KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 3, Juli (2024), pp. 1346-1354

Analisis Pengelompokan Jumlah Penumpang Bus Trans Jateng Menggunakan Metode Clustering K-Means Dan Agglomerative Hierarchical Clustering (AHC)

Muhammad Ryqo Jallu Wicaksono[1], Sri Yulianto Joko Prasetyo[2]
[1,2]Informatics Engineering, Universitas Kristen Satya Wacana, Salatiga, Indonesia
E-mail: [1]672020301@student.uksw.edu, [2]sri.yulianto@uksw.edu

Abstract

This study aims to analyze the clustering of Trans Jateng bus passenger numbers using the K-means clustering method and Agglomerative Hierarchical Clustering (AHC). Data on the number of Trans Jateng bus passengers from the Central Java Transportation Department is used to group passengers based on bus routes and shelters. The K-means method produces 3 clusters with different data quantities for each cluster, while AHC also produces 3 clusters with different patterns. The analysis results show that the K-means method provides more measurable results compared to the AHC method in clustering the Trans Jateng bus passenger data. The information obtained from this analysis can be used to improve Trans Jateng bus transportation services and provide useful insights for decision-makers. Thus, this research contributes to the field of public transportation data analysis and clustering techniques.

Keywords: *clustering k-means, agglomerative hierarchical clustering*

1. Introduction

Population growth in a country is directly proportional to the need for transportation facilities. Trans Jateng is a fast, affordable, air-conditioned bus transportation system around the City of Jateng, initiated by the Department of Transportation. Unlike most city buses, Trans Jateng buses require special shelters as meeting points between prospective passengers and the bus. Over time, the number of passengers on Trans Jateng buses has increased steadily.

Various algorithms can be used for clustering. Broadly, clustering is divided into hierarchical clustering and partitional clustering. This study will compare the performance of these two clustering methods, specifically hierarchical and partitional K-means clustering. Internal validation using the Dunn index, silhouette index, and connectivity index will be conducted to determine the best-performing algorithm, and the selected algorithm will be the basis for clustering. The K-Means clustering method is also widely used in solving clustering problems, such as sales.

Additionally, at certain shelters, there are only a few prospective passengers at specific times. Many residents have complained about the lack of shelters in some areas along the routes. As a result, many people want to use Trans Jateng bus services but cannot due to the absence of shelters. Clustering passenger data at Hang Nadim Batam Airport using the K-means clustering algorithm resulted in three clusters where the data within each cluster have similar characteristics, specifically in terms of the average number of passengers per month.

Therefore, this study aims to cluster the number of Trans Jateng bus passengers based on bus routes and shelters using clustering techniques and the K-means and Agglomerative Hierarchical Clustering (AHC) methods. K-means clustering is a non-hierarchical data clustering method that attempts to partition existing data into one or more clusters so that data with similar characteristics are grouped into the same cluster, while data with different characteristics are grouped into other clusters. The objective of this study is to determine the clustering results using the

ISSN: 2720-992X

K-means and AHC methods and identify the most effective method for clustering Trans Jateng bus passenger data. The findings from this study are expected to provide valuable information to relevant parties in improving the Trans Jateng bus transportation services.

Clustering is a technique for grouping data based on the similarities or differences among the data. The goal of clustering is to divide a dataset into groups (clusters) where data within one group share similar characteristics and are different from data in other groups. Clustering does not require training data to group or classify data objects, making it an unsupervised machine learning technique. Instead, clustering identifies intrinsic patterns in the data and groups them based on certain similarities. For example, consider a dataset consisting of a collection of fruits like apples, bananas, and oranges. A clustering technique will group these fruits based on their type, so all apples will be in one cluster, all bananas in another, and so on. There are several methods to measure the distance of data to the cluster center, one of which is the Euclidean distance, using the formula as follows:

$$D\left(x_2 - x_1\right) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p} |x_{2j} - x_{1j}|^2} \quad (1)$$

Agglomerative Hierarchical Clustering (AHC) is a hierarchical clustering method that follows a bottom-up approach, combining nnn clusters into a single cluster. This method starts by treating each data object as its own individual cluster (atomic cluster) and then successively merges these clusters into larger and larger clusters until all data objects are combined into a single cluster. The key to the AHC method is the calculation of proximity between two clusters. This calculation is divided into three types: Single Linkage (minimum distance), Complete Linkage (maximum distance), and Average Linkage (average distance). Since hierarchical methods do not directly produce clusters, the cophenet distance method is used to analyze the resulting hierarchy. The findings indicate that Agglomerative Hierarchical Clustering (AHC) can be effectively used for data clustering.

