LANCAH
JURNAL INOVASI DAN TREN
Vol. 3, No. 1, Mei 2025 pp. 227~234
e-ISSN: 3025-8189 ; p-ISSN: 3025-8197
DOI: https://doi.org/10.35870/ljit.v3i1.4028

Perbandingan Metode K-Means dan Hierarchical Clustering dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur

Ihsan Pratama Putra[1], Avram Fadhillah[2]
[1,2] Teknik Informatika, STMIK-IM, Indonesia

Info Artikel

Riwayat Artikel:

Diterima 12-04-2025
Disetujui 13-04-2025
Diterbitkan 14-04-2025

Kata kunci:

Clustering;
Hierarchical Clustering;
K-Means;
Kemiskinan;
Pengambilan Kebijakan.

ABSTRACT

Penelitian ini membahas perbandingan antara dua metode klasterisasi, yaitu K-Means dan Hierarchical Clustering, dalam mengelompokkan data penduduk miskin di Kabupaten Cianjur pada periode 2018-2021. K-Means digunakan untuk membagi data ke dalam jumlah klaster tertentu yang telah ditentukan sebelumnya, sementara Hierarchical Clustering memungkinkan pembentukan klaster berbasis hierarki tanpa memerlukan jumlah klaster awal. Hasil analisis menunjukkan bahwa K-Means memiliki keunggulan dalam hal kecepatan dan efisiensi, terutama untuk dataset besar dengan distribusi data yang seragam. Di sisi lain, Hierarchical Clustering lebih efektif dalam mengungkap pola yang kompleks dan menawarkan visualisasi dendrogram untuk analisis mendalam. Dengan menggunakan kedua metode ini, pengambilan keputusan terkait kebijakan sosial dapat lebih terarah, berdasarkan analisis data yang akurat dan mendalam. Penelitian ini memberikan kontribusi penting untuk mendukung program pengentasan kemiskinan secara lebih optimal.

Penulis Korespodensi:

Ihsan Pratama Putra
Teknik Informatika, STMIK-IM, Indonesia
Email: iicanpp6@gmail.com

Cara Sitasi Artikel ini dalam APA:

Putra, I. P., & Fadhillah, A. (2025). Perbandingan Metode K-Means dan Hierarchical Clustering dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur. LANCAH: Jurnal Inovasi Dan Tren, 3(1), 227~234. https://doi.org/10.35870/ljit.v3i1.4028

PENDAHULUAN

Pentingnya Pengelompokan Data Penduduk Miskin untuk Analisis Kebijakan Sosial (Sudibyo et al., n.d.). Pengelompokan data penduduk miskin memiliki peran krusial dalam analisis kebijakan sosial. Dengan data yang akurat dan rinci, pemerintah serta organisasi sosial dapat mengidentifikasi kelompok paling rentan yang memerlukan bantuan (Novitasari et al., 2023). Data dari survei rumah tangga sering kali gagal mencakup kelompok marginal seperti tunawisma, populasi institusi, dan komunitas nomaden, sehingga menimbulkan bias dalam analisis ketimpangan pendapatan dan distribusi sumber daya (Carr-Hill, 2015; Edmiston, 2023; Sonang et al., 2019). Selain itu, ketidakseimbangan kelas dalam data dapat menghasilkan kesalahan klasifikasi, yang berpotensi memengaruhi kebijakan sosial yang diambil(Santoso et al., 2018).

Penelitian ini bertujuan membandingkan dua metode pengelompokan data penduduk miskin, yaitu K-Means dan Hierarchical Clustering, berdasarkan akurasi, kemudahan interpretasi, dan penerapan praktisnya. Metode K-Means sering digunakan untuk mengelompokkan penduduk miskin ke dalam kategori seperti miskin, sederhana, dan mampu, meskipun hasilnya terkadang tidak menunjukkan perubahan signifikan antar kelompok (Tarigan et al., 2023). Sebaliknya, metode hierarkis seperti Ward's Method terbukti lebih efektif dalam mengidentifikasi pola penyebab kemiskinan, dengan nilai Root Mean Square Standard Deviation (RMSSTD) yang lebih rendah, menunjukkan kualitas klaster yang lebih baik (Mongi et al., 2019).

