



Analisa Perbandingan Complate Linkage AHC dan K-Medoids Dalam Pengelompokkan Data Kemiskinan di Indonesia

Rifqi Habibi Sachrrial, Agus Iskandar*

Fakultas Teknologi Komunikasi dan Informatika, Program Studi Informatika, Universitas Nasional, Jakarta, Indonesia

Email: ¹rifqyhabibi06@gmail.com,^{2,*}iskandaragus1005@gmail.com

Email Penulis Korespondensi: iskandaragus1005@gmail.com

Submitted: 20/09/2023; Accepted: 30/09/2023; Published: 30/09/2023

Abstrak—Tingkat kemiskinan di Indonesia mengalami peningkatan dari 9,54 persen pada bulan Maret 2022 menjadi 9,57 persen pada bulan September 2022 akibat inflasi dan rendahnya upah serta pendapatan masyarakat. Untuk mengatasi masalah ini, diperlukan langkah-langkah seperti pemberian bantuan sosial, penciptaan lapangan kerja yang layak, dan peningkatan standar pengupahan guna meningkatkan daya beli masyarakat dan mengurangi kemiskinan di masa depan. Pemerintah perlu memberikan perhatian khusus pada provinsi-provinsi dengan tingkat kemiskinan tinggi melalui program-program khusus dan upaya peningkatan pendapatan serta perekonomian di wilayah tersebut. Data Mining adalah solusi dalam penyelesaian masalah tersebut dengan memanfaatkan metode clustering yang disebut sebagai metode pengelompokan. Metode clustering yang digunakan pada penelitian ini adalah metode AHC dan metode K-Medoids. Dalam rangka menentukan provinsi-provinsi dengan jumlah penduduk miskin terbanyak, metode pengelompokan (clustering) AHC dan K-Medoids akan diterapkan secara terpisah sehingga memiliki hasil akhir masing-masing yang akan dianalisa. Hasil analisis menunjukkan terbentuknya tiga cluster dengan letak cluster yang berbeda. Penerapan metode AHC menghasilkan cluster 2 dengan jumlah provinsi terbanyak, yaitu 22 provinsi, diikuti oleh cluster 0 dengan 9 provinsi, dan cluster 1 dengan hanya 3 provinsi. Sedangkan penerapan metode K-Medoids menghasilkan cluster 1 dengan jumlah provinsi terbanyak, yaitu 22 provinsi, diikuti oleh cluster 0 dengan 9 provinsi, dan cluster 2 dengan hanya 3 provinsi. Meskipun letak clusternya berbeda antara kedua metode tersebut namun jumlah provinsi dalam cluster sama sehingga cluster dengan jumlah 3 provinsi dinyatakan sebagai provinsi dengan jumlah penduduk miskin terbanyak.

Kata Kunci: Data Mining; Clustering; Algoritma AHC; Algoritma K-Medoids; Data Kemiskinan

Abstract—The poverty rate in Indonesia has increased from 9.54 percent in March 2022 to 9.57 percent in September 2022 due to inflation and low wages and people's incomes. To overcome this problem, steps such as providing social assistance, creating decent jobs, and increasing wage standards are needed to increase people's purchasing power and reduce poverty in the future. The government needs to pay special attention to provinces with high poverty rates through special programs and efforts to increase income and the economy in these areas. Data Mining is a solution in solving this problem by utilizing the clustering method which is known as the clustering method. The clustering method used in this study is the AHC method and the K-Medoids method. In order to determine the provinces with the highest number of poor people, the AHC and K-Medoids clustering methods will be applied separately so that the final results of each will be analyzed. The results of the analysis show the formation of three clusters with different cluster locations. The application of the AHC method resulted in cluster 2 with the largest number of provinces, namely 22 provinces, followed by cluster 0 with 9 provinces, and cluster 1 with only 3 provinces. While the application of the K-Medoids method resulted in cluster 1 with the largest number of provinces, namely 22 provinces, followed by cluster 0 with 9 provinces, and cluster 2 with only 3 provinces. Although the location of the clusters is different between the two methods, the number of provinces in the cluster is the same so that a cluster with a total of 3 provinces is declared the province with the largest number of poor people.

Keywords: Data Mining; Clustering; AHC Algorithm; K-Medoids Algorithm; Poverty Data

1. PENDAHULUAN

Kemiskinan merujuk pada situasi di mana seseorang atau sekelompok individu tidak memiliki sumber daya material yang mencukupi, seperti pendapatan, harta, atau akses ke layanan dasar yang diperlukan untuk memenuhi kebutuhan dasar mereka. Kebutuhan dasar ini mencakup makanan, tempat tinggal, pakaian, pendidikan, kesehatan, serta akses ke air bersih, sanitasi, dan energi. Ada berbagai faktor kompleks yang dapat menyebabkan kemiskinan, termasuk ketidakadilan sosial, ketimpangan ekonomi, kurangnya peluang pendidikan dan pekerjaan, konflik, perubahan iklim, dan faktor struktural lainnya. Dampak kemiskinan pada kehidupan individu dan masyarakat seringkali negatif, termasuk kesulitan dalam memenuhi kebutuhan dasar, terbatasnya akses ke layanan kesehatan dan pendidikan yang memadai, serta keterbatasan peluang ekonomi.

