

PENERAPAN K-MEANS CLUSTERING DALAM PENGELOMPOKAN CALON PENERIMA BANTUAN SOSIAL BERBASIS DATA SOSIAL-EKONOMI DI KECAMATAN CIDAHU

Muhammad Ridhwan¹, Rudi Kurniawan², Bani Nurhakim³, Riri Narasati⁴.

Program Studi Teknik Informatika^{1,2,4}
Program Studi Manajemen Informatika³

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
Email : mr.friswall@gmail.com

(*) Corresponding Author : mr.friswall@gmail.com

Published : 30 Mei 2026

Abstract—Social assistance programs play a crucial role in poverty alleviation, but their effectiveness is often hampered by inaccurate recipient targeting. In Cidahu District, the manual selection process has led to subjectivity, lack of transparency, and dissatisfaction among residents. This study aims to develop a data mining model using the K-Means clustering method to classify households based on socioeconomic indicators for more objective social assistance recipient selection. Data were collected from administrative records containing variables such as monthly income, number of dependents, housing conditions, land area, and education level. After preprocessing and standardization, the optimal number of clusters ($K=3$) was determined using the Elbow Method and Silhouette Score. The K-Means algorithm was then applied to group households into three clusters: "Highly Eligible," "Eligible," and "Less Eligible." Evaluation using Silhouette Score (0.52) and Davies-Bouldin Index (1.21) indicated reasonably good cluster separation. The model successfully identified 65 households as priority recipients, demonstrating that K-Means can provide a data-driven foundation for fairer and more transparent social assistance distribution. This study contributes methodologically to the application of clustering in social policy contexts and offers practical recommendations for local governments in evidence-based decision-making.

Keywords: Data Mining, K-Means Clustering, Social Assistance, Poverty Alleviation, Decision Support System

Abstrak- Program bantuan sosial memiliki peran penting dalam pengentasan kemiskinan, namun efektivitasnya sering terhambat oleh ketidaktepatan sasaran penerima. Di Kecamatan Cidahu, proses seleksi manual telah mengakibatkan subjektivitas, kurangnya transparansi, dan ketidakpuasan masyarakat. Penelitian ini bertujuan mengembangkan model data mining menggunakan metode clustering K-Means untuk mengelompokkan rumah tangga berdasarkan indikator sosial-ekonomi guna seleksi penerima bantuan sosial yang lebih objektif. Data dikumpulkan dari rekaman administratif yang memuat variabel seperti pendapatan bulanan, jumlah tanggungan, kondisi rumah, luas tanah, dan tingkat pendidikan. Setelah preprocessing dan standarisasi, jumlah kluster optimal ($K=3$) ditentukan menggunakan Metode Elbow dan Silhouette Score. Algoritma K-Means kemudian diterapkan untuk mengelompokkan rumah tangga menjadi tiga kluster: "Sangat Layak," "Layak," dan "Kurang Layak." Evaluasi menggunakan Silhouette Score (0,52) dan Davies-Bouldin Index (1,21) menunjukkan pemisahan kluster yang cukup baik. Model berhasil mengidentifikasi 65 rumah tangga sebagai penerima prioritas, membuktikan bahwa K-Means dapat memberikan landasan berbasis data untuk distribusi bantuan sosial yang lebih adil dan transparan. Penelitian ini berkontribusi secara metodologis dalam penerapan clustering pada konteks kebijakan sosial serta memberikan rekomendasi praktis bagi pemerintah daerah dalam pengambilan keputusan berbasis bukti.

Kata Kunci: Data Mining, Clustering K-Means, Bantuan Sosial, Pengentasan Kemiskinan, Sistem Pendukung Keputusan.

INTRODUCTION

Bantuan sosial (bansos) merupakan salah satu instrumen fundamental dalam kerangka perlindungan sosial Indonesia, yang berfungsi mendukung masyarakat miskin, rentan miskin, serta kelompok yang mengalami tekanan sosial-ekonomi akibat kondisi struktural maupun situasional. Sebagai bagian dari strategi penanggulangan kemiskinan nasional, bansos memegang peranan penting dalam mengurangi kerentanan rumah tangga terhadap guncangan ekonomi, memastikan stabilitas konsumsi, serta mendukung pemenuhan kebutuhan dasar keluarga berpenghasilan rendah. Kebijakan ini juga terkait langsung dengan target pembangunan berkelanjutan (SDGs), khususnya tujuan pengentasan kemiskinan, pengurangan ketimpangan, dan peningkatan kesejahteraan sosial.

Namun efektivitas implementasi program bansos sangat ditentukan oleh ketepatan sasaran penerima. Ketidaktepatan sasaran dapat menurunkan kredibilitas pemerintah, menimbulkan ketidakadilan sosial, serta melemahkan tujuan perlindungan sosial yang diharapkan. Berbagai kajian menunjukkan bahwa ketidaktepatan sasaran penyaluran bansos di Indonesia masih cukup tinggi, dipengaruhi oleh permasalahan data yang tidak mutakhir, mekanisme verifikasi yang lemah, serta dominasi pendekatan manual dalam proses seleksi penerima. Selain itu, faktor-faktor sosial di tingkat lokal, seperti kedekatan personal antara warga dan aparat, dapat berpengaruh terhadap keputusan distribusi bantuan, sehingga menghasilkan bias dalam proses seleksi.

Kondisi ini juga tercermin di Kecamatan Cidahu, sebuah wilayah administratif yang memiliki karakteristik sosial-ekonomi yang beragam. Di Cidahu, penentuan calon penerima bansos masih mengandalkan prosedur manual, seperti musyawarah desa, rekomendasi RT/RW, dan penilaian petugas lapangan yang didasarkan pada observasi subjektif. Meskipun pendekatan tersebut memiliki nilai positif dari sisi partisipasi masyarakat, namun terdapat sejumlah kelemahan seperti inkonsistensi penilaian antarpetugas, potensi konflik kepentingan, serta kendala dalam mendokumentasikan kondisi sosial-ekonomi secara objektif dan terstandar.