K-Means Clustering adalah metode partisi yang mengelompokkan data ke dalam sejumlah klaster (K) yang telah ditentukan sebelumnya. Metode ini bekerja dengan menentukan centroid untuk setiap klaster, kemudian setiap data dikelompokkan ke klaster dengan centroid terdekat menggunakan jarak Euclidean. Keunggulan utama K-Means adalah efisiensinya dalam menangani dataset besar dan kemampuannya untuk mengelompokkan data dengan jumlah klaster yang jelas. Namun, metode ini memiliki kelemahan, seperti sensitivitas terhadap pemilihan awal centroid dan ketidakmampuannya dalam menangani klaster dengan bentuk yang tidak reguler atau ukuran yang berbeda. K-Means banyak digunakan dalam segmentasi pelanggan, analisis pasar, dan pengelompokan berdasarkan wilayah geografis (*FullBook Pengenalan Data Mining*, n.d.).

Hierarchical Clustering adalah metode klasterisasi yang menghasilkan hierarki data dalam bentuk dendrogram, menggunakan pendekatan agglomerative (bottom-up) atau divisive (top-down). Tidak seperti K-Means, metode ini tidak memerlukan jumlah klaster yang ditentukan di awal, sehingga dapat mengungkapkan hubungan hierarkis antar data. Kelebihan utamanya adalah fleksibilitas dalam menentukan klaster dan kemudahan interpretasi melalui visualisasi dendrogram. Namun, metode ini memiliki keterbatasan dalam hal efisiensi komputasi untuk dataset besar dan rentan terhadap outlier. Hierarchical Clustering sering digunakan dalam taksonomi, bioinformatika, serta analisis sosial-ekonomi. Penelitian sebelumnya menunjukkan bahwa Hierarchical Clustering lebih unggul dalam mengidentifikasi struktur data yang kompleks, sementara K-Means lebih efektif untuk dataset besar dengan klaster berbentuk bulat dan seragam.

Penelitian ini menawarkan implikasi penting bagi pengambilan keputusan di tingkat pemerintah dan organisasi sosial. Dengan metode pengelompokan yang lebih akurat dan dapat diinterpretasikan, kebijakan sosial dapat dirancang lebih tepat sasaran dan efektif dalam mengurangi kemiskinan. Sebagai contoh, metode over-sampling sintetis telah terbukti meningkatkan akurasi klasifikasi rumah tangga miskin, yang dapat membantu perencanaan serta evaluasi kebijakan sosial (Santoso et al., 2018; Amaliyah & Rianti Agustini, n.d.). Selain itu, analisis data yang lebih mendalam dapat mengungkapkan ketidakadilan struktural yang menjadi akar masalah kemiskinan (Klasen, 1997).

Penelitian ini bertujuan untuk memberikan wawasan yang lebih baik melalui perbandingan

metode K-Means dan Hierarchical Clustering, sehingga dapat mendukung pengambilan keputusan kebijakan sosial yang lebih efektif dan membantu mengurangi ketimpangan.

METODE PELAKSANAAN

Penelitian ini menggunakan metode berikut. Data sekunder mengenai tingkat kemiskinan dan penerima bantuan di Kabupaten Cianjur dari tahun 2018 hingga 2021 dikumpulkan sebagai dasar analisis. Selanjutnya, dilakukan teknik analisis data dengan dua pendekatan klasterisasi. Pada proses K-Means Clustering, jumlah klaster optimal ditentukan menggunakan metode Elbow dan Silhouette Index untuk memastikan hasil klasterisasi yang paling representatif. Sementara itu, pada proses Hierarchical Clustering, digunakan pendekatan single linkage untuk mengukur jarak terpendek antara dua klaster. Hasil dari kedua metode divisualisasikan dengan cara berbeda. Untuk K-Means, scatter plot digunakan untuk menggambarkan distribusi data dalam klaster. Sedangkan untuk Hierarchical Clustering, dendrogram dipakai untuk mempermudah analisis hubungan hierarkis antar data. Evaluasi hasil dari kedua metode ini dilakukan untuk membandingkan efektivitasnya dalam menghasilkan klaster yang relevan dan representatif, sehingga dapat mendukung pengambilan keputusan kebijakan yang lebih tepat.

