Jurnal Ilmiah MEDIA SISFO
Vol. 19, No. 1, April 2025
ISSN: 1978-8126
e-ISSN: 2527-7340

Clustering Nilai Mahasiswa Menggunakan Algoritma K-Means

Azwar Anas[1], Ade Jermawinsyah Zebua[2], Akhmadi[3]*

[1]Program Studi Sistem Informasi,Universitas Graha Karya Muara Bulian[1]
[2,3]Program Studi Manajemen, Universitas Graha Karya Muara Bulian[2]
Jl. Gajah Mada Muara Bulian-Indonesia
azwarzayn@gmail.com[1], adejermawinsyahzebua9@gmail.com[2], h.akhmadijambi@gmail.com[3]

Submitted : **08/03/2025**; Reviewed : **07/04/2025**; Accepted : **30/04/2025**; Published : **30/04/2025**

Abstract

The values obtained by each student will vary. This depends on the student's ability to absorb the lecture material given. The more students there are, the more variety of values that are seen. These values will only appear as meaningless numbers, if no in-depth analysis is carried out. However, when explained using the data mining method, these values will be presented in new knowledge that has been hidden so far. The purpose of this study is to divide student values into 3 (three) clusters based on their similarities. The method used is the K-Means algorithm. The results of the study showed that Cluster 2 (two) has the most members with 18 (eighteen), then cluster 1 (one) with 8 (eight) and cluster 3 (three) has the least, namely 2 (two) members. The distribution of cluster 2 data is more diverse because it has the most members, the distribution of cluster 1 is slightly piled up at one point, while cluster 3 only has 2 (two) members and has a fairly large distance.

Keywords: mining, cluster, k-means, value.

Abstrak

Nilai-nilai yang diperoleh setiap mahasiswa akan berbeda. Hal ini tergantung dari kemampuan mahasiswa menyerap materi perkuliahan yang diberikan. Semakin banyak jumlah mahasiswa, semakin banyak pula ragam nilai yang terlihat. Nilai-nilai tersebut hanya akan tampak seperti angka-angka yang tak bermakna, jika tidak dilakukan suatu analisis mendalam. Namun ketika dianalisis menggunakan metode *data mining* (penambangan data), nilai tersebut akan tersaji dalam suatu pengetahuan baru yang selama ini tersembunyi. Tujuan penelitian ini adalah membagi nilai mahasiswa ke dalam 3 (tiga) *cluster* berdasarkan kemiripannya. Metode yang digunakan adalah algoritma K-Means. Hasil penelitian menunjukkan Cluster 2 (dua) memiliki anggota paling banyak dengan 18 (delapan belas), kemudian cluster 1 (satu) dengan 8 (delapan) dan cluster 3 (tiga) paling sedikit yaitu 2 (dua) anggota. Sebaran data *cluster* 2 lebih beragam karena anggota yang paling banyak, sebaran *cluster* 1 sedikit menumpuk pada satu titik, sedangkan *cluster* 3 hanya 2 (dua) anggota dan memiliki jarak yang agak jauh.

Kata kunci: penambangan, cluster, k-means, nilai.

1. Pendahuluan

Tridharma Perguruan Tinggi adalah tiga acuan yang menjadi tugas pokok dalam perguruan tinggi, baik oleh dosen maupun mahasiswa. Melalui Tridharma inilah proses pembelajaran, penelitian dan pengabdian kepada masyarakat dilaksankan secara terukur dan terencana. Pembelajaran adalah dharma pertama yang dilakukan. Secara terjadwal mahasiswa dan dosen berinteraksi dalam rangka transfer ilmu pengetahuan baik di kelas maupun di luar kelas. Dharma kedua yaitu penelitian dilaksanakan pada tahap akhir semester sebagai impelementasi dari hasil pembelajaran. Sedangkan dharma ketiga yakni pengabdian kepada masyarakat dilaksanakan sebagai bentuk pengabdian ilmu pengetahuan yang telah diperoleh dari hasil pembelajaran dan penelitian sebelumnya kepada masyarakat.

