Jurnal Elektronika dan Teknik Informatika Terapan
Volume 4, Nomor 3, September 2026
E-ISSN : 2988-0874, P-ISSN : 2988-0866, Hal. 469-481
DOI: https://doi.org/10.59061/jentik.v4i3.1620
Tersedia: https://e-journal.poltek-kampar.ac.id/index.php/JENTIK

Pengelompokan Kabupaten/Kota Berdasarkan Jumlah dan Persentase Penduduk Miskin di Provinsi Sumatera Utara menggunakan Algoritma K-Means

Fahmi Firzada[1*], Saifullah[2], Eka Irawan[3], Wendi Robiansyah[4]
[1-4]Jurusan Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar, Indonesia
Email: fahmifirzada@amiktunasbangsa.ac.id[1], saifullah@amiktunasbangsa.ac.id[2], eka.irawan@amiktunasbangsa.ac.id[3], wendirobiansyah@amiktunasbangsa.ac.id[4]
**Penulis Korespondensi: fahmifirzada@amiktunasbangsa.ac.id*

Abstract*. Poverty is a social problem that has widespread impacts on society and remains a major concern for governments in their efforts to improve public welfare. This study aims to cluster districts and cities in North Sumatra Province based on the number and percentage of poor people using the K-Means Clustering algorithm. The data were obtained from Statistics Indonesia (BPS) and cover the period from 2019 to 2023. The K-Means algorithm was applied to group regions according to similarities in their poverty characteristics, particularly the number and percentage of poor people. The clustering process was implemented using RapidMiner software to facilitate data processing, analysis, and visualization of the resulting clusters. The findings indicate that K-Means can group districts and cities with similar poverty characteristics, providing an overview of the distribution of poverty across North Sumatra Province. The resulting clusters can serve as supporting information for the regional government in identifying areas that require greater attention and developing more targeted poverty alleviation policies. This study is expected to support more effective resource allocation and contribute to the formulation of sustainable poverty reduction strategies tailored to regional conditions and development priorities.*

Keywords*: Clustering; K-Means; North Sumatra; Poverty; RapidMiner.*

Abstrak. Kemiskinan merupakan permasalahan sosial yang berdampak luas terhadap masyarakat dan menjadi perhatian utama pemerintah dalam upaya meningkatkan kesejahteraan masyarakat. Penelitian ini bertujuan untuk mengelompokkan kabupaten dan kota di Provinsi Sumatera Utara berdasarkan jumlah dan persentase penduduk miskin menggunakan algoritma K-Means Clustering. Data penelitian diperoleh dari Badan Pusat Statistik (BPS) untuk periode 2019–2023. Algoritma K-Means diterapkan untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik kemiskinan, khususnya jumlah dan persentase penduduk miskin. Proses pengelompokan dilakukan menggunakan perangkat lunak RapidMiner untuk mempermudah pengolahan data, analisis, dan visualisasi hasil klasterisasi. Hasil penelitian menunjukkan bahwa algoritma K-Means dapat mengelompokkan kabupaten dan kota yang memiliki karakteristik kemiskinan serupa sehingga memberikan gambaran mengenai persebaran kemiskinan di Provinsi Sumatera Utara. Hasil pengelompokan tersebut dapat menjadi informasi pendukung bagi pemerintah daerah dalam mengidentifikasi wilayah yang memerlukan perhatian lebih serta menyusun kebijakan pengentasan kemiskinan secara tepat sasaran. Penelitian ini diharapkan dapat mendukung alokasi sumber daya yang lebih efektif serta membantu perumusan strategi pengurangan kemiskinan yang berkelanjutan, sesuai dengan kondisi wilayah dan prioritas pembangunan daerah, sehingga pelaksanaan program penanggulangan kemiskinan dapat direncanakan secara lebih terarah berdasarkan karakteristik masing-masing wilayah.

Kata Kunci: Kemiskinan; Klasterisasi; K-Means; Rapidminer; Sumatera Utara.

1. PENDAHULUAN

Kemiskinan merupakan salah satu isu pembangunan paling fundamental dan kompleks yang dihadapi oleh banyak negara berkembang, termasuk Indonesia. Fenomena kemiskinan tidak hanya berkaitan dengan keterbatasan pendapatan ekonomi semata, tetapi juga mencakup keterbatasan akses terhadap kebutuhan dasar seperti pangan, pendidikan, layanan kesehatan, perumahan yang layak, dan lapangan kerja. Dalam upaya mewujudkan kesejahteraan masyarakat yang merata, pemerintah Indonesia menempatkan penanggulangan kemiskinan sebagai salah satu prioritas utama dalam Rencana Pembangunan Jangka Menengah Daerah

Naskah Masuk: 29 Juli 2026; Revisi: 18 Agustus 2026; Diterima: 27 September 2026; Terbit: 30 September 2026

(RPJMD) maupun Nasional. Oleh karena itu, ketersediaan informasi mengenai pemetaan dan persebaran penduduk miskin secara akurat dan tepat sasaran menjadi kebutuhan mendesak agar intervensi kebijakan yang dirancang dapat berjalan efektif.

