



Clustering Data Penduduk Menggunakan Algoritma K-Means

TomI Ikhsan, Elin Haerani* , Fitri Wulandari, Fadhilah Syafria

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru , Indonesia

Email: ¹111950111745@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin-suska.ac.id, ³fitri_wulandari@yahoo.com,

⁴fadhilah.syafria@uin-suska.ac.id,

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Ketidakmerataan ekonomi masih menjadi faktor krusial yang di hadapi Indonesia pada saat ini, mulai dari kota-kota besar sampai ke pelosok desa ketidakmerataan ekonomi masih menjadi masalah utama. Tidak terkecuali di desa bina baru, desa yang dihuni sebanyak 5.760 jiwa dengan total 1.742 keluarga, yang tersebar dalam 30 Rukun Tetangga (RT) dan 8 Rukun Warga (RW). Berbagai upaya di lakukan untuk mengatasi masalah Ketidak merataan ekonomi, salah satunya adalah dengan menyalurkan bantuan atau pemberian kebijakan yang tepat sasaran. Salah satu langkah untuk mengatasi permasalahan tersebut adalah dengan melakukan pengelompokan data penduduk di desa Bina Baru dengan menggunakan metode K-Means Clustering yang bertujuan untuk mengetahui tingkat ekonomi keluarga di wilayah tersebut, sehingga pemerintah daerah dapat lebih akurat untuk mengambil kebijakan tentang masalah ketidakmerataan ekonomi yang terjadi. Data yang digunakan berasal dari kuesioner sebanyak 1.005 data keluarga dengan 64 atribut serta 1.005 data individu dengan 84 atribut. Penerapan algoritma k-means dilakukan menggunakan python, juga menggunakan DBI (Davies-Bouldin Index) untuk menentukan nilai k yang optimum. Pada penelitian ini nilai k yang optimal adalah 3 cluster. Berdasarkan pengujian didapat bahwa Cluster 0 mewakili rumah tangga dengan ekonomi menengah, cluster 1 mewakili kelompok dengan kondisi ekonomi lebih baik dan Cluster 2 adalah kelompok rumah tangga dengan kondisi ekonomi rendah. Dengan melakukan pengelompokan ekonomi penduduk diharapkan dapat membantu pemangku jabatan untuk memberikan kebijakan yang tepat sasaran.

Kata Kunci: Ketidakmerataan Ekonomi; K-Means Clustering; Pengelompokan Data; Davies-Bouldin Index (DBI); Kebijakan Tepat Sasaran

Abstract—Economic inequality is still a crucial factor facing Indonesia today, from big cities to remote villages, economic inequality is still a major problem. Bina Baru Village is no exception, a village inhabited by 5,760 people with a total of 1,742 families, spread across 30 neighbourhood associations (RT) and 8 community associations (RW). Various efforts are made to overcome the problem of economic inequality, one of which is by channeling assistance or providing policies that are right on target. One of the steps to overcome this problem is to group population data in Bina Baru village using the K-Means Clustering method which aims to determine the economic level of families in the region, so that local governments can more accurately make policies on the problem of economic inequality that occurs. The data used comes from a questionnaire of 1,005 family data with 64 attributes and 1,005 individual data with 84 attributes. The application of the k-means algorithm is carried out using python, also using DBI (Davies-Bouldin Index) to determine the optimum k value. In this study, the optimal k value is 3 clusters. Based on testing, it is found that Cluster 0 represents households with medium economic conditions, cluster 1 represents groups with better economic conditions and Cluster 2 is a group of households with low economic conditions. By clustering the population's economy, it is hoped that it can help stakeholders to provide targeted policies.

Keywords: Economic Inequality; K-Means Clustering; Data Clustering; Davies-Bouldin Index (DBI); Targeted Policies

1. PENDAHULUAN

Ketidakmerataan ekonomi masih menjadi faktor yang menakutkan bagi sebagian besar masyarakat dunia dan tidak terkecuali di Indonesia. Meskipun telah banyak upaya dilakukan untuk mendorong pertumbuhan ekonomi, kenyataannya hasil pertumbuhan tersebut tidak selalu dinikmati secara merata oleh seluruh lapisan masyarakat. Akses terhadap pendidikan, layanan kesehatan, lapangan pekerjaan yang layak, serta kesempatan berusaha sering kali masih didominasi oleh kelompok-kelompok tertentu saja. Akibatnya jurang antara kelompok kaya dan kelompok miskin semakin melebar. Bahkan di negara maju seperti Amerika Serikat, ketimpangan ekonomi tetap menjadi persoalan yang belum terselesaikan. Meskipun pemerintah setempat telah mengalokasikan anggaran besar dalam bentuk subsidi dan program bantuan sosial bagi masyarakat berpenghasilan rendah, distribusi sumber daya yang tidak merata masih terlihat jelas. Fenomena seperti ini menunjukkan bahwa pertumbuhan ekonomi yang tinggi tidak serta-merta menjamin pemerataan kesejahteraan. Di Indonesia sendiri, masalah ketidakmerataan juga semakin kompleks karena dipengaruhi oleh faktor geografis, kepadatan penduduk, perbedaan akses infrastruktur, serta kualitas sumber daya manusia yang belum merata antar wilayah. Kawasan perkotaan sering kali lebih cepat berkembang, sementara daerah-daerah pedesaan atau terpencil masih tertinggal jauh dari segi pembangunan dan kesejahteraan. Salah satu indikator penting yang mencerminkan ketidakmerataan ini adalah angka kemiskinan. Jumlah penduduk miskin yang masih tinggi menunjukkan bahwa masih banyak masyarakat yang belum dapat memenuhi kebutuhan dasar secara layak. Tidak hanya berdampak pada kehidupan sehari-hari, tingginya angka kemiskinan juga dapat menjadi penghambat utama bagi pertumbuhan ekonomi nasional secara keseluruhan. Hal ini karena penduduk miskin memiliki keterbatasan dalam berkontribusi secara produktif terhadap ekonomi, baik sebagai tenaga kerja maupun sebagai konsumen. Oleh karena itu, pengentasan kemiskinan dan upaya pemerataan ekonomi harus menjadi fokus utama dalam kebijakan pembangunan nasional. Pemerintah perlu memperkuat program-program yang mendukung inklusi sosial dan ekonomi, seperti peningkatan kualitas pendidikan, pelatihan keterampilan, pemberdayaan usaha mikro dan kecil, serta penyediaan akses permodalan dan teknologi. Di samping itu, penting juga untuk memastikan bahwa kebijakan ekonomi yang dibuat benar-benar



menyasar kelompok yang paling membutuhkan, agar pertumbuhan yang dicapai tidak hanya menguntungkan segelintir pihak, tetapi membawa manfaat nyata bagi seluruh rakyat Indonesia (Sari, 2021). Ketimpangan ekonomi tercermin dari wilayah-wilayah yang tertinggal secara ekonomi, yang tersebar di seluruh Indonesia, mulai dari dusun-dusun di dataran tinggi, komunitas masyarakat di sekitar hutan, desa-desa kecil yang terisolir, permukiman nelayan di tepi laut, sampai kawasan kumuh yang ada di perkotaan.