Proses pembelajaran dilaksanakan selama 16 (enam belas) kali pertemuan termasuk Ujian Tengah Semester (UTS) dan Ujian AKhir Semester (UAS). UTS dan UAS merupakan kegiatan untuk mengukur daya serap mahasiswa terhadap materi pembelajaran yang telah dilaksanakan sebelumnya. Setiap

ISSN: 1978-8126
e-ISSN: 2527-7340

mahasiswa akan memiliki nilai yang bervariasi. Nilai akhir yang diperoleh merupakan kombinasi dari berbagai nilai sesuai dengan ketentuan nasional maupun internal institusi. Setidaknya 3 (tiga) basis nilai yang akan menghasilkan nilai akhir, yaitu nilai harian, UTS dan UAS. Nilai harian merupakan nilai rata-rata mahasiswa dari unsur sikap, pengetahuan dan keterampilan selama proses pembelajaran dilaksanakan. Nilai UTS adalah hasil ujian pada rentang pertemuan 8 (delapan) dan 9 (Sembilan), sedangkan nilai UAS adalah hasil ujian setelah seluruh proses pembelajaran dilaksanakan dengan minimal jumlah perkuliahan sebanyak 14 (empat belas) kali.

Nilai-nilai yang diperoleh setiap mahasiswa akan berbeda. Hal ini tergantung dari kemampuan mahasiswa menyerap materi perkuliahan yang diberikan. Semakin banyak jumlah mahasiswa, semakin banyak pula ragam nilai yang terlihat. Nilai-nilai tersebut hanya akan tampak seperti angka-angka yang tak bermakna, jika tidak dilakukan suatu analisis mendalam. Namun ketika dianalisis menggunakan metode *data mining* (penambangan data), nilai tersebut akan tersaji dalam suatu pengetahuan baru yang selama ini tersembunyi [1]. Perlunya dilakukan clusterisasi adalah untuk mendapatkan Gambaran sebaran nilai secara umum maupun per individu mahasiswa. Berdasarkan clusterisasi ini pada akhirnya akan memberikan informasi berharga bagi dosen dalam mengevaluasi proses pembelajaran maupun penilaiannya.

Data mining menggunakan banyak algoritma tergantung pada jenis data dan pengetahuan yang ingin digali. *Pattern Recognition* adalah istilah tambahan untuk bidang ini [2]. Beberapa algoritma yang digunakan seperti algoritma K-Means untuk melakukan *clustering data* [3], algoritma *Market Bascet Analysis* (MBA) [4], *Frequent Pattern Tree* [5], dan Apriori [6] untuk menggali kaidah asosiasi (*Association Rule*).

Penelitian ini akan menggunakan algortima K-Means untuk melakukan *clustering data* nilai mahasiswa Universitas Graha Karya Muara Bulian (UGKMB). Hal ini sejalan dengan pengetahuan yang ingin dihasilkan dari bongkahan data nilai tersebut.

Formula yang digunakan dalam algoritma K-Means [7]:

$$E_{(m,n)} = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2}$$

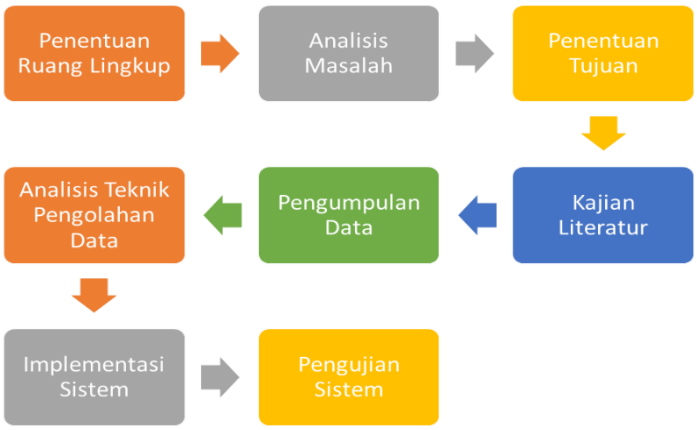
E = *Euclidean* atau jarak data dengan titik pusat (*centroid*) yang ditetapkan [8].

