

Pengelompokan Tingkat Kelayakan Penerima Bantuan Sosial Menggunakan Algoritma K-Means Clustering

Yusa Virginiawan¹⁾; Dede Fico²⁾

^{1,2)} Program Studi Teknik Informatika Fakultas Teknik, Universitas Muhamadiyah Bengkulu

Email: ¹⁾ yusavirginiawang@gmail.com

ARTICLE HISTORY

Received [07 Juli 2026]

Revised [19 Oktober 2026]

Accepted [21 Oktober 2026]

KEYWORDS

Social Assistance, Clustering, Data Mining, K-Means, Rapidminer.

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Program bantuan sosial (bansos) merupakan upaya pemerintah untuk meningkatkan kesejahteraan masyarakat berpenghasilan rendah. Namun, penentuan calon penerima bantuan masih sering mengalami kendala, terutama terkait akurasi dan ketepatan sasaran akibat minimnya proses analisis data yang sistematis. Oleh karena itu, pemanfaatan teknologi data mining menjadi penting untuk mendukung pengambilan keputusan yang lebih objektif dan terukur. Penelitian ini menggunakan algoritma K-Means Clustering untuk mengelompokkan tingkat penghasilan masyarakat guna mengidentifikasi pola penerima bansos berdasarkan kedekatan nilai atribut penghasilan. Data penelitian berjumlah 1.666 entri penerima bansos di Kota Malang dengan atribut nama, alamat, dan penghasilan. Proses analisis dilakukan menggunakan aplikasi RapidMiner melalui tahapan data preprocessing, meliputi pemilihan atribut dan normalisasi data menggunakan z-transformation agar seluruh variabel berada pada skala seragam. Selanjutnya, algoritma K-Means diterapkan untuk membentuk tiga cluster utama. Hasil penelitian menunjukkan bahwa cluster 0 merepresentasikan kelompok berpenghasilan rendah, cluster 1 kelompok berpenghasilan menengah, dan cluster 2 kelompok berpenghasilan tinggi. Temuan ini menunjukkan bahwa algoritma K-Means Clustering efektif dalam memetakan kondisi ekonomi masyarakat serta mendukung penyaluran bantuan sosial agar lebih tepat sasaran dan berbasis analisis data terstruktur. Penelitian ini diharapkan dapat menjadi referensi bagi instansi pemerintah dalam meningkatkan transparansi, akurasi, dan efektivitas kebijakan distribusi bantuan sosial berbasis data yang lebih adil dan berkelanjutan nasional.

ABSTRACT

Social assistance programmes (bansos) are government initiatives aimed at improving the welfare of low-income communities. However, the identification of potential beneficiaries often faces challenges, particularly regarding accuracy and targeting, due to a lack of systematic data analysis. Therefore, the use of data mining technology is essential to support more objective and measurable decision-making. This study utilises the K-Means Clustering algorithm to group the population by income level in order to identify patterns among social assistance recipients based on the proximity of their income attribute values. The research data comprised 1,666 entries of social assistance recipients in Malang City, with attributes including name, address and income. The analysis was conducted using the RapidMiner application through data pre-processing stages, including attribute selection and data normalisation using z-transformation to ensure all variables were on a uniform scale. Subsequently, the K-Means algorithm was applied to form three main clusters. The research findings indicate that cluster 0 represents the low-income group, cluster 1 the middle-income group, and cluster 2 the high-income group. These findings demonstrate that the K-Means Clustering algorithm is effective in mapping the economic conditions of the population and supports the distribution of social assistance in a manner that is more targeted and based on structured data analysis. It is hoped that this research will serve as a reference for government agencies in improving the transparency, accuracy and effectiveness of data-driven social assistance distribution policies that are fairer and more sustainable at the national level.

