

**KLASTERISASI PROVINSI DI INDONESIA MENGGUNAKAN METODE
K-MEANS DENGAN *CENTROID LINKAGE* BERDASARKAN JUMLAH
TENAGA KESEHATAN TAHUN 2023**

SKRIPSI

Diajukan untuk memenuhi syarat untuk memperoleh gelar Sarjana Matematika
Program Studi Matematika



Oleh:

Barqy Muhammad Ilhan

NIM 2102924

**PROGRAM STUDI MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA**

2025

LEMBAR HAK CIPTA

KLASTERISASI PROVINSI DI INDONESIA MENGGUNAKAN METODE *K-MEANS* DENGAN *CENTROID LINKAGE* BERDASARKAN JUMLAH TENAGA KESEHATAN TAHUN 2023

Oleh

Barqy Muhammad Ilhan

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Matematika pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan
Alam

© Barqy Muhammad Ilhan 2025

Universitas Pendidikan Indonesia

April 2025

Hak Cipta dilindungi undang-undang.

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difotokopi, atau cara lainnya tanpa izin dari penulis.

LEMBAR PENGESAHAN

BARQY MUHAMMAD ILHAN

**KLASTERISASI PROVINSI DI INDONESIA MENGGUNAKAN METODE
K-MEANS DENGAN *CENTROID LINKAGE* BERDASARKAN JUMLAH
TENAGA KESEHATAN TAHUN 2023**

Disetujui dan disahkan oleh pembimbing:

Pembimbing I,



Prof. Dr. Dadan Dasari, M.Si.

NIP. 196407171991021001

Pembimbing II,



Hj. Dewi Rachmatin, S.Si., M.Si.

NIP. 196909291994122001

Mengetahui,

Ketua Program Studi Matematika,



Dr. Kartika Yulianti, S.Pd., M.Si.

NIP. 198207282005012001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini, dengan ini menyatakan bahwa skripsi yang berjudul “**KLASTERISASI PROVINSI DI INDONESIA MENGGUNAKAN METODE K-MEANS DENGAN CENTROID LINKAGE BERDASARKAN JUMLAH TENAGA KESEHATAN TAHUN 2023**” ini beserta seluruh isinya adalah benar-benar karya sendiri. Saya tidak melakukan penjiplakan dan pengutipan dengan cara yang tidak sesuai dengan etika keilmuan. Atas pernyataan ini, saya siap menanggung risiko ataupun sanksi yang dijatuhkan kepada saya apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini atau ada klaim dari pihak lain terhadap keaslian karya saya.

Demikian pernyataan ini saya sampaikan.

Bandung, 28 April 2025



Barqy Muhammad Ilhan

NIM. 2102924

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasterisasi Provinsi di Indonesia Menggunakan Metode *K-Means* dengan *Centroid Linkage* Berdasarkan Jumlah Tenaga Kesehatan Tahun 2023”. Penulisan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Matematika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia. Dalam proses penyusunan skripsi ini, penulis menghadapi berbagai tantangan, baik secara teknis maupun nonteknis. Namun, berkat doa, dukungan, serta bantuan dari berbagai pihak, hambatan-hambatan tersebut dapat dilalui hingga skripsi ini terselesaikan dengan baik.

Skripsi ini merupakan hasil dari upaya terbaik yang telah penulis lakukan selama masa studi. Meskipun demikian, penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan maupun isi skripsi ini. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi penyempurnaan di masa mendatang. Akhir kata, semoga karya ini dapat memberikan manfaat, baik bagi penulis sendiri maupun bagi pembaca yang tertarik pada topik yang dibahas.

Bandung, 28 April 2025

Barqy Muhammad Ilhan

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala, Tuhan semesta alam, atas segala limpahan rahmat, karunia, serta ridho-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Klasterisasi Provinsi di Indonesia Menggunakan Metode *K-Means* dengan *Centroid Linkage* Berdasarkan Jumlah Tenaga Kesehatan Tahun 2023”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Matematika, Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam, Universitas Pendidikan Indonesia.