HASIL DAN PEMBAHASAN

Data tingkat kemiskinan di Kabupaten Cianjur dari 2018 sampai dengan 2021 diambil dari Badan Pusat Statistik (BPS, 2022) merupakan jumlah penduduk miskin (ribu). Berikut adalah tabel kemiskinan dan penerima bantuan di Kabupaten Cianjur dari 2018 sampai dengan 2021 tersaji pada Gambar 1.

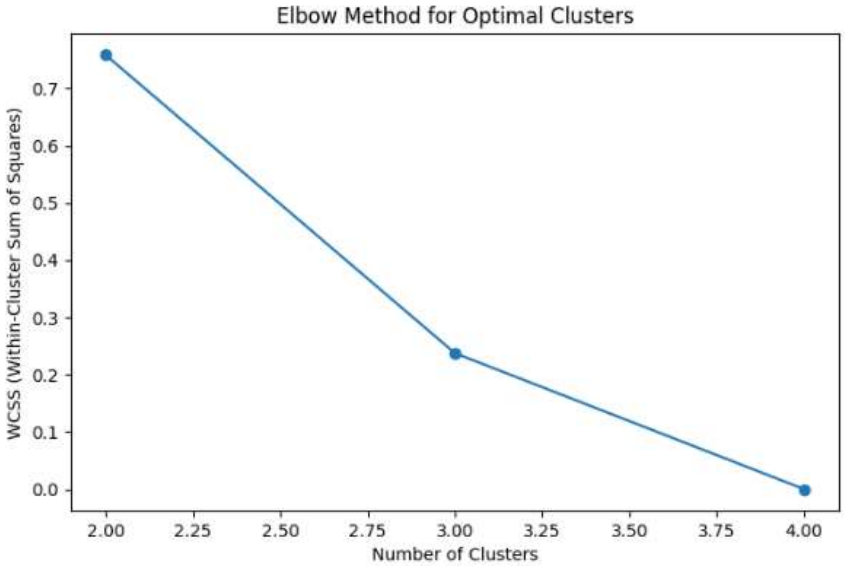
Wilayah Kabupaten	Jumlah Penduduk Miskin di Kabupaten Cianjur (Ribu Jiwa)	
	2020	2021
Cianjur	234,5	260,0

Wilayah Kabupaten	Jumlah Penerima Bantuan di Kabupaten Cianjur					
	Rencana			Realisasi		
	2020	2019	2018	2020	2019	2018
Cianjur	212.715	189.959	189.959	212.715	189.959	1.899.590

