

Klasterisasi Penerima Bantuan Sosial Berdasarkan Karakteristik Sosial-Ekonomi Menggunakan K-Means

Nala Vinelza¹, Delsi Kariman^{2*}, Nia Febriyani³

^{1,2,3} Program Studi Sains Data – Universitas PGRI Sumatera Barat

Jl. Gunung Pangilun – Padang Sumatera Barat - Indonesia

nalavinelza05@gmail.com, delsik@upgrisba.ac.id, febriyaninia@gmail.com

Abstrak— Pemanfaatan data penerima bantuan sosial di tingkat desa saat ini masih terbatas sebagai data administratif, sehingga karakteristik sosial-ekonomi penerima belum terpetakan secara komprehensif. Penelitian ini bertujuan melakukan segmentasi penerima bantuan sosial di Desa Silungkang Tigo menggunakan algoritma K-Means dengan pendekatan CRISP-DM. Dataset yang digunakan terdiri atas 156 kepala keluarga dengan atribut desil kesejahteraan, penghasilan, jumlah tanggungan, jenis pekerjaan, serta status penerimaan bantuan sosial (PKH dan BPNT). Penentuan jumlah kluster dilakukan melalui evaluasi menggunakan Davies-Bouldin Index (DBI), Silhouette Coefficient, dan distribusi ukuran kluster. Hasil penelitian menunjukkan bahwa konfigurasi dua kluster (K=2) merupakan konfigurasi yang paling layak diinterpretasikan dengan nilai Silhouette sebesar 0,2848 dan DBI sebesar 1,5897; nilai Silhouette tersebut mengindikasikan struktur kluster yang masih tergolong lemah sehingga hasil dimaknai sebagai gambaran awal yang bersifat eksploratif. Hasil klasterisasi mengidentifikasi dua kelompok utama, yaitu kelompok rentan (71,79%) yang didominasi rumah tangga berpendapatan rendah dengan pekerjaan informal dan kelompok relatif lebih sejahtera (28,21%) yang didominasi pekerjaan kategori produksi, pertanian, dan buruh (59,09%) dengan pekerjaan bertenun sebagai subkelompok yang menonjol dan kondisi ekonomi yang lebih stabil. Kontribusi utama penelitian ini adalah karakterisasi sosial-ekonomi per kluster yang rinci (mencakup rata-rata penghasilan, komposisi pekerjaan, dan cakupan bantuan), serta sebuah *dashboard* operasional yang dapat dimanfaatkan perangkat desa sebagai informasi pendukung dalam pengelolaan dan evaluasi program bantuan sosial berbasis data.

Kata Kunci— K-Means, klasterisasi, bantuan sosial, sosial-ekonomi, data mining.

Abstract— The utilization of social assistance recipient data at the village level is currently limited to administrative records, resulting in a lack of a comprehensive understanding of recipients' socio-economic characteristics. This study aims to segment social assistance recipients in Silungkang Tigo Village using the K-Means algorithm within the CRISP-DM framework. The dataset consisted of 156 households with attributes including welfare decile, income, number of dependents, occupation type, and social assistance status (PKH and BPNT). The number of clusters was determined using the Davies-Bouldin Index (DBI), Silhouette Coefficient, and cluster size distribution. The results indicate that a two-cluster configuration (K=2) is the most interpretable, achieving a Silhouette score of 0.2848 and a DBI value of 1.5897; this Silhouette value reflects a relatively weak cluster structure, so the findings are interpreted as a preliminary, exploratory overview. The clustering process identified two main groups: a vulnerable group (71.79%) characterized by lower household income and a relatively more prosperous group

(28.21%) with more stable economic conditions. The main contribution of this study is a detailed per-cluster socio-economic characterization (covering average income, occupational composition, and assistance coverage) together with an operational dashboard that village officials can use as supporting information for data-driven management and evaluation of social assistance programs.

Keywords— K-Means, clustering, social assistance, socio-economic, data mining.

I. PENDAHULUAN

Data penerima bantuan sosial (bansos) merupakan sumber informasi krusial yang menggambarkan kondisi sosial-ekonomi masyarakat pada tingkat rumah tangga. Data tersebut mencakup variabel seperti tingkat kesejahteraan (desil), penghasilan, jumlah tanggungan, jenis pekerjaan, serta status penerimaan bantuan. Informasi ini tidak hanya berfungsi sebagai dasar administratif, tetapi juga berpotensi besar untuk diolah menjadi wawasan analitis guna memetakan pola karakteristik masyarakat secara sistematis [1]. Namun, pengelolaan data di tingkat desa saat ini masih sering bersifat konvensional, sehingga karakteristik penerima bantuan belum terpetakan secara mendalam untuk mendukung pengambilan kebijakan desa yang lebih akurat [2]. Secara global, akurasi penetapan sasaran bantuan (*poverty targeting*) menjadi isu sentral, dan *proxy means testing* telah banyak dikaji sebagai pendekatan untuk mengestimasi kesejahteraan rumah tangga berdasarkan atribut yang teramati [3], [4], termasuk dalam konteks Indonesia.

Kondisi serupa ditemukan dalam manajemen data kesehatan, di mana sistem informasi yang tersedia belum mampu menyajikan analisis pengelompokan yang dapat dimanfaatkan untuk perencanaan strategis [5]. Padahal, pengelompokan data sangat penting untuk mendukung pengambilan keputusan berbasis data, baik dalam bidang kesehatan maupun kebijakan bantuan sosial, karena mampu mentransformasikan data mentah yang bersifat administratif menjadi informasi strategis yang dapat mengungkap pola tersembunyi di balik distribusi bantuan [6]. Kompleksitas dan heterogenitas data pada penerima bansos di Desa Silungkang Tigo menuntut penerapan teknik data mining, khususnya metode klasterisasi yang mampu mengelompokkan data berdasarkan kemiripan karakteristik tanpa label kelas sebelumnya [7].

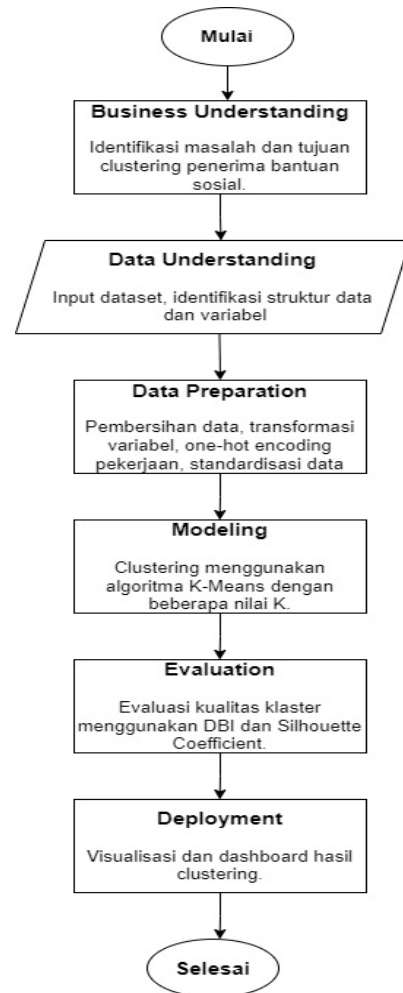
Algoritma K-Means merupakan metode klasterisasi yang paling sering digunakan karena memiliki struktur algoritma yang sederhana, efisien secara komputasi, serta mampu

memberikan hasil pengelompokan yang memadai pada data berskala besar [8]. K-Means bekerja dengan mempartisi data ke dalam kluster berdasarkan kedekatan terhadap pusat kluster (*centroid*), sehingga objek serupa berada pada kelompok yang sama sementara objek berbeda dipisahkan ke kluster lainnya [9]. Kelebihan K-Means antara lain skalabilitas yang baik, implementasi yang mudah, serta kemampuan deteksi anomali pada data yang membantu dalam mengidentifikasi profil karakteristik secara spesifik [10]. Selain itu, kompleksitas waktunya yang relatif rendah menjadikan K-Means pilihan utama dalam berbagai aplikasi pengelompokan.