Beberapa penelitian terkait diantaranya:
1. Penelitian yang dilakukan dengan meningkatkan performa algoritma K-Means menggunakan metode metrik jarak yang berbeda [9]. Hasilnya menunjukkan bahwa eksekusi waktu lebih baik.
2. Penelitian tentang peningkatan algoritma K-Means berdasarkan hasil pasca cluster [10]. Artikel ini focus pada peningkatan kualitas cluster yang dihasilkan.
3. Penelitian yang dilakukan untuk rumah sakit di Jakarta dengan metode Fuzzy K-Means [11]. Hasilnya menunjukkan terdapat perbedaan area cluster dari metode biasa.
4. Penelitian tentang sebaran Covid-19 dari berbagai Provinsi di Indonesia menggunakan algoritma K-Means [12]. Hasil penelitian menunjukkan terdapat 3 (tiga) cluster.
5. Penelitian yang dilakukan di Lampung Selatan tentang kelayakan area penanaman jagung [13]. Hasil penelitian menunjukkan terdapat 2 (dua) cluster besar berdarakan data 2 (dua) tahun.

2. Metode Penelitian

Penelitian ini adalah penelitian pustaka (*library research*). Penulis memilih algoritma K-Means karena merupakan komponen dari algoritma data mining yang sesuai dengan jenis data dan pengetahuan yang akan diproses. Untuk menghasilkan penelitian yang berkualitas, setiap pelaksana penelitian harus menggunakan metode yang tepat dan sesuai. Penelitian diharapkan berjalan dengan sistematis dan ilmiah berdasarkan metode yang disusun. Berikut adalah kerangka penelitian yang dilakukan.

ISSN: 1978-8126
e-ISSN: 2527-7340



Gambar 1. ***Kerangka Penelitian***

Tahapan penelitian yang dilakukan sebagai berikut :

1. Penentuan Ruang Lingkup Penelitian
   Untuk memastikan bahwa penelitian terkonsentrasi pada masalah yang akan dibahas, langkah pertama adalah menentukan ruang lingkup dan batasan masalah.
2. Analisis Masalah
   Analisis masalah dilakukan untuk menetapkan masalah penelitian atau batasannya. Proses penggunaan algoritma K-Means untuk mendapatkan *cluster* dari nilai mahasiswa dilakukan pada tahap ini.
3. Penentuan Tujuan
   Setelah masalah dipahami, tujuan penelitian ini ditetapkan. Tujuannya adalah untuk menjawab masalah-masalah sebelumnya. Seperti menggali tumpukan data nilai mahasiswa untuk kemudian dihasilkan suatu pengetahuan.
4. Kajian Literatur
   Untuk mencapai tujuan penelitian, berbagai referensi yang terkait dengan topik penelitian diperiksa. Referensi ini kemudian dipilih sesuai dengan kebutuhan penelitian dan dipilih dari jurnal ilmiah, buku, dan sumber internet yang relevan.
5. Pengumpulan Data
   Untuk memahami masalah yang ada, proses pengumpulan data dimulai dengan melakukan observasi langsung pada objek penelitian. Studi kepustakaan dilakukan dengan membaca berbagai buku yang relevan dengan masalah penelitian. Data utama yang digunakan adalah nilai mahasiswa.
6. Penentuan Teknik yang digunakan
   Penentuan teknik pengolahan data adalah langkah yang dilakukan dalam mengolah data. Dalam penelitian ini, penulis menggunakan algoritma K-Means.
7. Implementasi Sistem
   Selanjutnya, implementasi dilakukan pada *software data mining* Orange.
8. Pengujian Sistem
   Untuk membandingkan perhitungan manual dan komputerisasi pada *software*, pengujian sistem dilakukan sebagai berikut:
   a. Menguji data nilai siswa dengan rumus algoritma K-Means;
   b. Kemudian, menggunakan algoritma K-Means sebagai metode analisis untuk menerapkan data pada *software* Orange.
   c. Terakhir, membandingkan data yang diolah secara manual dan yang diolah secara komputerisasi. Jika hasilnya tidak jauh berbeda, hasilnya dapat dianggap benar.

ISSN: 1978-8126
e-ISSN: 2527-7340

3. Hasil Penelitian

3.1. Data Nilai Mahasiswa

Penulis menggunakan data nilai mahasiswa Program Studi Sistem Informasi UGKMB. Variabel yang digunakan adalah Nama Mahasiswa, Nilai UTS, UAS dan Nilai Akhir. Khusus nama mahasiswa pada tabel hanya diisi dengan huruf abjad.