Oleh karena itu, pemerintah provinsi, kabupaten, dan kota perlu memberikan perhatian khusus terhadap permasalahan ketidakmerataan ekonomi di tengah masyarakat. Masalah ini bukan hanya soal angka dan statistik, tetapi menyangkut kualitas hidup jutaan orang yang belum merasakan hasil dari pembangunan secara adil dan merata. Salah satu faktor utama yang turut memperparah kondisi ini adalah rendahnya kualitas sumber daya manusia di beberapa daerah, terutama di wilayah terpencil dan kurang berkembang. Ketika masyarakat tidak memiliki akses yang memadai terhadap pendidikan, pelatihan keterampilan, maupun layanan kesehatan yang berkualitas, mereka akan kesulitan untuk bersaing di dunia kerja maupun mengembangkan potensi ekonomi secara mandiri. Akibatnya, kesenjangan antara daerah maju dan tertinggal semakin melebar, menciptakan siklus kemiskinan yang sulit diputus. Oleh karena itu, dibutuhkan langkah konkret dari pemerintah daerah dalam membangun dan meningkatkan kapasitas sumber daya manusia melalui kebijakan yang berorientasi pada pemerataan. Misalnya, dengan memperluas akses pendidikan yang terjangkau dan bermutu, menyediakan program pelatihan kerja berbasis kebutuhan lokal, serta memastikan adanya infrastruktur penunjang yang layak seperti transportasi, internet, dan fasilitas publik lainnya. Lebih dari itu, kolaborasi antara pemerintah pusat dan daerah juga harus diperkuat agar program-program pembangunan tidak berjalan sendiri-sendiri, melainkan saling melengkapi dan menysasar akar masalah secara menyeluruh. Dengan begitu, ketimpangan ekonomi bisa perlahan dikurangi, dan masyarakat di seluruh pelosok negeri memiliki kesempatan yang setara untuk tumbuh dan berkembang (Saputra et al., 2023).

Situasi ini juga terjadi di Desa Bina Baru. Salah satu penyebab tidak meratanya ekonomi masyarakat di Desa Bina Baru adalah kebijakan yang kurang tepat sasaran. Desa Bina Baru terletak di Kecamatan Kampar Kiri Tengah, Kabupaten Kampar – Riau, dan terdiri atas 4 dusun: Dusun Sukajadi, Dusun Suka Damai, Dusun Suka Maju, dan Dusun Suka Baru. Desa ini dihuni oleh 5.760 jiwa yang terbagi dalam 1.742 keluarga, tersebar dalam 30 Rukun Tetangga (RT) dan 8 Rukun Warga (RW). Salah satu langkah untuk meningkatkan efektivitas kebijakan ekonomi adalah dengan melakukan pengelompokan data keluarga berdasarkan tingkat ekonomi menggunakan metode K-Means Clustering. Tujuan dari metode ini adalah untuk mengidentifikasi kelompok-kelompok ekonomi masyarakat di wilayah tersebut sehingga pemerintah daerah dapat mengambil keputusan yang lebih tepat dan terarah dalam upaya peningkatan kesejahteraan masyarakat.

K-Means Clustering adalah salah satu metode yang digunakan untuk mengelompokkan objek-objek ke dalam satu atau lebih grup, dengan tujuan untuk menganalisis dan mengekstrak pengetahuan secara otomatis. Beberapa pendekatan dalam data mining meliputi prediksi, perbandingan, klasifikasi, clustering, dan perkiraan (Waruwu et al., 2023). Metode K-Means clustering adalah salah satu teknik data clustering non-hierarki yang membagi data ke dalam satu atau lebih cluster. Dalam proses ini, data yang memiliki karakteristik serupa akan dikelompokkan dalam satu cluster yang sama, sementara data dengan karakteristik yang berbeda akan dimasukkan ke dalam cluster lainnya (Sitorus & Suhartika, 2024). Metode ini sering digunakan untuk memecahkan masalah dengan cara mengelompokkan data ke dalam beberapa kelompok, yang dikenal sebagai 'K', berdasarkan jarak terkecil antar cluster. Alasan Metode k-means sering digunakan dikarenakan proses K-Means cukup sederhana untuk diimplementasikan dan dijalankan. Selain kemampuan dalam melakukan klusterisasi dan kemudahan adaptasi, operasi matematis yang digunakan dalam metode ini juga relatif lebih mudah (Bahtiar, 2023). Dalam analisis data mining, algoritma K-Means berperan dalam membangun model yang dapat mempengaruhi proses pengambilan keputusan (Taryadi, 2022). Data mining adalah suatu teknik yang digunakan untuk mengidentifikasi informasi secara menyeluruh dari berbagai dataset besar, dengan tujuan untuk memperoleh pengetahuan baru (R et al., 2022).

Clustering adalah salah satu metode yang paling sering digunakan dan efektif dalam data mining. Teknik ini bertujuan untuk mengelompokkan objek-objek data ke dalam beberapa cluster berdasarkan kesamaan karakteristiknya (Amelia et al., 2023). Cluster sendiri adalah kumpulan data atau objek yang dikelompokkan bersama berdasarkan tingkat kemiripan atau kesamaan karakteristik tertentu. Dalam konteks clustering, kluster merupakan hasil dari proses pengelompokan data di mana objek-objek yang berada dalam satu kluster memiliki kemiripan yang tinggi, sedangkan objek-objek di antara kluster yang berbeda memiliki perbedaan yang signifikan.