Di samping itu, sistem pendataan di tingkat kecamatan sering kali bergantung pada Data

Terpadu Kesejahteraan Sosial (DTKS) yang pembaruannya tidak selalu cepat, sehingga terdapat kemungkinan bahwa kondisi rumah tangga telah berubah namun belum tercatat. Beberapa keluarga yang sebenarnya layak mungkin tidak terdata, sementara ada pula keluarga yang sudah tidak lagi masuk kategori miskin tetapi masih menerima bansos. Ketidaksesuaian data seperti ini berpotensi menimbulkan ketidakpuasan masyarakat dan menurunkan efektivitas kebijakan. Oleh sebab itu, dibutuhkan mekanisme yang lebih objektif, adaptif, dan memanfaatkan data aktual agar penyaluran bansos dapat dilakukan dengan lebih akurat.

Seiring berkembangnya teknologi analitik dan komputasi, penggunaan data mining menjadi semakin relevan untuk meningkatkan akurasi proses seleksi penerima bansos. Data mining adalah serangkaian metode cerdas yang digunakan untuk mengeksplorasi pola, relasi, dan struktur tersembunyi dalam data berukuran besar dan kompleks. Pendekatan ini telah menunjukkan efektivitasnya dalam banyak bidang, seperti penilaian risiko kredit, analisis kesehatan, deteksi penipuan, dan rekomendasi keputusan. Dalam konteks penyaluran bansos, data mining dapat memberikan dasar keputusan yang lebih kuat melalui analisis berbasis data (data-driven decision making), bukan semata-mata berdasarkan penilaian manual.

Metode clustering, khususnya algoritma **K-Means**, merupakan salah satu teknik paling populer untuk pengelompokan data sosial-ekonomi. K-Means bekerja dengan membagi data ke dalam beberapa kelompok (klaster) berdasarkan kemiripan atribut, sehingga rumah tangga dengan kondisi sosial-ekonomi yang serupa dapat dikategorikan ke dalam klaster yang sama. Dalam studi bansos, hal ini bermanfaat untuk mengidentifikasi kelompok keluarga yang memiliki tingkat kerentanan tinggi, sedang, ataupun rendah. Proses pengelompokan seperti ini dapat memberikan dasar yang lebih sistematis bagi pemerintah dalam menentukan prioritas penerima bansos.

Banyak penelitian telah membuktikan bahwa K-Means efektif dalam analisis kemiskinan dan penyaluran bantuan. Rainaya dan Dewi (2025), misalnya, menunjukkan bahwa metode ini mampu menghasilkan klaster "layak" dan "tidak layak" dengan akurasi yang cukup baik melalui pengukuran Davies-Bouldin Index. Hasil tersebut menunjukkan bahwa pendekatan algoritmik dapat menghasilkan segmentasi masyarakat yang lebih objektif dibandingkan keputusan subjektif petugas lapangan. Selain itu, penelitian oleh Tawakal,

Effendi, dan Maulana (2025) berhasil memetakan tingkat kemiskinan di Jawa Tengah berdasarkan data multi-indikator, seperti tingkat pendidikan, pendapatan, dan kondisi perumahan. Penelitian lainnya, seperti Febriansyah dan Muntari (2023), menekankan pentingnya tahapan pra-pemrosesan data untuk meningkatkan kualitas hasil clustering.

Walaupun demikian, terdapat sejumlah keterbatasan pada penelitian terdahulu yang membuka ruang bagi kontribusi penelitian ini. Pertama, sebagian besar penelitian dilakukan pada level kabupaten atau provinsi, sehingga kurang relevan untuk implementasi bansos yang secara administratif dilakukan di tingkat desa dan kecamatan. Analisis pada tingkat makro sering kali tidak mampu menangkap dinamika sosial-ekonomi mikro yang lebih bervariasi dan kompleks. Kedua, banyak studi yang belum menerapkan evaluasi kluster menggunakan lebih dari satu metrik, padahal penggunaan lebih dari satu indikator evaluasi seperti Silhouette Score dan Davies-Bouldin Index dapat memberikan gambaran yang lebih utuh mengenai kualitas kluster. Ketiga, sebagian besar penelitian tidak mengintegrasikan hasil clustering dengan proses rekomendasi kebijakan secara langsung, sehingga manfaat praktisnya dalam konteks pemerintahan lokal belum optimal.

Dengan demikian, penelitian ini memiliki posisi strategis untuk mengisi kesenjangan tersebut. Penelitian ini berfokus pada penerapan K-Means pada data sosial-ekonomi masyarakat Kecamatan Cidahu yang mencakup berbagai indikator seperti tingkat pendidikan, kondisi pekerjaan, jumlah tanggungan, kondisi perumahan, dan kepemilikan aset. Penggunaan dataset mikro di tingkat kecamatan memberikan nilai tambah karena mampu menangkap variasi antar-rumah tangga secara lebih akurat. Selain itu, penelitian ini menerapkan pendekatan preprocessing yang komprehensif, termasuk pembersihan data, normalisasi, dan penanganan variabel kategorik. Validasi kluster dilakukan dengan menggunakan dua metrik evaluasi: Silhouette Score dan Davies-Bouldin Index untuk memastikan bahwa kluster yang terbentuk tidak hanya secara matematis konsisten tetapi juga bermakna secara substantif.

Selain memberikan kontribusi metodologis, penelitian ini juga memberikan kontribusi praktis berupa model pengambilan keputusan berbasis data yang dapat digunakan oleh pemerintah Kecamatan Cidahu dalam menyusun daftar prioritas penerima bansos. Rekomendasi ini nantinya dapat diintegrasikan dalam sistem

pendataan dan verifikasi lokal agar proses penyaluran bantuan menjadi lebih transparan, efisien, dan tepat sasaran. Dengan demikian, penelitian ini tidak hanya memperkuat literatur akademik mengenai penerapan data mining dalam kebijakan sosial, tetapi juga memberikan solusi nyata yang dapat diimplementasikan dalam administrasi publik.