Tabel 1. ***Data Nilai Mahasiswa***

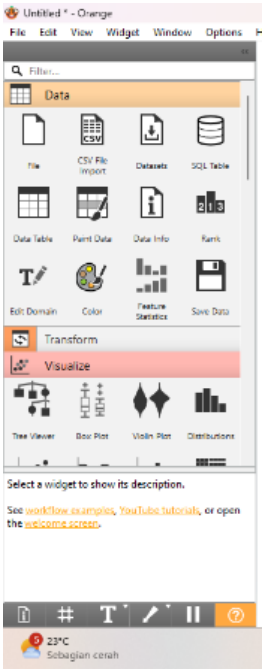
No.	Nama	Nilai UTS	Nilai UAS	Nilai AKHIR
1	A	70	100	81
2	B	70	100	81
3	C	95	100	89
4	D	70	100	81
5	E	80	80	79
6	F	90	100	97
7	G	85	90	87
8	H	70	90	77
9	I	95	100	88
10	J	95	100	84
11	K	95	90	84
12	L	80	100	83
13	M	70	100	81
14	N	70	50	60
15	O	70	100	81
16	P	95	100	89
17	Q	70	80	76
18	R	70	100	81
19	S	70	90	76
20	T	70	90	76
21	U	75	100	82
22	V	80	80	76
23	W	75	90	78
24	X	70	90	80
25	Y	65	70	60
26	Z	70	80	77
27	AA	70	80	76
28	AB	90	100	92

Tabel 1 di atas menyajikan data nilai mahasiswa yang diperoleh dari 3 (tiga) kategori yaitu nilai UTS, UAS dan nilai Akhir. Data tersebut akan diproses lebih lanjut dengan menentukan berapa cluster yang akan dibangun dan data mana yang akan menjadi titik pusat (*centroid*).

3.2. Analisis Algoritma K-Means

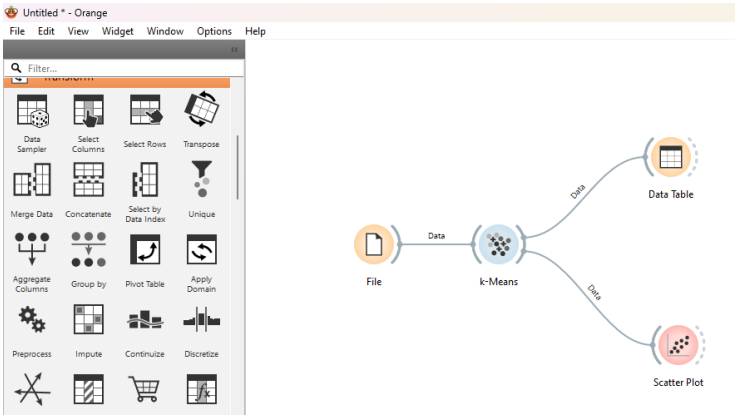
Jumlah *cluster* yang akan dibuat pada penelitian ini adalah 3 (tiga). Dengan menggunakan rumus algoritma K-Means, setiap nilai UTS, UAS dan Akhir mahasiswa akan tersebar dalam tiga *cluster* sesuai dengan jarak nilai terdekat dengan titik pusat data (*centroid*).

ISSN: 1978-8126
e-ISSN: 2527-7340



Gambar 2. ***Interface Software Orange***

Setelah tampil seperti gambar di atas, maka langkah selanjutnya adalah memilih menu Data → New, Data Table, Simple K-Means, Scatter Plot lalu hubungkan, sebagaimana gambar berikut.