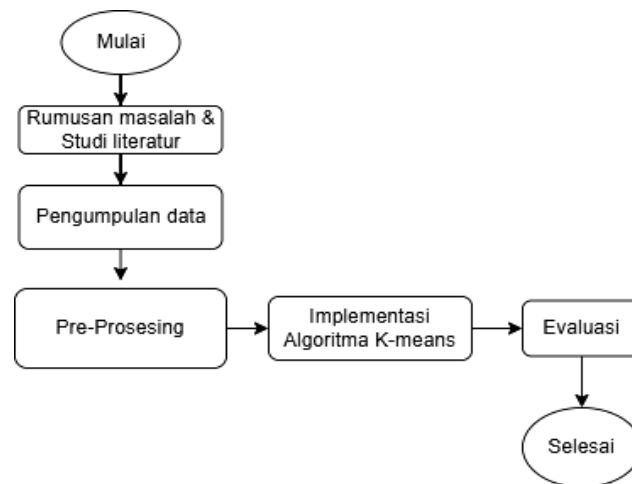
Dalam penelitian sebelumnya yang dilakukan oleh Luchia dkk bahwa algoritma K-Means lebih baik dalam melakukan cluster data penduduk di bandingkan algoritma K-Medoids (Luchia et al., 2022). Selanjutnya Dalam penelitian Fitriyadi & kurniawai bahwa dalam pengujian tingkat akurasi K-Means lebih tinggi daripada K-Medoids (Fitriyadi, 2021). Berdasarkan penelitian Nirwana Hendrastuty menunjukkan bahwa algoritma K-Means mampu mengelompokkan siswa menjadi dua kelompok, yaitu 63 siswa yang tergolong Rajin (C0) dan 91 siswa yang tergolong Sangat Rajin (C1). Hasil ini diperkuat dengan nilai silhouette score sebesar 0,9168, yang menandakan bahwa proses pengelompokan tersebut sangat akurat dan berkualitas baik (Hendrastuty, 2024). Dalam penelitian Wargijono Utomo menunjukkan bahwa algoritma K-Means lebih unggul dibandingkan K-Medoids dalam mengelompokkan penyebaran virus corona di Indonesia. Hal ini terlihat dari nilai indeks Davies-Bouldin yang lebih rendah, yaitu 0,064 pada K=5, yang menandakan kualitas pengelompokan yang lebih baik. Sebaliknya, algoritma K-Medoids mencatat nilai terkecil 0,411 pada K=2. Oleh karena itu, K-Means dinilai lebih efektif untuk digunakan dalam studi ini (Utomo, 2021). Berdasarkan penelitian yang dilakukan Aceng Supriyadi dkk didapatkan hasil penelitian, validitas cluster yang

dihasilkan oleh algoritma K-Means—diukur dengan Davies–Bouldin Index—terbukti secara konsisten lebih rendah daripada K-Medoids pada percobaan dengan jumlah kluster 3 hingga 10, sehingga K-Means dipilih untuk diimplementasikan dalam aplikasi web berbasis clustering ini (Supriyadi et al., 2021). Dalam penelitian Riva dkk di peroleh hasil, K-Means maupun K-Medoids berhasil mengelompokkan obat menjadi kategori cepat, sedang, dan lambat habis dengan K-Means menunjukkan performa sedikit lebih baik dengan nilai silhouette 0,627 dan 0,536 berturut-turut sehingga metode ini dapat diandalkan untuk merekomendasikan manajemen stok obat di puskesmas (Farissa et al., 2021). Berdasarkan penjelasan di atas, maka penelitian ini menggunakan metode K-Means Clustering untuk mengelompokkan data ekonomi penduduk di Desa Bina Baru, sehingga diharapkan dapat membantu pemangku kebijakan di wilayah tersebut dalam meningkatkan kesejahteraan masyarakat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian langkah yang harus diikuti oleh peneliti untuk melaksanakan suatu penelitian secara sistematis dan terstruktur. Pada tahapan pada penelitian ini dapat di lihat dari Gambar 1 berikut.



Gambar 1. Tahapan penelitian

Terlihat pada Gambar 1 terdapat beberapa tahapan dalam melakukan penelitian ini seperti; rumusan masalah, pengumpulan data, selection data, pre-prosesing, implementasi algoritma k-means, evaluasi.

2.2 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data yang diperoleh dari data penelitian sebelumnya yakni penelitian yang dilakukan oleh Felina Amelia jurusan Teknik Informatika UIN Suska Riau pada tahun 2022 dengan judul penelitian “Clustering Keluarga Miskin Desa Bina Baru dengan Metode K-Medoids”. Penelitian ini menggunakan data penduduk desa Bina Baru dengan jumlah 1000 data dan 16 variabel.

2.3 Pre-processing

Pada tahapan ini dilakukan perubahan suatu proses yang melibatkan beberapa tahap untuk mengubah data awal menjadi bentuk yang lebih sederhana dan dapat dipahami.

a. Cleaning Data

Tahap pembersihan (cleaning data) merupakan bagian dari preprocessing data. Tujuannya untuk membersihkan data mentah dari masalah nilai kosong (missing values), data duplikat, dan lainnya yang dapat mengganggu keakuratan hasil analisis

b. Transformasi Data

Pada tahapan transformasi data dilakukan perubahan struktur ataupun format dari data mentah agar lebih sesuai untuk analisis atau pemodelan. Jika terdapat data yang berbentuk teks, maka data dapat diubah menjadi representasi numerik. Namun jika sudah berbentuk numerik maka tidak perlu dilakukan perubahan atau transformasi data

c. Normalisasi Data

Normalisasi data merupakan proses transformasi data dalam rangkaian sehingga semua nilai datanya berada dalam rentang tertentu

2.4 Data Mining

Data mining merupakan istilah yang merujuk pada proses menemukan pengetahuan tersembunyi dalam suatu database. Data mining merupakan proses analisa data untuk menemukan suatu pola dari kumpulan suatu data (Febriansyah &



Muntari, 2023). Jadi dapat disimpulkan bahwa Data mining adalah proses penggalian data secara mendalam untuk mengetahui hal yang berarti dan tidak diketahui keberadaannya. Data Mining dengan KDD (Knowledge Discovery in Databases) adalah pendekatan sistematis untuk menemukan pengetahuan yang bermanfaat dari data besar. Proses ini melibatkan beberapa tahapan yang berurutan dan saling terkait.