$$DUV = \min\{dUV\}\ ,\ dUV \in D \quad (2)$$

K-Means clustering is an unsupervised machine learning algorithm used for data clustering and pattern recognition. It works by randomly selecting several initial data points (k) and then iteratively moving them until the most optimal clustering is found. K-Means clustering is widely used in various fields, such as image segmentation, customer segmentation, market research, and document clustering. During the iterative process, the value of the cluster centroid is determined by calculating the average of the data points within the same cluster using the following formula:

$$v_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (3)$$

where: i : index of the cluster
j : index of variable.
k : index of data.
vij : the centroid value of the i-th cluster for the j-th variable.
xkj : the kth data value in the cluster for the jth variable.
Ni : The number of data that are members of cluster i.

2. Reseach Methodology

A. R. Verdwiansyah Subandi, H. Gunawan, and K. Lorenza Supriyanto, “Penerapan Teknik Data Mining Untuk Memprediksi Jumlah Penumpang Pesawat Pada Bandara Indonesia,” *J. Digit. Ecosyst. Nat. Sustain.*, vol. 1, no. 2, pp. 2798–

6179, 2021. This article discusses the concepts and techniques of data mining, including clustering techniques. Clustering is a technique for grouping data based on similarities or differences among the data. The purpose of clustering is to divide a dataset into groups (clusters), where the data in one group have similar characteristics and are simultaneously different from the data in other groups.

W. Widyawati, W. L. Y. Saptomo, and Y. R. W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *J. Ilm. SINUS*, vol. 18, no. 1, p. 75, 2020, doi: 10.30646/sinus.v18i1.448. This article discusses clustering algorithms. It can serve as a reference for understanding the Agglomerative Hierarchical Clustering (AHC) method used in this research. Agglomerative Hierarchical Clustering is a bottom-up hierarchical clustering method that combines n clusters into a single cluster.

Hikmah, D. Ariyanti, and M. Sugesti, "Penerapan Teknik Data Mining untuk Clustering Armada pada PT. Siaga Transport Indonesia Menggunakan Metode k-Means," *Explore*, vol. 9, no. 1, p. 8, 2019, doi: 10.35200/explore.v9i1.116. This article discusses clustering techniques and algorithms used in data mining. Cluster analysis is a statistical method in research that allows researchers to group or cluster a set of objects into small but distinct clusters with characteristics that differ from other clusters. The underlying theme in exploratory data analysis helps brands, organizations, and researchers gain insights from visual data to identify trends and validate explicit hypotheses and assumptions.

The methods or stages that will be used in creating this scientific work can be seen in Figure 1.

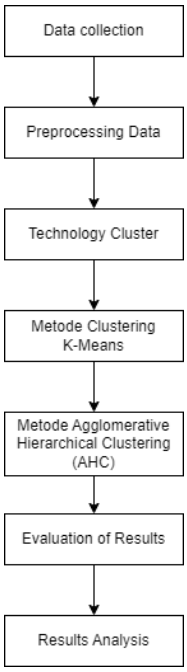

Figure 1. Research Stages Stages

The research stages in Figure 1 can be explained as follows:

1) Data CollectionThe data used in this research is the number of passengers on Trans Jateng buses at various shelters for three bus routes. This data was obtained from the Department of Transportation, Communication, and Informatics (Dishubkominfo) of Central Java. The data used is daily data collected over two weeks.

2) Data PreprocessingThe collected data is processed to remove irrelevant data and fill in missing data. The data is also normalized to avoid bias in the analysis results.
3) Clustering TechniqueClustering techniques are used to group the number of Trans Jateng bus passengers based on bus routes and shelters. The clustering techniques used are K-Means clustering and Agglomerative Hierarchical Clustering (AHC).
4) K-Means Clustering MethodThe K-Means clustering method is used to group the number of Trans Jateng bus passengers based on bus routes and shelters. This method divides the data into several clusters based on the distance between the data points and the cluster centroids. The number of clusters used is determined based on the elbow method.
5) Agglomerative Hierarchical Clustering (AHC) MethodThe Agglomerative Hierarchical Clustering (AHC) method is used to group the number of Trans Jateng bus passengers based on bus routes and shelters. This method builds a cluster hierarchy by merging clusters with the closest distance. The number of clusters used is determined based on the dendrogram.
6) Evaluation of ResultsThe clustering results using K-Means and AHC methods are evaluated using clustering evaluation metrics such as SSE, silhouette coefficient, and purity.
7) Analysis of ResultsThe clustering results using K-Means and AHC methods are analyzed to determine which method is more effective in grouping the number of Trans Jateng bus passengers. The analysis results are also used to provide useful information to relevant parties in improving Trans Jateng bus transportation services.

