 ISSN:2302-8149
e-ISSN:2540-9719

Penerapan Algoritma *k-Means* pada *Clustering* Data Penduduk Miskin untuk Penghapusan Kemiskinan Ekstrem

Application of k-Means Algorithm on Clustering Poor Population Data for Extreme Poverty Elimination

[1]Widya Syaharani*, [2]Sriani
[1,2]Program Studi Ilmu Komputer, Fakultas Sains Dan Teknologi, Universitas Islam Negeri Sumatera Utara, Kota Medan, Sumatera Utara, Indonesia.
*e-mail: widiasyaharani53@gmail.com

(***received*:** 28 June 2024, ***revised*:** 1 July 2024, ***accepted*:** 23 July 2024)

Abstrak

Kemiskinan merupakan salah satu permasalahan sosial yang hampir dihadapi oleh setiap negara didunia. Salah satu Faktor penyebab kemiskinan belum tertuntaskan yaitu dalam sebuah pelaksanaan kebijakan bantuan sosial survey yang dilakukan pemerintah terhadap masyarakat masih dilakukan secara manual sehingga tidak tepat sasaran. Maka, penelitian ini bertujuan untuk mengidentifikasi kriteria yang dimiliki oleh setiap kelompok masyarakat miskin yang dihasilkan dari pengelompokan data menggunakan algoritma *K-Means clustering*. Dengan menerapkan algoritma K-Means *clustering* pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik dan memodelkan *clustering* data penduduk miskin Desa Sei Litur Tasik. Hasil pengujian dan evaluasi model K-Means *Clustering* pada data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) ditetapkan menjadi 2 klaster yang optimal dengan nilai interia 0,40 menggunakan metode pengujian *Silhouette Score* dimana *cluster* 1 kategori kaya sebanyak 366 keluarga dan *cluster* 2 kategori miskin sebanyak 60 keluarga. Pemodelan rancangan sistem pengelompokan data dengan menggunakan metode *K-Means clustering* dilakukan pada *Google Colaboratory* serta dibantu dengan pustaka-pustaka pendukung. Hasil penelitian menunjukkan *accuracy K-Means clustering* sebesar 85,92% yang berarti bahwa *accuracy* dari data yang dianalisis dapat dengan tepat dikelompokkan ke dalam kategori *cluster* yang sesuai.

Kata kunci: algoritma *k-means clustering*, data penduduk, penghapusan kemiskinan

Abstract

Poverty is one of the social problems faced by almost every country in the world. One of the factors causing poverty has not been resolved, namely in an implementation of social assistance policies, the government's survey of the community is still carried out manually so that it is not right on target. So, this research aims to identify the criteria possessed by each group of poor people resulting from data grouping using the K-Means clustering algorithm. By applying the K-Means clustering algorithm to the data of the Targeting for the Acceleration of the Elimination of Extreme Poverty (P3KE) of Sei Litur Tasik Village and modeling the data clustering of the poor population of Sei Litur Tasik Village. The results of testing and evaluating the K-Means Clustering model on the data of the Acceleration of the Elimination of Extreme Poverty (P3KE) are determined to be 2 optimal clusters with an interia value of 0.40 using the Silhouette Score testing method where cluster 1 rich category is 366 families and cluster 2 poor category is 60 families. Modeling of the data clustering system design using the K-Means clustering method was carried out on Google Collaboratory and assisted by supporting literature. The results showed the accuracy of K-Means clustering of 85.92% which means that the accuracy of the analyzed data can be correctly grouped into the appropriate cluster category.

Keywords: *k-means clustering algorithm, population data, poverty elimination*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Kemiskinan merupakan salah satu permasalahan sosial yang hampir dihadapi oleh setiap negara didunia. Kemiskinan tidak hanya menjadi permasalahan serius yang dihadapi oleh pemerintahan pusat tetapi pemerintahan daerah juga menghadapi permasalahan tersebut [1][2]. Bagi pemerintah Indonesia masalah kemiskinan menjadi salah satu masalah yang sudah lama terjadi dan belum tertuntaskan hingga saat ini. Masalah kemiskinan ini hampir terjadi disetiap wilayah di Indonesia. Salah satu wilayah yang menghadapi masalah kemiskinan yaitu Desa Sei Litur Tasik [3][4].

Desa Sei Litur Tasik merupakan salah satu desa yang ada di kecamatan Sawit Seberang, kabupaten Langkat, Sumatera Utara, dengan jumlah penduduk sebanyak ± 5.864 jiwa dan ± 1.817 KK. Mayoritas suku penduduk di Desa Sei Litur Tasik yaitu suku jawa dan berprofesi sebagai petani, wiraswasta, dan buruh harian lepas. Tingkat kemiskinan yang sedang dihadapi Masyarakat di Desa Sei Litur Tasik dipengaruhi oleh beberapa faktor diantaranya adalah lapangan kerja yang tidak memadai serta pekerjaan yang dimiliki penghasilannya tidak sesuai dengan kebutuhan dan tanggungan hidup. Selain itu, yang menjadi faktor penyebab kemiskinan masih belum bisa tertuntaskan lainnya yaitu dalam sebuah pelaksanaan kebijakan bantuan sosial survey yang dilakukan pemerintah terhadap masyarakat masih dilakukan secara manual sehingga tidak tepat sasaran [5][6].

Salah satu kesulitan yang dihadapi pemerintah khususnya di Desa Sei Litur Tasik dalam proses penanganan masalah kemiskinan yaitu proses pembagian bantuan sosial yang banyak salah sasaran karena masih dilakukan secara manual, sehingga sebagian masyarakat yang mengalami masalah kemiskinan akan tetap memiliki kesulitan dalam pemenuhan kebutuhan kehidupan sehari-hari [7]. sedangkan tujuan bantuan sosial untuk membantu masyarakat yang mengalami kemiskinan ekstrem [8]. Penyebab hal ini terjadi karena validasi data sering diabaikan sehingga menimbulkan data yang tidak akurat dan keterbatasan dana yang akan disalurkan kepada masyakarat serta terjadinya kemiripan data keadaan ekonomi dalam penerimaan bantuan yang dilihat berdasarkan kriteria yang telah ditentukan, permasalahan lainnya adalah adanya subjektivitas pengambilan keputusan distribusi bantuan kepada masyarakat yang tidak merata [9][10].

Berdasarkan latar belakang permasalahan tersebut penelitian ini dilakukan agar mempermudah pemerintah Desa Sei Litur Tasik dalam mengambil keputusan terhadap masyarakat yang lebih berhak mendapatkan bantuan sosial sehingga bantuan yang disalurkan pemerintah tepat sasaran serta tidak menimbulkan konflik antar masyarakat desa, dengan hal ini salah satu kesulitan yang dihadapi pemerintah Desa Sei Litur Tasik dalam proses penanganan masalah kemiskinan dapat tertangani [11][12]. Data yang digunakan dalam penelitian ini adalah data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik tahun 2022, sumber data ini berasal dari Kementerian Koordinator Pembangunan Manusia dan Kebudayaan. Dalam data tersebut pemerintah menerapkan kebijakan dan program untuk membantu meringankan beban pengeluaran kelurga kelompok miskin melalui program seperti bantuan sosial dan subsidi [13][14].