Dengan penuh rasa hormat dan cinta, skripsi ini penulis persembahkan kepada kedua orang tua tercinta, yang senantiasa mengiringi langkah penulis dengan doa, kasih sayang, serta dukungan tiada henti dalam setiap keadaan. Penulis juga menyadari bahwa penyusunan skripsi ini tidak terlepas dari dukungan, bantuan, dan kontribusi berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih dan apresiasi setinggi-tingginya kepada:

1. Prof. Dr. Dadan Dasari, M.Si. dan Hj. Dewi Rachmatin, S.Si., M.Si., selaku Dosen Pembimbing I dan II, atas waktu, perhatian, bimbingan, serta arahan yang sangat berharga selama proses penyusunan skripsi ini.
2. Dra. Hj. Rini Marwati, M.Si., selaku Dosen Pembimbing Akademik, atas bimbingan dan perhatian yang senantiasa diberikan selama masa studi penulis di Program Studi Matematika FPMIPA UPI.
3. Seluruh dosen di lingkungan Program Studi Matematika FPMIPA UPI yang telah berbagi ilmu, wawasan, dan pengalaman berharga kepada penulis selama masa perkuliahan.
4. Bunga Cintya Dewi, Felicitia Ifianasa Lathifah, Hastialisna Hurul Aeni Setiawan, dan Humaira Rizqi Tridianti yang telah membantu secara langsung maupun tidak langsung dalam penyusunan skripsi dan perjalanan akademik penulis.

5. Mohammad Fariz Arrasyid dan Sarah Aufa Zahra yang telah bersedia menjadi bagian dari keseharian penulis selama masa perkuliahan.
6. Diva Asavahani Rahayu yang selalu menjadi tempat berbagi cerita, memberikan semangat, serta menjadi bagian penting dalam perjalanan ini.
7. Gloria Angelica Abolla yang telah menjadi tempat berbagi cerita dan berkeluh kesah, serta turut membantu penulis dalam menjalani kehidupan akademik.
8. Diri penulis sendiri karena dapat melewati berbagai rintangan dan hambatan hingga akhirnya sampai di titik ini.

ABSTRAK

Klasterisasi adalah metode untuk mengelompokkan data berdasarkan kemiripan antarunit data tanpa adanya label awal yang disediakan. Metode ini merupakan salah satu teknik *unsupervised learning* dalam *Knowledge Discovery in Databases* (KDD). Dalam konteks jumlah tenaga kesehatan secara geografis, teknik ini dapat digunakan untuk mengelompokkan wilayah berdasarkan karakteristiknya masing-masing. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan jumlah tenaga kesehatan pada tahun 2023 dengan menggunakan metode *K-Means*, di mana pemilihan pusat awal (*centroid*) dilakukan dengan metode *Centroid Linkage*. Dataset yang digunakan mencakup 12 variabel kuantitatif diskrit terkait jenis-jenis tenaga kesehatan di tiap provinsi di Indonesia. Hasil analisis menunjukkan bahwa terdapat dua klaster terbaik. Klaster pertama (Klaster 1) terdiri dari 34 provinsi yang membentuk kelompok yang relatif homogen dengan jumlah tenaga kesehatan yang cenderung rendah, nilai di bawah rata-rata pada semua variabel, serta tingkat variasi data yang lebih rendah dibandingkan dengan klaster kedua. Klaster kedua (Klaster 2) mencakup 4 provinsi dengan jumlah tenaga kesehatan yang lebih banyak, nilai di atas rata-rata untuk seluruh variabel yang dianalisis, serta tingkat variasi yang lebih tinggi. Berdasarkan nilai koefisien silhouette rata-rata, hasil klasterisasi menunjukkan skor sebesar 0,73, dengan skor individu sebesar 0,77 untuk Klaster 1 dan 0,46 untuk Klaster 2. Analisis terhadap profil klaster menunjukkan bahwa kedua klaster memiliki profil yang tidak sejajar, yang mengindikasikan adanya perbedaan karakteristik yang signifikan di antara keduanya. Hasil ini diharapkan dapat menjadi dasar bagi pengambilan kebijakan yang lebih tepat dalam rangka mendorong distribusi tenaga kesehatan yang lebih merata di seluruh wilayah Indonesia.

Kata kunci: Klasterisasi, Tenaga Kesehatan, *K-Means*, Centroid Linkage, Analisis Profil

ABSTRACT

Clustering is a method of grouping data based the on similarity within the record without any prior labels provided and it is a technique of an unsupervised learning under Knowledge Discovery in Databases (KDD). In the context of geographical health workforce distribution planning, this technique can then be used to cluster regions according to their characteristics. The purpose of this study is to group provinces in Indonesia based on the number of health workers in 2023 using the K-Means method with initial centroids selected using the Centroid Linkage method. The dataset contains 12 discrete quantitative variables that indicate the different types of health workers by province in Indonesia. The analysis showed two optimal clusters. The first cluster (Cluster 1) is a relatively homogenous group of 30 provinces that have comparatively low numbers of health workers, with below-average values on all variables and lower variability of data compared to the second cluster. The 2nd cluster represents 4 provinces with many health workers, above-average values for each of the variables assessed, and greater variability. We note that in terms of the average silhouette coefficient, the clustering result has the average of 0.73, with individual scores of 0.77 and 0.46 for cluster 1 and 2, respectively. Analysis of the profiles suggests that the two clusters have nonparallel profiles, indicating that the two clusters have substantially different characteristics. The results are anticipated to facilitate better policy making for fairer distribution of health workers population proportion across Indonesia.