Provinsi Sumatera Utara, sebagai salah satu wilayah terbesar dan berpenduduk padat di Indonesia, memiliki dinamika sosial-ekonomi yang sangat beragam di 33 kabupaten dan kota yang ada di dalamnya. Ketimpangan pertumbuhan ekonomi, perbedaan potensi sumber daya alam, serta perbedaan tingkat aksesibilitas infrastruktur antarwilayah menyebabkan distribusi tingkat kemiskinan tidak tersebar secara merata. Beberapa daerah mengalami penurunan jumlah penduduk miskin yang cukup signifikan, sementara wilayah lain masih berkutat dengan angka kemiskinan yang tinggi baik dari segi jumlah jiwa maupun persentase populasi. Kondisi keterpaduan data yang beragam ini menuntut adanya suatu pendekatan analitis yang dapat mengelompokkan wilayah-wilayah tersebut berdasarkan tingkat kerentanan kemiskinannya.

Badan Pusat Statistik (BPS) Provinsi Sumatera Utara secara berkala menerbitkan data mengenai jumlah dan persentase penduduk miskin per kabupaten/kota. Namun, data yang disajikan dalam bentuk tabel statistik konvensional sering kali sulit diinterpretasikan secara langsung oleh para pembuat kebijakan untuk menentukan prioritas penanganan wilayah. Volume data yang terus berkembang dari tahun ke tahun (historical data) membutuhkan pengolahan berbasis teknologi informasi agar dapat mentransformasikan sekumpulan data mentah (raw data) menjadi pengetahuan yang bernilai strategis (knowledge discovery). Dalam konteks inilah penerapan teknik data mining menjadi penting untuk mengekstraksi pola tersembunyi dari himpunan data kemiskinan tersebut.

Salah satu metode data mining yang sangat efektif dan populer digunakan dalam analisis pengelompokan (clustering) adalah algoritma K-Means. Algoritma K-Means merupakan metode pembelajaran tanpa pengawasan (unsupervised learning) yang bekerja dengan cara membagi himpunan data ke dalam sejumlah k klaster berdasarkan tingkat kemiripan karakteristiknya (similarity). Algoritma ini mengukur jarak antar data menggunakan rumus Euclidean Distance dan memperbarui titik pusat (centroid) secara berulang hingga mencapai kondisi konvergen. Keunggulan K-Means terletak pada efisiensi komputasinya yang tinggi, fleksibilitas dalam menangani data kuantitatif, serta kemampuannya dalam menghasilkan kelompok-kelompok daerah yang memiliki homogenitas internal yang tinggi namun memiliki perbedaan yang jelas antar kelompok.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengaplikasikan algoritma K-Means dalam mengelompokkan daerah kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator jumlah dan persentase penduduk miskin. Selain pengujian dan

perhitungan secara manual untuk memastikan keakuratan matematis algoritma, penelitian ini juga memanfaatkan perangkat lunak RapidMiner sebagai sarana verifikasi dan visualisasi data. Hasil dari pengelompokan ini diharapkan dapat memberikan pemetaan yang jelas mengenai wilayah mana saja yang masuk dalam kategori tingkat kemiskinan tinggi, sedang, maupun rendah, sehingga dapat dijadikan acuan dan landasan pertimbangan bagi pemerintah daerah dalam menetapkan prioritas alokasi bantuan dan program pengentasan kemiskinan secara lebih tepat sasaran.

2. KAJIAN TEORITIS

Data Mining

Data Mining merupakan bidang ilmu yang mengintegrasikan berbagai metode analitis seperti pembelajaran mesin, pengenalan pola, statistika, basis data, dan *visualisasi*, untuk mengeksplorasi dan mengekstraksi informasi penting dari sejumlah basis data. Tujuannya adalah menemukan relasi yang tidak terduga dan menyajikan data secara lebih informatif serta aplikatif bagi pemiliknya. Data Mining memiliki berbagai manfaat, diantaranya digunakan untuk Deskripsi, estimasi, prediksi, klasifikasi, pengelompokan (*clustering*), dan penemuan hubungan (*asosiasi*) adalah beberapa aspek dari Data Mining yang juga dikenal sebagai *Knowledge Discovery in Databases* (KDD). Proses ini bertujuan untuk mengekstrak pengetahuan dari kumpulan data.

Clustering

Clustering merupakan teknik yang digunakan untuk mengelompokkan data dalam suatu kumpulan berdasarkan karakteristik yang dimiliki oleh masing masing data. Dalam penelitian ini, metode *clustering* yang diterapkan adalah *K-means*, yang berfungsi untuk mengelompokkan data berdasarkan jarak antara data tersebut dan pusat setiap klaster. Tujuannya adalah untuk meminimalkan kesamaan dalam kelompok yang berbeda dan memaksimalkan kesamaan di dalam satu kelompok. Salah satu keunggulan *K-Means* adalah kemudahannya dalam implementasi, kemampuannya untuk menangani data dalam jumlah besar, dan kecepatannya dalam melakukan komputasi. Namun, metode ini memiliki keterbatasan, yaitu jumlah klaster harus ditentukan secara manual. Jika jumlah klaster tidak ditentukan dengan tepat, hasil pengelompokan data bisa kurang optimal.