Gambar 1. Tabel tingkat kemiskinan dan penerima bantuan

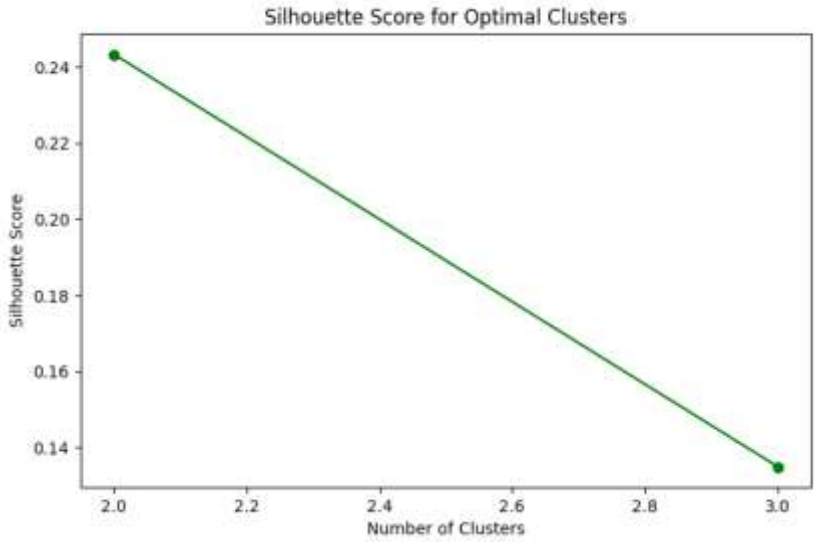
Dari Gambar 1 terlihat bahwa data jumlah penduduk miskin dan penerima bantuan di Kabupaten Cianjur. Pada tahun 2020, jumlah penduduk miskin tercatat 234,5 ribu jiwa, meningkat menjadi 260,0 ribu jiwa pada tahun 2021. Sementara itu, data penerima bantuan menunjukkan bahwa pada tahun 2020, rencana dan realisasi penerima bantuan sama, yaitu 212.715 orang, sedangkan pada tahun 2019 dan 2018, rencana penerima bantuan tetap di angka 189.959 orang. Realisasi pada tahun 2019 juga sesuai dengan rencana, tetapi pada tahun 2018 terdapat lonjakan signifikan, dengan realisasi mencapai 1.899.590 orang, jauh melebihi rencana. Data ini mengilustrasikan adanya peningkatan jumlah penduduk miskin serta fluktuasi signifikan dalam penerimaan bantuan pada

tahun-tahun tertentu. Selanjutnya data dimasukkan dalam dimasukan ke dalam VSCode. Dengan bantuan VSCode diperoleh hasil sebagai berikut.



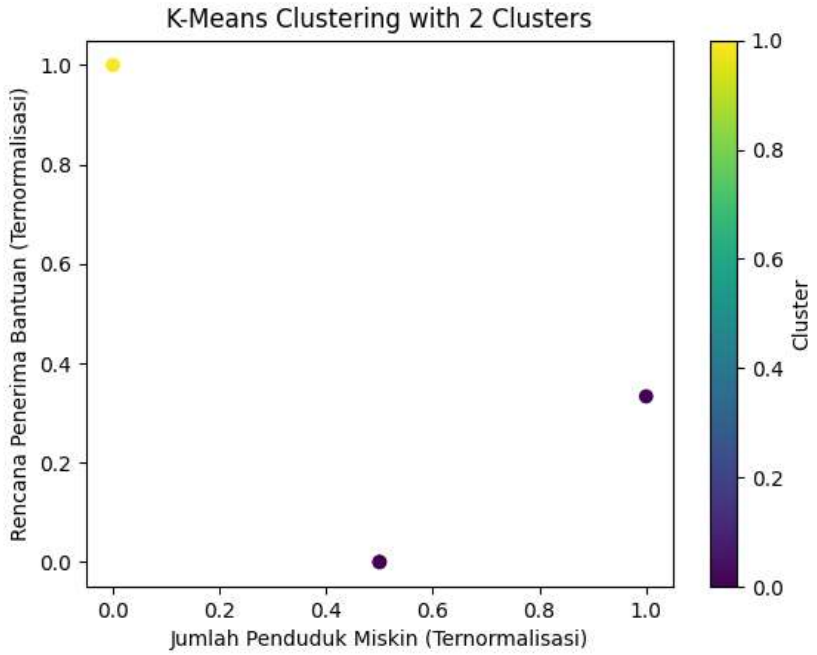
Gambar 2. *Analisis cluster metode Elbow*