Keywords: Clustering, Health workers, *K-Means*, Centroid Linkage, Profile Analysis

DAFTAR ISI

LEMBAR HAK CIPTA.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN.....	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I.....	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	8
1.3. Tujuan Penelitian	8
1.4. Manfaat Penelitian	9
BAB II.....	10
TINJAUAN PUSTAKA	10
2.1. <i>Data Mining</i>	10
2.1.1. Tahapan <i>Data Mining</i>	10
2.1.2. Metode-metode dalam <i>Data Mining</i>	12
2.2. Metode Klasterisasi.....	13
2.2.1. Klasterisasi Hirarki.....	14

2.2.2.	Klasterisasi Nonhirarki.....	20
2.3.	Mengukur Kesamaan Antarobjek	22
2.4.	Metode Optimasi Klaster	23
2.5.	Analisis Validasi Klaster.....	24
2.5.1.	<i>Silhouette Coefficient</i>	25
2.6.	Profilisasi Klaster	26
2.7.	Analisis Profil	26
BAB III		28
METODOLOGI PENELITIAN		28
3.1.	Jenis Penelitian dan Pengumpulan Data	28
3.2.	Variabel Penelitian	28
3.3.	Prapemrosesan Data	31
3.4.	Standardisasi Variabel.....	31
3.5.	Menentukan Ukuran Jarak	32
3.6.	Algoritma <i>Centroid Linkage</i>	33
3.7.	Metode <i>Elbow</i>	34
3.8.	Algoritma <i>K-Means</i>	35
3.9.	Validasi Klaster.....	36
3.9.1.	<i>Silhouette Coefficient</i>	37
3.10.	Interpretasi dan Profilisasi Klaster	40
3.11.	Analisis Profil	41
3.12.	Diagram Alir	45
BAB IV		47
HASIL DAN PEMBAHASAN		47
4.1.	Prapemrosesan Data	47
4.2.	Standardisasi Variabel.....	47
4.3.	Hasil Dendogram menggunakan Algoritma <i>Centroid Linkage</i>	48
4.4.	Penentuan Banyak Klaster Optimal	48

4.5.	Hasil Pengelompokkan dengan Algoritma <i>Centroid Linkage</i>	49
4.6.	Penentuan Nilai Pusat Klaster	51
4.7.	Hasil Pengelompokkan dengan Algoritma <i>K-Means</i>	51
4.8.	Validasi Klaster	53
4.9.	Interpretasi dan Profilisasi Klaster	54
4.10.	Analisis Profil	60
4.11.	Keterbatasan Penelitian	60
BAB 5		61
KESIMPULAN DAN SARAN		61
5.1.	Kesimpulan	61
5.2.	Saran	62
DAFTAR PUSTAKA		64
LAMPIRAN		75

DAFTAR TABEL

Tabel 3.1 Variabel Dan Definisi Operasional	29
Tabel 3.2 Kriteria Pengukuran Pengelompokkan	40
Tabel 4.1 Standardisasi Variabel.....	47
Tabel 4.2 Klaster	50
Tabel 4.3 Nilai Pusat Klaster Untuk Data Jumlah Tenaga Kesehatan.....	51
Tabel 4.4 Klaster Untuk Pengelompokkan Algoritma <i>K-Means</i>	52
Tabel 4.5. Nilai Silhouette Coefficient	53
Tabel 4.6. Nilai Rata-Rata Klaster	56
Tabel 4.7. Nilai Simpangan Baku Klaster.....	56

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Penelitian	45
Gambar 3.2 Diagram Alir Penelitian (Lanjutan).....	46
Gambar 4.1 Dendogram Klasterisasi Dengan Algoritma <i>Centroid Linkage</i>	48
Gambar 4.2 Metode Elbow	48
Gambar 4.3 Dendogram <i>Centroid Linkage</i> Dengan $K = 2$	49
Gambar 4.4 Visualisasi Klaster.....	53
Gambar 4.5 Plot Nilai Silhouette Coefficient	54
Gambar 4.6 Grafik Nilai Rata-Rata Klaster.....	55