Sudah banyak penelitian terdahulu yang relevan dengan penelitian yang akan dilakukan penulis, yaitu penelitian yang berjudul "Penerapan *Data mining* Untuk *Clustering* Data Penduduk Miskin Menggunakan Metode k-Means" tahun 2021 dengan hasil penelitian bahwa algoritma k-Means mampu mengelompokkan 812 data penduduk miskin ke dalam 2 *cluster* dengan nilai 0,1643, yang mana *cluster* 1 memberikan golongan menjadi penduduk miskin berjumlah 103 data dengan persentase 13%, dan *cluster* 2 memberikan golongan menjadi tidak miskin berjumlah 709 data dengan persentase 83% dan dilakukan iterasi sebanyak 4 kali. Dalam pengujian ini peneliti menggunakan 2 aplikasi, yang pertama dengan menggunakan aplikasi pengolah angka Microsoft Excel dan dengan menggunakan tools weka, sedangkan validitas yang dilakukan menggunakan metode DBI [8].

Tujuan dari penelitian ini adalah Untuk mengidentifikasi kriteria yang dimiliki oleh setiap kelompok masyarakat miskin yang dihasilkan dari pengelompokan data menggunakan algoritma k-Means *clustering*. Untuk menerapkan algoritma k-Means *clustering* pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik. Untuk memodelkan *clustering* data penduduk miskin Desa Sei Litur Tasik. Beberapa manfaat penelitian ini adalah Penelitian ini memberikan kontribusi dalam pengelolaan data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) ke dalam kelompok-kelompok yang serupa sehingga pemerintah dapat menyalurkan bantuan sosial yang lebih tepat sasaran kepada masyarakat Desa Sei Litur Tasik. Memberikan kontribusi

 ISSN:2302-8149
e-ISSN:2540-9719

dalam memperoleh pemahaman tentang kriteria yang dimiliki oleh setiap kelompok masyarakat miskin yang dihasilkan dari pengelompokan data menggunakan algoritma k-Means *clustering*. Dengan demikian penelitian ini dapat memberikan informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan pada penyaluran bantuan sosial kepada Masyarakat dalam upaya menangani salah satu kesulitan yang dihadapi pemerintah dalam proses penanganan masalah kemiskinan. Memberikan kontribusi untuk peneliti selanjutnya sebagai gambaran tentang penggunaan algoritma k-Means *clustering* dalam mengelompokkan data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE). Hal ini dapat membantu peneliti selanjutnya untuk memahami dasar-dasar algoritma k-Means *clustering* dan mengaplikasikannya dalam konteks tingkat kemiskinan pada masyarakat.

2 Tinjauan Literatur

Penelitian terdahulu merupakan acuan penulis dalam melakukan penelitian sehingga penulis mendapat banyak informasi yang digunakan untuk mengkaji penelitian ini. Penelitian yang dilakukan oleh Febriansyah & Muntari [15] dengan judul "Penerapan Algoritma k-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam" menghasilkan sistem algoritma K- Means untuk mengklasifikasikan penduduk miskin di Kota Pagar Alam. Metode pengujian sistem menggunakan black box testing dengan metode alpha dan didapatkan hasil pengujian database dengan nilai 4, antarmuka dengan nilai 4, fungsionalitas 4,42, dan algoritma dengan nilai 4. Kemudian penelitian Siahaa [16] dengan judul " Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma k-Means" Hasil perhitungan *cluster* C1 terdapat 3 kecamatan (Belakang Padang, Bulang dan Galang). Hasil *cluster* C2 terdapat 4 kecamatan (Sungai Beduk, Sagulung, Lubuk Baja dan Batu Ampar). Hasil *cluster* C3 terdapat 3 kecamatan (Sekupang, Batu Aji, dan Bengkong). Sedangkan perhitungan *cluster* C4 terdapat 2 kecamatan (Nongsa dan Batam Kota).

Penelitian Muthmainnah et al, [17] dengan judul "Penerapan Algoritma k-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat" Penerapan Algoritma k-Means menghasilkan penerimaan 10 kabupaten/kota yang memiliki jumlah pengangguran tingkat rendah, lalu ada 15 kabupaten/kota yang memiliki jumlah pengangguran tingkat sedang dan terdapat 2 kabupaten/kota yang memiliki jumlah pengangguran dengan tingkat tinggi. Hasil dari pengujian dengan menggunakan Davies Bouldin Index *cluster* = 3 mempunyai kualitas *cluster* terbaik. Penelitian Rosmini et al, [18] dengan judul " Implementasi Metode k-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah" Menghasilkan 2 *cluster*, yaitu *Cluster* A adalah mahasiswa yang lulus tepat waktu sedangkan *cluster* B adalah mahasiswa yang lulusnya tidak tepat waktu. Data pengelompokan mahasiswa ini merupakan masukan bagi dosen wali dalam membimbing dan mengawasi proses belajar mahasiswa agar bisa lulus tepat waktu.

Penelitian Harahap et al, [19] dengan judul "k-Means *Cluster*ing Algorithm Approach in *Cluster*ing Data on Cocoa Production Results in the Sumatra Region" Hasil pengujian algoritma k-Means *Cluster*ing dan Gaussian Mixture Model pada data produksi kakao di empat provinsi yaitu Sumatera Utara, Sumatera Barat, Lampung dan Aceh yang dioptimasi dengan metode Silhouette menghasilkan nilai *cluster* 2, 3 dan 4. Kedua dengan nilai 59,8%. Penelitian Nursia et al., [20] dengan judul "Analisis Kelayakan Penerima Bantuan Covid-19 Menggunakan Metode K–Means" Implementasi Metode K-means untuk menentukan kelayakan penerima bantuan Covid-19 Pada Kantor Kepala Desa Punggulan Kecamatan Air Joman menggunakan metode K-means pada penelitian ini memberikan hasil yaitu calon penerima bantuan Covid-19 dengan 2 *cluster* yaitu layak (C1) dan tidak layak (C2). Penelitian Ali [21] dengan judul " Klasterisasi Data Rekam Medis Pasien Menggunakan Metode k-Means *Cluster*ing di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo" Hasil dari penelitian ini menghasilkan kolom 4 *cluster* yang terdiri dari kecamatan, diagnosa penyakit, usia dan jenis kelamin. Pengelompokan data rekam medis pasien ini menghasilkan informasi baru mengenai pola pengelompokan penyebaran penyakit di setiap kecamatan berdasarkan data rekam medis pasien dari rumah sakit anwar medika sebanyak 534 data dengan waktu penyelesaian sebanyak 0.06 detik.