3. Results and Discussion

3.1. Using the K-Means Algorithm

From the results of processing using the K-Means method, it can be obtained that the number of cluster 0 is 15 data, cluster 1 is 44 data, and cluster 2 is 54 data. With the center point value in Table 1.

Table 1. K-Means Cluster Center Points

Atribut	Cluster			
	Full data	0	1	2
	113	15	44	54
Number of arrivals	3.4248	4	3.3636	3.3148
Bus lane	19.115	24.2667	29.8864	8.9074
Total passenger	12.3363	39.0667	6.6667	

And for the data included in cluster 0, it is displayed in Table 2, cluster 1 is displayed in Table 3, and cluster 2 is displayed in Table 4. In cluster 0, the data represent a high number of passengers, cluster 1 represents a moderate number of passengers, and cluster 2 represents a low number of passengers. The distribution of data across the three clusters results in the following number of data points in each cluster: cluster 0 = 15 data points, cluster 1 = 44 data points, and cluster 2 = 54 data points. The clustering results show that the density of data in each cluster appears balanced, indicating that the data have good similarity.

Table 2. Data on cluster 0 of the K-Means method

Bus	Time	Bus Lane	Shelter	Total Passenger
49	15-17	3B	Terminal Bawen Penurunan Permanan	74
53	15-17	3B	Dusun Smilir	21
55	15-17	1B	Kecamatan Bawen	46
63	15-17	1A	Harjosari 2	53
65	15-17	2A	Bergas 2	33
69	15-17	1A	SMA 1 Bergas	21
71	15-17	2A	Sub Terminal Karangjati	37
73	15-17	2B	Ngobo	39
77	15-17	2B	Wuji 2	32
79	15-17	3A	Langensari 2	22
81	15-17	2A	Undaris 2	55
83	15-17	3A	Mang Engking	39
75	15-17	3A	SMAN 1 Ungaran 2	34
97	15-17	2A	Bateng 2	44
99	15-17	3A	Pramuka	36

Table 3. Data on cluster method 1 K-Means

Id bus	Time	Bus lane	Shelter	total passenger
57	13-14	3B	BPK 2	11
59	15-17	3B	SD Pudak Payung 2	11
60	13-14	3B	Makodam 2	9
61	15-17	3B	Terminal Bayumanik	13
62	13-14	1A	Sukun 2	11
64	13-14	2A	Ada Setiabudi 2	6
66	13-14	3A	Srondol	19
67	15-17	3A	Tembalang 2	14
68	13-14	1A	Gombel 2	19
70	13-14	2A	Pasar Jatingaleh 2	13
72	13-14	2B	Ksatrian PLN 2	16
74	13-14	3A	Don Bosco 2	22
75	15-17	3A	Akpol 2	34
76	13-14	2B	Papandayan	8
78	13-14	3A	Elisabeth 2	6
80	13-14	2A	Merapi 2	10
82	13-14	3A	Gajahmungkur 2	6
86	13-14	3A	Srondol	2
88	13-14	3A	Nganglik 2	3
89	15-17	3A	RS kariadi 2	12
90	13-14	3A	RST	5
91	15-17	3A	Undaris 2	8
92	13-14	1A	BPK 2	12

Id bus	Time	Bus lane	Shelter	total passenger
92	13-14	1A	Undinus	12
93	15-17	1A	Undinus	18
94	13-14	3A	Undinus	11
95	15-17	3A	Undinus	16
96	13-14	2A	Beringin	26
97	15-17	2A	Beringin	44
98	13-14	3A	Beringin	27
99	15-17	3A	Beringin	36
100	13-14	3A	Satsiun Poncol	7
101	15-17	3A	Petek	24
102	13-14	3A	Layur	4
103	15-17	3A	Layur	11
104	13-14	3A	DAMRI	11
105	15-17	3A	DAMRI	16
106	13-14	3A	Tawang	3
107	15-17	3A	Tawang	7
108	13-14	2A	Bergas 2	2
109	15-17	2A	Bergas 2	6
110	13-14	3A	Bergas 2	3
111	15-17	3A	Undaris 2	11
112	13-14	3A	Pramuka	3
113	15-17	3A	BPK 2	6