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian	75
Lampiran 2 Standardisasi Variabel Data Jumlah Tenaga Kesehatan Tahun 2023....	80
Lampiran 3 Nilai <i>Elbow</i> Untuk Data Jumlah Tenaga Kesehatan Tahun 2023	84
Lampiran 4 <i>Syntax Software R</i> Untuk Standardisasi Variabel Dan Jarak.....	85

DAFTAR PUSTAKA

- Ackerman, M., Ben-David, S., Brânzei, S., & Loker, D. (2021). Weighted clustering: Towards solving the user's dilemma. *Pattern Recognition*, 120. <https://doi.org/10.1016/j.patcog.2021.108152>
- Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. Dalam *Jurnal Sistem dan Informatika* (Vol. 3).
- Ahmed, M. H., Tiun, S., Omar, N., & Sani, N. S. (2023). Short Text Clustering Algorithms, Application and Challenges: A Survey. Dalam *Applied Sciences (Switzerland)* (Vol. 13, Nomor 1). MDPI. <https://doi.org/10.3390/app13010342>
- Aldahdooh, R. T., & Ashour, W. (2013). DIMK-means "Distance-based Initialization Method for K-means Clustering Algorithm." *International Journal of Intelligent Systems and Applications*, 5(2), 41–51. <https://doi.org/10.5815/ijisa.2013.02.05>
- Arai, K., & Ridho Barakbah, A. (2007). Hierarchical K-means: an algorithm for centroids initialization for K-means. Dalam *Rep. Fac. Sci. Engrg. Reports of the Faculty of Science and Engineering* (Vol. 36, Nomor 1).
- Archika Jain, Devendra Somwanshi, Kapil Joshi, & Siddharth Shankar Bhatt. (2022). A Review: Data Mining Classification Techniques. *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, 636–642.
- Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.
- Atthina, N., & Iswari, L. (2014, Juni 21). Klasterisasi Data Kesehatan Penduduk untuk Menentukan Rentang Derajat Kesehatan Daerah dengan Metode K-Means. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*.
- Bacher, J., Wenzig, K., & Vogler, M. (2004). *SPSS TwoStep Cluster-a first evaluation*. <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-327153>
- Badan Pusat Statistik Indonesia. (2025, Maret 12). *Jumlah Tenaga Kesehatan Menurut Provinsi, 2023*.
- Bano, S., & Khan, M. N. A. (2018). A Survey of Data Clustering Methods. *International Journal of Advanced Science and Technology*, 113, 133–142. <https://doi.org/10.14257/ijast.2018.113.14>

- Bao, C. (2021). K-means clustering algorithm: a brief review. *Academic Journal of Computing & Information Science*, 4(5). <https://doi.org/10.25236/ajcis.2021.040506>
- Bateni, M. H., Dhulipala, L., Fletcher, W., Gowda, K. N., Hershkowitz, D. E., Jayaram, R., & Łacki, J. (2024). *Efficient Centroid-Linkage Clustering*. <http://arxiv.org/abs/2406.05066>
- Bholowalia, P., & Kumar, A. (2014). EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, 105(9), 975–8887.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bulut, O., & Desjardins, C. D. (2020). *Profile Analysis of Multivariate Data: A Brief Introduction to the profileR Package*. <https://doi.org/10.31234/osf.io/sgy8m>
- Candra, Y., Goejantoro, R., & Dani, A. T. R. (2024). Pengelompokan Provinsi Berdasarkan Indikator Ekonomi, Pendidikan, Kesehatan, dan Kriminalitas di Indonesia Menggunakan Algoritma Centroid Linkage. *Euler: Jurnal Ilmiah Matematika, Sains dan Teknologi*, 12(1), 9–15. <https://doi.org/10.37905/euler.v12i1.24887>
- Capó, M., Pérez, A., & Lozano, J. A. (2018). *An efficient K-means clustering algorithm for massive data*. <http://arxiv.org/abs/1801.02949>
- Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2018). *Deep Clustering for Unsupervised Learning of Visual Features*. <http://arxiv.org/abs/1807.05520>
- Cheng, J., Chen, X., Yang, H., & Leng, M. (2012). *AN ENHANCED K-MEANS ALGORITHM USING AGGLOMERATIVE HIERARCHICAL CLUSTERING STRATEGY*.
- Coates, A., & Ng, A. Y. (2012). Learning Feature Representations with K-means. Dalam *Neural Networks: Tricks of the Trade* (hlm. 561–580). Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-35289-8_30
- Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., & Mathieu, C. (2019). Hierarchical clustering. *Journal of the ACM*, 6. <https://doi.org/10.1145/3321386>
- Creswell, J. W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (4 ed.). Pearson.