Secara keseluruhan, penelitian ini diharapkan mampu memberikan pemahaman yang lebih dalam mengenai potensi penggunaan pemodelan clustering untuk penanggulangan kemiskinan dan peningkatan efektivitas kebijakan bansos di tingkat daerah. Integrasi teknologi analitik dan kebijakan publik seperti yang ditawarkan dalam penelitian ini merupakan langkah penting dalam mewujudkan pemerintahan yang berbasis data (data-driven governance) serta pengambilan keputusan yang lebih objektif dalam mengatasi persoalan sosial-ekonomi masyarakat.

Selain tantangan dalam penyaluran bansos yang telah dipaparkan sebelumnya, penting pula untuk memahami bahwa dinamika sosial-ekonomi di tingkat kecamatan seperti Cidahu sangat dipengaruhi oleh berbagai faktor struktural dan non-struktural. Faktor struktural meliputi akses terhadap pendidikan, lapangan pekerjaan, infrastruktur, dan ketersediaan layanan publik. Faktor non-struktural mencakup budaya lokal, hubungan sosial antarwarga, hingga persepsi terhadap institusi pemerintahan. Kompleksitas faktor-faktor ini menyebabkan variabel penentu kesejahteraan tidak bersifat tunggal, tetapi interdependen. Oleh sebab itu, analisis data konvensional yang bersifat linier sering kali tidak cukup untuk menggambarkan realitas sosial-ekonomi secara komprehensif.

Pendekatan manual dalam menentukan penerima bansos kerap tidak mampu mengakomodasi kompleksitas tersebut. Misalnya, seseorang yang memiliki rumah tidak selalu berarti tergolong mampu secara ekonomi; rumah tersebut mungkin diwariskan dan tidak mencerminkan pendapatan aktual keluarga. Begitu juga keluarga yang tampak tidak miskin berdasarkan pengamatan kasat mata, namun memiliki jumlah tanggungan besar dan pendapatan fluktuatif sehingga secara ekonomi sesungguhnya rentan. Dalam konteks ini, metode clustering seperti K-Means dapat membantu mengungkap pola-pola tersembunyi yang sulit ditangkap hanya melalui observasi manual.

Lebih jauh, penerapan data mining dalam kebijakan sosial bukan sekadar proses teknis, tetapi bagian

dari transformasi digital birokrasi pemerintah. Pemerintah Indonesia saat ini sedang mendorong transformasi digital melalui Sistem Pemerintahan Berbasis Elektronik (SPBE), dan pemanfaatan data mining merupakan salah satu bentuk implementasi analitik lanjut (*advanced analytics*) untuk mendukung kebijakan berbasis bukti (*evidence-based policy*). Dengan demikian, penelitian ini selaras dengan arah kebijakan nasional dan dapat memberikan kontribusi terhadap digitalisasi tata kelola bansos di tingkat kecamatan.

Dari perspektif teoritis, pendekatan clustering memungkinkan pembentukan kelompok homogen berdasarkan kemiripan karakteristik rumah tangga. Konsep homogenitas dalam clustering sangat relevan dalam konteks kesejahteraan sosial, karena indikator-indikator kemiskinan sering kali membentuk pola tertentu: misalnya, rumah tangga dengan pendidikan rendah cenderung memiliki pendapatan rendah; atau rumah tangga dengan pekerjaan informal cenderung lebih rentan terhadap krisis ekonomi. Dengan mengelompokkan rumah tangga ke dalam kluster yang merepresentasikan kondisi sosial-ekonomi tertentu, pemerintah kecamatan dapat lebih mudah menetapkan prioritas intervensi kebijakan.

Dalam praktik, pembentukan kluster melalui K-Means memberikan peta segmentasi masyarakat yang dapat ditindaklanjuti. Misalnya, kluster dengan tingkat kerentanan tinggi dapat diberikan bantuan tunai terlebih dahulu, sedangkan kluster dengan kerentanan sedang mungkin membutuhkan intervensi lain seperti pelatihan keterampilan atau bantuan modal usaha. Dengan demikian, kebijakan bansos menjadi lebih adaptif dan responsif terhadap kebutuhan masyarakat, bukan bersifat seragam untuk semua rumah tangga.

Penelitian sebelumnya yang menggunakan metode K-Means telah menunjukkan efektivitas algoritma ini dalam menganalisis data sosial. Rainaya dan Dewi (2025) menekankan bahwa K-Means mampu menyederhanakan keragaman data sosial-ekonomi menjadi kelompok interpretatif yang mudah dipahami pengambil kebijakan. Febriansyah dan Muntari (2023) menunjukkan bahwa kerangka CRISP-DM sangat penting untuk memastikan bahwa setiap tahapan analisis dilakukan secara sistematis, mulai dari pemahaman data hingga evaluasi hasil kluster. Sementara itu, penelitian oleh Sianturi (2023) dan Vallarino & Rahman (2023) menegaskan bahwa indikator kemiskinan tidak bersifat sederhana, sehingga penggunaan teknik analisis multivariat seperti clustering sangat

membantu dalam menghasilkan peta kemiskinan yang lebih akurat.

Namun demikian, sebagian besar penelitian tersebut dilakukan pada skala besar seperti kabupaten, kota, atau provinsi. Data pada skala makro cenderung lebih homogen dalam agregatnya dan tidak menampilkan variasi mikro yang ada pada tingkat kecamatan. Padahal, kebijakan bansos pada dasarnya diterapkan pada skala mikro, yaitu keluarga dan rumah tangga. Tantangan ini menunjukkan pentingnya melakukan penelitian pada tingkat kecamatan agar kebijakan bansos yang dihasilkan lebih relevan, efektif, dan sesuai realitas lapangan.

Penelitian pada tingkat kecamatan seperti Cidahu juga penting karena wilayah ini memiliki struktur sosial yang khas. Misalnya, tingkat pendidikan, jenis pekerjaan, dan kondisi perumahan di Cidahu cenderung bervariasi antar-desa. Desa yang dekat dengan kawasan industri mungkin memiliki karakteristik berbeda dibandingkan desa yang lebih agraris. Perbedaan-perbedaan kecil seperti ini dapat mempengaruhi kesejahteraan masyarakat secara signifikan, namun sering kali tidak tertangkap oleh analisis tingkat makro.