Gambar 3. ***Input dan Proses Data Nilai Mahasiswa***

ISSN: 1978-8126
e-ISSN: 2527-7340

	MHS	Cluster	Silhouette	UTS	UAS	AKHIR
1	A	C2	0.678812	70	100	81
2	B	C2	0.678812	70	100	81
3	C	C1	0.704004	95	100	89
4	D	C2	0.678812	70	100	81
5	E	C2	0.613529	80	80	79
6	F	C1	0.661984	90	100	97
7	G	C1	0.587094	85	90	87
8	H	C2	0.692408	70	90	77
9	I	C1	0.701677	95	100	88
10	J	C1	0.674788	95	100	84
11	K	C1	0.650654	95	90	84
12	L	C2	0.539452	80	100	83
13	M	C2	0.678812	70	100	81
14	N	C3	0.670506	70	50	60
15	O	C2	0.678812	70	100	81
16	P	C1	0.704004	95	100	89
17	Q	C2	0.658101	70	80	76
18	R	C2	0.678812	70	100	81
19	S	C2	0.692408	70	90	77
20	T	C2	0.690063	70	90	76
21	U	C2	0.638946	75	100	82
22	V	C2	0.624663	80	80	76
23	W	C2	0.67522	75	90	78
24	X	C2	0.687014	70	90	80
25	Y	C3	0.635074	65	70	60
26	Z	C2	0.66374	70	80	77
27	AA	C2	0.658101	70	80	76
28	AB	C1	0.684317	90	100	92

Gambar 4. ***Cluster Nilai Mahasiswa***

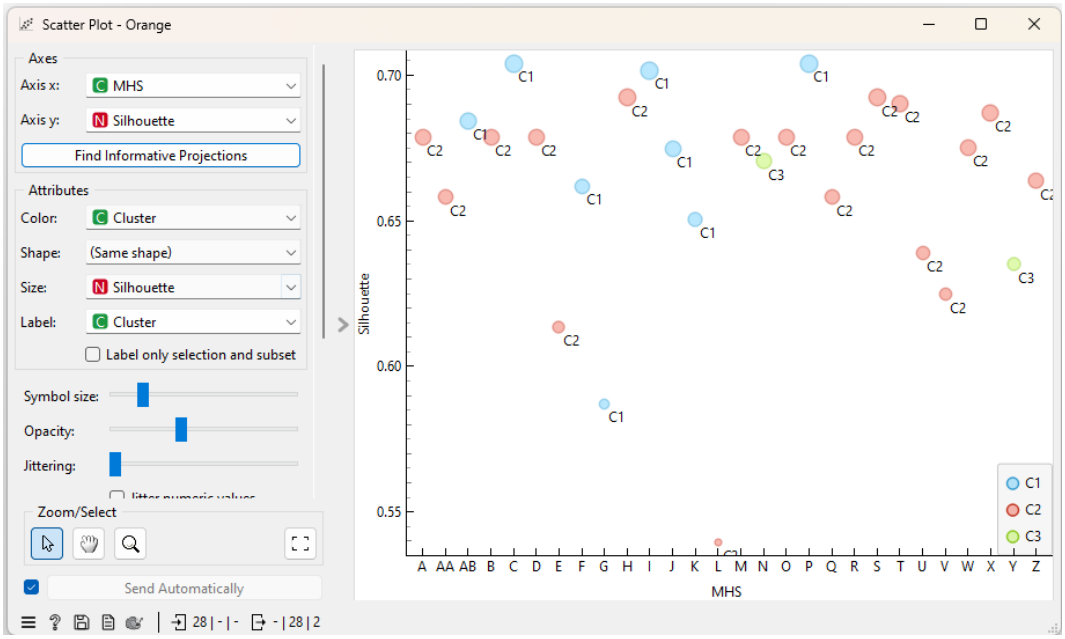
Pada gambar di atas, merupakan tahapan input data nilai mahasiswa dalam format csv (*comma delimited*), kemudian dihubungan dengan menu Simple K-Means untuk ditentukan berapa *cluster* yang akan dibuat, kemudian dihubungkan dengan menu Data Table untuk menampilkan hasil *cluster*, lalu Simple K-Means dihubungkan dengan menu Scatter Plot untuk menampilkan diagram sebaran nilai berdasarkan *cluster* masing-masing.

Berdasarkan gambar di atas, *cluster* (C) diperoleh untuk setiap nilai mahasiswa.

a. C1 dengan anggota C, F, G, I, J, K, P dan AB
b. C2 dengan anggota A, B, D, E, H, L, M, O, Q, R, S, T, U, V, W, X, Z dan AA
c. C3 dengan anggota N dan Y.

Cluster 2 (dua) memiliki anggota paling banyak dengan 18 (delapan belas), kemudian cluster 1 (satu) dengan 8 (delapan) dan cluster 3 (tiga) paling sedikit yaitu 2 (dua) anggota.