Table 4. Data on cluster 2 K-Means method

Id bus	Time	Bus lane	Shelter	total passenger
1	6-8	1B	Undaris 2	7
2	13-14	1B	Undaris 2	0
3	15-17	1B	Undaris 2	0
4	17-19	1B	Undaris 2	0
5	6-9	2B	Undaris 2	17
6	13-14	2B	BPK 2	15
7	15-17	2B	BPK 2	20
8	13-14	3B	BPK 2	8
9	15-17	3B	BPK 2	33
10	13-14	3B	Akpol 2	1
11	15-17	3B	Akpol 2	6
12	13-14	3B	Srondol	5
13	15-17	3B	Srondol	2
14	13-14	3B	Bergas 2	3
15	15-17	3B	Bergas 2	25
16	13-14	3A	Ngabean	8
17	15-17	3A	Ngabean	12
18	13-14	3B	Ngabean	6
19	15-17	3B	Ngabean	18
20	13-14	2B	Ngabean	1
21	15-17	2B	Ngabean	6
22	13-14	1A	Gombel 2	6
23	15-17	1A	Gombel 2	9
24	13-14	1B	Gombel 2	5
25	15-17	1B	Gombel 2	0
26	13-14	3B	Gombel 2	0
27	15-17	3B	Gombel 2	4
28	13-14	1B	Don Bosco 2	2
29	15-17	1B	Don Bosco 2	9
30	13-14	3B	Don Bosco 2	0
31	15-17	3B	Don Bosco 2	3
32	13-14	3B	Tembalang 2	9
33	15-17	3B	Tembalang 2	11
34	13-14	2B	Tembalang 2	6
35	15-17	2B	Tembalang 2	12
36	13-14	1B	DAMRI	0
37	15-17	1B	DAMRI	3
38	13-14	3B	DAMRI	1
39	15-17	3B	DAMRI	3
40	13-14	3B	Beringin	0
41	15-17	3B	Beringin	2
42	13-14	3B	Makodam 2	5
43	15-17	3B	Makodam 2	9
44	13-14	3B	Sukun 2	3
45	15-17	3B	Sukun 2	7
46	13-14	3B	MERAPI 2	0
47	15-17	3B	MERAPI 2	0
48	13-14	3B	Sukun 2	17
50	13-14	2B	Tawang	3
51	15-17	2B	Tawang	12
52	13-14	3B	Petek	8
54	13-14	1B	Satsiun Poncol	6
56	13-14	3B	Satsiun Poncol	2
57	15-17	3B	Satsiun Poncol	10

3.2. Use of AHC Clustering Algorithm

From the processing results using the AHC method, it was found that the number of data points in cluster 0 is 32, in cluster 1 is 57, and in cluster 2 is 54. The data included in cluster 0 are displayed in Table 4.6, cluster 1 in Table 4.7, and cluster 2 in Table 4.8. The distribution of data across the three clusters results in the following number of data points in each cluster: cluster 0 = 2 data points, cluster 1 = 57 data points, and cluster 2 = 54 data points. The clustering results show that the density of data in each cluster appears balanced, indicating that the data have poor similarity.

Table 5. Data on cluster 0 AHC method

Id bus	Time	Bus line	Shelter	total passenger
1	6-8	1B	Ksatrian PLN 2	7
5	6-9	2B	Ksatrian PLN 2	17

Table 6. Data on cluster 1 AHC method

Id bus	Time	Bus line	Shelter	total passenger
2	13-14	1B	Ksatrian PLN 2	0
3	15-17	1B	Ksatrian PLN 2	0
6	13-14	2B	Srondol	15
7	15-17	2B	Srondol	20
8	13-14	3B	Srondol	8
10	13-14	3B	Gajahmungkur 2	1
12	13-14	3B	DAMRI	5
14	13-14	3B	Bergas 2	3
16	13-14	3A	Ngabean	8
18	13-14	3B	Ngabean	6
20	13-14	2B	Ngabean	1
22	13-14	1A	Gombel 2	6
24	13-14	1B	Gombel 2	5
26	13-14	3B	Gombel 2	0
28	13-14	1B	Don Bosco 2	2