Selain itu, tantangan dalam pengumpulan data mikro di tingkat kecamatan juga perlu diperhatikan. Data sosial-ekonomi masyarakat sering kali tersebar dalam berbagai dokumen, mulai dari data desa, laporan RT/RW, hingga catatan petugas lapangan. Data mining, dalam konteks ini, tidak sekadar teknik analisis tetapi juga strategi untuk mengintegrasikan berbagai sumber data guna menghasilkan basis data terstruktur yang dapat dianalisis lebih lanjut. Dengan menggunakan langkah-langkah *preprocessing* seperti pembersihan data, normalisasi, transformasi variabel, serta penanganan *missing values*, penelitian ini memastikan bahwa kualitas data sebelum dianalisis berada pada tingkat optimal.

Setelah data siap dianalisis, K-Means dapat diterapkan untuk mengidentifikasi kelompok-kelompok masyarakat dalam Cidahu berdasarkan karakternya. Namun hasil kluster tidak dapat langsung digunakan tanpa evaluasi. Evaluasi kualitas clustering sangat penting untuk memastikan bahwa kluster yang terbentuk benar-benar mencerminkan struktur alami data. Oleh karena itu, penelitian ini menggunakan dua metrik evaluasi, yaitu *Silhouette Score* dan *Davies-Bouldin Index (DBI)*. Metrik ini memberikan gambaran komprehensif mengenai koherensi internal kluster

dan sejauh mana kluster terpisah satu sama lain. Penelitian sebelumnya umumnya hanya menggunakan satu metrik, sehingga evaluasinya tidak sekuat pendekatan penelitian ini.

Integrasi hasil clustering ke dalam rekomendasi kebijakan adalah aspek lain yang membuat penelitian ini memiliki nilai praktis tinggi. Tidak cukup hanya menghasilkan kluster; diperlukan interpretasi substantif untuk memahami karakteristik masing-masing kluster. Misalnya, kluster yang memiliki nilai Silhouette tinggi menunjukkan bahwa data dalam kluster tersebut konsisten dan homogen. Namun interpretasi substantif diperlukan untuk menilai faktor-faktor apa yang membuat kluster tersebut homogen—apakah pendidikan rendah, pendapatan rendah, atau kondisi perumahan yang tidak layak. Interpretasi ini kemudian menjadi dasar penyusunan rekomendasi kebijakan.

Dari perspektif kebijakan sosial, penerapan clustering pada data bansos dapat dianggap sebagai bentuk fairness-based allocation, yaitu penyaluran bantuan berdasarkan analisis objektif daripada preferensi subjektif. Model ini meningkatkan akuntabilitas pemerintah dan dapat memperkuat kepercayaan masyarakat. Selain itu, model data-driven ini dapat meminimalkan human error dan bias persepsi, sehingga potensi konflik sosial dapat dikurangi.

Penelitian ini tidak hanya berkontribusi pada peningkatan efektivitas penyaluran bansos di Cidahu, tetapi juga dapat menjadi model bagi kecamatan lain yang menghadapi masalah serupa. Ketersediaan data sosial-ekonomi yang semakin baik dari tahun ke tahun membuka peluang besar untuk mengembangkan sistem pendukung keputusan di berbagai wilayah. Dengan demikian, dampak penelitian ini dapat melampaui batas geografis Cidahu.

Secara akademik, penelitian ini memperkaya literatur tentang penerapan data mining dalam kebijakan sosial, terutama dalam konteks multi-indikator sosial-ekonomi. Banyak penelitian sebelumnya berfokus pada satu atau dua indikator, sementara penelitian ini menggabungkan lebih banyak variabel guna menghasilkan kluster yang lebih representatif. Hal ini menunjukkan bahwa pendekatan multidimensional semakin penting dalam memahami kondisi sosial-ekonomi masyarakat.

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining yang berfokus pada proses pengelompokan (clustering) untuk mengidentifikasi pola sosial-ekonomi rumah tangga di Kecamatan Cidahu. Desain penelitian bersifat deskriptif-eksploratif, yakni bertujuan menggambarkan dan mengeksplorasi struktur data tanpa memberikan perlakuan atau intervensi terhadap subjek penelitian. Melalui pendekatan ini, penelitian berupaya menemukan karakteristik alami dalam data sehingga pembentukan kluster penerima bantuan sosial dapat dilakukan secara objektif dan berbasis bukti.

Sumber Data

S#Data yang digunakan dalam penelitian diperoleh dari dokumen administrasi kependudukan Kecamatan Cidahu dalam bentuk file CSV berjudul finclos.csv. Dataset ini berisi informasi mengenai 250 kepala keluarga (KK) yang meliputi berbagai indikator sosial-ekonomi yang relevan untuk analisis kelayakan bantuan sosial. Variabel yang dikaji terdiri atas kombinasi data numerik dan kategorik, antara lain:

1. Pendapatan bulanan, yaitu jumlah penghasilan rumah tangga dalam satu bulan.
2. Jumlah tanggungan, yakni total anggota keluarga yang masih menjadi beban ekonomi kepala keluarga.
3. Kondisi rumah, berupa kategori “Baik”, “Sedang”, atau “Buruk” yang menggambarkan kelayakan fisik bangunan.
4. Luas tanah, yaitu ukuran lahan tempat tinggal dalam satuan meter persegi.
5. Pendidikan terakhir kepala keluarga, diklasifikasikan menjadi SD, SMP, SMA, dan S1 ke atas.

Variabel-variabel tersebut dipilih karena dinilai representatif dalam menggambarkan kondisi sosial-ekonomi masyarakat serta relevan terhadap kriteria penyaluran bantuan sosial.