PENDAHULUAN

Kemiskinan masih menjadi permasalahan sosial yang kompleks dan berkelanjutan di berbagai wilayah, khususnya di negara berkembang. Kondisi ini ditandai dengan keterbatasan kemampuan masyarakat dalam memenuhi kebutuhan dasar hidup, seperti pangan, kesehatan, pendidikan, dan perumahan yang layak. Ketimpangan ekonomi, keterbatasan lapangan pekerjaan, serta rendahnya pendapatan masyarakat menjadi faktor utama yang memperparah permasalahan kemiskinan. Jika tidak ditangani secara tepat, kemiskinan dapat menghambat proses pembangunan nasional dan menurunkan kualitas sumber daya manusia secara menyeluruh.

Sebagai upaya penanggulangan kemiskinan, pemerintah menyelenggarakan berbagai program bantuan sosial (bansos) yang bertujuan untuk meningkatkan kesejahteraan masyarakat berpenghasilan rendah. Namun, dalam implementasinya, penyaluran bantuan sosial masih sering menghadapi permasalahan ketidaktepatan sasaran. Proses penentuan penerima bantuan yang belum sepenuhnya berbasis analisis data yang sistematis menyebabkan terjadinya kesalahan klasifikasi penerima, distribusi bantuan yang tidak merata, serta potensi duplikasi data. Kondisi ini berdampak pada menurunnya

efektivitas program bantuan sosial yang seharusnya menjadi instrumen perlindungan sosial bagi masyarakat miskin.

Perkembangan teknologi informasi membuka peluang untuk mengatasi permasalahan tersebut melalui pemanfaatan teknik data mining. Data mining memungkinkan pengolahan data dalam jumlah besar untuk menemukan pola dan informasi tersembunyi yang dapat digunakan sebagai dasar pengambilan keputusan. Salah satu teknik data mining yang relevan untuk permasalahan penentuan kelayakan penerima bantuan sosial adalah clustering. Teknik clustering mampu mengelompokkan data ke dalam beberapa kelompok berdasarkan tingkat kemiripan karakteristik tertentu tanpa memerlukan label kelas sebelumnya, sehingga menghasilkan pengelompokan yang lebih objektif dan terukur.

Algoritma K-Means merupakan salah satu metode clustering yang paling banyak digunakan karena memiliki konsep yang sederhana, efisien dalam komputasi, serta mudah diimplementasikan. Beberapa penelitian terdahulu menunjukkan bahwa algoritma K-Means efektif dalam pengelompokan data bantuan sosial. Rainaya dan Dewi (2025) menyatakan bahwa K-Means Clustering mampu mempercepat proses penentuan prioritas penerima bantuan sosial di tingkat desa. Penelitian lain oleh Saputra dan Thamrin (2024) menunjukkan bahwa penerapan K-Means dapat meminimalkan kesalahan dalam penentuan intervensi prioritas penerima bansos. Selain itu, Hidayatullah dan Prihartono (2025) membuktikan bahwa algoritma K-Means dapat meningkatkan efektivitas program sosial di tingkat kota dan kabupaten melalui pengelompokan data ekonomi masyarakat yang lebih sistematis.

Beberapa penelitian sebelumnya seperti Penelitian oleh [4] menggunakan algoritma K-Means untuk mengelompokkan penerima bantuan sosial di Desa Gunung, yang menunjukkan bahwa metode ini efektif dalam membagi data penerima berdasarkan karakteristiknya sehingga dapat membantu meningkatkan ketepatan sasaran bantuan sosial. Selanjutnya penelitian oleh [1] menunjukkan bahwa algoritma K-Means Clustering mampu mengelompokkan data penerima bantuan sosial ke dalam beberapa cluster yang mencerminkan tingkat kebutuhan masyarakat.

Pengelompokan ini dinilai dapat membantu meningkatkan ketepatan sasaran penyaluran bantuan sosial serta mengurangi potensi kesalahan akibat penilaian secara manual. Dengan demikian, metode K-Means terbukti efektif sebagai alat bantu dalam proses pengambilan keputusan penyaluran dana bantuan sosial di tingkat desa. Berdasarkan penelitian terdahulu maka penelitian ini bertujuan untuk mengelompokkan tingkat kelayakan penerima bantuan sosial menggunakan algoritma K-Means Clustering berdasarkan atribut penghasilan masyarakat. Proses analisis dilakukan melalui tahapan prapemrosesan data, normalisasi, dan pemodelan menggunakan aplikasi RapidMiner. Hasil penelitian ini diharapkan dapat memberikan gambaran kondisi ekonomi masyarakat secara objektif serta menjadi dasar pendukung bagi instansi pemerintah dalam meningkatkan akurasi, transparansi, dan efektivitas penyaluran bantuan sosial agar lebih tepat sasaran.