2.5 Clustering

Clustering adalah salah satu metode yang digunakan dalam data mining, di mana proses kerjanya berfokus pada pencarian dan pengelompokan data berdasarkan kemiripan karakteristik yang dimiliki antar data. Dengan kata lain, data yang memiliki kesamaan fitur atau sifat akan dikelompokkan ke dalam satu kelompok atau cluster yang sama, sementara data yang berbeda akan ditempatkan di kelompok lain (Syaputri et al., 2021). Metode ini sangat bermanfaat dalam menganalisis pola tersembunyi dalam kumpulan data yang besar dan kompleks. Contohnya bisa kita lihat dalam dunia pemasaran, di mana clustering digunakan untuk mengelompokkan konsumen berdasarkan preferensi belanja mereka. Konsumen yang sering membeli produk serupa akan dimasukkan dalam satu kelompok, sehingga perusahaan dapat merancang strategi promosi yang lebih tepat sasaran. Selain itu, clustering juga sering digunakan dalam bidang kesehatan, pendidikan, dan bahkan keamanan siber—misalnya untuk mendeteksi anomali atau perilaku yang tidak biasa dalam sistem jaringan. Dengan kemampuan untuk mengorganisasi data secara otomatis tanpa label sebelumnya, metode ini menjadi alat yang sangat berguna dalam eksplorasi data dan pengambilan keputusan berbasis informasi. Secara keseluruhan, clustering membantu menyederhanakan pemahaman terhadap data yang besar dengan menyusun informasi ke dalam kelompok-kelompok yang lebih bermakna, sehingga proses analisis menjadi lebih efisien dan terarah (Triyana et al., 2022). Clustering dikatakan efektif apabila mampu membentuk kelompok yang anggotanya memiliki kemiripan tinggi satu sama lain dalam satu cluster, namun berbeda secara signifikan dengan anggota dari cluster lainnya.

2.6 K-Means

K-Means pertama kali diperkenalkan oleh MacQueen J.B. pada tahun 1976. Sejak saat itu, algoritma ini menjadi salah satu metode yang paling populer dan banyak digunakan dalam proses pengelompokan data, terutama dalam dunia data mining dan machine learning. K-means termasuk dalam metode non-hierarki, yang artinya pengelompokan data dilakukan tanpa membentuk struktur bertingkat seperti pada metode hierarki. Cara kerja algoritma ini cukup sederhana namun efektif. K-means akan membagi sekumpulan data ke dalam sejumlah kelompok atau cluster, yang jumlahnya ditentukan di awal (disebut sebagai nilai 'k'). Tujuannya adalah untuk mengelompokkan data sedemikian rupa sehingga setiap objek dalam satu kelompok memiliki karakteristik yang serupa, sedangkan objek yang berada di kelompok lain memiliki karakteristik yang berbeda. Proses ini dimulai dengan memilih titik pusat (centroid) secara acak untuk setiap cluster. Kemudian, setiap data akan dihitung jaraknya terhadap masing-masing centroid, dan data tersebut akan dimasukkan ke dalam cluster yang centroid-nya paling dekat. Setelah semua data dikelompokkan, centroid akan diperbarui berdasarkan rata-rata posisi data dalam cluster tersebut. Proses ini diulang hingga tidak ada lagi perubahan signifikan pada pembagian cluster. Karena sifatnya yang cepat dan efisien, K-means banyak digunakan dalam berbagai aplikasi praktis, seperti segmentasi pelanggan, analisis citra, pengelompokan dokumen, bahkan dalam sistem rekomendasi. Meskipun tergolong sederhana, hasil dari K-means dapat memberikan wawasan yang sangat berguna dalam memahami struktur dan pola dalam data (Bustami et al., 2022).

K-Means Clustering adalah salah satu metode dalam analisis data atau data mining yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kemiripan karakteristiknya. Istilah 'K' merujuk pada jumlah kelompok atau cluster yang ingin dibentuk, sedangkan 'Means' mengacu pada nilai rata-rata dari data dalam setiap cluster tersebut. Metode ini termasuk dalam kategori unsupervised learning, yang artinya proses pengelompokannya dilakukan tanpa label atau panduan sebelumnya. K-Means bekerja dengan cara membagi data ke dalam sejumlah cluster yang telah ditentukan di awal (nilai K), dan setiap cluster memiliki pusat (centroid) yang dihitung berdasarkan rata-rata data di dalamnya. Metode ini banyak digunakan karena sederhana, cepat, dan efektif dalam menemukan pola tersembunyi dalam kumpulan data yang besar. (Siregar et al., 2023). Berikut adalah langkah-langkah dalam metode k-means:

1. Menentukan jumlah cluster yang mau di bentuk
2. Menentukan centeroid (titik pusat Cluster) awal secara acak
3. Menghitung jarak setiap data ke masing-masing centroid dari masing-masing Clustering. Untuk menghitung jarak antara objek dengan centroid menggunakan rumus jarak Euclidean (1).

$$\sqrt{(x_i - x_j)^2 + y_i - y_j^2} \quad (1)$$

Keterangan dari rumus 1 diatas diantaranya x_i merupakan objek x ke- i , y_i merupakan daya y ke- i

4. Mengkelompokkan masing-masing data kedalam centroid yang terdekat
5. Menentukan posisi Centroid baru diperoleh dengan menghitung nilai rata-rata dari data yang ada dalam masing-masing kelompok cluster.
6. Lanjutkan proses dari langkah ketiga apabila masih ditemukan data yang berganti kelompok atau jika posisi centroid mengalami perubahan.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Algoritma K-Means Menggunakan Python

Pada tahapan penerapan algoritma K-means akan diterapkan dengan memanfaatkan bahasa pemrograman Python, yang dikenal luas karena kemampuannya dalam mengelola dan menganalisis data secara efisien. Untuk menentukan jumlah cluster (nilai k) yang paling tepat atau optimal, peneliti menggunakan metode evaluasi yang disebut Davies-Bouldin Index (DBI). Metode ini sangat berguna dalam menilai kualitas dari hasil pengelompokan, di mana semakin kecil nilai DBI, hal ini mencerminkan bahwa cluster yang terbentuk lebih baik karena memiliki tingkat pemisahan yang jelas antar cluster dan konsistensi dalam cluster itu sendiri. Setelah dilakukan proses evaluasi menggunakan DBI, diperoleh bahwa nilai k yang paling optimal adalah 3. Artinya, data yang digunakan dalam penelitian ini paling ideal jika dikelompokkan ke dalam tiga cluster. Keputusan ini diambil berdasarkan perhitungan dan interpretasi terhadap nilai indeks DBI yang dihasilkan pada berbagai kemungkinan jumlah cluster. Dengan demikian, proses analisis selanjutnya akan difokuskan pada tiga kelompok utama yang dianggap mewakili karakteristik data swcara menyeluruh seperti yang terdapat dalam Gambar 2 berikut:

```
Jumlah klaster: 2, DBI: 1.0319561674337059
Jumlah klaster: 3, DBI: 0.9326538702758231
Jumlah klaster: 4, DBI: 0.9734688875738688
Jumlah klaster: 5, DBI: 0.945072179484083
Jumlah klaster optimal: 3, dengan DBI: 0.9326538702758231
```