Keefektifan K-Means telah dibuktikan melalui berbagai penelitian, mulai dari analisis keterlambatan pembayaran SPP [6] hingga pengelompokan data rekam medis pasien [11]. Dalam bidang bantuan sosial, metode serupa juga berhasil diterapkan pada segmentasi penerima BLT Dana Desa yang mampu memberikan klasifikasi kelompok ekonomi secara transparan sebagai dasar pengambilan keputusan bagi perangkat desa [12] serta pada pengelompokan penerima bantuan sosial masyarakat yang efektif memetakan profil kesejahteraan rumah tangga secara lebih akurat dan objektif [13]. Beberapa penelitian terdahulu telah memetakan profil kesejahteraan rumah tangga pada hasil klusterisasi penerima bantuan sosial [13], namun karakterisasi sosial-ekonomi yang rinci pada tingkat desa, mencakup rata-rata penghasilan, komposisi pekerjaan, dan cakupan bantuan per kluster, serta penyajiannya dalam bentuk *dashboard* operasional untuk perangkat desa masih jarang dilakukan. Oleh karena itu, penelitian ini bertujuan melakukan segmentasi penerima bantuan sosial di Desa Silungkang Tigo menggunakan algoritma K-Means dengan pendekatan CRISP-DM. Penelitian ini diposisikan sebagai studi kasus aplikatif, bukan sebagai kontribusi metodologis baru, karena penggunaan beberapa metrik validitas internal seperti *Davies-Bouldin Index* dan *Silhouette Coefficient* merupakan praktik yang lazim dalam evaluasi klusterisasi. Nilai tambah penelitian terletak pada tiga hal: (1) karakterisasi sosial-ekonomi per kluster yang rinci dan dapat ditindaklanjuti; (2) temuan empiris spesifik wilayah, misalnya pekerjaan bertenun sebagai subkelompok pekerjaan yang menonjol penanda kluster yang relatif lebih sejahtera di Desa Silungkang Tigo; dan (3) penyajian hasil dalam bentuk *dashboard* operasional bagi perangkat desa. Hasil penelitian diharapkan dapat menjadi informasi pendukung bagi pemerintah desa dalam memahami kondisi penerima bantuan sosial secara lebih terstruktur, objektif, dan berbasis data.

II. METODE PENELITIAN

Penelitian ini menerapkan pendekatan data mining dengan teknik klusterisasi. Algoritma K-Means dipilih karena mampu mengelompokkan data berdasarkan kemiripan karakteristik, memiliki efisiensi komputasi yang tinggi, serta telah teruji dalam berbagai penelitian data mining [13]. Tahapan penelitian disusun mengikuti kerangka kerja CRISP-DM yang terdiri atas enam tahap, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [12]. Diagram alir tahapan penelitian tersebut ditunjukkan pada Gambar 1.



Gbr. 1 Diagram Alir CRISP-DM

A. Business Understanding

Tahap ini menetapkan konteks dan tujuan analitis penelitian. Permasalahan yang diangkat adalah ketimpangan informasi dalam penyaluran bantuan sosial di Desa Silungkang Tigo, di mana data penerima selama ini hanya dimanfaatkan sebagai catatan administratif sehingga karakteristik sosial-ekonomi penerima belum terpetakan secara komprehensif. Tujuan data mining yang ditetapkan adalah melakukan segmentasi penerima bantuan sosial berdasarkan kemiripan karakteristik sosial-ekonomi guna menghasilkan profil per kelompok yang dapat dimanfaatkan perangkat desa sebagai informasi pendukung dalam pengelolaan dan evaluasi program bantuan berbasis data.

B. Data Understanding

Tahap ini merupakan tahap yang dilakukan untuk memahami karakteristik awal data yang digunakan dalam penelitian. Pada tahap ini, peneliti mengidentifikasi sumber data, jenis data, atribut yang tersedia, serta struktur data yang akan dianalisis. Data yang digunakan dalam penelitian ini merupakan data penerima bantuan sosial di Desa Silungkang

Tigo yang memuat informasi terkait karakteristik sosial-ekonomi penerima bantuan.

Pada tahap ini dilakukan pemeriksaan awal terhadap data untuk mengetahui kelengkapan, konsistensi, tipe data, serta potensi permasalahan yang dapat memengaruhi proses analisis. Pemeriksaan data meliputi identifikasi jumlah data, atribut yang digunakan, nilai kosong, data duplikat, serta kesesuaian format data. Selain itu, dilakukan eksplorasi awal terhadap setiap atribut untuk memperoleh gambaran umum mengenai kondisi data sebelum masuk ke tahap prapemrosesan.

Tahap *Data Understanding* menjadi dasar penting dalam penelitian ini karena kualitas dan karakteristik data akan menentukan ketepatan proses segmentasi penerima bantuan sosial. Hasil dari tahap ini digunakan sebagai acuan untuk menentukan kebutuhan pembersihan data, pemilihan atribut, serta persiapan data sebelum algoritma K-Means diterapkan.

C. Data Preparation

Tahap preparasi data dilakukan untuk meningkatkan kualitas dataset sebelum proses klasterisasi menggunakan algoritma K-Means. Konsistensi kualitas data merupakan aspek yang krusial karena turut memengaruhi akurasi hasil pengelompokan. Pada tahap ini dilakukan empat proses utama, yaitu pembersihan data, transformasi, encoding variabel kategorikal, dan standarisasi data.

Pembersihan data (*data cleaning*) dilakukan dengan menstandarkan penulisan teks pada atribut kategorikal, memperbaiki ketidakkonsistenan penamaan jenis pekerjaan, serta memeriksa kewajaran nilai pada atribut numerik. Nilai penghasilan yang tidak valid, yaitu bernilai nol atau negatif, diperlakukan sebagai data hilang, sedangkan nilai jumlah tanggungan yang bernilai negatif dikoreksi menjadi nol.

Transformasi dilakukan dengan mengubah status penerimaan PKH dan BPNT menjadi nilai biner, serta mengelompokkan beragam jenis pekerjaan ke dalam tiga kategori utama agar lebih ringkas dan mudah dipahami.