- Dalal, M. A., & Harale, N. D. (2011). A Survey on Clustering In Data Mining. *International Conference and Workshop on Emerging Trends in Technology (ICWET 2011)*, 559–562.
- Dasgupta, S. (2008). The Hardness of K-Means Clustering. Dalam *Technical Reports*.
- Davies, P. B. (2004). *Database Systems* (3 ed.). Red Globe Press London.
- Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications*. Wiley.
- Dong, G. (2002). Knowledge Discovery in Databases. Dalam *In Encyclopedia of Software Engineering* (2nd ed.). John Wiley & Sons.
- Dunham, M. H. (2003). *Data Mining: Introductory and Advanced Topics*. Prentice Hall.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis* (5 ed.). John Wiley & Sons, Ltd.
- Fang, C., & Liu, H. (2021). Research and application of improved clustering algorithm in retail customer classification. *Symmetry*, 13(10). <https://doi.org/10.3390/sym13101789>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3), 37. <https://doi.org/https://doi.org/10.1609/aimag.v17i3.1230>
- Gan, G., Ma, C., & Wu, J. (2007). Data Clustering: Theory, Algorithms, and Applications. Dalam *Data Clustering: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718348>
- García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. *Knowledge-Based Systems*, 98, 1–29. <https://doi.org/10.1016/j.knosys.2015.12.006>
- Goel, A. (2014). A Study of Different Partitioning Clustering Technique. Dalam *IJSRD-International Journal for Scientific Research & Development* (Vol. 2). www.ijsrd.com
- Gordon, A. D. (1996). Hierarchical Classification. Dalam P. Arabie, L. Hubert, & G. De Soete (Ed.), *Clustering and Classification* (hlm. 65–121). World Scientific.

- Gower, J. C. (1967). A Comparison of Some Methods of Cluster Analysis. *Biometrics*, 23(4), 623–637.
- Hamzah, N. A., Kek, S. L., & Saharan, S. (2017). The Performance of K-Means and K-Modes Clustering to Identify Cluster in Numerical Data. Dalam *Journal of Science and Technology* (Vol. 9, Nomor 3).
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3 ed.). Elsevier.
- Harnanto, Y. I., Rusgiyono, A., & Wuryandari, T. (2017). PENERAPAN ANALISIS KLASSTER METODE WARD TERHADAP KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN PENGGUNA ALAT KONTRASEPSI. *JURNAL GAUSSIAN*, 6, 528–537. <http://ejournal-s1.undip.ac.id/index.php/gaussian>
- Hartigan, J. A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., & Abdalla, N. N. (2024). From A-to-Z Review of Clustering Validation Indices. *Neurocomputing*, 601(C).
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Data Mining, Inference, and Prediction. Dalam *The Elements of Statistical Learning* (2 ed.). Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Herdiana, I., Kamal, M. A., Triyani, Estri, M. N., & Renny. (2025). *A More Precise Elbow Method for Optimum K-means Clustering*. <http://arxiv.org/abs/2502.00851>
- Hermawati, F. A. (2013). *Data Mining* (1 ed.). <https://www.researchgate.net/publication/317692865>
- Hidayanti, H. (2019). Pemerataan Tenaga Kesehatan di Kabupaten Lamongan. *CAKRAWALA*, 12(2), 162–177. <https://doi.org/10.32781/cakrawala.v12i2.272>
- Hidayat, F. P., & Hakim, R. B. F. (2021). IMPLEMENTASI METODE CLUSTERING K-MEDOIDS DALAM MENGELOMPOKKAN JUMLAH ADUAN DI KABUPATEN SLEMAN. *Prosiding Sendika*.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>