Dari grafik di atas menggambarkan hasil dari metode elbow yang digunakan untuk menentukan jumlah cluster optimal dalam sebuah dataset. Pada grafik ini, sumbu x menunjukkan jumlah cluster yang diuji, sementara sumbu y menggambarkan nilai WCSS (Within-Cluster Sum of Squares), yang merupakan jumlah kuadrat jarak antara setiap titik data dengan centroid cluster terdekat. Secara umum, grafik ini menunjukkan penurunan nilai WCSS seiring dengan peningkatan jumlah cluster. Namun, terdapat titik belok yang jelas di sekitar 3 cluster, yang menunjukkan bahwa penambahan cluster lebih lanjut setelah angka tersebut tidak menghasilkan penurunan WCSS yang signifikan. Oleh karena itu, metode elbow menunjukkan bahwa jumlah cluster yang paling optimal untuk dataset ini adalah 3. Berikut adalah hasil analisi cluster metode Silhouette Score disajikan pada Gambar 3.



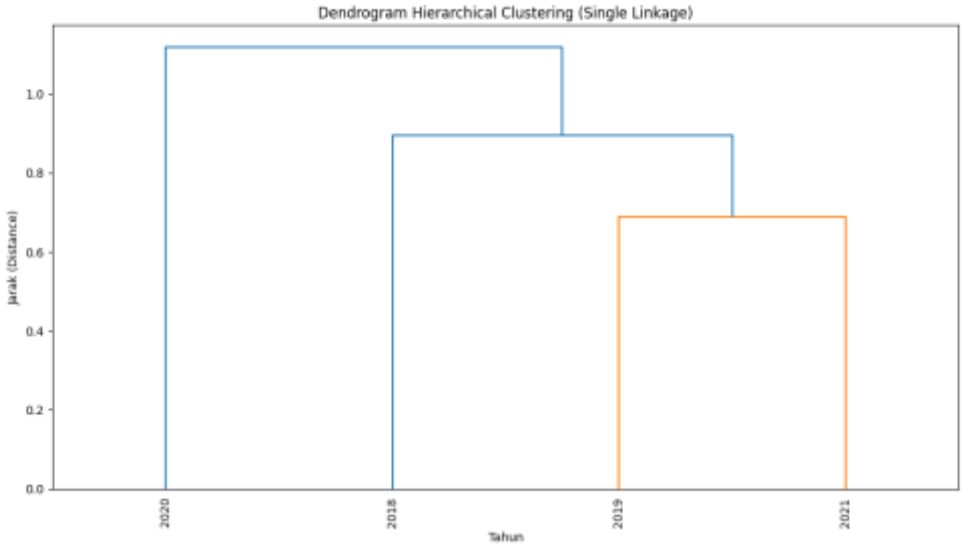
Gambar 3. Analisis cluster metode Silhouette Score

Pada grafik ini sumbu x menunjukkan jumlah cluster yang diuji, sementara sumbu y menunjukkan nilai silhouette score. Nilai silhouette score mengukur sejauh mana setiap titik data cocok dengan cluster yang ditetapkan dibandingkan dengan cluster lainnya. Semakin tinggi nilai silhouette score, semakin baik data point tersebut tergabung dalam cluster yang sesuai. Grafik ini menunjukkan bahwa nilai silhouette score umumnya menurun seiring dengan bertambahnya jumlah cluster, dengan penurunan paling tajam terjadi antara 2 dan 3 cluster. Oleh karena itu, berdasarkan metode silhouette, jumlah cluster yang optimal untuk dataset ini adalah 2. Selanjutnya akan dilakukan clustering algoritma K-Means tersaji pada gambar 4.



Gambar 4. Visualisasi Scatter plot Algoritma K-Means

Hasil clustering menggunakan algoritma K-Means dengan jumlah cluster sebanyak 2. Data yang digunakan telah dinormalisasi, meliputi jumlah penduduk miskin dan rencana penerima bantuan. Setiap titik pada grafik merepresentasikan satu data, dengan warna titik yang menunjukkan cluster yang sesuai. Sumbu X menggambarkan jumlah penduduk miskin yang telah dinormalisasi, sementara sumbu Y menggambarkan rencana penerima bantuan yang juga telah dinormalisasi. Dari visualisasi ini, dapat terlihat bahwa data terbagi menjadi dua kelompok yang terpisah dengan jelas. Hasil dari Hierarchical Clustering tersaji pada gambar 5.