Penelitian Nurdiyansyah & Akbar [22] dengan judul "Implementasi Algoritma k-Means untuk Menentukan Persediaan Barang pada *Poultry Shop"* Penerapkan Algoritma K-means *cluster*ing pada CV. Muria PS menghasilkan kelompok persediaan barang yang laris, cukup laris dan kurang laris dari

data jumlah stok barang dan jumlah barang terjual pada setiap item barang yang ada. Penelitian Hardiani [23] dengan judul "Analisis *Cluster*ing Kasus Covid 19 di Indonesia Menggunakan Algoritma k-Means" Penerapan algoritma K Means pada penelitian ini menghasilkan 3 *cluster* yang optimal berdasarkan perhitungan Davies Bouldin Index. Nilai perhitungan DBI terendah sebesar 0,474. Penelitian Yolanda & Suhardi [14] dengan judul "Penerapan Algoritma k-Means *Cluster*ing Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba" Menghasilkan tiga kelompok (*cluster*) berdasarkan kriteria mereka. *Cluster* pertama merupakan kelompok umur 26- 45 tahun dengan lama penggunaan 1->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, inex, dan extacy. *Cluster* kedua adalah kelompok umur 55 tahun dengan lama penggunaan 3->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, dan extacy.

Penelian terdahulu dengan penelitian ini memiliki persamaan yaitu menggunakan algoritma yang sama dalam mengelompokkan data. Tetapi ada perbedaan penelitian terdahulu dengan penelitian ini yaitu pada penelitian yang dilakukan [8]peneliti hanya berfokus pada variabel kepemilikan aset yang dikelompokkan menjadi 2 *cluster*. Sedangkan penelitian yang akan dilakukan berfokus pada data penduduk Desa Sei Litur Tasik yang akan dikelompokkan menjadi 2 *cluster*, maka penulis akan melakukan penelitian menggunakan metode yang sama dengan penelitian terdahulu namun data yang digunakan berbeda.

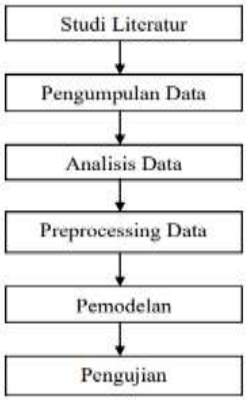
3 Metode Penelitian

Penelitian ini dilakukan di Desa Sei Litur Tasik Kec. Sawit Seberang, Kabupaten Langkat, Sumatera Utara 20811. Pada penelitian ini menggunakan metode penelitian kuantitatif, yaitu penelitian ilmiah yang terdiri dari variabel-variabel yang diukur dengan angka atau numerik, dimulai dengan data yang dikumpulkan, data yang diolah serta tampilan dan hasilnya. Pada penelitian ini pengujian dilakukan dengan menggunakan Bahasa pemrograman python dan memanfaatkan tools Google collaboratory. Penggunaan keduanya sebagai pengujian bertujuan untuk membandingkan hasil perhitungan yang diperoleh secara manual dan agar mengetahui apakah fungsinya berjalan dengan baik atau tidak. Dalam tahap pengujian ini, peneliti menggunakan library yang ada pada python sebagai pembantu dalam mengolah dan eksekusi data. Rencana pembahasan ini memuat daftar isi pembahasan dan hasil penelitian, yang mana rencana pembahasan ini merupakan jawaban dari rumusan masalah pada penelitian ini yaitu sebagai berikut:

1. Kriteria setiap kelompok penduduk miskin yang dihasilkan dari pengelompokan data menggunakan algoritma k-Means *clustering*. Pada tahap ini akan dilakukan analisis deskriptif terhadap kriteria setiap kelompok penduduk miskin yang terbentuk setelah dilakukan pengelompokan menggunakan algoritma k-Means *clustering*.
2. Penerapan algoritma k-Means *clustering* pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik. Pada tahapan ini penulis akan menjelaskan penerapan algoritma k-Means *clustering* pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik. Tahapan penerapan algoritma k-Means *clustering* pada penelitian ini meliputi *preprocessing data*, penentuan nilai k, inisialisasi *centroid*, dan iterasi hingga konvergen.
3. Pemodelan *clustering* data penduduk miskin Desa Sei Litur Tasik menggunakan algoritma k-Means *clustering*. Pada tahapan ini penulis akan menjelaskan mengenai pemodelan *clustering* pada data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik menggunakan algoritma k-Means *clustering*. Pemodelan ini meliputi normalisasi data, pemilihan nilai k, pemilihan metode inisialisasi *centroid*, dan evaluasi hasil pengelompokan. Selain itu akan dibahas juga mengenai interpretasi hasil pengelompokan dan kesimpulan dari pemodelan *clustering* yang telah dilakukan.

Tahapan penelitian yaitu sebuah langkah yang akan dilaksanakan dalam penelitian ini yang terdiri dari studi literatur, pengumpulan data, analisis data, *preprocessing data*, pemodelan, dan pengujian, Tahapan dalam penelitian ini dapat dilihat pada Gambar 1.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 1. tahap-tahap kerangka penelitian

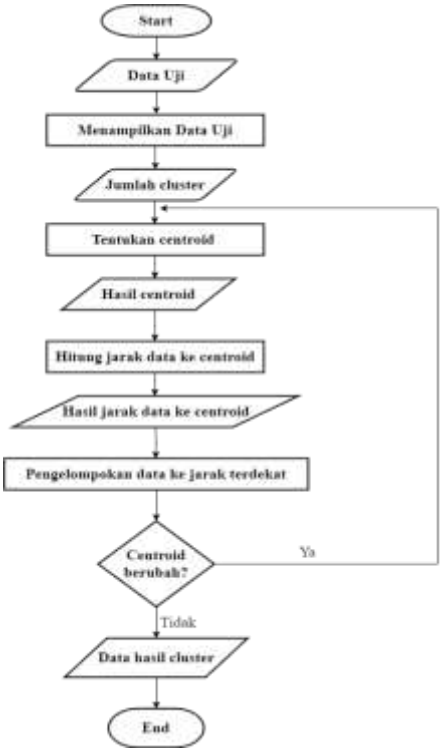
Studi literatur dilakukan untuk mengumpulkan pengetahuan dengan cara mempelajari sumber kepustakaan untuk digunakan sebagai acuan data yang berhubungan dengan penelitian yang akan dilakukan. Pada tahapan ini peneliti menggunakan sumber literatur berupa buku-buku, jurnal dan karya ilmiah lainnya. Pada penelitian ini tahap pengumpulan data yang akan dilakukan yaitu:

1. Wawancara, merupakan metode pengumpulan data yang dilakukan secara langsung dengan narasumber. Wawancara yang dilakukan dengan sekretaris Desa Sei Litur Tasik, yaitu Bapak Didi Darmawan mendapatkan data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) Desa Sei Litur Tasik yang akan digunakan untuk pengelompokan data.
2. Observasi, adalah metode pengumpulan data yang dilakukan dengan pengamatan langsung ke lapangan. Pada penelitian ini penulis melakukan pengamatan langsung terhadap penduduk Desa Sei Litur Tasik untuk mendapatkan data berupa informasi yang akan diteliti.

Dalam penelitian ini dibutuhkan Analisa data agar dapat terlaksana dengan baik. Pada tahap ini meliputi kebutuhan apa saja yang digunakan dalam penelitian, data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) menjadi fokus utama. Data yang diperlukan sebanyak 426 Data keluarga penduduk Desa Si Litur Tasik meliputi profil penduduk, pekerjaan kepala keluarga, memiliki simpanan uang, kepemilikan rumah, dinding rumah, dan resiko stunting.