Gambar 2. Hasil Metode DBI Menentukan Nilai k Yang Optimum

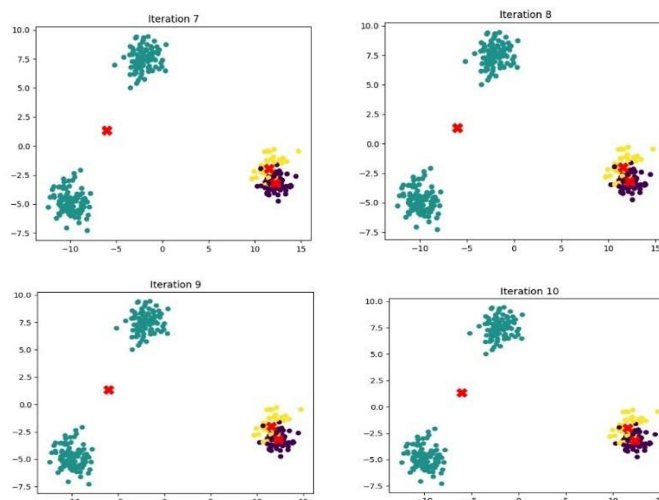
Gambar 2 menampilkan hasil pengujian nilai Davies-Bouldin Index (DBI) untuk berbagai pilihan k (2–5), di mana nilai DBI terendah yaitu 0,93265 dicapai pada k = 3, sehingga k = 3 ditetapkan sebagai jumlah klaster yang paling optimal.

Setelah menentukan nilai k yang optimal, data diskalakan menggunakan MinMaxScaler untuk menormalkan nilai-nilai agar tidak terlalu besar. Selanjutnya, dilakukan inisialisasi centroid sebanyak k secara acak. Kemudian, jarak antara setiap data dengan masing-masing centroid dihitung, dan setiap objek dikelompokkan ke dalam klaster berdasarkan jarak terdekat dengan centroid.

Tabel 1. Jumlah Anggota Tiap Cluster

Cluster	Jumlah
Cluster 0	201
Cluster 1	208
Cluster 2	592

Pada Tabel 1 jumlah anggota tiap cluster di dapat nilai cluster 0 sebanyak 201, cluster 1 sebanyak 208, dan cluster 2 sebanyak 592.



Gambar 3. Persebaran Data Iterasi

Pada Gambar 3 terlihat proses K-Means dari iterasi 7 sampai 10: tanda silang merah (pusat klaster) awalnya masih jauh dari kumpulan titik, kemudian perlahan bergerak mendekati ke kelompok data di iterasi 8 dan 9 sehingga lebih banyak titik masuk klaster yang benar, hingga di iterasi 10 pusat klaster hampir tidak berubah lagi dan semua titik sudah berada di klasternya masing-masing—menunjukkan proses telah selesai dengan pembagian klaster yang stabil.

Proses perhitungan berhenti pada iterasi ke-9, ketika tidak lagi terjadi perubahan dalam persebaran data. Pada iterasi ini, hasil pengelompokan menunjukkan bahwa cluster 0 berisi 201 data, cluster 1 terdiri dari 208 data, dan cluster 2 mencakup 529 data, seperti yang ditampilkan pada gambar

3.2 Centroid Hasil Pengujian

Pada tahap ini, temuan atau wawasan yang diperoleh dari dataset disajikan dengan menggunakan algoritma K-means. Selanjutnya, dijelaskan langkah-langkah keputusan atau tindakan yang dapat diambil berdasarkan hasil yang telah didapatkan

Tabel 2. Tabel Centroid

Variabel	Custer 0	Cluster 1	Cluster 2
Jumlah anggota keluarga	2,338308458	4,480769231	2,983108
Jenis pendapatan utama	1,621890547	1,725961538	1,493243
Dinding rumah	1,039800995	1,048076923	1,199324
Lantai rumah	1,472636816	1,846153846	1,403716
Kepemilikan rumah	1,124378109	1,100961538	1,201014
Atap rumah	1	1	2
Fasilitas sanitasi	1	1	1,001689
Fasilitas mandi dan mencuci	1	1	1,001689
Sumber air konsumsi	1,019900498	1,052884615	1,035473
Sumber penerangan	1	1	1,001689
Jenis bahan bakar masak	1,014925373	1,043269231	1,016892
Pembuangan sampah	1,07960199	1,072115385	1,070946
Pembuangan limbah cair	1,004975124	1,048076923	1,040541
Kelayakan hunian	1,995024876	1,990384615	1,998311
Penghasilan	4246019,9	4519711,538	2985659
Pengeluaran	2997512,438	3511538,462	2124493

Tabel 2 memperlihatkan nilai centroid tiap cluster berdasarkan variabel. Untuk memberikan gambaran yang lebih nyata mengenai hasil klasterisasi, titik centroid dari masing-masing cluster kemudian diterjemahkan ke dalam bentuk data asli yang lebih mudah dipahami. Proses ini dilakukan agar karakteristik tiap kelompok tidak hanya terlihat sebagai angka statistik, tetapi juga mencerminkan kondisi rumah tangga dalam kehidupan sehari-hari.