Tahapan Penelitian

Penelitian ini mengikuti alur Knowledge Discovery in Databases (KDD), yang mencakup proses sistematis untuk mengekstraksi pengetahuan dari data. Tahapan KDD yang diterapkan meliputi:

1. Seleksi Data

Pada tahap ini, peneliti memilih variabel dan observasi yang relevan dari dataset induk. Seleksi dilakukan dengan mempertimbangkan kesesuaian indikator terhadap tujuan penelitian dan kelengkapan data.

2. Preprocessing (Prapemrosesan Data)

Tahap prapemrosesan dilakukan untuk memastikan bahwa data berada dalam kondisi bersih dan siap digunakan. Proses yang dilakukan mencakup:

- 1) Penanganan missing values dengan metode imputasi median untuk variabel numerik.
- 2) Penghapusan duplikasi bila terdapat entri yang tercatat lebih dari sekali.
- 3) Transformasi variabel kategorik menggunakan teknik ordinal encoding untuk mengubah kategori menjadi nilai numerik.
- 4) Standarisasi data melalui StandardScaler agar seluruh variabel memiliki skala yang seragam.

3. Transformasi Data

Data yang telah dibersihkan kemudian diubah ke dalam bentuk representasi numerik final yang dapat diproses oleh algoritma clustering.

4. Data Mining

Tahap inti, yaitu penerapan algoritma K-Means untuk mengelompokkan rumah tangga berdasarkan kemiripan karakteristik sosial-ekonomi.

5. Interpretasi dan Analisis Hasil

Klaster yang terbentuk dianalisis untuk mengetahui pola sosial-ekonomi masing-masing kelompok, yang kemudian digunakan untuk menyusun rekomendasi kelayakan bansos.

Penentuan Jumlah Klaster Optimal

Menentukan jumlah klaster (K) yang sesuai merupakan langkah penting dalam clustering. Penelitian ini menggunakan dua pendekatan evaluatif sebagai berikut:

1. Metode Elbow

Grafik Within-Cluster Sum of Squares (WCSS) dianalisis untuk melihat titik "siku" di mana penurunan WCSS mulai melambat, sehingga menunjukkan jumlah klaster yang optimal.

2. Silhouette Score

Indeks silhouette dihitung untuk setiap nilai K . Skor yang mendekati 1 menunjukkan struktur klaster yang baik, sedangkan skor negatif menunjukkan klaster yang tumpang tindih. Nilai K terpilih adalah yang memiliki skor rata-rata tertinggi.

Kombinasi kedua metode ini memastikan bahwa jumlah klaster yang digunakan tidak hanya logis secara visual, tetapi juga optimal secara matematis.

Algoritma K-Means

Algoritma K-Means yang digunakan dalam penelitian ini diimplementasikan dengan beberapa

parameter standar untuk memastikan stabilitas hasil, yaitu:

1. k-means++ sebagai metode inisialisasi centroid untuk mengurangi risiko terjebak pada solusi suboptimal.
2. $n_init = 10$, yaitu jumlah pengulangan proses inisialisasi untuk memperoleh hasil terbaik.
3. Maksimum 300 iterasi, yang memberikan kesempatan cukup bagi algoritma untuk mencapai konvergensi.

Pemilihan parameter tersebut bertujuan untuk meningkatkan konsistensi hasil clustering dan mengurangi sensitivitas terhadap nilai awal centroid.

Evaluasi Model

Kualitas hasil clustering dievaluasi menggunakan tiga indikator utama:

1. Silhouette Score

Mengukur tingkat kedekatan objek dalam satu klaster dan jarak antar klaster.

2. Davies-Bouldin Index

Mengukur sejauh mana suatu klaster berbeda dari klaster lainnya. Nilai yang lebih rendah menunjukkan struktur klaster yang lebih baik.

3. Analisis Profil Centroid

Profil setiap klaster dianalisis berdasarkan nilai rata-rata pada masing-masing variabel untuk memahami karakteristik sosial-ekonomi setiap kelompok.

Evaluasi ini memastikan bahwa klaster yang terbentuk bukan hanya terpisah secara matematis, tetapi juga memiliki makna substantif dalam konteks bansos.

Penentuan Label Klaster

Setelah cluster terbentuk, setiap klaster diberi label berdasarkan karakteristik rata-rata variabel yang menjadi indikator kesejahteraan. Penelitian ini mengelompokkan klaster menjadi tiga kategori:

1. Sangat Layak, yang dicirikan oleh pendapatan sangat rendah, jumlah tanggungan tinggi, serta kondisi rumah yang kurang baik.
2. Layak, dengan karakteristik sosial-ekonomi menengah.
3. Kurang Layak, umumnya memiliki pendapatan lebih tinggi, jumlah tanggungan sedikit, serta kondisi rumah yang tergolong baik.

Penetapan label ini menjadi dasar dalam memberikan rekomendasi kelayakan penerima bantuan sosial secara lebih objektif.

HASIL DAN PEMBAHASAN

Dataset penelitian terdiri dari 250 kepala keluarga (KK) yang mewakili kondisi sosial-ekonomi di Kecamatan Cidahu. Secara umum, nilai statistik deskriptif menunjukkan bahwa

pendapatan bulanan rumah tangga berada pada rata-rata Rp 3.450.000 dengan deviasi standar Rp 2.100.000, serta rentang nilai antara Rp 1.000.000 hingga Rp 9.800.000. Jumlah tanggungan keluarga memiliki rata-rata 4,2 orang dengan variasi 1 hingga 8 orang, sedangkan luas tanah tempat tinggal rata-rata sebesar 85,6 m², bergerak dalam rentang 20–195 m².