Selain data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE), algoritma k-Means dipilih dalam pengelompokan dikarenakan algoritma k-Means mampu mengelompokkan data yang memiliki kriteria sama kedalam satu *cluster* yang sama dan data yang memiliki kriteria berbeda di kelompokkan ke dalam *cluster* yang lain. Pengelompokan data yang diperoleh yaitu tingkat kemiskinan penduduk menjadi 2 golongan (Kaya dan miskin). Hasil yang diperoleh digunakan sebagai pengetahuan atau informasi yang bermanfaat bagi pengguna kebijakan pada proses pengambilan Keputusan dalam pembagian bantuan sosial kepada masyarakat agar tepat sasaran. Setelah pengumpulan data dan analisis data, tahapan selanjutnya yaitu *preprocessing data* yang berguna untuk membuang data yang tidak konsisten sehingga diperoleh data bersih, data yang akan digunakan untuk klusterisasi dengan algoritma k-Means. Pada tahap ini juga dilakukan seleksi data dan transformasi data menjadi data *numerik*. Pada penelitian ini dilakukan pemodelan dimana *preprocessing data* dan penerapan algoritma k-Means untuk menggali data dan pengelompokan dengan kriteria-kriteria (parameter) yang telah ditentukan. Adapun parameter yang digunakan adalah kepemilikan rumah, pekerjaan, resiko stunting, memiliki simpanan uang/perhiasan dan lainnya, jenis lantai, jenis dinding, fasilitas jamban, serta sumber penerangan. Teknik *data mining* yang digunakan adalah *clustering*. Adapun model yang telah dirancang dalam bentuk *flowchart* dapat dilihat pada Gambar 2.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 2. flowchart k-means clustering

4 Hasil dan Pembahasan

Dalam bagian ini peneliti akan membahas mengenai hasil analisis dan implementasi sistem yang di hasilkan dalam penelitian ini sebagai berikut:

4.1 Analisis Data

Hasil dari tahapan ini adalah pengumpulan dan eksplorasi data. Data yang digunakan dalam penelitian ini adalah data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrean tahun 2022. Eksplorasi dilakukan untuk melihat apakah ada data yang kosong atau duplikat. Data yang digunakan dalam penelitian ini adalah sebanyak 426 data dengan sampel yang digunakan untuk pengujian manual sebanyak 20 data.

4.2 Representasi Data

Dalam penerapan algoritma K-means untuk pengelompokan data penduduk miskin, data direpresentasikan dalam bentuk atribut-atribut yang relevan. Atribut yang digunakan dalam penelitian ini meliputi pekerjaan, memiliki simpanan uang, kepemilikan rumah, dinding rumah, dan resiko stunting. Adapun representasi data pada penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Representasi data sampel

No	K. Keluarga	Pekerjaan	Kepemilikan Rumah	Memiliki Simpanan	Jenis Dinding	Jenis Lantai	Sumber Penerangan	Fasilitas Jamban	R. Stunting
1	HP	Wiraswasta	Menumpang	Ya	Tembok	Keramik/Granit/Marmer/Ubin/Tegel/Teraso	Listrik Pribadi s/d 900 Watt	Ya, dengan Septic Tank	2
2	SD	Pegawai Swasta	Menumpang	Ya	Tembok	Keramik/Granit/Marmer/Ubin/Tegel/Teraso	Listrik Pribadi s/d 900 Watt	Ya, dengan Septic Tank	2

 ISSN:2302-8149
e-ISSN:2540-9719

3	MD	Wiraswasta	Kontrak/Sewa	Tidak	Kayu/ Papan	Semen	Listrik Pribadi > 900 Watt	Ya, tanpa Septic Tank	1
4	BI	Wiraswasta	Milik Sendiri	Ya	Tembok	Keramik/Granit/ Marmer/Ubin/Tegel/Teraso	Listrik Pribadi s/d 900 Watt	Ya, dengan Septic Tank	1
...	...	...	...	...	...	...	...	...	...
20	S	Wiraswasta	Milik Sendiri	Tidak	Tembok	Semen	Listrik Pribadi > 900 Watt	Ya, dengan Septic Tank	1

4.3Analisis Algoritma k-Means Clustering

Dalam clustering data, langkah awal yang dilakukan untuk pembentukan cluster adalah mentransformasikan data kedalam bentuk numerik dengan kode-kode yang telah ditentukan, lalu tentukan jumlah group (K), hitung centroid, hitung jarak objek ke centroid, kemudian clusterkan ke jarak terdekat, jika cluster tidak berubah maka iterasi selesai. Untuk menentukan cluster dari suatu objek, pertama yang harus dilakukan adalah mengukur jarak Eulidean antara atribut-atribut yang didefinisikan dapat dilihat pada Tabel 2,3,4,5,6,7,8 dan 9.

Tabel 2. Pekerjaan

Bobot	Pekerjaan (P)
1	Tidak/Belum bekerja
2	Pensiunan
3	Pekerja lepas
4	Petani
5	Wiraswasta
6	Pedagang
7	Pegawai Swasta

Tabel 3. Kepemilikan rumah

Bobot	Kepemilikan Rumah (KR)
1	Kontrak/Sewa
2	Bebas Sewa
3	Menumpang
4	Milik sendiri

Tabel 4. Memiliki simpanan

Bobot	Memiliki Simpanan (MS)
1	Tidak
2	Ya

Tabel 5. Jenis dinding

Bobot	Jenis Dinding (JD)
1	Bambu
2	Kayu/papan
3	Tembok

Tabel 6. Jenis lantai

Bobot	Jenis Lantai (JL)
1	Tanah
2	Kayu/papan

 ISSN:2302-8149
e-ISSN:2540-9719

3	Semen
4	Keramik/Granit/Marmer/Ubin/Tegel/Teraso

Tabel 7. Sumber penerangan

Bobot	Sumber Penerangan (SP)
1	Non-Listrik
2	Listrik bersama
3	Listrik pribadi s/d 900 watt
4	Listrik pribadi > 900 watt

Tabel 8. Fasilitas jamban

Bobot	Memiliki Fasilitas Jamban (FJ)
1	Tidak, jamban umum/bersama
2	Ya, tanpa septic tank
3	Ya, dengan septic tank

Tabel 9. Resiko stunting

Kode	Resiko Stunting (RS)
1	2 orang
2	1 orang
3	Tidak ada

Dari kriteria-kriteria diatas telah dilakukan pengkodean selanjutnya yaitu transformasi data kriteria menjadi data numerik untuk dihitung menggunakan algoritma k-means clustering dan teknik *Euclidean Distance* dijelaskan dalam Tabel 10.