K-Means

K-Means adalah salah satu metode yang digunakan dalam proses clustering. Algoritma *K-Means* menghasilkan titik centroid yang sesuai dengan tujuannya. Setelah iterasi *K-Means* berakhir, setiap objek dalam dataset akan tergabung dalam suatu klaster. Penentuan klaster

dilakukan dengan mencari objek yang memiliki jarak paling dekat satu sama lain, algoritma ini berfungsi dengan mengelompokkan data ke dalam klaster berdasarkan jarak terpendek antara elemen-elemen dalam dataset.

Langkah-langkah yang terdapat dalam algoritma K-Means Clustering dapat diuraikan sebagai berikut:

a. Tentukan jumlah cluster yang ingin dibentuk, yang dilambangkan dengan k.
b. Tetapkan Menentukan nilai secara acak untuk pusat awal cluster (centroid) sesuai dengan jumlah cluster (k).
c. Langkah selanjutnya adalah menghitung jarak antara setiap data dengan masing-masing centroid menggunakan rumus Euclidean Distance. Perhitungan ini dilakukan untuk menentukan kedekatan relatif antara data dan centroid, sehingga dapat diperoleh pasangan data dan centroid dengan jarak terdekat. Adapun rumus Euclidean Distance yang digunakan adalah sebagai berikut:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

d. Kelompokkan setiap data berdasarkan jarak terdekatnya dengan centroid.
e. Perbarui nilai centroid dengan menghitung rata-rata (mean) dari semua data pada setiap fitur dalam satu cluster. Apabila M merupakan total jumlah data dalam sebuah kluster, di mana i menunjukkan fitur ke-i , dan p adalah dimensi data, maka pusat untuk fitur ke-i dapat dihitung dengan rumus berikut:

$$C_i = \frac{1}{M} \sum_{j = i}^{M} x_j$$

f. Lakukan langkah 2 hingga 5 berulang kali sampai tidak ada lagi perubahan anggota di setiap cluster.
g. Ketika kondisi pada langkah 6 terpenuhi, nilai centroid dari iterasi terakhir akan digunakan sebagai acuan untuk menetapkan pengelompokan data akhir.

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam studi ini adalah pendekatan kuantitatif dengan menerapkan teknik data mining khusus untuk tugas pengelompokan (clustering). Penelitian ini mengacu pada kerangka kerja penyelesaian masalah secara sistematis yang dimulai dari pengumpulan data, pra-pemrosesan data (data preprocessing), penerapan algoritma komputasi, hingga pengujian dan evaluasi hasil klaster. Pendekatan ini dipilih karena mampu mengolah himpunan data numerik berukuran besar dari indikator kependudukan dan

mentransformasikannya menjadi informasi spasial dan kelompok wilayah yang terstruktur tanpa membutuhkan variabel target (unsupervised learning).

Tahap awal penelitian dimulai dengan proses pengumpulan data sekunder yang diperoleh dari laporan resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data yang digunakan mencakup dua variabel utama, yaitu jumlah penduduk miskin (dalam ribuan jiwa) dan persentase penduduk miskin di 33 kabupaten/kota se-Sumatera Utara selama rentang periode waktu tertentu. Setelah data diperoleh, dilakukan pembersihan data (data cleaning) untuk memastikan tidak ada nilai yang hilang (missing values) atau data anomali (outliers). Selanjutnya, dilakukan proses transformasi atau normalisasi data jika terdapat rentang skala yang berjarak sangat jauh antar variabel, sehingga seluruh atribut memiliki bobot pengaruh yang seimbang dalam proses perhitungan jarak.

Tahap inti dari metodologi ini adalah penerapan algoritma K-Means untuk membentuk kelompok wilayah berdasarkan tingkat kemiskinan. Langkah awal algoritma diawali dengan menentukan jumlah k klaster yang ingin dibentuk (misalnya 3 klaster: tingkat kemiskinan tinggi, sedang, dan rendah). Setelah itu, dilanjutkan dengan menginisialisasi atau memilih titik pusat klaster awal (centroid) secara acak atau berdasarkan representasi data tertentu. Setiap data kabupaten/kota kemudian dihitung jaraknya terhadap masing-masing titik centroid menggunakan rumus jarak Euclidean (Euclidean Distance). Berdasarkan nilai jarak terdekat tersebut, setiap daerah akan dialokasikan ke dalam klaster yang sesuai.

Setelah seluruh data terdistribusi ke dalam kelompoknya pada iterasi pertama, dilakukan pembaruan titik pusat klaster (centroid update). Titik centroid baru diperoleh dengan menghitung nilai rata-rata (mean) dari seluruh anggota data yang berada di dalam klaster tersebut. Proses perhitungan jarak Euclidean, pengelompokan ulang, dan pembaruan centroid ini dilakukan secara berulang (iteration) hingga tidak ada lagi perubahan keanggotaan data pada setiap klaster atau telah mencapai kondisi konvergen (convergent). Perhitungan manual ini dilakukan secara cermat untuk memahami alur komputasi matematis algoritma secara mendalam.