TINJAUAN PUSTAKA

Data Mining

Kepuasan Data mining adalah proses mengekstraksi informasi berharga dari kumpulan data yang besar. Fungsi data mining tidak hanya menemukan pola tersembunyi, namun juga membantu pengambilan keputusan berdasarkan fakta. Data mining mencakup berbagai teknik seperti klasifikasi, clustering, asosiasi, prediksi, dan deteksi anomali. Data mining merupakan salah satu cabang ilmu komputer yang cukup baru yang banyak digunakan dan dipelajari oleh para ahli ilmu komputer dan programmer [10]. Data Mining atau penambangan data merupakan proses pencarian pola yang menarik dan tersembunyi dari Kumpulan data yang berukuran besar. [11]

Dalam konteks sosial, data mining banyak digunakan untuk menganalisis data kemiskinan, kesehatan masyarakat, pendidikan, serta bantuan sosial. Teknik ini dapat menghasilkan model yang lebih objektif dibandingkan seleksi manual karena berbasis data numerik yang terukur.

Clustering

Clustering pada suatu data adalah suatu tahapan untuk menggolongkan himpunan data yang atribut kelasnya belum dideskripsikan, secara konsep clustering adalah untuk memaksimalkan dan meminimalkan kemiripan intra antar kelas. sebagai contoh, ada suatu himpunan obyek, proses pertama dapat di klasterisasi menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompokklasifikasi tertentu. Cluster juga dapat diartikan sebagai kelompok. Maka analisa clustering pada dasarnya akan menghasilkan sejumlah cluster (kelompok). [12]

Algoritma K-Means

K-Means adalah algoritma clustering yang bekerja berdasarkan jarak antar data. Algoritma ini membagi data menjadi K kelompok yang masing-masing memiliki centroid (titik pusat). kemudian menghitung jarak setiap data terhadap centroid terdekat menggunakan ukuran jarak, misalnya Euclidean Distance[13]. Dengan tujuan meminimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain.

Tahapan K-Means:

1. Menentukan jumlah cluster (K).
2. Menginisialisasi centroid awal secara acak.
3. Menghitung jarak setiap data ke centroid menggunakan jarak Euclidean.
4. Mengelompokkan data ke centroid terdekat.
5. Memperbarui centroid berdasarkan rata-rata masing-masing cluster.
6. Mengulangi proses sampai centroid stabil.
7. Dihitung dengan rumus persamaan (1) jarak Euclidean:

$$d(x, c) = \sqrt{d(x - c)^2} = |x - c| \quad (1)$$

“Jarak antara x dan c sama dengan akar dari (x dikurangi c) kuadrat, atau sama dengan nilai absolut dari (x dikurangi c).”

Keterangan :

1. $d(x, c)$ → jarak antara data (x) dan centroid (c)
2. x → nilai data
3. c → nilai centroid
4. $(x - c)$ → selisih data dan centroid
5. $(x - c)^2$ → selisih dikuadratkan agar tidak ada nilai negative
6. $\sqrt{(\dots)}$ → mengambil akar untuk mendapatkan jarak Euclidean
7. $|x - c|$ → nilai absolut (sama hasilnya untuk 1 variabel)

Normalisasi Data

Normalisasi digunakan agar semua atribut berada dalam rentang nilai yang sama, sehingga tidak ada satu variabel pun mendominasi perhitungan jarak. Pada penelitian ini digunakan metode *range transformation*. Dihitung dengan rumus persamaan (2) normalisasi :

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2)$$

“x aksen sama dengan x dikurangi nilai minimum, dibagi nilai maksimum dikurangi nilai minimum.”