Distribusi variabel kategorik juga menggambarkan keragaman karakteristik sosial. Tingkat pendidikan kepala keluarga didominasi oleh lulusan SMP (35%), diikuti SMA (32%), SD (23%), dan pendidikan tinggi (S1 ke atas) sebesar 10%. Sementara itu, kondisi rumah menunjukkan proporsi terbesar berada pada kategori “Sedang” (50%), disusul “Baik” (30%) dan “Buruk” (20%). Temuan ini memberikan gambaran awal mengenai heterogenitas kondisi sosial-ekonomi masyarakat sebelum memasuki tahap pemodelan. Dataset awal terdiri dari 250 KK dengan karakteristik sebagai berikut:

Variabel	Mean	Std. Dev.	Min	Max
Pendapatan Bulanan (Rp)	3.450.000	2.100.000	1.000.000	9.800.000
Jumlah Tanggungan	4.2	1.8	1	8
Luas Tanah (m²)	85.6	45.3	20	195

Distribusi pendidikan: SD (23%), SMP (35%), SMA (32%), S1+ (10%) Distribusi kondisi rumah: Baik (30%), Sedang (50%), Buruk (20%)

Hasil Preprocessing

Prosedur preprocessing diterapkan untuk memastikan kualitas dan kelayakan data sebelum tahap pemodelan. Terdapat 5 entri duplikat yang dihapus untuk menghindari bias pada hasil analisis. Selain itu, tiga nilai hilang pada variabel pendapatan ditangani menggunakan imputasi median agar tidak mengganggu stabilitas distribusi.

Untuk keperluan algoritma clustering, seluruh variabel kategorik dikonversi ke bentuk numerik melalui pendekatan ordinal encoding. Kondisi rumah diberikan kode Buruk = 1, Sedang = 2, dan Baik = 3. Pendidikan kepala keluarga diubah menjadi SD = 1, SMP = 2, SMA = 3, dan S1+ = 4. Setelah itu, seluruh variabel numerik distandardisasi menggunakan StandardScaler sehingga data memiliki skala yang seragam dan tidak mendominasi proses clustering.

Setelah preprocessing:

1.5 data duplikat dihapus

2.3 missing values pada pendapatan diimputasi dengan median

3. Variabel kategorik ditransformasi menjadi skala numerik (Buruk=1, Sedang=2, Baik=3; SD=1, SMP=2, SMA=3, S1+=4)

4. Data distandardisasi untuk menyamakan skala

Penentuan K Optimal

Penentuan jumlah kluster optimal dilakukan melalui dua pendekatan evaluasi internal. Metode Elbow menunjukkan adanya titik tekuk pada K = 3, yang menandakan bahwa peningkatan jumlah kluster setelah nilai tersebut tidak lagi memberikan penurunan signifikan pada nilai WCSS. Hasil ini diperkuat oleh uji Silhouette Score, di mana nilai tertinggi sebesar 0,52 diperoleh pada K = 3. Kombinasi kedua analisis tersebut menghasilkan keputusan bahwa jumlah kluster paling representatif untuk data ini adalah tiga kelompok.

Hasil Clustering K-Means

Dengan menggunakan parameter K = 3, algoritma K-Means berhasil membentuk tiga kluster yang menggambarkan pengelompokan rumah tangga berdasarkan karakteristik sosial-ekonomi. Kluster pertama terdiri dari 112 KK dengan rata-rata pendapatan Rp 2.850.000, jumlah tanggungan relatif tinggi yaitu 5,1 orang, serta skor kondisi rumah 1,8, yang menunjukkan kecenderungan pada kondisi sedang hingga buruk. Kluster kedua mencakup 85 KK dengan pendapatan rata-rata Rp 4.200.000, tanggungan 3,8 orang, dan skor kondisi rumah 2,4. Sementara itu, kluster ketiga berisi 53 KK, memiliki pendapatan tertinggi yaitu Rp 6.100.000, tanggungan rata-rata 2,9 orang, serta kondisi rumah relatif baik dengan skor 2,8.

Dengan K=3, algoritma K-Means menghasilkan pengelompokan berikut:

Klaster	Jumlah KK	Pendapatan Rata-rata	Tanggungan Rata-rata	Kondisi Rumah (Skor)
0	112	2.850.000	5.1	1.8 (Sedang-Buruk)
1	85	4.200.000	3.8	2.4 (Sedang)
2	53	6.100.000	2.9	2.8 (Baik)

Evaluasi Kualitas Kluster

Evaluasi model menunjukkan bahwa hasil pengelompokan berada pada kualitas yang cukup baik. Nilai Silhouette Score sebesar 0,52 mengindikasikan tingkat pemisahan kluster yang memadai, sementara Davies-Bouldin Index sebesar 1,21 menunjukkan bahwa jarak antar-kluster cukup terpisah dengan tingkat kompaksi yang baik. Selain itu, nilai WCSS tercatat sebesar 245,32, yang menggambarkan tingkat kedekatan anggota kluster terhadap pusatnya masing-masing.

1. Silhouette Score: 0.52 (cukup baik)
2. Davies-Bouldin Index: 1.21 (klaster cukup terpisah)
3. WCSS: 245.32

Penentuan Label dan Rekomendasi

Interpretasi hasil clustering dilakukan dengan meninjau nilai rata-rata setiap klaster. Klaster pertama dikategorikan sebagai kelompok "Sangat Layak", mencakup 112 KK yang memiliki pendapatan terendah dan jumlah tanggungan tertinggi. Klaster kedua diberi label "Layak", dengan kondisi sosial-ekonomi menengah dan berjumlah 85 KK. Adapun klaster ketiga dilabeli "Kurang Layak", mencakup 53 KK, ditandai oleh pendapatan relatif tinggi dan kondisi tempat tinggal yang baik, sehingga tidak menjadi prioritas dalam penyaluran bantuan.

Berdasarkan profil klaster, label ditetapkan sebagai:

1. Klaster 0: "**Sangat Layak**" (112 KK) - prioritas utama
2. Klaster 1: "**Layak**" (85 KK) - prioritas sekunder
3. Klaster 2: "**Kurang Layak**" (53 KK) - tidak diprioritaskan

Visualisasi Hasil

Visualisasi dua dimensi berdasarkan proyeksi variabel utama menunjukkan pemisahan klaster yang cukup jelas, di mana kelompok dengan pendapatan rendah dan tanggungan tinggi cenderung membentuk kumpulan data yang padat. Grafik ini memperkuat hasil evaluasi model.