Encoding variabel kategorikal dilakukan dengan menerapkan teknik *One-Hot Encoding* pada variabel pekerjaan hasil pengelompokan untuk menghindari terbentuknya hubungan ordinal antarkategori.

Standarisasi data dilakukan menggunakan *StandardScaler* sehingga seluruh atribut berada pada skala yang sebanding dan tidak ada atribut yang mendominasi pembentukan *centroid* pada algoritma K-Means.

D. Modeling

Pada tahap pemodelan, proses klasterisasi dilakukan menggunakan algoritma K-Means dengan menguji beberapa variasi jumlah klaster (K), yaitu K=2, K=3, dan K=4. Pengujian ini bertujuan untuk memperoleh pola pengelompokan yang paling representatif terhadap karakteristik sosial-ekonomi penerima bantuan sosial. Pemodelan diimplementasikan menggunakan pustaka *scikit-learn* dengan inisialisasi *centroid* `k-means++`, jumlah inisialisasi ulang sebanyak 50 (`n_init=50`), dan `random_state=42` untuk menjaga konsistensi serta reproduktibilitas hasil klasterisasi.

Algoritma K-Means bekerja dengan mengelompokkan data ke dalam sejumlah klaster berdasarkan kedekatan jaraknya terhadap pusat klaster (*centroid*) [12]. Ukuran jarak yang digunakan dalam penelitian ini adalah *Euclidean Distance* karena merupakan ukuran yang lazim diterapkan untuk menilai kemiripan data numerik dalam proses klasterisasi, sejalan dengan pendekatan yang digunakan oleh Dwiguna dan Bahtiar [14]. Seluruh atribut, termasuk fitur biner dan hasil *one-hot*, terlebih dahulu distandarasi agar berada pada skala yang sebanding sebelum jarak dihitung. Konsekuensi penggunaan jarak *Euclidean* pada data bertipe campuran ini dibahas sebagai keterbatasan pada bagian Saran. Tahapan proses klasterisasi menggunakan algoritma K-Means dijelaskan sebagai berikut:

1. Menentukan jumlah klaster (K).
2. Menginisialisasi *centroid* awal sebagai pusat klaster.
3. Menghitung jarak setiap data terhadap *centroid* menggunakan *Euclidean Distance*.
4. Mengelompokkan data ke dalam klaster menggunakan jarak terdekat.
5. Memperbarui *centroid* berdasarkan rata-rata anggota klaster.
6. Mengulangi proses hingga *centroid* stabil atau tidak mengalami perubahan yang signifikan.

Proses ini menghasilkan pengelompokan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik sosial-ekonomi.

E. Evaluation

Evaluasi klaster merupakan tahapan penting dalam proses klasterisasi untuk menilai tingkat separasi dan kedekatan antarklaster, sekaligus menentukan jumlah klaster yang paling layak diinterpretasikan. Pada penelitian ini digunakan dua metode evaluasi, yaitu *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI), dilengkapi pemeriksaan distribusi ukuran klaster, mengikuti pendekatan pada studi evaluasi model klasterisasi [15].

1. *Silhouette Coefficient* [15] digunakan untuk mengukur seberapa dekat suatu data dengan klaster asalnya dibandingkan dengan klaster lain. Nilai *Silhouette* berada pada rentang -1 hingga 1, dan nilai yang mendekati 1 menunjukkan bahwa data sangat mirip dengan anggota klaster asalnya serta terpisah dengan jelas dari klaster lain.
2. *Davies-Bouldin Index* (DBI) [16] digunakan untuk mengukur kualitas klasterisasi berdasarkan tingkat separasi dan kepadatan antarklaster. Nilai DBI yang semakin kecil menunjukkan kualitas klasterisasi yang semakin baik, karena setiap klaster terpisah dengan jelas dari klaster lain dan memiliki kepadatan internal yang tinggi.
3. Distribusi ukuran klaster diperiksa melalui jumlah anggota klaster terkecil (*minimum cluster size*) pada setiap konfigurasi K. Kriteria ini digunakan untuk menolak konfigurasi yang menghasilkan klaster degeneratif, yaitu klaster dengan jumlah anggota yang sangat sedikit (misalnya hanya satu data atau di bawah ambang minimum yang ditetapkan), karena klaster semacam itu tidak merepresentasikan kelompok yang bermakna.

Dengan demikian, pemilihan jumlah kluster tidak hanya bergantung pada nilai *Silhouette Coefficient* dan DBI, tetapi juga pada kelayakan distribusi anggota antarkluster.

Evaluasi dilakukan dengan membandingkan beberapa variasi jumlah kluster (K), lalu hasilnya digunakan untuk menentukan konfigurasi kluster yang paling representatif berdasarkan nilai *Silhouette Coefficient*, *Davies-Bouldin Index*, dan distribusi anggota kluster.

F. Deployment

Tahap *deployment* menyajikan hasil klusterisasi dalam bentuk yang dapat dimanfaatkan secara operasional. Pada penelitian ini, setiap rumah tangga dilabeli dengan kluster dan kategorinya, kemudian disusun profil sosial-ekonomi per kluster yang mencakup jumlah dan proporsi anggota, rata-rata penghasilan, cakupan bantuan (PKH dan BPNT), serta komposisi pekerjaan. Profil tersebut divisualisasikan melalui sebuah *dashboard* dan diekspor sebagai berkas tabel agar dapat ditindaklanjuti oleh perangkat desa sebagai informasi pendukung dalam pengelolaan dan evaluasi program bantuan sosial.

III. HASIL DAN PEMBAHASAN

Bagian ini menyajikan dan membahas hasil yang diperoleh dari proses klusterisasi penerima bantuan sosial.

A. Business Understanding

Pada tahap ini ditetapkan bahwa permasalahan utama penelitian adalah belum terpetakannya karakteristik sosial-ekonomi penerima bantuan sosial di Desa Silungkang Tigo, karena data penerima selama ini hanya dimanfaatkan sebagai catatan administratif. Berdasarkan kondisi tersebut, tujuan analitis yang ditetapkan adalah melakukan segmentasi penerima bantuan sosial berdasarkan kemiripan karakteristik sosial-ekonomi. Penetapan tujuan ini menjadi dasar bagi keseluruhan tahapan analisis berikutnya, dengan keluaran akhir berupa profil tiap kelompok penerima yang dapat dimanfaatkan perangkat desa sebagai informasi pendukung dalam pengelolaan dan evaluasi program bantuan.

B. Data Understanding

Tahap *Data Understanding* dilakukan untuk memperoleh gambaran awal mengenai karakteristik, kualitas, dan struktur data yang digunakan dalam penelitian. Dataset yang digunakan merupakan data penerima bantuan sosial di Desa Silungkang Tigo pada tahun 2023. Data tersebut mencakup 156 rumah tangga dan diperoleh dari instansi yang berwenang dalam pengelolaan data bantuan sosial di wilayah penelitian.