Untuk menjamin keakuratan dan keandalan hasil perhitungan manual, tahap akhir metodologi penelitian ini melibatkan validasi dan pengujian sistem menggunakan perangkat lunak data mining RapidMiner. Data mentah yang telah dirapi dikirimkan ke dalam alur kerja (workflow) RapidMiner menggunakan operator Read Excel, yang kemudian dihubungkan ke operator Clustering (K-Means) dengan parameter k yang disesuaikan. Hasil keluaran dari RapidMiner, baik berupa cluster model, deskripsi statistik centroid, maupun grafik visualisasi klaster, akan dibandingkan dengan hasil perhitungan manual untuk memastikan ketiadaan

presisi yang terlewat serta memberikan gambaran visual yang jelas mengenai pemetaan tingkat kemiskinan wilayah di Sumatera Utara.

Desain Penelitian

Desain penelitian ini adalah deskriptif kuantitatif, yang berfokus pada pengumpulan dan analisis data numerik untuk menggambarkan fenomena tingkat kemiskinan wilayah di kabupaten/kota Provinsi Sumatera Utara. Penelitian ini dilakukan dalam dua tahap:

a. Pengumpulan Data Awal: ada tahap ini, data terkait jumlah penduduk miskin dan persentase penduduk miskin untuk setiap kabupaten/kota dikumpulkan dari sumber data sekunder. Data ini digunakan untuk menganalisis karakteristik serta pola persebaran kemiskinan daerah.
b. Eksperimen Implementasi Penjadwalan: Pada tahap kedua, algoritma K-Means diterapkan melalui perhitungan manual dan komputasi software untuk mengelompokkan wilayah ke dalam beberapa klaster tingkat kemiskinan (tinggi, sedang, dan rendah).

Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh wilayah kabupaten dan kota yang berada di wilayah Provinsi Sumatera Utara. Sampel penelitian diambil menggunakan teknik total sampling, yaitu mencakup keseluruhan 33 kabupaten/kota di Provinsi Sumatera Utara yang memiliki data indikator kemiskinan lengkap pada periode rentang waktu penelitian (2019–2023). Total sampel yang digunakan dalam penelitian ini adalah 33 kabupaten/kota.

Teknik Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan metode dokumentasi dan pengumpulan data sekunder yang diperoleh dari laporan resmi Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data yang dikumpulkan meliputi:

a. Nama Kabupaten/Kota: Identitas wilayah administratif di Provinsi Sumatera Utara yang menjadi objek pengelompokan.
b. Jumlah Penduduk Miskin: Jumlah masyarakat berpenghasilan di bawah garis kemiskinan yang diukur dalam satuan ribuan jiwa.
c. Persentase Penduduk Miskin: Proporsi jumlah penduduk miskin dibandingkan dengan total jumlah penduduk di masing-masing wilayah kabupaten/kota.

Instrumen Penelitian

Instrumen utama dalam penelitian ini adalah Spreadsheet Excel, yang digunakan untuk mengolah dan menganalisis data terkait waktu kedatangan, durasi pemrosesan, dan waktu selesai. Alat ini juga digunakan untuk menerapkan metode penjadwalan FCFS dan SPT serta

untuk menghitung efisiensi operasional, pemborosan bahan baku, dan produktivitas tenaga kerja.

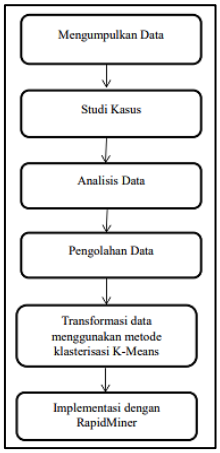
Teknik Analisis Data

Data yang dikumpulkan akan dianalisis menggunakan statistik deskriptif dan analisis inferensial untuk mengukur dampak metode penjadwalan terhadap efisiensi operasional. Analisis regresi digunakan untuk menentukan hubungan antara teknik penjadwalan dan variabel-variabel operasional seperti waktu penyelesaian pesanan, produktivitas tenaga kerja, dan pemborosan bahan baku Secara spesifik, langkah-langkah analisis data yang dilakukan adalah:

a. Deskriptif Statistik: Menyajikan ringkasan data berupa rata-rata, median, dan standar deviasi untuk menggambarkan kinerja produksi berdasarkan metode penjadwalan.
b. Uji Hipotesis: Menggunakan uji t untuk membandingkan perbedaan efisiensi antara metode FCFS dan SPT dalam hal waktu penyelesaian, pemborosan bahan baku, dan produktivitas tenaga kerja.
c. Regresi Linier: Untuk menguji pengaruh durasi pemrosesan dan metode penjadwalan terhadap waktu penyelesaian dan produktivitas tenaga kerja.