Keterangan:

1. x' (x prime) → nilai baru setelah dinormalisasi.
2. x → nilai asli.
3. $\min(x)$ → nilai terkecil dari kolom data.
4. $\max(x)$ → nilai terbesar dari kolom data.

Rapid Miner

RapidMiner merupakan suatu platform perangkat lunak yang sangat kuat untuk pembelajaran mesin dan ilmu data, menyediakan berbagai alat untuk pemodelan, evaluasi, implementasi, dan persiapan data. Meskipun dirancang untuk mudah digunakan, RapidMiner memungkinkan individu tanpa pengalaman pemrograman untuk dengan cepat membangun dan menguji berbagai model. Dengan antarmuka drag-and-drop, pengguna dapat membuat alur kerja untuk memproses dan menganalisis data. Keunggulan perangkat lunak ini terletak pada dukungan terhadap berbagai sumber data, seperti file datar, basis data, dan platform big data seperti Hadoop dan Spark. RapidMiner juga menyertakan berbagai operator bawaan yang mencakup semua langkah dalam proses pemrosesan data, seperti pemodelan, pembersihan data dan pemilihan fitur[15].

METODE PENELITIAN

Jenis Penelitian

Penelitian ini merupakan penelitian deskriptif kuantitatif menggunakan metode data mining dengan algoritma K-Means Clustering untuk mengelompokkan masyarakat berdasarkan tingkat penghasilan.

Sumber dan Bentuk Data

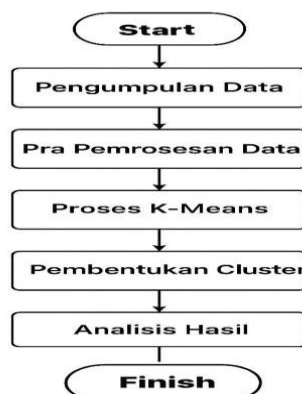
Data yang digunakan dalam penelitian ini merupakan data sekunder penerima bantuan sosial yang diperoleh dari sumber terbuka dan dimanfaatkan untuk kepentingan akademik. Data tersebut bersumber dari instansi pemerintah terkait dengan jumlah keseluruhan sebanyak 1.666 entri. Dataset disajikan dalam format Microsoft Excel (.xlsx) dan memuat beberapa atribut utama, yaitu nama, alamat, dan penghasilan. Atribut-atribut tersebut dipilih karena mampu merepresentasikan informasi dasar penerima bantuan sosial serta relevan dalam mendukung proses analisis pengelompokan data. Meskipun jumlah atribut yang digunakan relatif terbatas, variabel penghasilan dipilih sebagai indikator utama karena secara kuantitatif mencerminkan kondisi ekonomi rumah tangga dan menjadi dasar dalam proses pengelompokan menggunakan metode K-Means Clustering.

Dataset penerima bantuan sosial ditampilkan pada Tabel 1. Data yang disajikan merupakan sebagian dari keseluruhan dataset dan digunakan untuk menunjukkan struktur data yang dianalisis. Sementara itu, seluruh dataset sebanyak 1.666 entri digunakan secara penuh dalam proses analisis menggunakan aplikasi RapidMiner.

Tabel 1. Dataset

No	Nama	Alamat	Kelurahan	Kecamatan	Penghasilan (Rp)
1	Abdul Rosyid	Teluk Cendrawasih 78 RT 07 RW 02	Arjosari	Blimbing	2.867.722
2	Asmuah	Teluk Cendrawasih 33 RT 04 RW 01	Arjosari	blimbing	3.824.430
3	Doipah	Teluk Cendrawasih RT 01 RW 01	Arjosari	Blimbing	2.530.978
4	Ferdi Ari Tomsius	Teluk Cendrawasih RT 02 RW 02	Arjosari	Blimbing	3.126.453
...	...	...	...	...	...
1665	Zainuddin Chamim	JL Ir Rais GG 9 NO 167 RT 02 RW 04	Tanjungrejo	Sukun	4.292.657
1666	Zainul Arifin	JL Mergan Keramat RT 03 RW 11	Tanjungrejo	Sukun	4.764.600