Atribut yang digunakan dalam proses klusterisasi meliputi desil kesejahteraan, penghasilan, jumlah tanggungan, jenis pekerjaan, status penerimaan Program Keluarga Harapan (PKH), dan status penerimaan Bantuan Pangan Non Tunai (BPNT). Atribut-atribut tersebut dipilih karena memiliki keterkaitan langsung dengan kondisi sosial-ekonomi rumah tangga penerima bantuan. Sementara itu, atribut administratif yang tidak relevan dengan proses pengelompokan tidak

digunakan agar proses klusterisasi lebih terfokus pada karakteristik sosial-ekonomi penerima bantuan.

Hasil pemeriksaan awal menunjukkan bahwa dataset relatif lengkap dan dapat digunakan untuk tahap analisis berikutnya. Pada pemeriksaan nilai kosong, tidak ditemukan atribut dengan data yang benar-benar hilang; hanya terdapat sejumlah nilai penghasilan yang tidak valid (bernilai nol) yang ditandai untuk ditangani pada tahap *Data Preparation*, sehingga tidak ada baris yang harus dibuang dan jumlah data tetap utuh sebanyak 156 rumah tangga. Pada pemeriksaan duplikasi, ditemukan 28 baris data yang memiliki nilai identik pada semua atribut. Meskipun demikian, baris-baris tersebut tetap dipertahankan karena setiap entri merepresentasikan kepala keluarga yang berbeda. Kesamaan nilai antarbaris dipandang wajar dalam konteks data sosial-ekonomi desa yang relatif homogen, sehingga jumlah data yang digunakan dalam penelitian tetap sebanyak 156 rumah tangga.

Berdasarkan eksplorasi awal terhadap atribut numerik, diketahui bahwa rata-rata penghasilan rumah tangga sebesar Rp574.487, dengan rentang penghasilan antara Rp0 dan Rp1.500.000. Temuan ini menunjukkan adanya variasi tingkat penghasilan antar rumah tangga penerima bantuan sosial. Variasi tersebut menjadi informasi penting karena penghasilan merupakan salah satu indikator sosial-ekonomi yang relevan dalam proses segmentasi penerima bantuan.

Pada atribut status penerimaan bantuan, ditemukan bahwa variasi status penerimaan PKH sangat rendah. Sebanyak 155 dari 156 rumah tangga, atau 99,36%, berstatus sebagai penerima PKH. Kondisi ini menunjukkan bahwa atribut PKH hampir konstan, sehingga kurang informatif untuk membedakan karakteristik antarkluster. Sebaliknya, status penerimaan BPNT menunjukkan variasi yang lebih besar, yaitu sebanyak 101 dari 156 rumah tangga atau 64,74% berstatus sebagai penerima BPNT. Dengan demikian, atribut BPNT berpotensi memberikan kontribusi yang lebih baik dalam membedakan kelompok rumah tangga dibandingkan dengan atribut PKH.

Secara umum, hasil *Data Understanding* menunjukkan bahwa dataset memiliki struktur dan kualitas yang memadai untuk dilanjutkan ke tahap *Data Preparation*. Namun, fitur dengan varians yang sangat rendah, khususnya status penerimaan PKH, perlu diperhatikan karena dapat memengaruhi kualitas pembentukan kluster. Implikasi rendahnya variasi atribut PKH terhadap hasil klusterisasi dibahas lebih lanjut pada bagian hasil analisis dan keterbatasan penelitian.

C. Data Preparation

Pembersihan data (*data cleaning*) dilakukan dengan menstandarkan penulisan teks pada atribut kategorikal, memperbaiki ketidakkonsistenan penamaan jenis pekerjaan, serta memeriksa kewajaran nilai pada atribut numerik. Nilai penghasilan yang tidak valid (bernilai nol atau negatif) diperlakukan sebagai data hilang, dan nilai jumlah tanggungan yang negatif dikoreksi menjadi nol. Setelah proses tersebut, tidak terdapat baris yang harus dibuang sehingga jumlah data tetap 156 dan tidak ditemukan duplikasi. Selanjutnya, status

penerimaan PKH dan BPNT dikodekan menjadi nilai biner (1 untuk penerima dan 0 untuk bukan penerima). Transformasi berikutnya diterapkan pada variabel pekerjaan dengan mengelompokkan beragam jenis profesi ke dalam tiga kategori utama, yaitu (1) produksi, pertanian, dan buruh, (2) jasa, transportasi, dan perdagangan, serta (3) tidak bekerja/rumah tangga. Pengelompokan ini bertujuan agar hasil analisis menjadi lebih ringkas dan mudah diinterpretasikan.

Variabel kategorikal PEKERJAAN_3 selanjutnya ditransformasikan menggunakan teknik *One-Hot Encoding* menjadi variabel biner untuk menghindari terbentuknya hubungan ordinal antarkategori. Setelah itu, seluruh atribut yang digunakan dalam proses klasterisasi distandarisasi menggunakan *StandardScaler* agar berada pada skala yang sebanding, sehingga tidak ada atribut tertentu yang mendominasi pembentukan *centroid* pada algoritma K-Means. Melalui proses *One-Hot Encoding*, variabel pekerjaan menghasilkan tiga fitur baru, sehingga total dimensi fitur yang digunakan dalam proses klasterisasi menjadi 8 fitur.

Struktur akhir dataset yang digunakan dalam penelitian terdiri atas atribut sosial-ekonomi awal, yaitu desil kesejahteraan, jumlah tanggungan, status penerimaan PKH dan BPNT, jenis pekerjaan, dan penghasilan, yang dilengkapi dengan hasil transformasi jenis pekerjaan ke dalam tiga kategori (PEKERJAAN_3). Pengelompokan pekerjaan menghasilkan kategori produksi/pertanian/buruh sebanyak 83 data, jasa/transportasi/perdagangan sebanyak 37 data, dan tidak bekerja/rumah tangga sebanyak 36 data. Setelah variabel PEKERJAAN_3 ditransformasikan melalui *One-Hot Encoding* dan seluruh atribut numerik distandarisasi menggunakan *StandardScaler*, dataset memuat delapan fitur yang siap digunakan pada proses klasterisasi, sebagaimana cuplikannya dapat dilihat pada Gambar 2.