Alur Metodologi

Alur penelitian ini terdiri dari enam tahapan utama yang dilakukan secara berurutan dan sistematis, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian.

Pada tahapan alur penelitian diatas mengenai Langkah Langkah dalam menyelesaikan masalah penelitian ini adalah sebagai berikut Pengumpulan data: Langkah awal dalam penelitian ini dilakukan dengan mengumpulkan data mengenai jumlah dan persentase penduduk miskin yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara. Data tersebut digunakan sebagai dasar untuk proses analisis selanjutnya dan dikumpulkan

selama periode pelaksanaan penelitian. Studi kasus: Setelah data dikumpulkan, dilakukan studi kasus untuk memahami konteks data tersebut. Studi ini bertujuan untuk mengetahui tujuan analisis dan bagaimana data dapat digunakan untuk menyelesaikan suatu permasalahan.

Analisis data: Setelah data terkumpul, langkah selanjutnya adalah melakukan analisis data awal untuk memahami distribusi dan karakteristik data. Analisis ini mencakup pemeriksaan kelengkapan data, identifikasi kesalahan atau anomali, serta statistik deskriptif dasar. Pengolahan data: Pada tahap ini, data yang telah dianalisis akan diolah untuk mempersiapkannya bagi proses klasterisasi. Pengolahan data mencakup pembersihan data, normalisasi, dan transformasi yang diperlukan untuk memastikan data siap untuk dianalisis menggunakan algoritma K-Means.

Tranformasi data melalui Penerapan Metode *K-Means*: Data yang telah diolah akan melalui proses klasterisasi menggunakan algoritma *K-Means*. Algoritma ini akan mengelompokkan daerah-daerah di Provinsi Sumatera Utara berdasarkan kesamaan dalam jumlah dan persentase penduduk miskin. Hasil klasterisasi ini akan memberikan insight mengenai pola distribusi kemiskinan. Implementasi dengan *RapidMiner*: Langkah terakhir adalah implementasi proses klasterisasi menggunakan perangkat lunak RapidMiner. RapidMiner akan membantu dalam visualisasi hasil klasterisasi dan analisis lebih lanjut. Hasil dari implementasi ini akan dianalisis dan diinterpretasikan untuk memberikan rekomendasi kebijakan penanggulangan kemiskinan yang lebih efektif.

4. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis dan mengelompokkan wilayah kabupaten/kota di Provinsi Sumatera Utara berdasarkan indikator jumlah dan persentase penduduk miskin menggunakan algoritma K-Means. Pada bagian ini, akan dibahas hasil yang diperoleh dari penerapan perhitungan algoritma secara manual serta validasi komputasi menggunakan perangkat lunak RapidMiner, diikuti dengan pembahasan mengenai pembentukan klaster tingkat kemiskinan (tinggi, sedang, dan rendah) serta implikasi hasil pengelompokan tersebut terhadap penentuan prioritas kebijakan penanggulangan kemiskinan oleh pemerintah daerah.

Perhitungan Manual Algoritma *K-Means*

Langkah-langkah yang diterapkan oleh peneliti dalam mengelompokkan data jumlah dan persentase penduduk miskin pada tingkat kabupaten/kota di Provinsi Sumatera Utara dilakukan dengan menggunakan algoritma *K-Means* sebagai metode klasterisasi.

Menentukan data yang akan dikelompokkan, dalam menentukan jumlah claster ini dibentuk dengan menggunakan nilai secara random dari data sebanyak 3 cluster. *Cluster* tersebut diataranya cluster yaitu *cluster* rendah (C1), *cluster* sedang C2 dan *cluster* tinggi (C3). Data yang digunakan pada proses *clustering* sebanyak 33 kabupaten/kota. Menentukan *Centroid* awal *cluster* yang ditentukan secara acak, maka dipilih *cluster* 1 yaitu kabupaten Simalungun dan *cluster* 2 yaitu Tapanuli Utara serta *cluster* 3 yaitu Medan seperti pada Tabel 1 berikut:

Tabel 1. *Centroid* Awal.

Centeroid	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
C1(A9)	39,33	8,81	73,64	8,46	39,99	8,81	72,47	8,46	69,21	7,87
C2(A5)	28,57	9,48	28,41	9,37	29,72	9,72	27,47	9,37	26,39	8,54
C3(A30)	183,79	8,08	183,54	8,01	193,03	8,34	20,53	8,01	187,28	8

Menghitung jarak dari nilai *Centroid* dengan rumus sebagai berikut:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \qquad \text{(i)}$$

Keterangan:

D_e = *Euclidean Distance*

(x,y) = Koordinat Objek

(s,t) = Koordinat *Centroid*

i = Banyaknya Objek

Maka perhitungan untuk jarak dari centroid ke-1 adalah sebagai berikut:

$$\begin{aligned} Dx1,C1 = \sqrt{} & (22{,}10-39{,}33)^2 + (15{,}94-8{,}81)^2 + (23{,}12-73{,}64)^2 + (16{,}60-8{,}46)^2 \\ & + (24{,}33-39{,}99)^2 + (16{,}82-8{,}81)^2 + (23{,}23-2{,}47)^2 + (16{,}60-8{,}46)^2 \\ & + (21{,}99-69{,}21)^2 + (16{,}82-7{,}87)^2 \\ & = 7796{,}66 \end{aligned} \qquad \text{(ii)}$$

Dan seterusnya sampai dengan Dx33,C1.