ISSN: 1978-8126
e-ISSN: 2527-7340



Gambar 5. ***Scatter Plot Cluster Nilai Mahasiswa***

Pada gambar di atas, warna biru untuk C1, merah untuk C2 dan hijau untuk C3. Sebaran data C2 lebih beragam karena anggota yang paling banyak, sebaran C1 sedikit menumpuk pada satu titik, sedangkan C3 hanya 2 (dua) anggota dan memiliki jarak yang agak jauh.

4. Penutup

Berdasarkan hasil penelitian menunjukkan mayoritas mahasiswa atau sebanyak 18 (delapan belas) orang masuk dalam *cluster* 1 (C1), *cluster* 2 (C2) terdapat 8 (delapan) anggota dan *cluster* 3 (C3) hanya memiliki 2 (dua) anggota. Sebaran data cukup merata untuk C2 hal ini karna jumlah data yang cukup banyak dan beragam, C1 cenderung menumpuk pada satu lokasi dan C3 hanya 2 (dua) anggota dan memiliki jarak yang berjauhan. Artinya secara umum mahasiswa telah mampu menyerap materi pembelajaran yang dibuktikan dengan pencapaian nilai evaluasi pembelajaran yang baik. Dengan begitu maka proses clusterisasi ini telah berhasil menyajikan informasi berharga dari sebaran nilai mahasiswa.

Daftar Pustaka

[1] B. Hasmaulina, "Penerapan Data Mining Untuk Membentuk Kelompok Belajar Menggunakan Metode Clustering Di SMK Negeri 3 Seluma," *JUKOMIKA (Jurnal Ilmu Komput. dan Inform.*, vol. 4, no. 2, pp. 57–71, 2022, doi: 10.54650/jukomika.v4i2.368.

[2] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.

[3] N. T. Suswanto, P. Chyan, and V. Putri, "Using K-Means Algorithm to Investigate Community Behavior in Treating Waste toward Smart City," vol. 11, no. 4, 2021.

[4] A. Anas, Azwar; Jermawinsyah Zebua, "Implementasi Data Mining Menggunakan Algoritma Apriori Untuk Meningkatkan Pola Penjualan Obat," *J. Ilm. MEDIA SISFO*, vol. 16, no. 1, pp. 54–61, 2022, doi: 10.35957/jatisi.v7i2.195.

[5] B. S. Hasugian, "Penerapan Metode Association Rule Untuk Menganalisa Pola Pemakaian Bahan Kimia Di Laboratorium Menggunakan Algoritma FP-Growth (Studi Kasus di Laboratorium Kimia PT . PLN (Persero) Sektor Pembangkitan Belawan Medan) Buyung Solihin Hasugian Universitas," *J. Ilmu Komput. dan Inform.*, vol. 6341, no. November, pp. 56–69, 2019.

[6] D. Ai, H. Pan, X. Li, Y. Gao, and D. He, "Association rule mining algorithms on high-dimensional datasets," *Artif. Life Robot.*, vol. 23, no. 3, pp. 420–427, 2018, doi: 10.1007/s10015-018-0437-y.

ISSN: 1978-8126
e-ISSN: 2527-7340

[7] D. Ariyanto, "Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut," *J. Sistim Inf. dan Teknol.*, pp. 13–18, 2022, doi: 10.37034/jsisfotek.v4i1.117.

[8] U. Stemmer, "Locally private k-means clustering," *J. Mach. Learn. Res.*, vol. 22, pp. 1–30, 2021.

[9] T. M. Ghazal *et al.*, "Performances of k-means clustering algorithm with different distance metrics," *Intell. Autom. Soft Comput.*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[10] I. D. Borlea, R. E. Precup, and A. B. Borlea, "Improvement of K-means Cluster Quality by Post Processing Resulted Clusters," *Procedia Comput. Sci.*, vol. 199, pp. 63–70, 2021, doi: 10.1016/j.procs.2022.01.009.

[11] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 356–363, 2022, doi: 10.1016/j.procs.2022.12.146.

[12] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual. Quant.*, vol. 56, no. 3, pp. 1283–1291, 2022, doi: 10.1007/s11135-021-01176-w.

[13] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012038.