DESIL	JUMLAH TANGGUNGAN	PKH	BPNT	PEKERJAAN	PENGHASILAN	PEKERJAAN_3	
0	2	8	1	1	BERTENUN	800000	PRODUKSI_PERTANIAN_BURUH
1	3	5	1	1	BURUH HARIAN LEPAS	680000	PRODUKSI_PERTANIAN_BURUH
2	3	1	1	1	PETANI	600000	PRODUKSI_PERTANIAN_BURUH
3	2	0	1	1	PETANI	900000	PRODUKSI_PERTANIAN_BURUH
4	5	2	1	1	BURUH HARIAN LEPAS	660000	PRODUKSI_PERTANIAN_BURUH
5	2	2	1	1	MENGURUS RUMAH TANGGA	260000	TIDAK_BEKERJA_RUMAH_TANGGA
6	4	0	1	1	BERTENUN	900000	PRODUKSI_PERTANIAN_BURUH
7	2	4	1	1	BENGGEL LAS	780000	PRODUKSI_PERTANIAN_BURUH
8	5	2	1	1	OJEK	300000	JASA_TRANSPORT_PERDAGANGAN
9	1	1	1	1	TIDAK BEKERJA	0	TIDAK_BEKERJA_RUMAH_TANGGA

Gbr 2. Cuplikan Dataset yang digunakan dalam Model

D. Modeling

Pada tahap pemodelan, algoritma K-Means diterapkan pada delapan fitur sosial-ekonomi hasil tahap *Data Preparation* untuk membentuk klaster rumah tangga penerima bantuan sosial. Pemodelan dijalankan menggunakan pustaka *scikit-learn* dengan inisialisasi *centroid* *k-means++*, jumlah inisialisasi ulang sebanyak 50 (*n_init=50*), dan *random_state=42* guna menjaga konsistensi serta reproduktibilitas hasil. Karena jumlah klaster optimal belum diketahui di awal, pemodelan tidak dilakukan untuk satu nilai K saja, melainkan diulang untuk beberapa konfigurasi jumlah klaster, yaitu K=2, K=3, dan K=4. Setiap konfigurasi menghasilkan satu model K-Means beserta label klaster untuk seluruh 156 rumah tangga, yang kemudian dinilai kualitasnya

pada Tahap Evaluation. Sesuai sifat iteratif CRISP-DM, model untuk masing-masing nilai K dibangun terlebih dahulu, lalu hasilnya dibandingkan pada tahap evaluasi sebelum konfigurasi akhir ditetapkan.

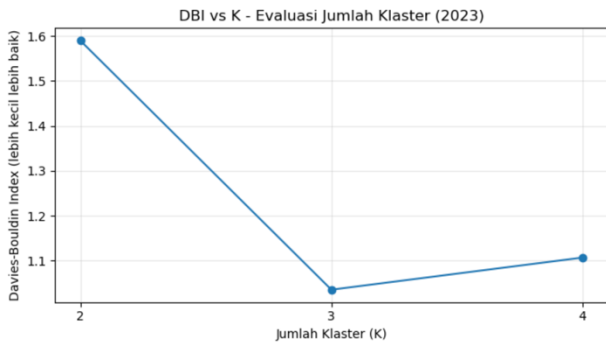
E. Evaluation

Pengujian dilakukan untuk menentukan jumlah klaster (K) yang paling representatif dalam proses klasterisasi penerima bantuan sosial, dengan kualitas klaster dinilai menggunakan *Silhouette Coefficient* dan DBI. Pengujian dibatasi pada rentang K=2 hingga K=4 agar diperoleh konfigurasi klaster yang representatif sekaligus menghindari terbentuknya klaster dengan jumlah anggota yang terlalu sedikit pada dataset penelitian. Hasil evaluasi untuk setiap jumlah klaster ditunjukkan pada Tabel 1.

TABEL I
HASIL EVALUASI JUMLAH KLASTER

K	DBI	Silhouette	Minimum Cluster Size
2	1,5897	0,2848	44
3	1,0352	0,2727	1
4	1,1067	0,2591	1

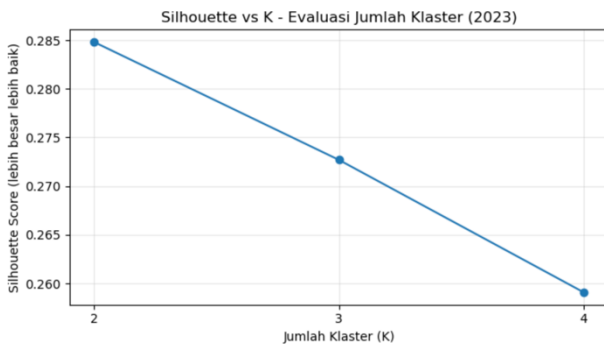
Berdasarkan Tabel 1, nilai DBI terendah diperoleh pada K=3 sebesar 1,0352, yang secara teoritis menunjukkan kualitas pemisahan klaster yang lebih baik. Namun, nilai *Silhouette* pada K=2 (0,2848) dan K=3 (0,2727) hampir setara, dengan selisih hanya 0,012. Kedua metrik tersebut tidak konsisten dalam menunjuk satu konfigurasi terbaik, dan keduanya berada pada rentang yang tergolong lemah. Pertimbangan yang lebih menentukan adalah bahwa konfigurasi K=3 dan K=4 menghasilkan klaster dengan jumlah anggota minimum hanya satu data (*singleton cluster*). *Singleton* semacam ini bukan kelompok yang bermakna, melainkan sinyal adanya satu rumah tangga pencilon (*outlier*) yang terpisah dari keseluruhan data. Penelusuran data menunjukkan bahwa fitur status PKH bersifat nyaris konstan, yaitu 155 dari 156 rumah tangga (99,36%) berstatus penerima PKH. Akibatnya, setelah standarisasi, satu rumah tangga tanpa PKH memperoleh nilai standar yang ekstrem dan berpotensi kuat menjadi penyebab terbentuknya *singleton* pada K=3. Dengan demikian, K=2 dipilih bukan semata karena kemudahan interpretasi, melainkan karena merupakan satu-satunya konfigurasi yang tidak menghasilkan klaster degeneratif (*singleton*) sekaligus menjaga distribusi anggota klaster tetap proporsional. Grafik DBI ditunjukkan pada Gambar 3.



Gbr 3. Grafik Davies-Bouldin Index

Gambar 3 memperlihatkan bahwa nilai DBI mencapai titik terendah pada K=3, yang secara teoritis mengindikasikan kualitas pemisahan kluster yang paling baik. Akan tetapi, sebagaimana dibahas sebelumnya, konfigurasi K=3 justru menghasilkan kluster dengan jumlah anggota yang sangat kecil sehingga kurang representatif untuk penelitian ini. Hal ini menegaskan bahwa pemilihan jumlah kluster tidak dapat hanya bergantung pada nilai DBI, tetapi juga harus mempertimbangkan distribusi anggota kluster serta kemudahan interpretasi hasilnya.

Selain DBI, kualitas kluster juga dinilai melalui nilai *Silhouette Coefficient* untuk setiap jumlah kluster yang diuji. Grafik *Silhouette Coefficient* ditunjukkan pada Gambar 4.



Gbr 4. Grafik Silhouette Coefficient

Berdasarkan Gambar 4, nilai *Silhouette* tertinggi diperoleh pada K=2 sebesar 0,2848. Nilai ini menunjukkan bahwa struktur kluster yang terbentuk masih tergolong lemah karena karakteristik antarkluster belum terpisah sepenuhnya. Kondisi semacam ini umum dijumpai pada data sosial-ekonomi yang memiliki tingkat heterogenitas dan tumpang tindih karakteristik yang cukup tinggi. Dengan mempertimbangkan hal tersebut, konfigurasi K=2 tetap dipilih karena memberikan nilai *Silhouette* tertinggi di antara seluruh konfigurasi, sekaligus menghasilkan distribusi anggota kluster yang lebih proporsional dan lebih mudah diinterpretasikan.