Tahapan Penelitian



Gambar 1. Tahapan penelitian

Tahap Pengumpulan Data

Tahap pengumpulan data pada penelitian ini dilakukan dengan mengakses data sekunder penerima bansos yang tersedia dalam bentuk file Microsoft Excel (.xlsx). Data tersebut berisi daftar penerima bantuan sosial tahap I dan digunakan sebagai bahan analisis dalam penelitian ini. Setelah data diperoleh, dilakukan pemeriksaan awal untuk memastikan kualitas data, meliputi pengecekan adanya data duplikat, memastikan tidak terdapat nilai kosong pada atribut penghasilan, serta memverifikasi bahwa nilai penghasilan tersimpan dalam format numerik. Tahap pemeriksaan ini penting dilakukan agar data yang digunakan pada proses analisis selanjutnya bersifat valid, konsisten, dan siap diolah.

Tahap Pra-pemrosesan

Pra-pemrosesan sangat penting dalam data mining. Tahapan meliputi:

1. Cleaning (Pembersihan Data)
Pada tahap pembersihan data, dilakukan verifikasi untuk memastikan setiap baris memiliki nilai penghasilan yang valid sehingga dapat digunakan dalam proses analisis. Selain itu, data yang dianggap tidak relevan atau tidak berkaitan dengan tujuan penelitian dihapus agar hasil pengolahan menjadi lebih akurat dan fokus.
2. Select Attributes
Hanya atribut "Penghasilan" yang digunakan untuk clustering.
3. Normalisasi
Normalisasi range 0–1 menggunakan operator Normalize pada RapidMiner. Proses normalisasi data penghasilan dilakukan untuk menyamakan skala nilai sebelum proses clustering, dan hasil normalisasi tersebut disajikan pada Tabel 2.

Tabel 2. Nilai Normalisasi

No	Penghasilan	Nilai Normalisasi
1	2.867.722	0,53
2	3.824.430	0,74
3	2.530.978	0,43
4	3.126.453	0,58
...	...	...
1665	4.292.657	0,84
1666	4.764.600	0,95

Pemodelan Menggunakan K-Means

Pada tahap pemodelan, proses pengelompokan data dilakukan menggunakan algoritma K-Means Clustering. Jumlah cluster ditetapkan sebanyak tiga untuk mengelompokkan data ke dalam kategori penghasilan rendah, menengah, dan lebih tinggi. Proses pemodelan dilakukan menggunakan aplikasi RapidMiner dengan pengukuran jarak Euclidean Distance. Parameter yang digunakan dalam pemodelan K-Means pada penelitian ini disajikan pada Table 3.

Tabel 3. Parameter Algoritma K-Means

Parameter	Nilai
Jumlah cluster (k)	3
Atribut	Penghasilan
Normalisasi	Min–Max (0–1)
Distance Measure	Euclidean Distance
Tools	RapidMiner

HASIL DAN PEMBAHASAN

Hasil Pengelompokan

Setelah menjalankan proses clustering pada RapidMiner, diperoleh tiga cluster berdasarkan nilai penghasilan. Adapun pembagian data ke dalam masing-masing cluster dapat dilihat pada Tabel 4.

Table 4. Hasil pengelompokkan

Cluster	Jumlah Data	Interpretasi
Cluster0	538 data	Penghasilan rendah (layak prioritas bantuan)
Cluster1	562 data	Penghasilan menengah (layak bersyarat)
Cluster2	566 data	Penghasilan lebih tinggi (bukan prioritas)

Pada Tabel 4. Perbandingan jumlah anggota antar cluster cukup seimbang sehingga menunjukkan distribusi penghasilan masyarakat Kota Malang relatif merata dalam tiga kategori tersebut.

