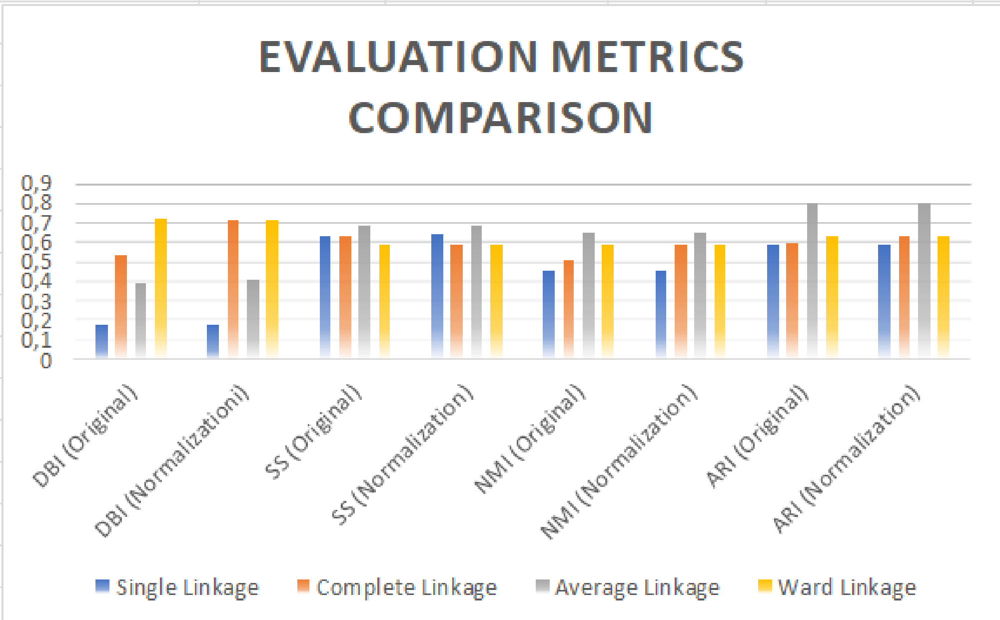


Fig 6. Evaluation Metrics Comparison

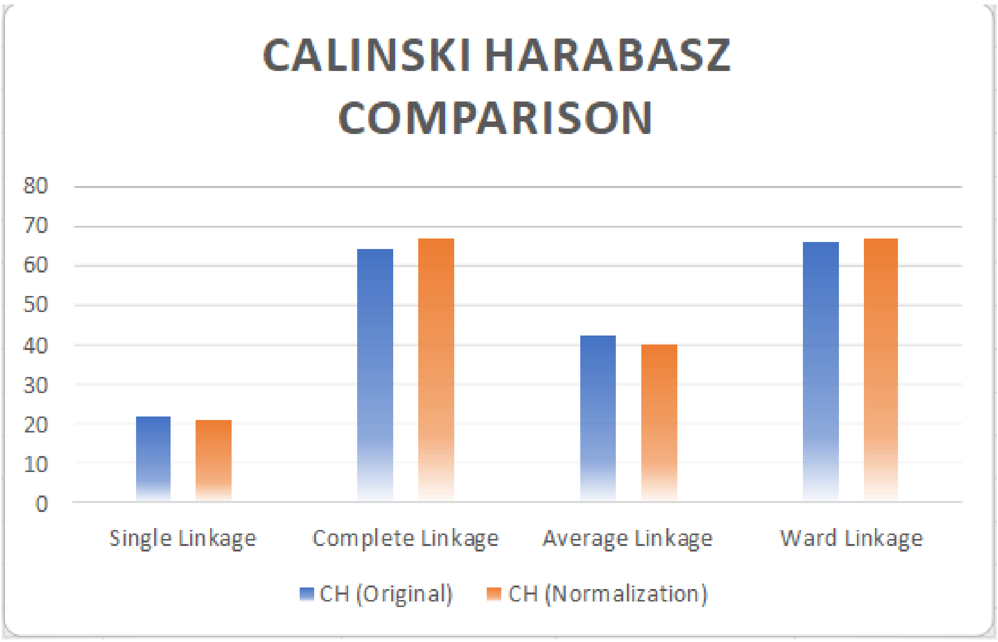


Fig 7. Calinski-Harabasz Evaluation Metrics Comparison

In the Single Linkage method, normalization had little impact on the clustering results. DBI decreased from 0.1793 to 0.1765, indicating a slight increase in cluster separation. SS increased from 0.6271 to 0.6400, indicating better data attachment to the cluster. NMI and ARI remained at 0.4537 and 0.5865, indicating a stable cluster structure. However, CH dropped from 21.731 to 21.072, indicating a slight weakening in cluster distribution and separation.

In the Complete Linkage method, normalization calls for DBI to increase from 0.5327 to 0.7116 and SS to decrease from 0.6336 to 0.5830, indicating a decrease in cluster quality. NMI and ARI remained at 0.5855 and 0.6298, indicating a stable cluster structure. CH increased from 64.244 to 66.873, marking a slight increase in cluster separation, although the overall quality declined.

In the Average Linkage method, the clustering results were relatively stable after normalization. DBI increased from 0.3874 to 0.4091, indicating a slight decrease in cluster separation. SS decreased slightly from 0.6839 to 0.6825, while NMI and ARI remained at 0.6481 and 0.7993, suggesting a consistent cluster structures. CH decreased from 42.358 to 40.251, indicating reduced cluster density. Overall, Average Linkage remained stable despite a slight decrease in quality.

In the Ward Linkage method, normalization did not pose a significant challenge. DBI dropped from 0.7190 to 07115, indicating a slight decrease in cluster separation. SS remained a... the cluster structure did not change. CH increased from 65.826 to 66.873, indicating a slight improvement in cluster density and separation. Overall, Ward Linkage remained stable after normalization.

5. Conclusion

The results of clustering with agglomerative hierarchical clustering show that the choice of linkage method significantly affects the results. Normalization improves clustering quality, especially in scale-sensitive methods such as Single Linkage. Ward Linkage is recommended because it produces more compact, stable, and precise clusters.
Evaluation of the metrics (DBI, SS, NMI, ARI, CH) shows that Single Linkage has the worst performance due to the challenging effect. Complete Linkage is better, but too aggressive in splitting the data. Average Linkage provides a balance between splitting and cluster closeness, although it is less dense than Ward Linkage. Ward Linkage excels with clearer splits and the best metric evaluation. Wald Linkage shows the best results with more transparent cluster structure and superior metric evaluation compared to other methods.