Gambar 5. Visualisasi Dendogram Hierarchical Clustering

Hasil clustering hierarkis dengan menggunakan metode single linkage. Sumbu X menunjukkan tahun (2020, 2018, 2019, 2021), sementara sumbu Y menggambarkan jarak atau dissimilaritas antar kelompok data. Garis vertikal yang ada menghubungkan kelompok data yang digabungkan pada setiap tahap clustering, dengan panjang garis vertikal menunjukkan jarak antara dua kelompok yang digabung. Semakin panjang garis, semakin besar jarak antara kedua kelompok tersebut. Berdasarkan dendrogram ini, dapat dilihat bahwa data terbagi menjadi dua kelompok utama: satu kelompok yang mencakup tahun 2020 dan 2018, serta kelompok lainnya yang terdiri dari tahun 2019 dan 2021.

Hasil Pembahasan dan Evaluasi Perbandingan Clustering

1. Hasil Clustering dengan K-Means

K-Means menggunakan jumlah cluster optimal berdasarkan metode Elbow dan Silhouette Score. Hasil analisis menunjukkan:

- **Metode Elbow**: Titik optimal ditemukan pada 3 cluster, tetapi untuk dataset ini, 2 cluster digunakan demi kesederhanaan dan interpretasi.
- **Scatter Plot Visualisasi**: Data terbagi dalam dua kelompok besar, mencerminkan pengelompokan data yang relatif seragam. Nilai centroid cluster mempermudah interpretasi data, tetapi hasil ini terbatas pada asumsi distribusi data berbentuk bulat.

Kelebihan:

- Cepat dalam komputasi.
- Efektif untuk dataset besar.

Kekurangan:

- Tidak efektif menangani data dengan struktur kompleks atau distribusi yang tidak seragam.

2. Hasil Clustering dengan Hierarchical Clustering

Metode hierarkis menggunakan pendekatan single linkage dan divisualisasikan melalui dendrogram.

- **Dendrogram**: Struktur hierarkis menunjukkan dua kelompok utama: satu dengan data tahun 2020 dan 2018, lainnya dengan 2019 dan 2021.

- Fleksibilitas metode ini membantu dalam mengungkap hubungan antar data secara lebih mendalam.

Kelebihan:

- Visualisasi dendrogram memudahkan analisis hubungan antar data.
- Tidak memerlukan jumlah cluster awal.

Kekurangan:

- Lambat untuk dataset besar.
- Rentan terhadap outlier.

3. Evaluasi Perbandingan

- **Akurasi Klasterisasi**: Hierarchical Clustering unggul dalam mengidentifikasi struktur kompleks data, sementara K-Means lebih efektif untuk data yang besar dengan cluster berbentuk bulat.
- **Kemudahan Interpretasi**: Hierarchical Clustering lebih baik karena dendrogram memberikan gambaran lengkap, tetapi K-Means tetap unggul dalam visualisasi cepat seperti scatter plot.
- **Efisiensi Komputasi**: K-Means lebih cepat dibandingkan Hierarchical Clustering, terutama untuk dataset besar.

Rekomendasi:

- Gunakan K-Means untuk aplikasi praktis yang memerlukan kecepatan dan pengelompokan data yang relatif sederhana.
- Pilih Hierarchical Clustering untuk analisis eksplorasi mendalam, terutama ketika data memiliki struktur yang rumit.

Dengan hasil ini, kedua metode memiliki keunggulan masing-masing tergantung pada konteks dan tujuan analisis.