F. Pembahasan dan Deployment

Proses klusterisasi menggunakan K-Means menghasilkan dua kluster yang dibentuk berdasarkan karakteristik sosial-ekonomi. Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa K-Means mampu menghasilkan

segmentasi data sosial secara efektif [17], [18]. Distribusi jumlah anggota pada masing-masing kluster ditunjukkan pada Tabel 2.

TABEL II
DISTRIBUSI ANGGOTA KLUSTER

Kluster	Kategori	Jumlah	Persentase (%)
1	Kelompok Rentan	112	71,79
2	Kelompok Relatif Lebih Sejahtera	44	28,21
Total		156	100

Berdasarkan Tabel 2, kluster pertama beranggotakan 112 data atau 71,79% dari total penerima bantuan sosial, sedangkan kluster kedua beranggotakan 44 data atau 28,21% dari total data.

Komposisi tersebut menunjukkan bahwa sebagian besar penerima bantuan sosial terhimpun pada kluster pertama, yang mengindikasikan bahwa mayoritas data memiliki karakteristik sosial-ekonomi yang relatif serupa dan cenderung rentan.

Perbedaan jumlah anggota antarkluster mencerminkan adanya variasi kondisi sosial-ekonomi di antara penerima bantuan sosial. Dengan demikian, hasil klusterisasi pada penelitian ini memberikan gambaran awal mengenai pola pengelompokan penerima bantuan sosial berdasarkan kemiripan karakteristik sosial-ekonominya.

Untuk memperjelas karakterisasi sosial-ekonomi setiap kluster, Tabel 3 merangkum profil kedua kluster berdasarkan jumlah anggota, proporsi, dan rata-rata penghasilan seluruh anggota, sebagaimana dihitung pada *dashboard* (Gambar 5). Rata-rata penghasilan seluruh anggota Kluster 1 (Rentan) sebesar Rp422.143, sedangkan Kluster 2 (Relatif Lebih Sejahtera) sebesar Rp962.273. Perlu dicatat bahwa nilai tersebut merupakan rata-rata seluruh anggota kluster, sehingga berbeda dari angka pada Tabel 5 yang hanya menghitung rata-rata penghasilan pada kategori pekerjaan dominan di setiap kluster.

TABEL III
PROFIL SOSIAL-EKONOMI SETIAP KLUSTER

Kluster	Kategori	Jumlah	Persentase (%)	Rata-rata Penghasilan (Rp)
1	Kelompok Rentan	112	71,79	422.143
2	Kelompok Relatif Lebih Sejahtera	44	28,21	962.273

*Catatan: Rata-rata penghasilan dihitung untuk seluruh anggota kluster (sumber: *dashboard* pada Gambar 5). Secara keseluruhan, rata-rata penghasilan dataset sebesar Rp574.487, dengan komposisi penerima PKH sebanyak 99,36% (155 dari 156 rumah tangga) dan penerima BPNT sebanyak 64,74% (101 dari 156 rumah tangga).

Analisis karakteristik pekerjaan menunjukkan adanya perbedaan kondisi sosial-ekonomi antarklaster. Klaster rentan didominasi oleh jenis pekerjaan berpendapatan tidak stabil, seperti buruh dan pekerjaan informal, yang mencerminkan keterbatasan akses ekonomi, sebagaimana juga ditemukan dalam penelitian sebelumnya. Rincian karakteristik pekerjaan pada setiap klaster ditunjukkan pada Tabel 4.

TABEL IV
KATEGORI PEKERJAAN PADA SETIAP KLASTER

Kategori Pekerjaan	Klaster Rentan (n=112)	Klaster Relatif Sejahtera (n=44)
Produksi/Pertanian/Buruh	57 (50,89%)	26 (59,09%)
Jasa/Transport/Perdagangan	19 (16,96%)	18 (40,91%)
Tidak Bekerja/Rumah Tangga	36 (32,14%)	0 (0,00%)

Berdasarkan Tabel 4, klaster rentan didominasi oleh kategori pekerjaan produksi, pertanian, dan buruh sebesar 50,89%, diikuti oleh kategori tidak bekerja atau rumah tangga sebesar 32,14%. Komposisi ini menunjukkan bahwa sebagian besar anggota klaster rentan memiliki aktivitas ekonomi yang terbatas atau bergantung pada pekerjaan informal dengan tingkat pendapatan yang cenderung tidak stabil. Keberadaan kelompok yang tidak bekerja semakin mengindikasikan keterbatasan rumah tangga dalam memperoleh sumber pendapatan secara mandiri, sehingga berpotensi meningkatkan ketergantungan pada program bantuan sosial.

Sementara itu, klaster relatif lebih sejahtera didominasi oleh kategori produksi, pertanian, dan buruh sebesar 59,09% serta kategori jasa, transportasi, dan perdagangan sebesar 40,91%, tanpa adanya anggota berkategori tidak bekerja atau rumah tangga. Komposisi tersebut menunjukkan bahwa anggota klaster ini memiliki partisipasi ekonomi yang lebih tinggi, sehingga peluang memperoleh pendapatan secara berkelanjutan juga lebih besar dibandingkan dengan klaster rentan.

Temuan ini menunjukkan bahwa jenis pekerjaan menjadi salah satu faktor penting yang membedakan karakteristik sosial-ekonomi antarklaster: rumah tangga dengan keterlibatan ekonomi yang lebih rendah cenderung terkonsentrasi pada kelompok rentan, sedangkan rumah tangga dengan aktivitas ekonomi yang lebih aktif cenderung berada pada kelompok yang relatif lebih sejahtera.

Analisis terhadap pekerjaan dominan dan rata-rata penghasilan dilakukan untuk mempertajam karakterisasi sosial-ekonomi masing-masing klaster. Hasilnya ditunjukkan pada Tabel 5.

TABEL V
PEKERJAAN DOMINAN DAN RATA-RATA PENGHASILAN PER KLASTER

Klaster	Pekerjaan Dominan	Jumlah	Persentase (%)	Rata-rata Penghasilan (Rp)
Rentan	Buruh Harian Lepas	18	31,58	554.444
Relatif Lebih Sejahtera	Bertenun	15	57,69	1.176.667

*Catatan: Persentase pekerjaan dominan dihitung dari jumlah anggota pada kategori pekerjaan terkait terhadap total anggota di setiap klaster.