Tabel 10. Data transformasi

NO	Kepala Keluarga	P	KR	MS	JD	JL	SP	FJ	RS
1	HP	5	3	2	3	4	3	3	1
2	SD	7	3	2	3	4	3	3	1
3	MD	5	1	1	2	3	4	2	2
4	BI	5	4	2	3	4	3	3	2
...	...	...	...	...	...	...	...	...	..
20	S	5	4	1	3	3	4	3	2

Setelah mentransformasi data selanjutnya membentuk cluster menjadi 2 kelompok

Setelah mentransformasi data selanjutnya membentuk cluster menjadi 2 kelompok (k=2) dan menentukan titik pusat *centroid*. Adapun proses perhitungan k-means clustering menggunakan rumus persamaan *Euclidean Distance* yaitu:

$$d_{Euclidean}(x,y) = \sum_i (x_i - y_i)^2 \text{.....(1)}$$

Menentukan titik K- 2 *Centroid*

C1 = (5, 1, 1, 2, 3, 4, 2, 2) diambil dari data ke- 3

C2 = (1, 3, 1, 2, 3, 3, 3, 3) diambil dari data ke- 16

Selanjutnya dilakukan proses perhitungan seperti dibawah ini.

1. Data ke-1 (5, 3, 2, 3, 4, 3, 3, 1)

$$C1=\sqrt{\begin{matrix}(5-5)^2+(3-1)^2+(2-1)^2+(3-2)^2+\\(4-3)^2+(3-4)^2+(3-2)^2+(1-2)^2\end{matrix}} = 3{,}162278$$

$$C2=\sqrt{\begin{matrix}(5-1)^2+(3-3)^2+(2-1)^2+(3-2)^2+\\(4-3)^2+(3-3)^2+(3-3)^2+(1-3)^2\end{matrix}} = 4{,}795832$$

2. Data ke-2 (7, 3, 2, 3, 4, 3, 3, 1)

 ISSN:2302-8149
e-ISSN:2540-9719

$$C1=\sqrt{\begin{matrix}(7-5)^2+(3-1)^2+(2-1)^2+(3-2)^2+\\(4-3)^2+(3-4)^2+(3-2)^2+(1-2)^2\end{matrix}}=3{,}741657$$

$$C2=\sqrt{\begin{matrix}(7-1)^2+(3-3)^2+(2-1)^2+(3-2)^2+\\(4-3)^2+(3-3)^2+(3-3)^2+(1-3)^2\end{matrix}}=6{,}557439$$

3. Data ke-3 (5, 1, 1, 2, 3, 4, 2, 2)

$$C1=\sqrt{\begin{matrix}(5-5)^2+(1-1)^2+(1-1)^2+(2-2)^2+\\(3-3)^2+(4-4)^2+(2-2)^2+(2-2)^2\end{matrix}}=0$$

$$C2=\sqrt{\begin{matrix}(5-1)^2+(1-3)^2+(1-1)^2+(2-2)^2+\\(3-3)^2+(4-3)^2+(2-3)^2+(2-3)^2\end{matrix}}=4{,}795832$$

Dari perhitungan diatas maka diperoleh hasil perhitungan iterasi 1, yaitu seperti pada tabel dibawah ini.

Tabel 11. Hasil iterasi 1

No.	KK	P	KR	MS	JD	JL	SP	FJ	RS	C1	C2	G
1	HP	5	3	2	3	4	3	3	1	3,162278	4,795832	1
2	SD	7	3	2	3	4	3	3	1	3,741657	6,557439	1
3	MD	5	1	1	2	3	4	2	2	0	4,795832	1
4	BI	5	4	2	3	4	3	3	2	3,741657	4,582576	1
...	...	...	...	...	...	...	...	...	...	...	...	...
20	S	5	4	1	3	3	4	3	2	3,316625	4,472136	1

Setelah dilakukan perhitungan menggunakan teknik *Euclidean Distance*, maka group cluster berdasarkan jarak terdekat centroid adalah:

Group lama: (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)

Group baru: (1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1)

Terjadi perubahan group cluster, maka akan dilakukan iterasi selanjtnya yaitu iterasi ke-2 dengan K=2 sebagai berikut:

Centroid 1 Group 1

$C1 = \left(\frac{5+7+5+5+5+5+5+5+5+5+5+5+5+5+5}{15}\right) = 5{,}133333$

$\left(\frac{3+3+1+4+3+4+3+1+1+4+3+4+4+4+4}{15}\right) = 3{,}066667$

$\left(\frac{2+2+1+2+2+2+2+1+1+1+1+1+2+1+1}{15}\right) = 1{,}466667$

$\left(\frac{3+3+2+3+3+3+3+2+2+2+3+3+2+3+3}{15}\right) = 2{,}666667$

$\left(\frac{4+4+3+4+4+4+4+3+3+3+3+4+3+3+3}{15}\right) = 3{,}466667$

$\left(\frac{3+3+4+3+4+3+4+3+4+3+3+3+4+4+4}{15}\right) = 3{,}466667$

$\left(\frac{3+3+2+3+3+3+3+3+2+3+3+3+3+3+3}{15}\right) = 2{,}866667$

$\left(\frac{1+1+2+2+1+1+1+1+2+2+1+2+2+3+2}{15}\right) = 1{,}6$

Centroid baru

C1= (5,133333; 3,066667; 1,466667; 2,666667; 3,466667; 3,466667; 2,866667; 1,6)

Centroid 2 Group 2

$C2 = \left(\frac{3+3+1+1+2}{5}\right) = 2$, $\left(\frac{4+4+4+3+4}{5}\right) = 3{,}8$, $\left(\frac{2+1+1+1+1}{5}\right) = 1{,}2$, $\left(\frac{3+2+2+2+3}{5}\right) = 2{,}4$,

$\left(\frac{4+3+3+3+3}{5}\right) = 3{,}2$, $\left(\frac{3+3+3+3+4}{5}\right) = 3{,}2$, $\left(\frac{3+3+3+3+3}{5}\right) = 3$, $\left(\frac{2+2+3+3+3}{5}\right) = 2{,}6$

Centroid baru

C2 = (2; 3,8; 1,2; 2,4; 3,2; 3,2; 3; 2,6)

Selanjutnya dilakukan proses perhitungan seperti dibawah ini

 ISSN:2302-8149
e-ISSN:2540-9719

1. Data ke-1 (5, 3, 2, 3, 1, 4, 3, 1)

$$C1=\sqrt{\begin{matrix}(5-5{,}133333)^2+(3-3{,}066667)^2+\\(2-1{,}466667)^2+(3-2{,}666667)^2+\\(1-3{,}466667)^2+(4-3{,}466667)^2+\\(3-2{,}866667)^2+(1-1{,}6)^2\end{matrix}}=1{,}1392$$

$$C2=\sqrt{\begin{matrix}(5-2)^2+(3-3{,}8)^2+(2-1{,}2)^2+\\(3-2{,}4)^2+(1-3{,}2)^2+(4-3{,}2)^2+\\(3-3)^2+(1-2{,}6)^2\end{matrix}}=3{,}725587$$