Langkah selanjutnya adalah menghitung kembali jarak Euclidean untuk masing-masing pusat cluster. Tabel hasil pada iterasi ke 2 seperti pada Tabel 2. sebagai berikut:

Tabel 2. Perhitungan Jarak Data Ke Titik Pusat Centroid Pada Iterasi 2.

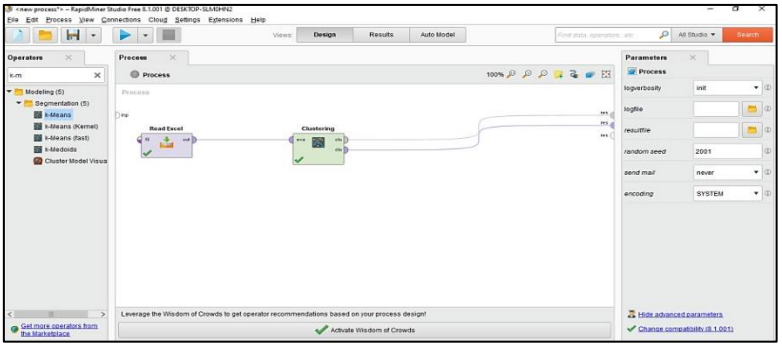
No	Kabupaten/Kota	*Cost* 1	*Cost* 2	*Cost* 3	Jarak Terdekat	*Cluster* Iterasi 2	*Cluster* Iterasi 1
1	Nias	10484,06	202,37	82043,09	202,37	2	2
2	Mandailing Natal	4579,59	935,09	64619,77	935,09	2	2
3	Tapanuli Selatan	10471,48	101,81	81816,58	101,81	2	2
4	Tapanuli Tengah	3195,84	1823,90	59646,61	1823,90	2	2
5	Tapanuli Utara	8622,59	27,15	76832,30	27,15	2	2
6	Toba Samosir	13380,92	497,49	89064,45	497,49	2	2
7	Labuhan Batu	4182,27	1161,03	63487,60	1161,03	2	2
8	Asahan	1092,97	6104,66	46876,45	1092,97	1	1

9	Simalungun	1317,47	6446,65	52283,18	1317,47	1	1
10	Dairi	10650,74	122,01	82137,35	122,01	2	2
11	Karo	5947,09	435,61	69011,09	435,61	2	2
12	Deli Serdang	2167,89	14786,61	34874,80	2167,89	1	1
13	Langkat	8908,01	28865,72	35258,90	8908,01	1	1
14	Nias Selatan	2186,19	3257,11	55134,94	2186,19	1	1
15	Humbang Hasundutan	12656,24	345,06	86948,78	345,06	2	2
16	Pakpak Bharat	19092,17	2013,85	101486,76	2013,85	2	2
17	Samosir	13277,07	477,29	89446,81	477,29	2	2
18	Serdang Bedagai	9027,32	1947,94	58565,44	1947,94	2	2
19	Batu Bara	7031,79	1789,87	57024,29	1789,87	2	2
20	Padang Lawas Utara	6523,21	522,85	78575,52	522,85	2	2
21	Padang Lawas	7042,15	625,62	81079,00	625,62	2	2
22	Labuhan Batu Selatan	8519,53	64,82	75434,67	64,82	2	2
23	Labuhan Batu Utara	7614,69	234,77	70041,07	234,77	2	2
24	Nias Utara	8034,53	1061,26	73044,21	1061,26	2	2
26	Nias Barat	10405,82	1178,65	85538,98	1178,65	2	2
26	Kota Sibolga	12721,36	892,60	94977,41	892,60	2	2
26	Kota Tanjung Balai	10692,76	91,83	81600,99	91,83	2	2
28	Kota Pematang Siantar	13064,53	408,61	84179,58	408,61	2	2
29	Kota Tebing Tinggi	12192,15	346,06	88224,40	346,06	2	2
30	Kota Medan	51524,37	78014,54	0	0	3	3
31	Kota Binjai	13681,22	656,33	89573,62	656,33	2	2
32	Kota Padangsidimpuan	15946,34	26675,52	116186,17	15946,34	1	1
33	Kota Gunungsitoli	11715,84	286,33	81982,12	286,33	2	2

Pada Tabel 2 titik *cluster* tidak mengalami perubahan maka iterasi diberhentikan maka diperoleh hasil *cluster* pada iterasi kedua dengan *cluster* 1 terdiri dari 6 Kabupaten/Kota, *cluster* 2 mencakup 26 kabupaten/kota, dan *cluster* 3 memiliki 1 Kabupaten/Kota.