Id bus	Time	Bus line	Shelter	total passenger
30	13-14	3B	Don Bosco 2	0
32	13-14	3B	Tembalang 2	9
34	13-14	2B	Tembalang 2	6
36	13-14	1B	DAMRI	0
38	13-14	3B	DAMRI	1
40	13-14	3B	Beringin	0
42	13-14	3B	Makodam 2	5
44	13-14	3B	Sukun 2	3
46	13-14	3B	MERAPI 2	0
48	13-14	3B	Bergas 2	17
50	13-14	2B	Tawang	3
52	13-14	3B	Petek	8
54	13-14	1B	Satsiun Poncol	6
...	...	..	...	...
58	13-14	3B	Bergas 2	11
60	13-14	3B	Tejokusuman	9
62	13-14	1A	Wuji 2	11
64	13-14	2A	Wuji 2	6
66	13-14	3A	Wuji 2	19
68	13-14	1A	Janti 2	19
70	13-14	2A	Termconcat	13
72	13-14	2B	Termconcat	16
74	13-14	3A	Undaris 2	22
76	13-14	2B	Undaris 2	8
78	13-14	3A	BPK 2	6
80	13-14	2A	Ngabean	10
82	13-14	3A	Ngabean	6
84	13-14	3A	Ksatrian PLN 2	0
86	13-14	3A	Ksatrian PLN 2	2
88	13-14	3A	BPK 2	3
90	13-14	3A	SMAN 1 Ungaran 2	5
92	13-14	1A	MANG ENGKING	12
94	13-14	3A	MANG ENGKING	11
96	13-14	2A	Ngobo	26
98	13-14	3A	Ngobo	27
100	13-14	3A	Janti Utara	7
102	13-14	3A	Disnaker	4
104	13-14	3A	Makodam 2	11
106	13-14	3A	UPN	3
108	13-14	2A	Manggung	2
110	13-14	3A	Manggung	3
112	13-14	3A	Bergas 2	3

Table 7. Data on cluster 1 AHC method

Id bus	time	Bus line	Shelter	total passenger
59	15-17	3B	Ksatrian PLN 2	11
61	15-17	3B	Tejokusuman	13
63	15-17	1A	Wuji 2	53
65	15-17	2A	Wuji 2	33
67	15-17	3A	Wuji 2	14
69	15-17	1A	Janti 3	21
71	15-17	2A	Termconcat	37
73	15-17	2B	Termconcat	39
75	15-17	3A	Ksatrian PLN 2	34
77	15-17	2B	Ksatrian PLN 2	32
79	15-17	3A	Undaris 2	22
81	15-17	2A	BPK 2	55
83	15-17	3A	BPK 2	39
85	15-17	3A	Undaris 2	0
87	15-17	3A	Srondol	0
89	15-17	3A	Bergas 2	12
91	15-17	3A	SMAN 1 Ungaran 2	8
93	15-17	1A	MANG ENGKING	18
95	15-17	3A	MANG ENGKING	16
97	15-17	2A	Ngobo	44
99	15-17	3A	Ngobo	36
101	15-17	3A	Janti Utara	24

Id bus	time	Bus line	Shelter	total passenger
103	15-17	3A	Disnaker	11
105	15-17	3A	Makodam 2	16
107	15-17	3A	UPN	7
109	15-17	2A	Manggung	6
111	15-17	3A	Manggung	11
113	15-17	3A	Ksatrian PLN 2	6

Id bus	time	Bus line	Shelter	total passenger
59	15-17	3B	SMAN 1 Ungaran 2	11
61	15-17	3B	Tejokusuman	13
63	15-17	1A	Wuji 2	53
65	15-17	2A	Wuji 2	33
67	15-17	3A	Wuji 2	14
69	15-17	1A	Janti 3	21
71	15-17	2A	Termconcat	37
73	15-17	2B	Termconcat	39
75	15-17	3A	SMAN 1 Ungaran 2	34
77	15-17	2B	SMAN 1 Ungaran 2	32
79	15-17	3A	BPK 2	22
81	15-17	2A	BPK 2	55
83	15-17	3A	Ngabean	39
85	15-17	3A	Srondol	0
87	15-17	3A	Bergas 2	0
89	15-17	3A	Undaris 2	12
91	15-17	3A	SMAN 1 Ungaran 2	8
93	15-17	1A	MANG ENGKING	18
95	15-17	3A	MANG ENGKING	16
97	15-17	2A	Ngobo	44
99	15-17	3A	Ngobo	36
101	15-17	3A	Janti Utara	24
103	15-17	3A	Disnaker	11
105	15-17	3A	Makodam 2	16
107	15-17	3A	UPN	7
109	15-17	2A	Manggung	6

From the clustering results using the K-Means and AHC methods, it can be seen that the K-Means clustering results have a more balanced similarity for each group compared to the AHC clustering results. Therefore, it can be concluded that although the K-Means clustering process requires determining the centroid or points first, in this case, it produces better clustering data. On the other hand, the AHC clustering results, although more efficient because they do not use centroids at the beginning and only calculate the level of similarity, have less favorable results.