Analisis Centroid

Centroid pada setiap cluster merepresentasikan nilai rata-rata penghasilan dari kelompok tersebut setelah melalui proses normalisasi. Semakin kecil nilai centroid, semakin rendah tingkat penghasilan kelompok, sehingga kelompok tersebut dapat dikategorikan sebagai penerima bantuan yang paling membutuhkan. Sementara itu, centroid dengan nilai menengah menggambarkan kelompok berpenghasilan sedang, dan centroid tertinggi menunjukkan kelompok dengan penghasilan relatif lebih tinggi yang umumnya tidak menjadi prioritas dalam penyaluran bantuan sosial. Interpretasi centroid ini penting karena menjadi dasar dalam memahami karakteristik tiap cluster dan dalam menentukan tingkat kelayakan penerima bantuan secara lebih objektif. Nilai centroid akhir dari masing-masing cluster ditampilkan pada Tabel 5.

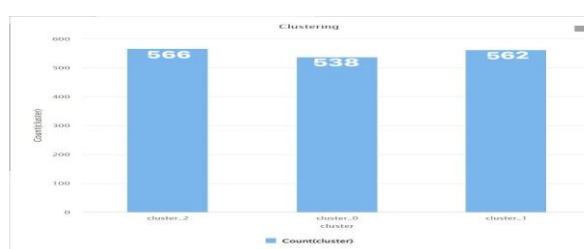
Tabel 5. Nilai Centroid

Cluster	Nilai Centroid (Penghasilan Ternormalisasi)
Cluster 0	0.849
Cluster 1	0.172
Cluster 2	0.512

Berdasarkan Tabel 5, nilai centroid tertinggi terdapat pada Cluster 0 sebesar 0,849 yang menunjukkan kelompok dengan penghasilan relatif paling tinggi sehingga tidak menjadi prioritas bantuan. Cluster 2 memiliki nilai centroid 0,512 yang merepresentasikan kelompok berpenghasilan menengah. Sementara itu, Cluster 1 memiliki nilai centroid terendah sebesar 0,172 sehingga merupakan kelompok dengan penghasilan paling rendah dan menjadi prioritas utama penerima bantuan sosial.

Visualisasi pada RapidMiner

Hasil pengelompokan data ke dalam tiga cluster divisualisasikan menggunakan aplikasi RapidMiner sebagaimana ditunjukkan pada Gambar 2.

**Gambar 2. Diagram visualisasi**

Visualisasi hasil clustering pada RapidMiner menampilkan jumlah anggota pada masing-masing cluster yang terbentuk melalui algoritma K-Means. Grafik batang pada tampilan menunjukkan distribusi data ke dalam tiga cluster dengan pembagian yang relatif seimbang. Cluster 0 berisi 538 data, cluster 1 berisi 562 data, dan cluster 2 berisi 566 data dari total 1.666 entri. Tinggi masing-masing batang menggambarkan berapa banyak data yang terkelompok dalam setiap cluster, sehingga memudahkan analisis terhadap pola distribusi penghasilan masyarakat. Dari visualisasi tersebut terlihat bahwa ketiga cluster memiliki jumlah anggota yang hampir sama, menunjukkan bahwa algoritma K-Means melakukan pembagian data secara cukup merata tanpa adanya cluster yang mendominasi ataupun yang berisi sangat sedikit data. Distribusi yang seimbang ini menunjukkan bahwa variasi penghasilan masyarakat dalam dataset cenderung terbagi ke dalam tiga kelompok yang memiliki kemiripan jumlah anggota, sehingga mempermudah proses interpretasi dan penentuan kategori tingkat kelayakan bantuan sosial.