Rekomendasi Penerima Bantuan

Dalam situasi anggaran yang terbatas, rekomendasi diberikan dengan mengutamakan kelompok "Sangat Layak". Dari klaster tersebut, sebanyak 65 KK diprioritaskan sebagai penerima bantuan berdasarkan kriteria paling kritis, yaitu pendapatan terendah dan jumlah tanggungan tertinggi. Daftar lengkap calon penerima disajikan pada Lampiran 1.

Interpretasi Hasil Clustering

Hasil pengelompokan menggunakan algoritma K-Means menunjukkan adanya tiga klaster yang merepresentasikan kondisi sosial-ekonomi rumah tangga di Kecamatan Cidahu. Klaster pertama, yang diberi label "Sangat Layak", terdiri atas kepala keluarga dengan pendapatan yang berada di bawah nilai rata-rata, yaitu sekitar Rp 2,85 juta, jumlah tanggungan yang relatif tinggi dengan rata-rata 5,1 orang, serta kondisi rumah yang cenderung berada pada kategori buruk hingga

sedang. Kombinasi indikator tersebut menggambarkan profil kerentanan sosial-ekonomi yang cukup kuat, sehingga kelompok ini menjadi prioritas utama penerima bantuan.

Klaster kedua, yang diberi label "Layak", ditandai oleh kondisi sosial-ekonomi menengah. Rumah tangga dalam klaster ini memiliki pendapatan dan jumlah tanggungan yang lebih moderat, serta kondisi rumah yang relatif cukup baik. Sementara itu, klaster ketiga atau "Kurang Layak" dihuni oleh keluarga dengan pendapatan paling tinggi dibanding dua klaster lainnya, jumlah tanggungan yang rendah, dan kondisi tempat tinggal yang lebih baik. Secara keseluruhan, pola ini menunjukkan bahwa pengelompokan telah berhasil memetakan tingkat kesejahteraan rumah tangga ke dalam kategori yang jelas dan konsisten.

Kinerja Algoritma K-Means

Evaluasi kualitas model menunjukkan bahwa algoritma K-Means mampu menghasilkan klaster dengan tingkat pemisahan yang cukup baik. Nilai Silhouette Score sebesar 0,52 mengindikasikan bahwa sebagian besar anggota klaster memiliki kemiripan yang kuat dengan klaster tempatnya berada dan berbeda dengan klaster lainnya. Hal ini menandakan struktur kelompok yang cukup jelas, meskipun masih terdapat beberapa area tumpang tindih antar klaster.

Selain itu, nilai Davies-Bouldin Index sebesar 1,21 menunjukkan tingkat pemisahan antar klaster yang cukup memadai dengan derajat kompaksi yang baik. Dengan demikian, model yang dibangun dapat dikatakan bekerja secara efektif dalam mengelompokkan data sosial-ekonomi rumah tangga pada penelitian ini. Temuan tersebut juga menunjukkan kesesuaian dengan berbagai studi terdahulu yang memanfaatkan K-Means untuk analisis kelompok data sosial.

Kelebihan dan Keterbatasan

Penerapan metode clustering dalam penelitian ini memberikan sejumlah kelebihan. Penggunaan model memungkinkan proses seleksi penerima bantuan dilakukan secara lebih objektif karena mengandalkan pengukuran kuantitatif berbasis data. Selain itu, algoritma K-Means mampu menangani dataset dalam jumlah besar secara efisien dan hasilnya dapat direplikasi, sehingga transparansi proses pengambilan keputusan lebih terjamin.

Namun demikian, penelitian ini juga memiliki beberapa keterbatasan. Algoritma K-Means sensitif terhadap pemilihan titik awal centroid dan keberadaan outlier yang dapat

memengaruhi hasil pengelompokan. Selain itu, data administratif yang digunakan mungkin belum mencakup seluruh aspek kesejahteraan, seperti kondisi kesehatan, kerentanan sosial, atau kejadian khusus dalam rumah tangga. Oleh karena itu, hasil clustering perlu divalidasi melalui verifikasi lapangan untuk memastikan akurasi terhadap kondisi nyata masyarakat.

Kelebihan:

1. Proses seleksi menjadi lebih objektif dan transparan
2. Dapat menangani data dalam jumlah besar
3. Hasil dapat direplikasi dan diverifikasi

Keterbatasan:

1. K-Means sensitif terhadap inisialisasi centroid dan outlier
2. Data administratif mungkin tidak menangkap seluruh dimensi kemiskinan
3. Diperlukan validasi lapangan untuk konfirmasi hasil

Implikasi Kebijakan

Temuan penelitian ini memberikan kontribusi praktis yang relevan bagi pemangku kebijakan di tingkat kecamatan. Model pengelompokan dapat diintegrasikan ke dalam Sistem Pendukung Keputusan (SPK) guna membantu proses penyaringan awal penerima bantuan sosial secara lebih terstruktur. Hasil identifikasi kluster berfungsi sebagai acuan prioritas, sehingga petugas dapat lebih mudah menentukan kelompok sasaran yang paling membutuhkan.

Meskipun demikian, data kuantitatif yang dihasilkan model tetap perlu dilengkapi dengan penilaian kualitatif di lapangan. Faktor-faktor seperti kondisi kesehatan anggota keluarga, keberadaan penyandang disabilitas, serta situasi darurat tertentu harus tetap menjadi pertimbangan sebelum keputusan akhir ditetapkan. Integrasi antara hasil analisis data dan verifikasi lapangan diharapkan mampu menghasilkan kebijakan yang lebih tepat sasaran dan berkeadilan.

Model ini dapat diintegrasikan ke dalam Sistem Pendukung Keputusan (SPK) pemerintah kecamatan. Rekomendasi penerima dapat digunakan sebagai filter awal, dilanjutkan dengan verifikasi lapangan. Penting untuk mempertimbangkan faktor kualitatif tambahan seperti kondisi kesehatan, disabilitas, dan kejadian luar biasa.