Tabel 3. Tabel karkteristik Tiap Cluster

Variabel	Custer 0	Cluster 1	Cluster 2
Jumlah anggota keluarga	2 orang	5 orang	3 orang
Jenis pendapatan utama	pemilik kebun sawit	pemilik kebun sawit	buruh tani
Dinding rumah	semen	Semen	semen
Lantai rumah	semen	Keramik	semen
Kepemilikan rumah	milik pribadi	milik pribadi	milik pribadi
Atap rumah	genteng	Genteng	seng
Fasilitas sanitasi	jamban pribadi	jamban pribadi	jamban pribadi
Fasilitas mandi dan mencuci	pribadi	Pribadi	pribadi
Sumber air konsumsi	air isi ulang	air isi ulang	air isi ulang
Sumber penerangan	listrik pln	listrik pln	listrik pln
Jenis bahan bakar masak	gas kota/LPG	gas kota/LPG	gas kota/LPG
Pembuangan sampah	dibakar	Dibakar	dibakar
Pembuangan limbah cair	lubang resapan	lubang resapan	lubang resapan
Kelayakan hunian	layak huni	layak huni	layak huni

Tabel 3 yang memperlihatkan bagaimana setiap cluster memiliki ciri khas tersendiri, mulai dari jumlah anggota keluarga, jenis pendapatan utama, hingga kelayakan hunian dan akses terhadap fasilitas dasar. Berikut adalah analisis karkteristik dari setiap cluster:

- Cluster 0 Terdiri dari rumah tangga dengan dua anggota keluarga dan jenis pendapatan utama sebagai pemilik kebun sawit. Rumah memiliki struktur dasar yang cukup baik dengan dinding dan lantai berbahan semen, serta atap dari genteng. Fasilitas sanitasi cukup memadai, termasuk kepemilikan jamban pribadi dan akses terhadap fasilitas mandi dan mencuci, air minum isi ulang, dan listrik PLN. Penghasilan rumah tangga ini sebesar Rp4.246.019,90 dengan pengeluaran Rp2.997.512,44, menghasilkan rasio pengeluaran terhadap penghasilan sebesar 70,60%. Hal ini menunjukkan bahwa meskipun jumlah penghasilan cukup tinggi, pengeluaran masih terkendali, menandakan efisiensi dalam pengelolaan keuangan rumah tangga tersebut.
- Cluster 1 Memiliki lima anggota keluarga dengan jenis pendapatan utama yang sama, yaitu sebagai pemilik kebun sawit. Kualitas tempat tinggal sedikit lebih baik dibanding Cluster 0, dengan lantai keramik dan kelayakan hunian



layak huni. Semua fasilitas dasar tersedia secara mandiri, seperti jamban, fasilitas mandi dan mencuci, air minum, listrik, serta pengelolaan limbah. Penghasilan rumah tangga mencapai Rp4.519.711,54 dengan pengeluaran Rp3.511.538,46, menghasilkan rasio pengeluaran sebesar 77,69%, tertinggi di antara semua cluster. Hal ini mengindikasikan bahwa meskipun penghasilan tinggi, beban pengeluaran keluarga yang lebih besar (karena jumlah anggota keluarga lebih banyak) menyebabkan tingkat efisiensi pengeluaran yang lebih rendah.

- c. Cluster 2 Memiliki tiga anggota keluarga dan jenis pendapatan utama sebagai buruh tani, berbeda dari dua cluster sebelumnya. Struktur fisik rumah cenderung lebih sederhana, dengan atap dari seng dan lantai semen. Namun, rumah masih dinilai layak huni dan memiliki akses yang sama terhadap fasilitas dasar seperti air isi ulang, listrik, dan sistem sanitasi. Penghasilan mereka adalah Rp2.985.658,78 dengan pengeluaran Rp2.124.493,24, menghasilkan rasio pengeluaran sebesar 71,16%. Meskipun penghasilan lebih rendah dibanding cluster lain, rasio pengeluaran cukup terkendali, menandakan adanya penyesuaian gaya hidup dan pengeluaran sesuai kemampuan.

Berdasarkan analisis karakteristik masing-masing cluster, dapat disimpulkan bahwa Cluster 0 merupakan kelompok dengan tingkat kemiskinan rendah, ditandai oleh penghasilan yang tinggi, pengeluaran yang terkontrol, dan akses terhadap fasilitas dasar yang memadai, sehingga tidak termasuk dalam kategori miskin. Cluster 1 berada pada tingkat kemiskinan menengah; meskipun memiliki penghasilan tertinggi, jumlah anggota keluarga yang besar menyebabkan rasio pengeluaran juga tinggi, menjadikannya kelompok yang rentan terhadap tekanan ekonomi. Sementara itu, Cluster 2 memiliki tingkat kemiskinan tertinggi, ditandai oleh penghasilan paling rendah, kelayakan hunian yang paling sederhana, dan pekerjaan sebagai buruh tani, meskipun pengeluarannya relatif terkendali. Dengan demikian, Cluster 2 menjadi prioritas utama bagi pemerintah dalam upaya pengentasan kemiskinan, melalui program bantuan sosial, peningkatan keterampilan, serta akses terhadap ekonomi dan infrastruktur dasar.

Berdasarkan data penghasilan dan pengeluaran yang diberikan, kita dapat mendukung analisis klaster dengan melihat perbedaan kemampuan ekonomi dari masing-masing kelompok seperti pada Tabel 4 berikut:

Tabel 4. Penghasilan Dan Pengeluaran

Cluster	Penghasilan(RP)	Pengeluaran(RP)
Cluster 0	4.246.020	2.997.512
Cluster 1	4.519.712	3.511.538
Cluster 2	2.985.659	2.124.493

Berdasarkan Tabel 4, di peroleh rasio pengeluaran terhadap penghasilan cluster 0 sebesar 70.6%, Cluster 1 sebesar 77.7% dan Cluster 2 sebesar 71.1%. Berdasarkan analisis data penghasilan dan pengeluaran, terdapat kecenderungan bahwa semakin tinggi penghasilan suatu rumah tangga, semakin besar pula pengeluarannya. Misalnya, rumah tangga dengan penghasilan 4.246.020 memiliki pengeluaran sebesar 2.997.512, sedangkan rumah tangga dengan penghasilan lebih tinggi, yaitu 4.519.712, memiliki pengeluaran yang juga lebih besar, yaitu 3.511.538. Sebaliknya, rumah tangga dengan penghasilan lebih rendah, yaitu 2.985.659, memiliki pengeluaran sebesar 2.124.493. Hal ini menunjukkan bahwa daya beli rumah tangga dengan penghasilan lebih tinggi lebih besar, sehingga kemungkinan mereka memiliki kondisi hunian yang lebih baik serta akses terhadap fasilitas yang lebih layak. Namun, rasio pengeluaran terhadap penghasilan cukup tinggi, yang menandakan bahwa sebagian besar pendapatan digunakan untuk kebutuhan pokok.