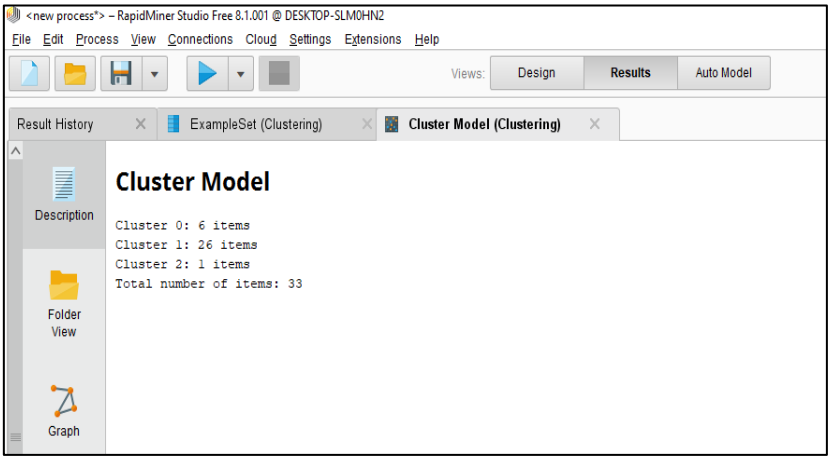
Hasil Perhitungan *RapidMiner*

Dalam pengujian yang dilakukan pada perhitungan manual ini menggunakan software *RapidMiner* versi 8.1. kemudian masukkan operator algoritma K-Means dan koneksikan semua operator seperti pada Gambar 2. Sebagai berikut.



Gambar 2. Menghubungkan *Read Excel* dengan Operator *K-Means*.

Setelah semua terkoneksi dan tidak ada yang *error* selanjutnya klik tombol *run* pada *toolbox* untuk mendapatkan hasil *clustering* terdapat pada Gambar 3 sebagai berikut:



Gambar 3. Hasil *Cluster* Model.

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian tentang pengelompokan jumlah dan persentase penduduk miskin di kabupaten/kota Provinsi Sumatera Utara menggunakan algoritma K-Means, dapat disimpulkan bahwa algoritma K-Means terbukti efektif dalam memetakan 33 kabupaten/kota ke dalam tiga kelompok utama berdasarkan tingkat kemiskinan, di mana hasil perhitungan manual dan simulasi aplikasi RapidMiner menunjukkan konsistensi yang sama persis. Pengelompokan ini menghasilkan Klaster 1 (tingkat kemiskinan rendah) dengan anggota 6 kabupaten/kota, Klaster 2 (tingkat kemiskinan sedang) yang mencakup 26 kabupaten/kota, serta Klaster 3 (tingkat kemiskinan tinggi) yang terdiri dari 1 wilayah yaitu Kota Medan. Hasil klasterisasi ini menyajikan gambaran peta kemiskinan daerah yang jelas sekaligus membuktikan efisiensi penggunaan perangkat lunak RapidMiner untuk analisis visualisasi data, sehingga dapat dijadikan landasan dan acuan strategis bagi pemerintah dalam merumuskan prioritas program penanggulangan kemiskinan yang lebih tepat sasaran di Sumatera Utara.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, disarankan agar Pemerintah Daerah Sumatera Utara memanfaatkan peta klasterisasi ini sebagai acuan strategis dalam memprioritaskan penanganan kemiskinan di wilayah klaster sedang dan tinggi, sekaligus memperluas penerapan metode *data mining* berbasis data (*evidence-based*) untuk isu-isu sosial lainnya. Bagi pengembangan penelitian selanjutnya, dianjurkan untuk menambahkan variabel

pendukung seperti tingkat pendidikan, pengangguran, dan akses layanan dasar agar pengelompokan menjadi lebih komprehensif, serta menerapkan metode evaluasi validitas klaster seperti *Silhouette Score* atau *Davies-Bouldin Index* guna menentukan jumlah klaster yang paling optimal dan meningkatkan akurasi hasil pengujian.

DAFTAR PUSTAKA

Agustina, I., & Dana, R. D. (2024). Implementasi K-Means clustering untuk pengelompokan tingkat kemiskinan di Pulau Jawa menggunakan bahasa pemrograman Python. JATI (Jurnal Mahasiswa Teknik Informatika). https://doi.org/10.36040/jati.v8i1.8858

Alifah, A. (2019). Pemetaan kemiskinan kabupaten/kota di Provinsi Jawa Tengah menggunakan algoritma K-Means. Seminar Nasional Edusainstek.

Anita, Y. (2024). Penerapan algoritma K-Means untuk clustering data penduduk miskin per kabupaten/kota di Pulau Sumatera. Repository Fakultas Ilmu Komputer Sriwijaya.