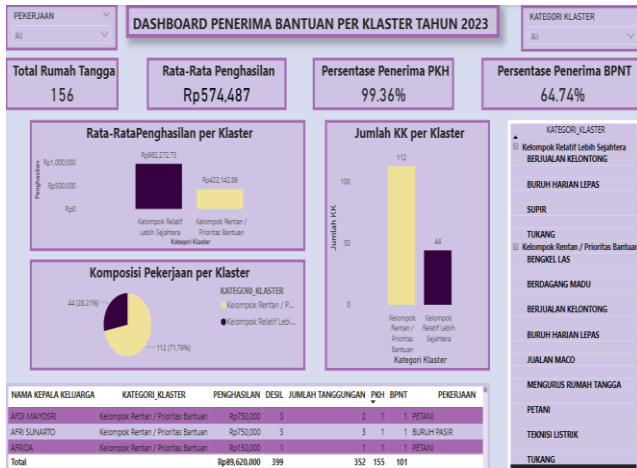
Berdasarkan Tabel 5, klaster rentan didominasi oleh pekerjaan buruh harian lepas sebanyak 18 data atau 31,58% dari kategori pekerjaan produksi, pertanian, dan buruh pada klaster tersebut. Terhadap total anggota klaster rentan (n=112), jumlah ini setara dengan 16,07%. Rata-rata penghasilan pada subkelompok pekerjaan dominan ini sebesar Rp554.444. Perlu ditegaskan bahwa angka tersebut hanya mencerminkan rata-rata penghasilan anggota dengan pekerjaan dominan, bukan rata-rata seluruh anggota klaster yang dilaporkan pada Tabel 3 (Rp422.143 untuk klaster rentan). Karakteristik ini menunjukkan bahwa sebagian besar anggota klaster rentan bergantung pada pekerjaan informal yang umumnya berpendapatan rendah dan tidak menentu, sehingga kemampuan rumah tangga dalam memenuhi kebutuhan ekonomi terbatas dan lebih berisiko mengalami kerentanan sosial-ekonomi.

Sementara itu, klaster relatif lebih sejahtera didominasi oleh pekerjaan bertenun sebanyak 15 data atau 57,69% dari kategori pekerjaan dominan pada klaster tersebut. Terhadap total anggota klaster relatif lebih sejahtera (n=44), jumlah ini setara dengan 34,09%. Rata-rata penghasilan pada subkelompok pekerjaan dominan ini mencapai Rp1.176.667, lebih tinggi dibandingkan klaster rentan. Sebagaimana pada klaster rentan, nilai tersebut merupakan rata-rata penghasilan pada pekerjaan dominan saja, bukan rata-rata seluruh anggota klaster yang dilaporkan pada Tabel 3 (Rp962.273 untuk klaster relatif lebih sejahtera). Hasil ini menunjukkan bahwa pekerjaan bertenun di wilayah penelitian memberikan kontribusi yang lebih baik terhadap pembentukan pendapatan rumah tangga, sehingga anggota klaster ini berada pada kondisi ekonomi yang relatif lebih stabil.

Perbedaan pekerjaan dominan dan rata-rata penghasilan antara kedua klaster menunjukkan bahwa tingkat pendapatan berperan penting dalam pembentukan kelompok penerima bantuan sosial. Semakin rendah pendapatan rumah tangga, semakin besar kecenderungannya berada pada kelompok rentan; sebaliknya, rumah tangga dengan sumber pendapatan yang relatif lebih baik cenderung terkonsentrasi pada kelompok relatif lebih sejahtera.

Visualisasi *dashboard* digunakan untuk memperjelas hasil pengelompokan penerima bantuan sosial berdasarkan

karakteristik sosial-ekonomi. *Dashboard* ini menampilkan distribusi anggota kluster, rata-rata penghasilan, persentase penerima bantuan sosial, serta komposisi pekerjaan pada masing-masing kluster, sebagaimana ditunjukkan pada Gbr 5.



Gbr 5. Dashboard Klasterisasi Penerima Bantuan Sosial

Berdasarkan Gambar 5, kluster rentan memiliki jumlah anggota yang lebih banyak sekaligus rata-rata penghasilan yang lebih rendah dibandingkan dengan kluster relatif lebih sejahtera.

Dashboard juga memperlihatkan perbedaan karakteristik pekerjaan antarkluster. Kluster relatif lebih sejahtera didominasi oleh pekerjaan kategori produksi, pertanian, dan buruh, dengan pekerjaan bertenun sebagai subkelompok yang menonjol, dan memiliki tingkat penghasilan yang relatif lebih tinggi, sedangkan kluster rentan didominasi oleh pekerjaan informal dengan tingkat penghasilan yang lebih rendah.

Penggunaan dashboard dalam penelitian ini membantu interpretasi hasil klasterisasi secara lebih visual dan informatif. Dengan demikian, hasil pengelompokan penerima bantuan sosial lebih mudah dipahami berdasarkan perbedaan karakteristik sosial-ekonomi setiap kluster.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Algoritma K-Means dapat digunakan untuk mengelompokkan penerima bantuan sosial berdasarkan karakteristik sosial-ekonomi, dan proses klasterisasi menghasilkan dua kluster utama, yaitu kelompok rentan dan kelompok yang relatif lebih sejahtera. Mengingat struktur kluster yang masih tergolong lemah (nilai Silhouette 0,2848), hasil ini perlu dimaknai sebagai gambaran awal yang bersifat eksploratif, bukan sebagai penetapan prioritas penyaluran bantuan yang bersifat final. Hal ini sejalan dengan hasil evaluasi menggunakan DBI dan *Silhouette Coefficient*, yang menunjukkan bahwa kedua metrik tidak konsisten dalam menunjuk satu konfigurasi terbaik dan keduanya berada pada rentang yang tergolong lemah. Konfigurasi K=2 dipilih karena merupakan satu-satunya konfigurasi yang tidak menghasilkan kluster degeneratif (*singleton*), sedangkan K=3 dan K=4

membentuk *singleton* yang ditengarai berasal dari satu rumah tangga pencila akibat fitur PKH yang nyaris konstan. Oleh karena itu, K=2 dipandang sebagai konfigurasi yang paling layak diinterpretasikan, bukan sebagai konfigurasi optimal yang final. Dari kedua kluster tersebut, karakteristik pekerjaan dan tingkat penghasilan menjadi faktor utama yang membedakannya: kelompok rentan didominasi oleh pekerjaan informal dengan rata-rata penghasilan yang lebih rendah, sedangkan kelompok relatif lebih sejahtera memiliki tingkat partisipasi ekonomi dan penghasilan yang lebih tinggi. Dengan demikian, analisis klasterisasi mampu mengidentifikasi variasi kondisi sosial-ekonomi penerima bantuan sosial secara lebih terstruktur. Visualisasi *dashboard* klasterisasi turut memperjelas hasil pengelompokan data secara lebih informatif dan mudah dipahami, sehingga hasil klasterisasi pada penelitian ini dapat menjadi gambaran awal untuk memahami karakteristik sosial-ekonomi penerima bantuan sosial berdasarkan kemiripan data yang dimiliki.