- Irwanto, Purwananto, Y., & Soelaiman, R. (2012). Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra. *Jurnal Teknik ITS*, 1(1), 197–02.
- Izzuddin, K. H., & Wijayanto, A. W. (2024). Pemodelan Clustering Ward, K-Means, Diana, dan PAM dengan PCA untuk Karakterisasi Kemiskinan Indonesia Tahun 2021. *Komputika : Jurnal Sistem Komputer*, 13(1), 41–53. <https://doi.org/10.34010/komputika.v13i1.10803>
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323. <https://doi.org/https://doi.org/10.1145/331499.331504>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (5th ed.). Pearson Prentice Hall.
- Kapita, S. N., & Abdullah, S. Do. (2020). Pengelompokan Data Mutu Sekolah Menggunakan Jaringan Syaraf Tiruan Kohonen-SOM. *JIKO (Jurnal Informatika dan Komputer)*, 3(1), 56–61.
- Kasiram, M. (2010). *Metodologi Penelitian: Kualitatif-Kuantitatif*. UIN Maliki Press.
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- Kaufmann, L., & Rousseeuw, P. (1987). Clustering by Means of Medoids. *Data Analysis based on the L1-Norm and Related Methods*, 405–416.
- Kaur, R., Jyoti, K., & Kumar, R. (2014). A Review Paper on Knowledge Discovery and Data Mining Techniques. Dalam *International Journal of Advanced Research in Computer Science and Software Engineering* (Vol. 4, Nomor 8). www.ijarcsse.com
- Kellom, M., & Raymond, J. (2017). Using cluster edge counting to aggregate iterations of centroid-linkage clustering results and avoid large distance matrices. *Journal of Biological Methods*, 4(1), 1. <https://doi.org/10.14440/jbm.2017.153>
- Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters*, 25(11), 1293–1302. <https://doi.org/10.1016/j.patrec.2004.04.007>

- Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6). www.ijarcsms.com
- Kotler, P., & Keller, K. L. (2016). *Marketing management*. Pearson Education Limited.
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Dalam *Informatica* (Vol. 31).
- Kusmanto, S., Watrianthos, R., & Suryadi, S. (2023). Distribusi Spasial Unmet Need Pelayanan Kesehatan dengan Algoritma K-Means untuk Pemetaan Provinsi di Indonesia. *Bulletin of Information Technology (BIT)*, 4(2), 361–368. <https://doi.org/10.47065/bit.v3i1>
- Lance, G. N., & Williams, W. T. (1967). A General Theory of Classificatory Corting Strategies 1. Hierarchical Systems. *The Computer Journal*, 9(4), 373–380.
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc.
- Liantoni, F., & Cahyani, L. (2017). PEMANFAATAN HIERARCHICAL CLUSTERING UNTUK PENGELOMPOKKAN DAUN BERDASARKAN FITUR MOMENT INVARIANT. Dalam *Jurnal Ilmiah Edutic* (Vol. 3, Nomor 2).
- Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of internal clustering validation measures. *IEEE International Conference on Data Mining, ICDM*, 911–916. <https://doi.org/10.1109/ICDM.2010.35>
- Macqueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observation. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281–297.
- Madhulatha, T. S. (2012). AN OVERVIEW ON CLUSTERING METHODS. *IOSR Journal of Engineering*, 2(4), 719–725. www.iosrjen.org
- Madрахimov, S., Makharov, K., & Lolaev, M. (2021). Data Preprocessing on Input. *INTERNATIONAL UZBEKISTAN-MALAYSIA CONFERENCE ON “COMPUTATIONAL MODELS AND TECHNOLOGIES (CMT2020)”*.
- Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>

- McQuitty, L. L. (1960). Hierarchical Syndrome Analysis. *Educational and Psychological Measurement*, 20(2), 293–304.
- Menéndez, H. D. (2021). Clustering: Finding Patterns in the Darkness. *Journal of Machine Learning*.
- Mitchell, T. M. . (1997). *Machine Learning*. McGraw-Hill.
- Muflihan, Y., Retnawati, H., & Kistian, A. (2022). Analisis Cluster dengan Metode Hierarki untuk Pengelompokan Sekolah Menengah Atas Berdasarkan Raport Mutu Sekolah di Kabupaten Nagan Raya. *Measurement In Educational Research (Meter)*, 2(1), 22. <https://doi.org/10.33292/meter.v2i1.155>
- Mukhopadhyay, S. (2018). Unsupervised Learning: Clustering. Dalam *Advanced Data Analytics Using Python* (hlm. 77–98). Apress. https://doi.org/10.1007/978-1-4842-3450-1_4
- Munawar, Z., Herdiana, Y., & Putri, N. I. (2021). Sistem Rekomendasi Hibrid Menggunakan Algoritma Apriori Mining Asosiasi. *TEMATIK - Jurnal Teknologi Informasi Dan Komunikasi*, 8(1), 84–95.
- Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *Computer Journal*, 26(4), 354–359. <https://doi.org/10.1093/comjnl/26.4.354>
- Muslikh, A. R., Andono, P. N., Marjuni, A., & Santoso, H. A. (2023, September 16). Systematic Literature Review of Data Distribution in Preprocessing Stage with Focus on Outliers. *2023 International Seminar on Application for Technology of Information and Communication (iSemantic)*.
- Nahdliyah, M. A., Widiharah, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index Dan C-Index (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018). *Jurnal Gaussian*, 8(2), 161–170. <http://ejournal3.undip.ac.id/index.php/gaussian>
- Nanda, A., & Rout, S. K. (2010). *Data Mining & Knowledge Discovery in Databases: An AI Perspective*.
- Narang, B., Verma, P., & Kochar, P. (2016). Application based, advantageous k-means algorithm. *INTERNATIONAL JOURNAL OF LATEST TRENDS IN ENGINEERING AND TECHNOLOGY*, 7(2). <https://doi.org/10.21172/1.72.520>