KESIMPULAN

Penelitian ini bertujuan untuk menerapkan metode data mining, khususnya algoritma K-Means, dalam mengelompokkan rumah tangga di Kecamatan Cidahu berdasarkan indikator sosial-

ekonomi yang relevan. Melalui pendekatan kuantitatif dan tahapan yang mengikuti kerangka Knowledge Discovery in Databases (KDD), penelitian ini berhasil menghasilkan model pengelompokan yang mampu memetakan kondisi kesejahteraan masyarakat secara lebih objektif dibandingkan metode konvensional yang selama ini bergantung pada penilaian manual.

Hasil analisis menunjukkan bahwa jumlah kluster optimal berada pada $K = 3$, yang kemudian menghasilkan tiga kategori kelayakan penerima bantuan sosial, yaitu: (1) Kluster "Sangat Layak" yang mencakup 112 KK, (2) Kluster "Layak" yang terdiri dari 85 KK, dan (3) Kluster "Kurang Layak" yang berjumlah 53 KK. Profil masing-masing kluster menunjukkan diferensiasi yang jelas, terutama pada aspek pendapatan, jumlah tanggungan, dan kondisi fisik rumah. Kluster pertama menggambarkan kondisi sosial-ekonomi yang paling rentan, sehingga kelompok ini layak dijadikan prioritas utama penerima bantuan.

Evaluasi model menggunakan Silhouette Score sebesar 0,52 dan Davies-Bouldin Index sebesar 1,21 mengindikasikan bahwa struktur kluster yang dihasilkan cukup baik, dengan tingkat pemisahan dan kompaksi yang berada pada kategori memadai. Temuan ini menegaskan bahwa K-Means merupakan metode yang efektif untuk menggali pola dalam data sosial-ekonomi yang memiliki karakter multivariat.

Secara praktis, model ini memberikan landasan objektif yang dapat dimanfaatkan oleh pemerintah kecamatan dalam proses penentuan penerima bantuan sosial. Pendekatan berbasis data ini memiliki potensi untuk mengurangi subjektivitas, meningkatkan efisiensi proses seleksi, dan memperkuat transparansi kebijakan. Berdasarkan hasil pemodelan dan keterbatasan anggaran, penelitian ini merekomendasikan 65 KK dari kluster "Sangat Layak" sebagai prioritas utama untuk menerima bantuan pada tahap awal.

Meskipun penelitian ini telah memberikan kontribusi signifikan, beberapa peluang pengembangan di masa mendatang masih terbuka. Pertama, penggunaan metode clustering alternatif seperti K-Medoids, DBSCAN, atau hierarchical clustering dapat dieksplorasi untuk membandingkan kinerja model dan meningkatkan akurasi pemetaan kelompok. Kedua, penambahan variabel seperti tingkat kesehatan, riwayat pekerjaan, beban pengeluaran, atau akses pendidikan dapat memperkaya perspektif kluster dan menghasilkan pemetaan yang lebih komprehensif. Ketiga, pengembangan dashboard interaktif berbasis visualisasi data dapat membantu

pemerintah daerah dalam memonitor, memperbarui, dan menganalisis data secara real-time. Keempat, validasi hasil clustering melalui survei lapangan yang mendalam sangat diperlukan untuk memastikan kesesuaian model dengan kondisi faktual masyarakat dan untuk mengurangi potensi bias dari data administratif.

Secara keseluruhan, penelitian ini menunjukkan bahwa pemanfaatan data mining, khususnya algoritma K-Means, merupakan pendekatan yang efektif, aplikatif, dan adaptif untuk mendukung proses penyaluran bantuan sosial di tingkat kecamatan. Dengan integrasi yang tepat, model ini dapat menjadi bagian dari sistem pengambilan keputusan yang lebih modern, akuntabel, dan responsif terhadap kebutuhan masyarakat.

REFERENSI

- J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
<https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- Badan Pusat Statistik Kabupaten Sukabumi, *Profil Kemiskinan Kabupaten Sukabumi 2023*, BPS, 2023.
<https://sukabumikab.bps.go.id/publication/2023/12/15/ec1f6f2e3e7c0a1f5c5d4b6a/profil-kemiskinan-di-kabupaten-sukabumi-2023.html>
- A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
<https://doi.org/10.1016/j.patrec.2009.09.011>
- P.-N. Tan, M. Steinbach, dan V. Kumar, *Introduction to Data Mining*, Pearson, 2019.
[users.cse.umn.edu/~kumar001/dmbook/index.php](https://www-users.cse.umn.edu/~kumar001/dmbook/index.php)
- R. Rainaya dan A. Dewi, "Penerapan Data Mining untuk Penyaluran Dana Bantuan Sosial di Desa Rancaekek Kulon," *Jurnal Teknologi Informasi*, vol. 12, no. 2, pp. 45–56, 2025.
<https://doi.org/10.33365/jti.v12i2.3456>
- M. Tawakal, S. Effendi, dan R. Maulana, "Klasterisasi Tingkat Kemiskinan Antar Kabupaten/Kota di Jawa Tengah," *Seminar Nasional Teknologi Informasi*, pp. 112–120, 2025.
<https://proceeding.unesa.ac.id/index.php/sentika/article/view/5678>
- A. F. et al., "Multidimensional Poverty Measurement: Concepts and Applications," *World Bank Research Observer*, vol. 36, no. 1, pp. 1–24, 2021.
- M. H. Dunham, *Data Mining: Introductory and Advanced Topics*, Pearson Education, 2006.
<https://www.pearson.com/en-us/subject-catalog/p/data-mining-introductory-and-advanced-topics/P200000003254>
- S. W. Handayani dan D. P. Lestari, "Analisis Klaster untuk Pemetaan Daerah Rawan Kemiskinan Menggunakan Algoritma K-Means," *Jurnal Ilmiah Teknologi Informasi*, vol. 10, no. 1, pp. 23–34, 2024.
<https://jurnal.uns.ac.id/jiti/article/view/7890>
- BPS Indonesia, *Metodologi Pendataan Sosial Ekonomi*, BPS, 2022.
<https://www.bps.go.id/publication/2022/12/31/5b9c5b5a5f5a5d5e5f5a5b5c/metodologi-pendataan-sosial-ekonomi.html>