KESIMPULAN

mengelompokkan data tingkat kemiskinan di Kabupaten Cianjur antara tahun 2018 hingga 2021. Berikut adalah kesimpulan utama:

1. **K-Means Clustering** lebih unggul dalam efisiensi komputasi dan cocok untuk dataset besar dengan cluster berbentuk bulat dan seragam. Metode ini menghasilkan visualisasi sederhana yang mempermudah interpretasi cepat, meskipun kurang efektif untuk data dengan struktur kompleks atau distribusi tidak merata.
2. **Hierarchical Clustering** lebih baik dalam mengidentifikasi struktur data yang kompleks dan memberikan hubungan hierarkis antar data melalui dendrogram. Namun, metode ini kurang efisien untuk dataset besar dan rentan terhadap outlier.
3. Berdasarkan hasil analisis, K-Means direkomendasikan untuk aplikasi praktis dengan fokus pada kecepatan dan efisiensi, sedangkan Hierarchical Clustering lebih sesuai untuk eksplorasi data yang mendalam dan mengungkap pola-pola tersembunyi dalam dataset.

Implementasi kedua metode memberikan wawasan yang berharga untuk mendukung pengambilan kebijakan sosial yang lebih tepat sasaran. Dengan memilih metode klasterisasi yang sesuai, pemerintah dan organisasi sosial dapat meningkatkan efektivitas program pengentasan kemiskinan.

DAFTAR PUSTAKA

Amaliyah, S., & Rianti Agustini, S. (n.d.). *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM) Penerapan Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan PKH Menggunakan Metode Clustering K-Means Pada Desa Kuala Dendang*. http://ejournal.unama.ac.id/index.php/jakakom

Carr-Hill, R. (2015). Non-Household Populations: Implications for Measurements of Poverty Globally and in the UK. *Journal of Social Policy*, *44*, 255–275. https://doi.org/10.1017/S0047279414000907

Edmiston, D. (2023). Who counts in poverty research? *The Sociological Review*. https://doi.org/10.1177/00380261231213233

FullBook Pengenalan Data Mining. (n.d.).

Klasen, S. (1997). Poverty, Inequality and Deprivation in South Africa: An Analysis of the 1993 SALDRU Survey. *Social Indicators Research*, *41*, 51–94. https://doi.org/10.1007/978-94-009-1479-7_3

Mongi, C., Langi, Y., Montolalu, C., & Nainggolan, N. (2019). Comparison of hierarchical clustering methods (case study: data on poverty influence in North Sulawesi). *IOP Conference Series: Materials Science and Engineering*, *567*. https://doi.org/10.1088/1757-899X/567/1/012048

Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Jurnal Informatika Terpadu PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA JUMLAH PENDUDUK MISKIN BERDASARKAN KOTA/KABUPATEN DI JAWA BARAT MENGGUNAKAN RAPIDMINER. *Jurnal Informatika Terpadu*, *9*(1), 68–73. https://journal.nurulfikri.ac.id/index.php/JIT

Santoso, B., Wijayanto, H., Notodiputro, K., & Sartono, B. (2018). A Comparative Study of Synthetic Over-sampling Method to Improve the Classification of Poor Households in Yogyakarta Province. *IOP Conference Series: Earth and Environmental Science*, *187*. https://doi.org/10.1088/1755-1315/187/1/012048

Sonang, S., Purba, A. T., & Pardede, F. O. I. (2019). PENGELOMPOKAN JUMLAH PENDUDUK BERDASARKAN KATEGORI USIA DENGAN METODE K-MEANS. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, *2*(2), 166. https://doi.org/10.37600/tekinkom.v2i2.115

Sudibyo, N. A., Iswardani, A., Sari, K., Suprihatiningsih, S., Duta, U., & Surakarta, B. (n.d.). *PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI INDONESIA*. *1*(3), 2020. https://doi.org/10.46306/lb.v1i3

Tarigan, N. M. Br., Tarigan, S. E. Br., & Simatupang, A. (2023). Implementation of Data Mining in Grouping Data of the Poor Using the K-Means Method. *Journal of Computer Networks, Architecture and High Performance Computing*. https://doi.org/10.47709/cnahpc.v5i2.2625