References

[1] A. S. Revi, "Peramalan Jumlah Tindak Pidana Menurut Kepolisian Daerah Dengan Algoritma Backpropagation," *Semin. Nas. Sains Teknol. Inf.*, pp. 246–250, 2019.
[2] E. Darnila, "Infant Routine Monitoring System During 0-12 Months Immunization Using Agglomerative Hierarchical Clustering Algorithm," vol. 1, no. 2, pp. 1–6, 2021, doi: 10.30596/jcositte.v1i1.xxxx.
[3] Y. Reinaldi, N. Ulinnuha, and M. Hafiyusholeh, "Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare Analysis in Cities and Regencies in East Java," *J. Mat. Stat. dan Komputasi*, vol. 18, no. 1, pp. 130–140, 2021, doi: 10.20956/j.v18i1.14228.
[4] S. Wulandari, "Clustering Indonesian Provinces on Prevalence of Stunting Toddlers Using Agglomerative Hierarchical Clustering," *Fakt. Exacta*, vol. 16, no. 2, 2023, doi: 10.30998/faktorexacta.v16i2.17186.
[5] A. Riani, Y. Susianto, and N. Rahman, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Mengunakan Metode Naive Bayes," *J. Innov. Inf. Technol. Appl.*, vol. 1, no. 01, pp. 25–34, 2019, doi: 10.35970/jinita.v1i01.64.
[6] N. Hendrastuty, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa," *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, [Online]. Available: https://doi.org/10.58602/jima-ilkom.v3i1.26
[7] A. Septianingsih, "PEMETAAN KABUPATEN KOTA DI PROVINSI JAWA TIMUR BERDASARKAN TINGKAT KASUS PENYAKIT MENGGUNAKAN," vol. 3, no. 2, 2022.
[8] N. H. Sujacka Retno , Rozzi Kesuma Dinata, "EVALUATING THE QUALITY OF K-MEDOIDS CLUSTERING ON CRIME DATA IN INDONESIA," vol. 8, no. 2, pp. 274–280, 2024.
[9] R. M. F. Lubis, J.-P. Huang, P.-C. Wang, K. Khoifin, Y. Elvina, and D. A. Kusumaningtyas, "Algglomerative Hierarchical Clustering (AHC) Method for Data Mining Sales Product Clustering," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3569.
[10] J. Mantik, M. Ula, A. Faridhatul Ulva, and I. Sahputra, "Implementation of Machine Learning in Determining Nutritional Status using the Complete Linkage Agglomerative Hierarchical Clustering Method," *J. Mantik*, vol. 5, no. 3, pp. 1910–1914, 2021.
[11] A. Radovanovic, J. Li, J. V. Milanovic, N. Milosavljevic, and R. Storchi, "Application of agglomerative hierarchical clustering for clustering of time series data," *IEEE PES Innov. Smart Grid Technology. Conf. Eur.*, vol. 2020-October, pp. 640–644, 2020, doi: 10.1109/ISGT-Europe47291.2020.9248759.
[12] Z. Muttaqin, "Implementasi Unsupervised Learning Pada Nilai Jasmani Kesamaptaan Sekolah Polisi Negara Dengan Metode Clustering Analysis," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 10, no. 1, pp. 18–23, 2023, doi: 10.30656/prosisko.v10i1.6269.
[13] R. D. Firdaus, T. G. Laksana, and R. D. Ramadhani, "Pengelompokan Data Persediaan Obat Menggunakan Perbandingan Metode K-Means Dengan Hierarchical Clustering Single LinkageFirdaus, Rahmatika Diana Laksana, Tri Ginanjar Ramadhani, Rima Dias," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 2, no. 1, pp. 33–48, 2019.
[14] L. M. R. D. J. Nabila Tri Amanda, Riska Yulianti, Anwar Fitrianto, Erfiani, "PERBANDINGAN METODE COMPLETE LINKAGE, AVERAGE LINKAGE DAN WARD'S UNTUK PENGELOMPOKAN KETAHANAN PANGAN DI PROVINSI JAWA TIMUR," *MATHunesa J. Ilm. Mat.*, vol. 13, no. 01, pp. 227–235, 2025.
[15] G. Abdurrahman and U. M. Jember, "Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage," no. December, 2021, doi: 10.32528/justindo.v4i1.2418.
[16] F. H. A. Putera, S. F. Mangitung, Madinawati, and L. Handayani, "Clustering of Province in Indonesia Based on Aquaculture Productivity Using Average Linkage Method," *Param. J. Stat.*, vol. 2, no. 1, pp. 16–20, 2021, doi: 10.22487/27765660.2021.v2.i1.15683.
[17] A. J. Isma Muthahharah, "Cluster Analysis with Complete Linkage and Ward's Method for Health Service Data in Makassar City," vol. 4, no. 2, pp. 109–116, 2021.
[18] S. Indra Maiyanti, O. Dwipurwani, and N. Aisyah, "Pengelompokkan Provinsi Di Indonesia Berdasarkan Konsumsi Kalori Per Kapita Sehari Menurut Kelompok Komoditas / Makanan Menggunakan Average Linkage Dan Ward's Method," *Azizah, dan Nurul Aisyah*, vol. 2, no. 5, pp. 1698–1713, 2024, [Online]. Available: https://journal.institercom-edu.org/index.php/multiple
[19] O. Eric U. and O. Michael O., "Overview of Algglomerative Hierarchical Clustering Methods," *Br. J. Comput. Netw. Inf. Technol.*, vol. 7, no. 2, pp. 14–23, 2024, doi: 10.52589/bjcnit-cv9poogw.
[20] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Callinski-Harabasz index," *IOP Conf. Ser.. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.