Cluster 0 (Penghasilan Menengah, Pengeluaran Seimbang, Akses Fasilitas Standar) Rumah tangga dalam cluster ini memiliki penghasilan sekitar 4.246.020 dengan pengeluaran yang juga cukup tinggi ($\pm 2.997.512$). Sebagian besar rumah tangga di cluster ini memiliki akses yang lebih baik dibandingkan cluster 0, misalnya pada aspek jenis lantai rumah, sumber air konsumsi, dan fasilitas sanitasi. Namun, masih terdapat beberapa keterbatasan, terutama dalam hal jenis bahan bakar masak dan cara pembuangan limbah cair. Cluster ini mencerminkan rumah tangga dengan kondisi ekonomi menengah yang masih rentan terhadap perubahan ekonomi dan inflasi, sehingga kebijakan yang relevan adalah subsidi bahan pokok dan program pelatihan ekonomi untuk meningkatkan pendapatan.

Cluster 1 (Penghasilan dan Pengeluaran Tinggi, Akses Fasilitas Baik) Cluster ini mencakup rumah tangga dengan penghasilan tertinggi ($\pm 4.519.712$) dan pengeluaran yang juga besar ($\pm 3.511.538$). Rumah tangga dalam cluster ini cenderung memiliki fasilitas hunian yang lebih baik, seperti jenis dinding dan lantai yang lebih berkualitas (semen, marmer, keramik), sumber air konsumsi yang lebih layak, dan sistem pembuangan limbah yang lebih tertata. Tingginya pengeluaran menunjukkan bahwa rumah tangga dalam cluster ini memiliki daya beli yang lebih besar, memungkinkan mereka untuk menikmati fasilitas rumah tangga yang lebih baik. Kebijakan yang cocok untuk cluster ini lebih berfokus pada pemberdayaan ekonomi jangka panjang, seperti investasi pada wirausaha atau peningkatan akses ke sektor keuangan formal.

Cluster 2 (Penghasilan dan Pengeluaran Rendah, Akses Fasilitas Terbatas) Cluster ini terdiri dari rumah tangga dengan penghasilan rata-rata lebih rendah dibandingkan cluster lainnya ($\pm 2.985.659$) dan pengeluaran yang juga lebih rendah ($\pm 2.124.493$). Rumah tangga dalam cluster ini cenderung memiliki keterbatasan akses terhadap sumber air konsumsi, fasilitas sanitasi dan fasilitas mandi dan mencuci, dan tempat pembuangan sampah. Selain itu, kelayakan hunian mereka lebih sederhana dengan bahan bangunan yang kurang berkualitas seperti dinding kayu dan atap seng atau ijuk. Intervensi yang dapat dilakukan pada cluster ini meliputi bantuan sosial ekonomi, peningkatan akses terhadap infrastruktur dasar, dan program perbaikan rumah layak huni.



Hasil analisis menunjukkan bahwa penghasilan rumah tangga sangat berpengaruh terhadap kondisi hunian dan akses terhadap fasilitas dasar. Cluster dengan penghasilan rendah menghadapi tantangan dalam memenuhi standar hunian layak, sedangkan cluster dengan penghasilan lebih tinggi menikmati fasilitas yang lebih baik. Oleh karena itu, kebijakan yang disarankan adalah memastikan pemerataan akses terhadap infrastruktur dasar, memberikan subsidi bagi kelompok berpenghasilan rendah, serta mendukung program peningkatan ekonomi agar kesejahteraan masyarakat dapat meningkat secara merata.

3.3 Perbandingan Temuan Hasil

Penelitian ini bertujuan untuk mengelompokkan rumah tangga berdasarkan karakteristik sosial-ekonomi menggunakan algoritma k-means clustering, yang menghasilkan tiga kluster dengan profil yang berbeda. Hasil pengelompokan menunjukkan perbedaan signifikan dalam hal pendapatan, pengeluaran, serta akses terhadap fasilitas dasar seperti air bersih, sanitasi, dan bahan bakar memasak. Setiap kluster merepresentasikan kondisi kesejahteraan yang berbeda, sehingga rekomendasi kebijakan yang diusulkan pun disesuaikan dengan karakteristik masing-masing kelompok.

Jika dibandingkan dengan penelitian sejenis, terdapat beberapa poin kesamaan dan perbedaan yang dapat dicermati. Penelitian pertama menggunakan algoritma k-means clustering dan menetapkan jumlah kluster optimal sebanyak dua, berdasarkan metode elbow dan validasi menggunakan nilai Silhouette Score sebesar 0,916. Hasil tersebut menunjukkan kualitas kluster yang sangat baik. Fokus penelitian tersebut adalah mengidentifikasi karakteristik penerima dan non-penerima Bantuan Pangan Non Tunai (BPNT), sehingga klasifikasi dilakukan berdasarkan kelayakan untuk menerima bantuan. Perbedaan utama terletak pada konteks dan tujuan analisis; penelitian ini menggunakan tiga kluster untuk mendapatkan pemetaan yang lebih rinci dan menyusun intervensi kebijakan sosial secara lebih terarah.

Sementara itu, penelitian kedua menerapkan algoritma K-Medoids dengan jumlah kluster yang terbentuk sebanyak tiga, yaitu kategori kaya, menengah, dan miskin, berdasarkan 1.000 data penduduk di Desa Bina Baru. Validasi hasil kluster menggunakan Davies-Bouldin Index (DBI) memperoleh nilai 0,991 yang menunjukkan kualitas pemodelan kluster yang cukup baik. Meskipun serupa dalam hal jumlah kluster, perbedaan terletak pada pendekatan algoritmik serta ruang lingkup indikator yang digunakan. Penelitian ini tidak hanya mengelompokkan berdasarkan indikator ekonomi, tetapi juga mencakup variabel-variabel yang berkaitan dengan kualitas hidup, seperti ketersediaan fasilitas dasar dan kondisi tempat tinggal.