2. Data ke-2 (7, 3, 2, 3, 4, 3, 3, 1)

$$C1=\sqrt{\begin{matrix}(7-5{,}133333)^2+(3-3{,}066667)^2+\\(2-1{,}466667)^2+(3-2{,}666667)^2+\\(4-3{,}466667)^2+(3-3{,}466667)^2+\\(3-2{,}866667)^2+(1-1{,}6)^2\end{matrix}}=2{,}182761$$

$$C2=\sqrt{\begin{matrix}(7-2)^2+(3-3{,}8)^2+(2-1{,}2)^2+\\(3-2{,}4)^2+(4-3{,}2)^2+(3-3{,}2)^2+\\(3-3)^2+(1-2{,}6)^2\end{matrix}}=5{,}46626$$

3. Data ke-3 (5, 1, 1, 2, 3, 4, 2, 2)

$$C1=\sqrt{\begin{matrix}(5-5{,}133333)^2+(1-3{,}066667)^2+\\(1-1{,}466667)^2+(2-2{,}666667)^2+\\(3-3{,}466667)^2+(4-3{,}466667)^2+\\(2-2{,}866667)^2+(2-1{,}6)^2\end{matrix}}=2{,}522785$$

$$C2=\sqrt{\begin{matrix}(5-2)^2+(1-3{,}8)^2+(1-1{,}2)^2+\\(2-2{,}4)^2+(3-3{,}2)^2+(4-3{,}2)^2+\\(2-3)^2+(2-2{,}6)^2\end{matrix}}=4{,}368066$$

Dari perhitungan yang telah dilakukan dengan *centroid* baru, maka diperoleh hasil perhitungan iterasi 2 yaitu seperti pada Tabel 12 dibawah ini.

Tabel 12. Hasil iterasi 2

NO	KK	P	KR	MS	JD	JL	SP	FJ	RS	C1	C2	C
1	HP	5	3	2	3	4	3	3	1	1,1392	3,725587	1
2	SD	7	3	2	3	4	3	3	1	2,182761	5,46626	1
3	MD	5	1	1	2	3	4	2	2	2,522785	4,368066	1
4	BI	5	4	2	3	4	3	3	2	1,401586	3,328663	1
...	...	...	...	...	...	...	...	...	..	...	...	1
20	S	5	4	1	3	3	4	3	2	1,377599	3,237283	1

Setelah dilakukan perhitungan menggunakan teknik *Euclidean Distance*, maka group cluster berdasarkan jarak terdekat centroid adalah:
Group lama: (1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1)
Group baru: (1 1 1 1 1 1 2 1 2 1 1 1 2 1 2 2 1 1 1 1)
Karena tidak terjadi perubahan pada cluster, maka proses iterasi dapat berhenti dan perhitungan selesai.

4.1 Pemodelan k-Means *Clustering*

1. Pengujian Model k-Means *Clustering*

Pada penelitian ini sistem yang telah dirancang dilakukan proses pemodelan yang menggunakan *Google Colaboraty* dan bahasa pemograman python serta pustaka – pustaka pendukung untuk mendapatkan hasil pengelompokan. Berikut ini adalah penjelasan tahapan – tahapan yang

 ISSN:2302-8149
e-ISSN:2540-9719

digunakan prose pemodelan Pengelompokan Data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) menggunkan metode K – Means Clustering.

1. Menginstal *Python Libraries*

Pada tahap ini dilakukan proses impor pustaka pendukung yang ditunjukkan pada Gambar 3 berikut ini.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from openpyxl import Workbook
from sklearn.metrics import silhouette_score
import warnings
```

Gambar 3. *Python libraries*

Terdapat beberapa pustaka-pustaka yang di instal untuk mempermudah proses pengelompokan data. Pustaka Pandas berguna untuk membersihkan, memanipulasi dan menganalisis data. Lalu, pustaka numpy berguna untuk melakukan perhitungan saintifik seperti matriks, aljabar, dan statistik. Kemudian Pustaka Matplotlib digunakan untuk membuat visualisasi data, seperti plot, histogram, diagram batang, dan plot sebar. Selanjutnya, pustaka Scikit – learn berguna sebagai penyedia modul–modul algoritma yang siap pakai serta pustaka Openpyxl berguna untuk dapat membaca dan menulis file berekstensi xlsx, xlsm, xltx, dan xltm.

2. Menampilkan Dataset yang Ditrasformasi

Pada penelitian ini digunakan data yang berekstensi xlsx dan dibaca menggunakan pustaka Openpyxl lalu, dilakukan data ditransformasi serta ditampilkan ke dalam bentuk tabel menggunakan pustaka Pandas. Berikut ini adalah hasil pembacaan data yang dapat dilihat pada Gambar 4.

Dataset yang Ditransformasi:

	PK	KR	S	JD	JL	SP	KM	RS
0	5	4	2	2	3	3	2	2
1	5	3	1	2	3	3	3	1
2	5	3	1	2	3	2	3	2
3	5	4	2	3	3	3	3	2
4	5	3	2	2	3	3	2	3
..	..	..	..	..	..	..	..	..
421	4	4	2	3	3	3	3	1
422	5	4	2	3	3	4	3	1
423	5	4	2	3	4	4	3	2
424	5	3	2	2	3	3	3	1
425	4	4	2	3	4	3	3	3

[426 rows x 8 columns]

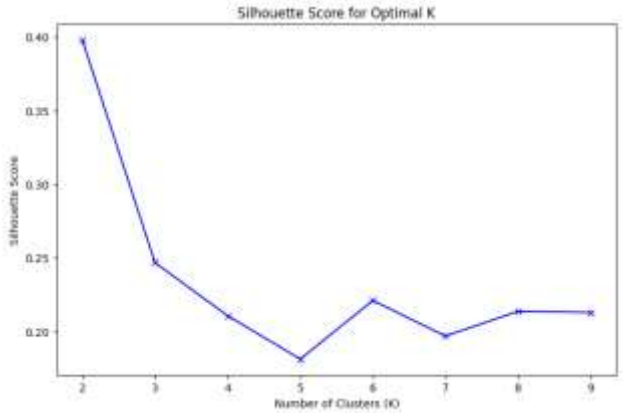
Gambar 4. Tampilan dataset yang ditransformasi

Pada data tersebut terdapat kolom – kolom yang berisikan penilaian dari setiap urutan data yaitu, PK, KR, S, JD, JL, SP, KM, dan RS.

3. Penentuan *Cluster* Optimal

Pada penelitian ini, digunakan *Silhouette Score* untuk menguji dan mengevaluasi model *K–Means Clustering* pada Pengelompokkan data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE). Metode *Silhouette Score* digunakan untuk menganalisis jumlah klaster optimal yang digunakan model pada proses pengelompokan data. *Silhouette Score* digunakan sebagai metode evaluasi kualitas clustering yang efektif dan efisien. grafik hasil pengujian model *K – Means Clustering* pada data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) yang di uji menggunakan *Google Colaboraty* dapat dilihat pada Gambar 5.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 5. Grafik optimal K

Pada Gambar 5 dijelaskan bahwa terdapat nilai klaster pengujian dan nilai interia dari setiap klaster. Pengujian model K–Means Clustering dilakukan dari klaster 2 sampai dengan klaster 10 dan hasil nilai interia yang didapat dari klaster 1 yaitu, 0,40 sampai dengan klaster terakhir yaitu 0,22. Pada Gambar 4.3 dapat dilihat bahwa, titik *Silhouette Score* berada di klaster 2 dengan nilai interia 0,40. Klaster 2 memiliki nilai yang paling besar artinya, klaster 2 dinilai sebagai klaster yang optimal untuk digunakan model K – Means Clustering dalam proses pengelompokan data pada penelitian ini.

4. Penentuan Titik *Centeroid*

Pada tahapan ini, ditentukan titik centeroid awal yang diambil secara acak dari dataset. Berikut ini adalah hasil penentuan centeroid yang dapat dilihat pada Gambar 6.