- Nazari, M., Hussain, A., & Musilek, P. (2023). Applications of Clustering Methods for Different Aspects of Electric Vehicles. Dalam *Electronics (Switzerland)* (Vol. 12, Nomor 4). MDPI. <https://doi.org/10.3390/electronics12040790>
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer dan Informatika*, 4(1), 61–72. <https://doi.org/10.54082/jiki.168>
- Hikmah, N. B., Rahman, H., & Puspitasari, A. (2020). Membandingkan Ketimpangan Ketersediaan Tenaga Kesehatan Puskesmas di Wilayah Indonesia Timur. *Window of Public Health Journal*, 1(1), 31–37.
- Pacovský, J. (2017). Data Processing. Dalam *Encyclopedia of Machine Learning and Data Mining*. Springer, Boston, MA.
- Piatetsky-Shapiro, G. (1991). Discovery, Analysis, and Presentation of Strong Rules. Dalam *Knowledge Discovery in Databases* (hlm. 255–264). MIT Press.
- Pireddu, A., Bedini, A., Lombardi, M., Ciribini, A. L. C., & Berardi, D. (2024). A Review of Data Mining Strategies by Data Type, with a Focus on Construction Processes and Health and Safety Management. Dalam *International Journal of Environmental Research and Public Health* (Vol. 21, Nomor 7). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/ijerph21070831>
- Pramudiono, I. (2007). *Pengantar Data Mining: Menambang Permata Pengetahuan di Gunung Data*. 4(2).
- Purwaningsih, E. (2023). Kebijakan Terkait Krisis Kesehatan: Analisa Kebutuhan Tenaga Kesehatan Selama Pandemi Covid-19 di Indonesia. *Jurnal Kebijakan Kesehatan Indonesia: JKKI*, 12(2), 66–73.
- Putriana, U., Setyawan, Y., & Noeryanti. (2016). Metode Cluster Analysis untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Variabel yang Mempengaruhi Kemiskinan pada Tahun 2013. *urnal Statistika Industri dan Komputasi*, 1(1).
- Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Bulletin on Data Engineering*, 7, 182–185. <http://dbs.uni-leipzig.de>
- Rahmawati, L. (2013). *Analisis Kelompok dengan Menggunakan Metode Hierarki untuk Pengelompokan Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kesehatan* [Skripsi]. Universitas Negeri Malang.

- Rani¹, Y., & Rohil, H. (2013). A Study of Hierarchical Clustering Algorithm. Dalam *International Journal of Information and Computation Technology* (Vol. 3, Nomor 11). <http://www.irphouse.com/ijict.htm>
- Rendón, E., & María Abundez, I. (2011). Internal Versus External Cluster Validation Indexes. *International Journal of Computers*. <https://www.researchgate.net/publication/284387933>
- Riduwan. (2013). *Metode dan Teknik Menyusun Tesis*. Alfabeta.
- Rivandi, A., Bu'ulolo, E., & Silalahi, N. (2019). Penerapan Metode Regresi Linier Berganda dalam Estimasi Biaya Pencetakan Spanduk (Studi Kasus : PT. Hansindo Setiaprata). *Jurnal Penelitian Pelita Informatika Budi Darma*, 7(3), 263–268.
- Santoso, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis* (1 ed.). Graha Ilmu.
- Santoso, S. (2010). *Statistik Multivariat dengan SPSS*. PT Elex Media Komputindo.
- Saraswi, E., Perdana, H., & Fakhrunnisa, A. (2024). Distribusi Tenaga Kesehatan di Kalimantan Barat Menggunakan Metode Ward. *Forum Analisis Statistik*, 4(1), 39–48.
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press. <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2020). A Quantitative Discriminant Method of Elbow Point for the Optimal Number of Clusters in Clustering Algorithm. *EURASIP Journal on Wireless Communications and Networking*.
- Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika "MANTIK,"* 4(1), 22–31. <https://doi.org/10.15642/mantik.2018.4.1.22-31>
- Simamora, B. (2005). *Analisis Multivariat Pemasaran*. PT. Gramedia Pustaka Utama.
- Sintiya, S., Laksana, T. G., & Tanjung, N. A. F. (2021). Kombinasi Single Linkage Dengan K-Means Clustering Untuk Pengelompokan Wilayah Desa Kabupaten Pematang. *Journal of Innovation Information Technology and Application (JINITA)*, 3(1), 17–27. <https://doi.org/10.35970/jinita.v3i1.589>