Dengan demikian, keunggulan penelitian ini dibandingkan dengan penelitian sejenis terletak pada dimensi analisis yang lebih komprehensif. Selain mempertimbangkan variabel ekonomi, penelitian ini juga memasukkan aspek aksesibilitas fasilitas dasar yang berkontribusi terhadap kualitas hidup rumah tangga. Hal ini memungkinkan perumusan rekomendasi kebijakan yang lebih holistik dan kontekstual, sesuai dengan kebutuhan masing-masing kluster. Pendekatan ini dinilai lebih aplikatif dalam upaya peningkatan kesejahteraan masyarakat secara menyeluruh.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengidentifikasi area yang membutuhkan bantuan atau intervensi, maka Cluster 2 dapat menjadi fokus utama untuk peningkatan kualitas hidup. Dari hasil analisis kluster menunjukkan perbedaan signifikan dalam kondisi sosial-ekonomi rumah tangga berdasarkan penghasilan, pengeluaran, dan akses fasilitas. Cluster 2 terdiri dari rumah tangga berpenghasilan rendah ($\pm 2,99$ juta) dan pengeluaran rendah ($\pm 2,12$ juta), dengan akses terbatas ke fasilitas dasar seperti air bersih dan sanitasi. Cluster 0 mencerminkan rumah tangga berpenghasilan menengah ($\pm 4,25$ juta) dan pengeluaran ± 3 juta, dengan akses fasilitas yang cukup baik meski ada keterbatasan dalam jenis bahan bakar masak dan pembuangan limbah. Cluster 1 mencakup rumah tangga berpenghasilan tinggi ($\pm 4,52$ juta) dan pengeluaran $\pm 3,51$ juta), dengan fasilitas rumah yang lebih baik seperti lantai keramik dan sistem sanitasi yang teratur. Penghasilan yang lebih tinggi cenderung diikuti dengan pengeluaran dan kualitas hidup yang lebih baik. Untuk meningkatkan kesejahteraan, Cluster 0 memerlukan bantuan sosial dan perbaikan infrastruktur dasar. Cluster 2 membutuhkan subsidi bahan pokok dan pelatihan ekonomi. Cluster 1 dapat didorong melalui investasi dan akses ke sektor keuangan. Kebijakan yang tepat pada setiap kluster akan meningkatkan kesejahteraan masyarakat.

REFERENCES

- Amelia, F., Iskandar, I., Kurnia Gusti, S., & Haerani, E. (2023). Clustering Keluarga Miskin Desa Bina Baru Dengan Metode K-Medoids. *Krea-Tif: Jurnal Teknik Informatika*, 11(1), 1–13. <https://doi.org/10.32832/krea-tif.v11i1.14104>
- Bahtiar, D. (2023). Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering. *Scientia Sacra: Jurnal Sains*, 3(2), 29–39. <https://doi.org/https://doi.org/10.31328/jointecs.v4i1.998>
- Bustami, B., Mahara, R., Ahmadian, H., Wahyuni, S., & AR, K. (2022). Analisis Clustering Penduduk Miskin Di Provinsi Aceh Menggunakan Algoritma K-Means Dan X-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 5(1), 26–35. <https://doi.org/10.32672/jnkti.v5i1.3961>
- Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsembung. *Journal of Applied Informatics and Computing*, 5(2), 109–116. <https://doi.org/10.30871/jaic.v5i1.3237>
- Febriansyah, F., & Muntari, S. (2023). Penerapan Algoritma K-Means untuk Klusterisasi Penduduk Miskin pada Kota



- Pagar Alam. *Jurnal Informatika Sunan Kalijaga*, 8(1), 66–77. <https://doi.org/https://doi.org/10.14421/jjska.2023.8.1.66-77>
- Fitriyadi, A. upi. (2021). Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional. *Kilat*, 10(1), 157–168. <https://doi.org/10.33322/kilat.v10i1.1174>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (Jima-Ilkom)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Luchia, N. T., Handayani, H., & Hamdi, F. S. (2022). Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia. *Indonesian Journal of Machine Learning and Computer Science*, 2(October), 35–41. <https://doi.org/https://doi.org/10.57152/malcom.v2i2.422>
- R, N. N. F., Anggraeni, D. S., & Enri, U. (2022). Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient. *Tematik*, 9(1), 29–35. <https://doi.org/10.38204/tematik.v9i1.901>
- Saputra, S. N., Haerani, E., Jasril, J., Oktavia, L., & Syafria, F. (2023). Penerapan Algoritma K - means Pada Cluster Penerima Bantuan Pangan Non Tunai (BPNT). *Journal of Computer Engineering, System and Science*, 8((2)), 438–449. <https://doi.org/https://doi.org/10.24114/cess.v8i2.48026>
- Sari, Y. A. (2021). Pengaruh Upah Minimum Tingkat Pengangguran Terbuka Dan Jumlah Penduduk Terhadap Kemiskinan Di Provinsi Jawa Tengah. *Equilibrium : Jurnal Ilmiah Ekonomi, Manajemen Dan Akuntansi*, 10(2), 121–130. <https://doi.org/10.35906/je001.v10i2.785>
- Siregar, H. A., Azlan, A., & Lumban Gaol, N. Y. (2023). Penerapan Data Mining Pada Penjualan Rumah Makan Kasih Ibu Menggunakan Metode K-Means Clustering. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(5), 750. <https://doi.org/10.53513/jursi.v2i5.8955>
- Sitorus, Z., & Suhartika. (2024). Penerapan Data Mining Untuk Clustering Penduduk Miskin Di Kota Tanjungbalai Menggunakan Metode Algoritma K-Means. *Journal of Science and Social Research*, 4307(1), 212–218. <https://doi.org/https://doi.org/10.54314/jssr.v7i1.1732>
- Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 229–240. <https://doi.org/10.29100/jupi.v6i2.2008>
- Syaputri, D., Noprita, P. H., & Romelah, S. (2021). Implementasi Algoritma K-Means untuk Pengelompokan Distribusi Sosial Ekonomi Masyarakat Berdasarkan Demografi Kependudukan. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 1–6. <https://doi.org/10.57152/malcom.v1i1.5>
- Taryadi, T. (2022). Klasterisasi Data Keluarga Pra Sejahtera Di Kota Pekalongan Dengan Metode K-Means Clustering. *Jurnal Litbang Kota Pekalongan*, 20(1), 70–76. <https://doi.org/10.54911/litbang.v20i1.180>
- Triyana, M., Juita, R., & Suhendra, C. D. (2022). Penerapan Metode K-Means dalam Pengelompokan Data Penduduk Tidak Mampu di Distrik Oransbari. *INFORMAL: Informatics Journal*, 7(3), 220. <https://doi.org/10.19184/isj.v7i3.34722>
- Utomo, W. (2021). The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. *ILKOM Jurnal Ilmiah*, 13(1), 31–35. <https://doi.org/10.33096/ilkom.v13i1.763.31-35>
- Waruwu, A., Yetri, M., & Setiawan, F. (2023). Implementasi Data Mining Dalam Mengelompokkan Data penduduk Kurang Mampu Menggunakan Metode K-Means Clustering. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(6), 945. <https://doi.org/10.53513/jursi.v2i6.8965>