Pembahasan

Hasil analisis menunjukkan bahwa algoritma K-Means bekerja secara efektif dalam mengelompokkan data penghasilan masyarakat. Dengan hanya menggunakan satu atribut, yaitu penghasilan, proses clustering dapat berjalan cepat dan efisien karena data telah dinormalisasi sehingga perhitungan jarak menjadi lebih optimal. Hasil pengelompokan menghasilkan tiga cluster utama dengan distribusi data yang relatif seimbang. Cluster 0 terdiri dari 538 data dan merepresentasikan kelompok masyarakat dengan penghasilan terendah sehingga menjadi pihak yang paling membutuhkan bantuan sosial. Cluster 1 berjumlah 562 data yang menggambarkan kelompok masyarakat berpenghasilan menengah dan masih dapat dipertimbangkan sebagai penerima bantuan sesuai kebijakan yang berlaku. Sementara itu, Cluster 2 terdiri dari 566 data yang merepresentasikan kelompok dengan penghasilan lebih tinggi sehingga tidak menjadi prioritas penerima bantuan sosial. Distribusi jumlah anggota cluster yang hampir seimbang menunjukkan bahwa data penghasilan masyarakat terbagi secara merata ke dalam tiga kategori. Meskipun demikian, penelitian ini memiliki keterbatasan karena hanya menggunakan satu atribut dalam proses analisis. Kondisi ekonomi masyarakat tidak sepenuhnya dapat direpresentasikan hanya berdasarkan penghasilan, sehingga penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti jumlah tanggungan, jenis pekerjaan, atau kondisi tempat tinggal agar hasil pengelompokan menjadi lebih akurat dan komprehensif.

KESIMPULAN

Kesimpulan

Penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan data masyarakat berdasarkan tingkat penghasilan secara efektif. Jumlah cluster ditetapkan sebanyak tiga cluster, yang disesuaikan dengan tujuan penelitian untuk mengelompokkan masyarakat ke dalam kategori penghasilan rendah, menengah, dan tinggi. Dengan menggunakan satu atribut utama, yaitu penghasilan yang telah dinormalisasi, proses clustering berjalan optimal dan menghasilkan pembagian kelompok yang jelas.

Cluster 0 merepresentasikan kelompok masyarakat dengan penghasilan rendah yang menjadi prioritas utama penerima bantuan sosial. Cluster 1 terdiri dari masyarakat berpenghasilan menengah yang masih layak menerima bantuan secara bersyarat, sedangkan Cluster 2 merupakan kelompok masyarakat dengan penghasilan lebih tinggi sehingga tidak termasuk prioritas penerima bantuan sosial.

Distribusi jumlah anggota pada setiap cluster yang relatif seimbang menunjukkan bahwa algoritma K-Means mampu memetakan kondisi ekonomi masyarakat secara objektif. Selain itu, penggunaan aplikasi RapidMiner mempermudah proses analisis melalui antarmuka yang intuitif, proses otomatis, serta visualisasi hasil yang jelas. Secara keseluruhan, algoritma K-Means terbukti dapat digunakan sebagai alat bantu yang cepat, akurat, dan efisien dalam mendukung penentuan kelayakan penerima bantuan sosial.

Saran

Penelitian selanjutnya disarankan untuk menambahkan lebih banyak atribut seperti jumlah tanggungan, kondisi rumah, aset, dan status pekerjaan agar hasil pengelompokan mencerminkan kondisi ekonomi masyarakat secara lebih menyeluruh. Penggunaan beberapa algoritma pembandingan seperti *K-Medoids*, *DBSCAN*, atau *Hierarchical Clustering* juga dapat dilakukan untuk melihat metode mana yang memberikan hasil paling akurat.

Selain itu, kualitas cluster dapat dievaluasi dengan metode validasi seperti *Silhouette Coefficient* atau *Davies-Bouldin Index*. Pengembangan sistem berbasis dashboard juga dapat membantu instansi pemerintah dalam menggunakan hasil clustering sebagai alat bantu pengambilan keputusan secara real-time. Pada penelitian mendatang, penggunaan dataset dari wilayah yang lebih luas akan memberikan model yang lebih general dan dapat diterapkan di berbagai daerah.