4. Conclusion

After conducting data mining using the K-Means and AHC clustering algorithms on the number of passengers of Trans Jateng buses, the following conclusions can be drawn:

a) Population growth in a country is directly related to the need for transportation infrastructure.
b) This study will compare the performance of hierarchical and partitional K-Means clustering methods. Internal validation will be conducted to determine the best algorithm, which will serve as the basis for classification.
c) K-Means clustering is widely used for grouping, such as in solving sales problems.
d) The K-Means data mining clustering algorithm is used to create 3 clusters for the number of passengers at Hang Nadim Batam Airport, based on similar characteristics in the number of passengers each month.
e) This study aims to categorize the number of Trans Jateng bus passengers based on routes and bus stops using clustering techniques, K-Means, and Agglomerative Hierarchical Clustering (AHC).

References

[1] F. Bimantoro Putra, A. Tri Rian Dani, and S. Wigantono, "Penerapan Algoritma Hierarchical Clustering Dalam Pengelompokkan Kabupaten/Kota Di Papua Berdasarkan Indikator Kemiskinan," *Math. Appl. J.*, vol. 5, no. 23, pp. 118–126, 2023.

[2] M. I. Sahputra, A. Triayudi, and A. Rubhasy, "Analisis Faktor Yang Mempengaruhi Penumpang Angkutan Umum Beralih Ke Transportasi Online Go-Jek Menggunakan Metode K-Means Clustering," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 6, no. 1, pp. 63–69, 2022, doi: 10.35870/jtik.v6i1.381.

[3] A. Permadi and Y. A. Wiyaja, "Pengelompokan Terbaik Menggunakan Algoritma K-Means Pada Dataset Bus Biskita Bogor," *Intern. (Information Syst. Journal)*, vol. 6, no. 1, pp. 88–100, 2023, doi: 10.32627/internal.v6i1.689.

[4] W. Widyawati, W. L. Y. Saptomo, and Y. R. W. Utami, "Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan," *J. Ilm. SINUS*, vol. 18, no. 1, p. 75, 2020, doi: 10.30646/sinus.v18i1.448.

[5] N. Hikmah, D. Ariyanti, and M. Sugesti, "Penerapan Teknik Data Mining untuk Clustering Armada pada PT. Siaga Transport Indonesia Menggunakan Metode k-Means," *Explore*, vol. 9, no. 1, p. 8, 2019, doi: 10.35200/explore.v9i1.116.

[6] A. F. Dewi and K. Ahadiyah, "Agglomerative Hierarchy Clustering Pada Penentuan Kelompok Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Pendidikan," *Zeta - Math J.*, vol. 7, no. 2, pp. 57–63, 2022, doi: 10.31102/zeta.2022.7.2.57-63.

[7] M. F. Daffauzan, "Analisa Penumpang Bus Transjakarta dengan Metode K-Means," *J. Comput. Sci. Inf. Syst. J-Cosys*, vol. 3, no. 2, pp. 61–64, 2023, doi: 10.53514/jco.v3i2.395.

[8] Ramadhana, Islamiyah, and A. P. A. Masa, "Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara," *Adopsi Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 35–42, 2023, doi: 10.30872/atasi.v2i1.595.

[9] A. Supriyatna, I. Carolina, S. Janti, and A. Haidir, "Clustering Koridor Transjakarta Berdasarkan Jumlah Penumpang Dengan Algoritma K-Means," *J. Sains Komput. Inform. (J-SAKTI*, vol. 4, no. 2, pp. 682–693, 2020.

[10] A. Azzahra and A. W. Wijayanto, "Perbandingan Agglomerative Hierarchical dan K-Means Dalam Pengelompokkan Provinsi Berdasarkan Pelayanan Kesehatan Maternal," *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 481–495, 2022, [Online]. Available: http://download.garuda.kemdikbud.go.id/article.php?article=2929100%5C&val=10006%5C&title=Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services.