Bantun, S., Sari, Y., Jayanti, Z., Noorhasanah, Syahrul, & Priambodo, D. F. (2022). Pengembangan sistem informasi pada Dinas Sosial Kabupaten Kolaka guna menunjang keberhasilan pelayanan sosial pada masyarakat berbasis web. Info Kripto, 16(2), 59–68. https://doi.org/10.56706/ik.v16i2.46

Haumahu, G., & Matdoan, M. Y. (2022). Algoritma K-Medoids clustering untuk mengelompokkan tingkat kemiskinan pada kabupaten dan kota di Kepulauan Maluku dan Papua. Variance: Journal of Statistics and Its Applications, 4(2), 81–87. https://doi.org/10.30598/variancevol4iss2page81-87

Isyarah, F., Hasan, A., & Wiza, F. (2020). Clustering daerah miskin di Provinsi Riau menggunakan metode K-Means. Prosiding Seminar Nasional Teknologi Informasi & Ilmu Komputer (SEMASTER).

Kasim, R. J., Bahri, S., & Amir, S. (2021). Implementasi metode K-Means untuk clustering data penduduk miskin dengan systematic random sampling. SISFOTEK (Sistem Informasi dan Teknologi).

Maulida, R., Haryadi, D., Counsela, S. R., & Zahra, N. R. A. (2024). Implementasi algoritma K-Means dan RapidMiner untuk clustering data kemiskinan Provinsi Banten. Journal of Informatics and Communications Technology (JICT), 6(1). https://doi.org/10.52661/j_ict.v6i1.246

Nugraha, T. S., Kurniawan, R., & Wijaya, Y. A. (2024). Analisis data Program Keluarga Harapan pada tingkat kemiskinan dengan menggunakan Fuzzy C-Means di Kecamatan Pabedilan. JATI (Jurnal Mahasiswa Teknik Informatika, 8(1). https://doi.org/10.36040/jati.v8i1.8371

Novitasari, N., Nuris, N. D., & Herdiana, R. (2023). Penerapan algoritma K-Means untuk clustering data jumlah penduduk miskin berdasarkan kota/kabupaten di Jawa Barat menggunakan RapidMiner. Jurnal Informatika Terpadu, 9(1). https://doi.org/10.54914/jit.v9i1.660

Nurohmah, Y., Mayasari, R., & Sari, B. N. (2023). Optimalisasi performa K-Means clustering dengan PCA dalam analisis tingkat kemiskinan di Jawa Barat. JATI (Jurnal Mahasiswa Teknik Informatika, 7(3). https://doi.org/10.36040/jati.v7i3.6884

Oktaviany, N., Suarna, N., & Prihartono, W. (2024). Implementasi algoritma K-Means clustering dalam mengelompokkan kepadatan penduduk di Provinsi DKI Jakarta. JATI (Jurnal Mahasiswa Teknik Informatika). https://doi.org/10.36040/jati.v8i1.8241

Rahadian, D. R., & Nurmalitasari. (2023). Analisis kemiskinan menggunakan metode algoritma clustering K-Means. Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB).

Safitri, A., Rahmawati, R., & Wandira, S. A. (2024). Pengelompokan data penduduk miskin di Sumatera Utara menggunakan K-Means. J-Com (Journal of Computer), 4(1). https://doi.org/10.33330/j-com.v4i1.2998

Saputra, A. K., Fahrizal, & Mico. (2021). Rancang bangun berbasis web CRM (Customer Relationship Management): Studi kasus PT Budi Berlian Motor Hajimena Bandar Lampung. Portaldata.org, 1–31.

Sembiring, Y. R., Saifullah, & Winanjaya, R. (2021). Implementasi data mining dalam mengelompokkan jumlah penduduk miskin berdasarkan provinsi menggunakan algoritma K-Means. KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen).

Sepriyanti, N., Nahampun, R. S., & Zikri, M. H. (2022). Penerapan K-Means clustering untuk mengelompokkan tingkat kemiskinan di Provinsi Riau. SENTIMAS (Seminar Nasional Penelitian dan Pengabdian Masyarakat).

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.

Thira, I. J. (2019). Peramalan data kunjungan wisatawan. Jurnal Edukasi dan Penelitian Informatika.

Wijaya, N. S., Jajuli, M., & Dermawan, B. A. (2024). Penerapan algoritma K-Means clustering dalam menentukan daerah prioritas penanganan kemiskinan di wilayah Jawa Timur. JATI (Jurnal Mahasiswa Teknik Informatika, 8(4). https://doi.org/10.36040/jati.v8i4.10248

Zalukhu, A., Swingly, Purba, Darma, & Dedi. (2023). Perangkat lunak aplikasi pembelajaran flowchart. Jurnal Teknologi, Informasi dan Industri.

Zilfi, E. M., Istiawan, D., Ngatimin, Zaenah, & Nahdluddin. (2021). Penentuan pusat awal klaster algoritma K-Means untuk pengelompokan kabupaten/kota di Jawa Timur berdasarkan tingkat kemiskinan. Journal of Applied Statistics & Data Mining, 2(1). https://doi.org/10.63229/jasdm.v2i1.18

Zunaida Sitorus, & Suhartika. (2024). Penerapan data mining untuk clustering penduduk miskin di Kota Tanjungbalai menggunakan metode algoritma K-Means. Journal of Science and Social Research.