- Sitinjak, D. K., Pangestu, B. A., & Sari, B. N. (2022). Clustering Jumlah Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means. *Journal of Applied Informatics and Computing (JAIC)*, 6(1), 46–54.
- Sivasankari, S., & Sukumaran, Dr. S. (2024). Supervised, Unsupervised Learning Methods in Data Mining: A Comprehensive Survey. *International Journal for Research in Applied Science and Engineering Technology*, 12(9), 507–517. <https://doi.org/10.22214/ijraset.2024.64224>
- Sneath, P. H. A. (1957). The Application of Computers to Taxonomy. *Journal General Microbiology*, 17, 201–226. [https://doi.org/https://doi.org/10.1099/00221287-17-1-201](https://doi.org/10.1099/00221287-17-1-201)
- Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of Numerical Taxonomy*. W. H. Freeman.
- Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1–34. <https://doi.org/10.1348/000711005X48266>
- Sukmawarni, Lisnawaty, & Hartoyo, A. M. (2022). Analisis Kebutuhan Tenaga Kesehatan dengan Metode Workload Indicator Staffing Need (WISN) di Puskesmas Tahun 2022. *Jurnal Administrasi dan Kebijakan Kesehatan*, 3(3), 113–122. <http://ojs.uho.ac.id/index.php/jakk->
- Sutomo, B., Nurdianto, H., & Saprudin, U. (2023). *Python Basic Programing*. Literasi Nusantara.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Pearson.
- Tan, S. C., Ting, K. M., & Teng, S. W. (2011). A general stochastic clustering method for automatic cluster discovery. *Pattern Recognition*, 44(10–11), 2786–2799. <https://doi.org/10.1016/j.patcog.2011.04.001>
- Tomar, D., & Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241–266. <https://doi.org/10.14257/ijbsbt.2013.5.5.25>
- Turnip, A. E. (2023). *PENGELOMPOKAN PROVINSI MENGGUNAKAN HYBRID CLUSTERING PADA PROVINSI DI INDONESIA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN KEBUDAYAAN* [Skripsi]. Universitas Padjadjaran.

- Vasudha, G. (2020). Web Mining: Knowledge Discovery in Data Base. *Int J Swarm Evol Comput*, 9(6), 100–194.
- Veeraiah, D., & Vasumathi, D. D. (2014). A link concerning various clusters using hierarchical clustering Techniques. *International Journal of Engineering Research & Technology (IJERT)*, 2. www.ijert.org
- Ward, J. H., & Hook, M. E. (1963). Application of an Hierarchial Grouping Procedure to a Problem of Grouping Profiles. *Educational and Psychological Measurement*, 23(1), 69–81.
- Wedel, M., & Kamakura, W. A. (2000). Profiling Segments. Dalam *Market Segmentation: Vol. Springer, Boston, MA*.
https://doi.org/https://doi.org/10.1007/978-1-4615-4651-1_9
- Winarta, A., & Kurniawan, W. J. (2021). OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON. *Jurnal Teknik Informatika Kaputama (JTik)*, 5(1).
- Wishart, D. (1969). An Algorithm for Hierarchical Classifications. *Biometrics*, 25(1), 165–170.
- Witten, I. H., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (2 ed.). Morgan Kaufmann Publisher.
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. Dalam *IEEE Transactions on Neural Networks* (Vol. 16, Nomor 3, hlm. 645–678).
<https://doi.org/10.1109/TNN.2005.845141>
- Zaki, M. J., & Meira, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.
- Zhao, X. Q. (2010). Clustering Validity Analysis Based on Silhouette Coefficient. *Journal of Computer Applications*.