```
Data Centroid dari Setiap Cluster:
                PK       KR        S        JD        JL        SP        KM  \
Cluster
1         5.073770  3.52459  1.745902  2.672131  3.401639  3.319672  2.806011
2         2.116667  3.55000  1.466667  2.583333  3.233333  3.383333  2.733333

                RS
Cluster
1         1.855191
2         2.466667
```

Gambar 6. Centroid cluster

Terdapat 2 titik centeroid yang digunakan sebagai acuan untuk melakukan proses pengelompokan data menjadi 2 kategori.

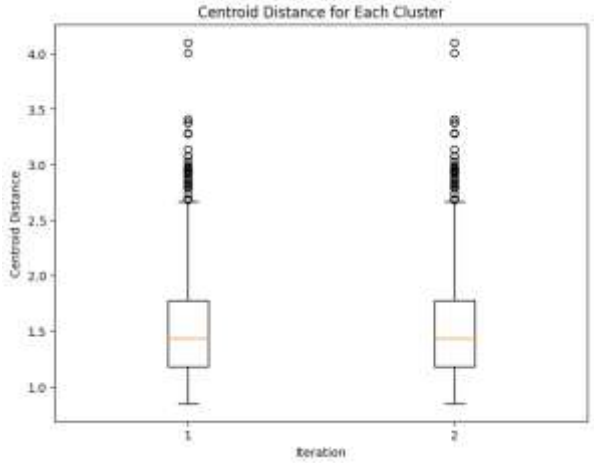
5. Perhitungan Jarak Data

Selanjutnya dilakukan perhitungan jarak antar setiap data dengan titik centeroid menggunakan pustaka Scikit – learn sehingga menghasilkan nilai jarak pada masing–masing data. Berikut ini adalah hasil perhitungan jarak data yang dapat dilihat pada gambar 7 dan 8.

```
[68] Jarak centroid per cluster pada Iterasi 2:
[1.29688491 1.52356645 1.80283568 0.85183703 1.7380185  1.66084714
 1.33426801 1.86100275 0.96038932 0.96038932 0.85183703 2.73590738
 3.41067884 1.60734247 1.36863703 1.778422   0.85183703 1.05786055
 0.85183703 1.19833586 1.0343542  2.26736121 2.15110801 2.10488583
```

Gambar 7. Jarak centroid setiap cluster

 ISSN:2302-8149
e-ISSN:2540-9719

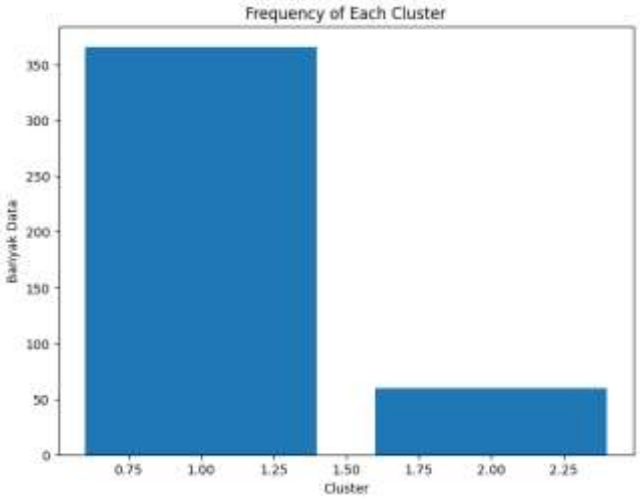


Gambar 8. Grafik jarak centroid setiap cluster

Pada Gambar 8 dapat diketahui terdapat nilai – nilai hasil perhitungan jarak kedekatan data ke centroid yang dilakukan pada iterasi pertama. Selanjutnya, dilakukan kembali serangkaian proses pengelompokan pada iterasi kedua dan menghasilkan jumlah anggota dalam setiap cluster dapat dilihat pada Gambar 9 dan 10.

```
Jumlah Anggota dalam Setiap Cluster:
 Cluster  Count
       1    366
       2     60
```

Gambar 9. Jumlah Anggota Setiap Cluster



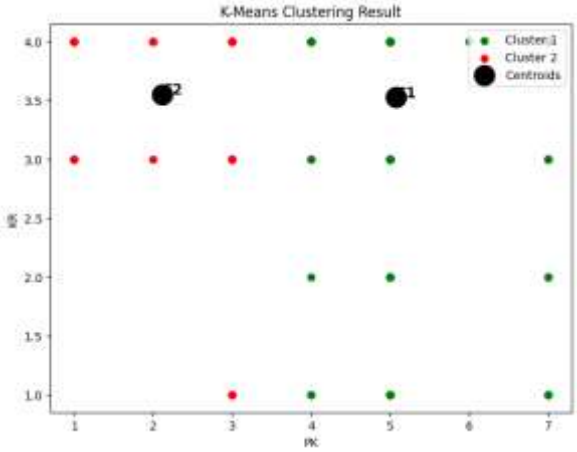
Gambar 10. Grafik jumlah anggota setiap cluster

Pada iterasi kedua diketahui nilai jarak kedekatan antara data dengan titik centeroid bernilai sama dengan hasil perhitungan jarak pada iterasi pertama sehingga, dari penjelasan tersebut ditetapkan nilai jarak kedekatan antar data pada iterasi kedua sebagai hasil akhir pengelompokan data pada penelitian ini. Data dikelompokan menjadi 2 kategori yaitu, kaya sebanyak 366 Keluarga dan miskin sebanyak 60 Keluarga.

6. Visualisasi Hasil Pengelompokan

Kemudian agar dapat mempermudah pembacaan hasil pengelompokan dilakukan visualisasi persebaran data yang dapat dilihat pada Gambar 10.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 10. Visualisasi hasil clustering

Pada Gambar 10 dapat diketahui bahwa, terdapat 3 titik dengan warna serta posisi yang berbeda. Titik berwarna hijau adalah data yang dikelompokan pada klaster 1 dengan kategori kelompok kaya. Lalu, titik berwarna merah adalah data yang dikelompokan pada klaster 2 dengan kategori kelompok miskin serta titik berwarna 2 titik hitam yang menandakan titik ketiga *centroid*.

7. Menghitung Akurasi *k-Means Clustering*

Pengujian adalah metode sistematis untuk mengevaluasi model *k-Means Clustering* dan memberikan pemahaman tentang seberapa akurat model tersebut dapat memberikan hasil. Akurasi pada model *k-Means Clustering* pada penelitian ini hasilnya dapat dilihat pada Gambar 11.