DAFTAR PUSTAKA

- S. Rainaya and S. Dewi, "Penerapan Data Mining Penyaluran Dana Bantuan Sosial Menggunakan Metode KMeans Clustering di Desa Rancaekek Kulon," vol. 4, no. 3, pp. 433–444, 2025.
DOI: <https://doi.org/10.32627/dimamu.v4i3.1618>
- R. Mcareavey and D. L. Brown, "inequality in the UK and the US," *Palgrave Commun.*, pp. 1–10, 2019, doi: 10.1057/s41599-019-0332-8.
- E. Mansi, E. Hysa, and M. Panait, "Poverty — A Challenge for Economic Development ? Evidences from Western Balkan Countries and the European Union," pp. 1–24, 2020.
<https://doi.org/10.3390/su12187754>

- T. P. Health, P. Families, and T. P. Health, "and future health status of children," vol. 12, no. 8, pp. 667–673, 2007. doi: [10.1093/pch/12.8.667](https://doi.org/10.1093/pch/12.8.667)
- F. V Shahidi, C. Ramraj, O. Sod-erdene, V. Hildebrand, and A. Siddiqi, "The impact of social assistance programs on population health : a systematic review of research in high-income countries," pp. 1–11, 2019. <https://doi.org/10.1186/s12889-018-6337-1>
- A. S. Ahmar, D. Napitupulu, and R. Rahim, "Using K-Means Clustering to Cluster Provinces in Indonesia Using K-Means Clustering to Cluster Provinces in Indonesia," 2018. DOI:[10.1088/1742-6596/1028/1/012006](https://doi.org/10.1088/1742-6596/1028/1/012006)
- N. A. Hidayatullah and W. Prihartono, "CLUSTERING ALGORITMA K-MEANS UNTUK MENINGKATKAN EFEKTIVITAS PROGRAM SOSIAL DI KOTA / KABUPATEN CIREBON," vol. 13, no. 1, pp. 629–636, 2025. <http://dx.doi.org/10.23960/jitet.v13i1.569>
- M. Sammour and Z. Othman, "An Agglomerative Hierarchical Clustering with Various Distance Measurements for Ground Level Ozone Clustering in Putrajaya , Malaysia," vol. 6, no. 6, pp. 1127–1133, 2016. DOI: <https://doi.org/10.18517/ijaseit.6.6.1482>
- A. K. Saputra and T. Thamrin, "Implementasi K-Means pada Klasterisasi Data Intervensi Prioritas Penerima Bantuan Sosial," pp. 2–7. DOI: <http://dx.doi.org/10.36448/expert.v14i1.3717>
- R. Sakit and S. Binjai, "Indonesian Journal of Education And Computer Science," vol. 1, no. 3, pp. 82–88, 2023. DOI: <https://doi.org/10.60076/indotech.v1i2.43>
- R. A. Manik and A. Ariesta, "Data Mining Untuk Klasterisasi Provinsi Di Indonesia Berdasarkan Data Pokok Pendidikan Nasional," vol. 11, pp. 159–164, 2023. DOI: <https://doi.org/10.70309/ticom.v11i3.115>
- R. Muliono and Z. Sembiring, "DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS UNTUK KLASERISASI TINGKAT," vol. 4, no. 2, pp. 272–279, 2019. DOI: <https://doi.org/10.24114/cess.v4i2.13620>
- M. Miftahul, U. Bengkak, K. Clustering, B. Indonesia, B. Inggris, and K. Clustering, "secara objektif dan efisien. Latar belakang penelitian ini adalah masih digunakannya metode manual dalam penentuan siswa berprestasi yang cenderung subjektif dan memakan waktu lama.," vol. 4, 2025. DOI: <https://doi.org/10.35473/jamastika.v4i2.4527>
- D. Apriandi, R. M. Sari, and M. I. Sarif, "Analisis Clustering Untuk Menentukan Siswa Berprestasi di SMK Swasta TI Panca Dharma Stungkit Menggunakan Metode K-Means," vol. 13, pp. 1117–1129, 2024. DOI:<https://doi.org/10.33395/jmp.v13i1.13959e-ISSN>
- D. F. Astri, M. Informatika, J. Barat, B. Langsung, A. Kecil, and A. S. Dasar, "CLUSTERING PENDUDUK MISKIN MENGGUNAKAN ALGORITMA K-MEANS PADA WILAYAH JAWA BARAT," vol. 8, no. 2, pp. 1548–1554, 2024. DOI:[10.36040/jati.v8i2.9012](https://doi.org/10.36040/jati.v8i2.9012)