B. Saran

Penelitian ini memiliki sejumlah keterbatasan yang perlu disampaikan secara terbuka. Pertama, jumlah data relatif terbatas dan hanya berasal dari satu wilayah penelitian, sehingga hasil yang diperoleh belum dapat digeneralisasi ke wilayah lain. Kedua, penentuan jumlah kluster belum melibatkan *Elbow method* yang lazim digunakan sebagai pelengkap metrik validitas internal, sehingga pemilihan K sepenuhnya bersandar pada DBI, Silhouette Coefficient, dan distribusi ukuran kluster. Ketiga, penelitian hanya menggunakan algoritma K-Means tanpa algoritma pembandingan (misalnya *K-Medoids* atau *Hierarchical Clustering*), sehingga kestabilan pemilihan K=2 belum divalidasi secara silang. Keempat, K-Means menggunakan jarak *Euclidean* pada data bertipe campuran - desil (ordinal), penghasilan dan jumlah tanggungan (kontinu/cacah), serta status PKH/BPNT dan hasil one-hot pekerjaan (biner) - sehingga setelah standarisasi bobot antartipe fitur menjadi kurang transparan; pendekatan seperti *Gower distance* atau *K-Prototypes* berpotensi lebih sesuai, namun belum diujikan pada penelitian ini. Kelima, fitur PKH yang nyaris konstan (99,36%) tidak dibuang dan tidak dilakukan *re-run* tanpa fitur tersebut, sehingga pengaruhnya terhadap pembentukan *singleton* pada K=3 masih bersifat dugaan yang beralasan dan belum diverifikasi secara eksperimental. Keenam, belum dilakukan uji stabilitas kluster antarinisialisasi (misalnya melalui *Adjusted Rand Index*) maupun analisis sensitivitas terhadap transformasi penghasilan yang distribusinya cenderung *skewed*. Berdasarkan keterbatasan tersebut, beberapa saran untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menambahkan variabel sosial-ekonomi lain, seperti tingkat pendidikan, kondisi tempat tinggal, dan kepemilikan aset, agar hasil klasterisasi menjadi lebih komprehensif.
2. Penelitian selanjutnya dapat membandingkan K-Means dengan metode *clustering* lain seperti Fuzzy C-Means, K-Medoids, atau *Hierarchical Clustering* untuk memvalidasi silang kestabilan pemilihan K=2.

3. Penentuan jumlah kluster sebaiknya dilengkapi dengan *Elbow method* sebagai pelengkap DBI dan Silhouette Coefficient.
 4. Perlakuan terhadap fitur bertipe campuran perlu dievaluasi, misalnya dengan menerapkan *Gower distance* atau *K-Prototypes* agar bobot antartipe fitur lebih transparan.
 5. Fitur yang nyaris konstan seperti status PKH sebaiknya dibuang, kemudian dilakukan *re-run* untuk menguji apakah singleton pada $K=3$ hilang.
 6. Uji stabilitas kluster antarinisialisasi (misalnya melalui *Adjusted Rand Index*) dan transformasi penghasilan (*log-transform*) perlu ditambahkan untuk menilai ketahanan hasil.
 7. Bagi instansi terkait, hasil klasterisasi dan visualisasi *dashboard* pada penelitian ini dapat digunakan sebagai informasi pendukung untuk memahami karakteristik penerima bantuan sosial secara lebih terstruktur. Namun, hasil pengelompokan tetap perlu dikombinasikan dengan verifikasi lapangan agar proses pengambilan keputusan menjadi lebih tepat dan objektif.
- DAFTAR PUSTAKA**
- [1] W. Sri Wahyuni and F. Prima Putra, "Analisis Karakteristik Rumah Tangga Miskin Penerima Program Keluarga Harapan dan Penerima Keluarga Harapan Sekaligus Bantuan Pangan Non-Tunai di Indonesia," *Jurnal Informatika Ekonomi Bisnis*, pp. 141–147, 2025, doi: 10.37034/infec.v7i2.1112.
 - [2] D. Widyarningsih, Ruhmaniyati, and N. Toyamah, "Mendorong Pemutakhiran Berkelanjutan terhadap Data Terpadu Kesejahteraan Sosial," Jan. 2022.
 - [3] D. Coady, M. Grosh, and J. Hoddinott, "Targeting Outcomes Redux," *World Bank Res. Obs.*, vol. 19, no. 1, pp. 61–85, 2004, doi: 10.1093/wbro/lkh016.
 - [4] A. Banerjee, R. Hanna, B. A. Olken, and S. Sumarto, "The (lack of) distortionary effects of proxy-means tests: Results from a nationwide experiment in Indonesia," *Journal of Public Economics Plus*, vol. 1, p. 100001, 2020, doi: <https://doi.org/10.1016/j.pubecp.2020.100001>.
 - [5] J. Osida Onunga, S. Anyango Onditi, and P. Onyango Odongo, "Health Information Management System Data Utilization in Strategic Management Decision Making in Hospitals: Systemic Review," *International Journal of Research and Scientific Innovation (IJRSI)*, vol. 12, no. 15, pp. 704–710, 2025, doi: 10.51244/IJRSI.
 - [6] E. Rosdianti and C. A. Sugianto, "Analisis Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) dengan Menggunakan Algoritma K-Means (Studi Kasus Politeknik TEDC Bandung)," *Journal Informatics and Electronics Engineering (JIEE)ERING*, vol. 03, no. 01, pp. 25–32, 2023.
 - [7] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *JANAPATI*, vol. 11, no. 2, pp. 156–165, 2022.
 - [8] M. Iqbal, Syaripuddin, and Moh. Nurul Huda, "Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor," *BASIS Jurnal Ilmiah Matematika*, vol. 2, no. 1, pp. 47–56, 2023.
 - [9] M. Orisa, "Optimasi Cluster pada Algoritma K-Means," in *Seniati 2022*, ITN Malang, 2022, p. 2022.
 - [10] I. Kamila and U. Khairunnisa, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi*, vol. 5, no. 1, pp. 119–125, 2019.
 - [11] D. Novaliendry, T. Wibowo, N. Ardi, T. Evi, and D. Admojo, "Optimizing Patient Medical Records Grouping through Data Mining and K-Means Clustering Algorithm: A Case Study at RSUD Mohammad Natsir Solok," *International journal of online and biomedical engineering*, vol. 19, no. 12, pp. 144–155, 2023, doi: 10.3991/ijoe.v19i12.42147.
 - [12] T. N. Muthmainnah, A. Voutama, and A. A. Hendriadi, "Klasterisasi Penerima Bantuan Langsung Tunai (BLT) Dana Desa Menggunakan Algoritma K-Means," *Jurnal Mahasiswa Teknik Informatika (JATI)*, vol. 8, no. 3, pp. 3916–3919, 2024.
 - [13] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.
 - [14] E. Dwiguna and A. Bahtiar, "Penerapan Data Mining Untuk Menentukan Penerima Bantuan BLT Menggunakan Metode Clustering K-Means Pada Desa Pamulihan," *Jurnal Mahasiswa Teknik Informatika (JATI)*, vol. 8, no. 2, pp. 1382–1388, 2024.
 - [15] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
 - [16] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, 1979, doi: 10.1109/TPAMI.1979.4766909.
 - [17] H. Rhomadhona, W. Kusriani, W. Aprianti, and J. Permadi, "Implementation of K-Means Clustering for Social Assistance Recipients with Silhouette Score Evaluation," *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 1, pp. 136–143, 2025, doi: 10.47709/brilliance.v5i1.5900.
 - N. Suharyanti, R. Rusdiansyah, H. Supendar, and T. Tuslaela, "Cluster Analysis of Food Social Assistance in DKI Jakarta: K-Means Approach to Identify Expenditure Patterns and Beneficiaries," *sinkron*, vol. 8, no. 4, pp. 2439–2446, 2024, doi: 10.33395/sinkron.v8i4.14095.