```
#Menghitung akurasi clustering
accuracy = (df['Group'].value_counts().max() / len(df['Group'])) * 100
print(f"Overall Accuracy: {accuracy:.2f}%")
```

```
Overall Accuracy: 85.92%
```

Gambar 11. Akurasi k-means clustering

5 Kesimpulan

Berdasarkan hasil dan pembahasan yang telah dihasilkan pada penelitian ini, dapat disimpulkan bahwa Pada data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) yang berjumlah 426 data kepala keluarga meliputi profil penduduk, pekerjaan kepala keluarga, kepemilikan uang simpanan, kepemilikan rumah, keadaan dinding rumah, dan resiko terkena stunting dilakukan pengelompokan data menggunakan algoritma pengelompokan *K – Means* menjadi 2 kategori atau klaster (*k*) yaitu *cluster* 1 kategori kaya sebanyak 366 keluarga dan *cluster* 2 kategori miskin sebanyak 60 keluarga. Hasil pengujian dan evaluasi model *K-Means Clustering* pada data Percepatan Penghapusan Kemiskinan Ekstrem (P3KE) ditetapkan 2 klaster adalah jumlah klaster yang optimal dengan nilai interia 0,40 yang menggunakan metode pengujian *Silhouette Score.* Pemodelan rancangan sistem pengelompokan data dengan menggunkan metode *K – Means Clustering* dilakukan pada *Google Colaboraty* serta dibantu dengan pustaka – pustaka pendukung. Hasil penelitian menunjukkan *accuracy K-Means Clustering* sebesar 85,92% yang berarti bahwa *accuracy* dari data yang dianalisis dapat dengan tepat dikelompokkan ke dalam kategori *cluster* yang sesuai.

Referensi

[1] Y. R. Sembiring, Saifullah, and R. Winanjaya, “Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma,” *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen) Vol.*, vol. 2, no. 2, pp. 125–132, 2021.

[2] N. Novitasari, N. D. Nuris, and R. Herdiana, “Penerapan Algoritma *k-Means* untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten Di Jawa Barat Menggunakan Rapidminer,” *J. Inform. Terpadu*, vol. 9, no. 1, pp. 68–73, 2023.

 ISSN:2302-8149
e-ISSN:2540-9719

[3] Suhartini and R. Yuliani, "Infotek : Jurnal Informatika dan Teknologi Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K- Means di Dusun Bagik Endep Sukamulia Timur Infotek : Jurnal Informatika dan Teknologi Pendahuluan masalah kemiskanan belum bis," *Infotek J. Inform. dan Teknol.*, vol. 4, no. 1, pp. 39–50, 2021.

[4] F. I. Putri, R. Damayanti, and Kismiantini, "Penerapan Algoritma *k-Means* Untuk Mengelompokkan Kecamatan Di Kabupaten Gunung Kidul Berdasarkan Program Keluarga Harapan," in *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, Kedua.Samarinda, 2022, pp. 408–418.

[5] H. Kurniawan, S. Defit, and Sumijan, "Data Mining Menggunakan Metode *k-Means Clustering* Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020.

[6] U. Syafiyah, D. P. Puspitasari, I. Asrafi, B. Wicaksono, and F. M. Sirait, *Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering Pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020*. 2022. doi: 10.34123/semnasoffstat.v2022i1.1221.

[7] W. Wahyu Pribadi, A. Yunus, and A. S. Wiguna, "Perbandingan Metode *k-Means* Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 6, no. 2, pp. 493–500, 2022, doi: 10.36040/jati.v6i2.4808.

[8] Parjito and Permata, "Penerapan Data Mining untuk Clustering Data Penduduk Miskin Menggunakan Metode k-Means," *J. Inform.*, vol. 3, no. 1, pp. 31–37, 2021.

[9] E. A. Putri, S. Muchsin, and Hayat, "Evaluasi Pelaksaan Program Bantuan Sosial Bagi Masyarakat Terdampak Di Era Pandemi Covid-19 (Di Desa Kersik Putih Kecamatan Batulicin Kabupaten Tanah Bumbu)," *J. Inov. Penelit.*, vol. 1, no. 12, pp. 2851–2860, 2021.

[10] I. D. Id, *MACHINE LEARNING: Teori, Studi Kasus dan Implementasi Menggunakan Python*. 2021. doi: 10.5281/zenodo.5113507.

[11] S. N. Mayasari and J. Nugraha, "Implementasi *k-Means Cluster* Analysis untuk Mengelompokkan Kabupaten / Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," *KONSTELASI Konvergensi Teknol. dan Sist. Inf. Vol.*, vol. 3, no. 2, pp. 317–329, 2023.

[12] N. I. Febianto and N. D. Palasara, "Analisis *Clustering k-Means* Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. SISFOKOM*, vol. 08, no. 02, pp. 130–140, 2019.

[13] A. Adji *et al.*, *Pemeringkatan Kesejahteraan Data Pensasaran Percepatan Penghapusan Kemiskinan Ekstrem (P3KE)*, Pertama. Jakarta Pusat: Tim Nasional Percepatan Penanggulangan Kemiskinan, 2022.

[14] E. Yolanda and Suhardi, "Penerapan Algoritma *k-Means Clustering* Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba," *J. Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 1, pp. 182–191, 2023, doi: 10.30865/klik.v4i1.1107.

[15] F. Febriansyah and S. Muntari, "Penerapan Algoritma *k-Means* untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.

[16] M. Siahaan, "Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma k-Means," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 316–324, 2022, doi: 10.32736/sisfokom.v11i3.1453.

[17] T. Nafsiah Muthmainnah, S. Indriyana, U. Enri, J. Sistem Informasi, F. Ilmu Komputer, and U. Singaperbangsa Karawang, "Penerapan Algoritme k-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 5, no. 2, pp. 122–129, 2023.

[18] R. Rosmini, A. Fadlil, and S. Sunardi, "Implementasi Metode *k-Means* Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, vol. 3, no. 1, pp. 22–31, 2018, doi: 10.25299/itjrd.2018.vol3(1).1773.

[19] M. Harahap, A. W. D. R. Zamili, M. A. Arvansyah, E. F. Saragih, S. Rajen, and A. M. Husein, "*k-Means Clustering Algorithm Approach in Clustering Data on Cocoa Production Results in the Sumatra Region*," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 6, pp. 905–910, 2022, doi: 10.29207/resti.v6i6.4199.

[20] A. N. Nursia, W. Ramdhan, and W. M. Kifti, "Analisis Kelayakan Penerima Bantuan Covid-19

 ISSN:2302-8149
e-ISSN:2540-9719

Menggunakan Metode K–Means," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 574–583, 2022, doi: 10.47065/bits.v3i4.1399.

[21] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode k-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.

[22] F. Nurdiyansyah and I. Akbar, "Implementasi Algoritma *k-Means* untuk Menentukan Persediaan Barang pada Poultry Shop," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 2, pp. 86–94, 2021, doi: 10.26905/jtmi.v7i2.6377.

[23] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma k-Means," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 156–165, 2022, doi: 10.23887/janapati.v11i2.45376.