Berdasarkan data yang dirilis oleh Badan Pusat Statistik, tingkat kemiskinan pada bulan September 2022 tercatat sebesar 9,57 persen, yang menunjukkan peningkatan dibandingkan dengan angka sebesar 9,54 persen pada bulan Maret 2022. Penyebab perubahan ini dapat dipengaruhi oleh beberapa faktor, termasuk kenaikan inflasi yang berdampak pada harga bahan bakar minyak. Upah dan pendapatan masyarakat yang rendah telah menjadi faktor yang berkontribusi terhadap tingginya tingkat kemiskinan di tengah inflasi yang terjadi di Indonesia. Pemberian bantuan sosial, penciptaan lapangan kerja yang layak dan peningkatan standar pengupahan menjadi sangat penting dalam meningkatkan daya beli masyarakat dalam menghadapi ketidakpastian ekonomi di masa depan. Misalnya, Provinsi Nusa Tenggara Timur merupakan salah satu provinsi di Indonesia yang memiliki tingkat kemiskinan tertinggi ketiga, mencapai 20,23%. Salah satu faktor yang memengaruhi kondisi ini adalah rata-rata upah di provinsi tersebut yang paling rendah secara nasional, yakni hanya Rp 2,1 juta per bulan. Selain itu, NTT juga termasuk provinsi dengan standar Upah Minimum Provinsi (UMP) terendah kelima di Indonesia, dengan jumlah UMP hanya Rp 1,97 juta per



bulan. Dari hal tersebut diperlukan tindak lanjut dari pemerintah agar provinsi yang dinyatakan sebagai provinsi termiskin dapat lebih diperhatikan pemerintah agar penduduk yang menempati provinsi tersebut merasa diperhatikan dan dipedulikan oleh pemerintah. Sehingga, untuk meningkatkan pendapatan dan perekonomian di provinsi tersebut dapat dilakukan dengan cara lain seperti membuat program khusus terhadap provinsi termiskin.

Berdasarkan permasalahan yang telah diuraikan, maka peneliti akan melakukan pengelompokan terhadap provinsi berdasarkan data kemiskinan yang telah ada. Pengelompokan tersebut dilakukan dengan memanfaatkan clustering dalam data mining. Clustering merupakan suatu teknik dalam analisis data yang bertujuan untuk mengelompokkan objek atau data berdasarkan kesamaan atau perbedaan karakteristik tertentu. Metode ini digunakan untuk menemukan struktur atau pola yang tidak terlihat sebelumnya dalam data, tanpa memerlukan label atau pengawasan sebelumnya. Algoritma yang dapat digunakan dalam proses pengelompokan diantaranya K-Means, K-Medoids, AHC, GMM, DBSCAN dan lainnya[1][2][3][4]. Pada penelitian ini, dalam pengelompokan data kemiskinan digunakan metode AH dan K-Medoids, dimana kedua metode tersebut akan dianalisa hasil akhir yang diperoleh.

Ada beberapa peneliti yang telah melakukan penelitian dengan metode ataupun pokok pembahasan yang sama dengan penelitian ini, sehingga hasil penelitian yang telah mereka lakukan dijadikan sebagai referensi dalam penyelesaian penelitian ini. Penelitian yang dialakukan oleh Siti Asmiyatun, dkk pada tahun 2019, mereka melakukan penelitian terhadap kondisi jalan sehingga nantinya akan dikelompokkan menjadi empat cluster berdasarkan kondisi jalan pada umumnya seperti rusak parah, kondisinya baik dan seterusnya. Dalam penelitian tersebut, mereka memanfaatkan metode K-Medoids dan juga dilakukan pengujian kualitas dengan memanfaatkan silhouette sehingga hasil akhir diperoleh cluster 4 sebagai cluster terbaik dengan nilai Silhouette Coefficient rata-ratanya sebesar 0,5743[5].

Pada tahun 2019, sebuah penelitian dilakukan oleh Nyoman Gde Prajnaviweka Ratmasa Taram dan rekannya untuk mengkategorikan tingkat kriminalitas di 32 provinsi menggunakan metode AHC dan K-Means. Penelitian tersebut membandingkan rasio simpangan baku antara ketiga metode AHC (Single, Complete, dan Average Linkage) dan K-Means sehingga menemukan bahwa metode terbaik untuk mengelompokkan tingkat kriminalitas adalah metode Single Linkage dalam AHC. Selanjutnya, dilakukan perbandingan antara metode K-Means dan Single Linkage AHC. Hasil penelitian menunjukkan adanya tiga kelompok dalam metode Single Linkage. Kelompok pertama terdiri dari 29 provinsi, sementara kelompok kedua dan ketiga terdiri dari Provinsi Sumatera Utara dan Provinsi Jawa Timur. Faktor-faktor yang membedakan ketiga kelompok tersebut adalah sebagai berikut: kelompok pertama ditandai oleh jenis kejahatan KDRT, kelompok kedua ditandai oleh jenis kejahatan penganiayaan ringan, dan kelompok ketiga ditandai oleh jenis kejahatan korupsi[6].

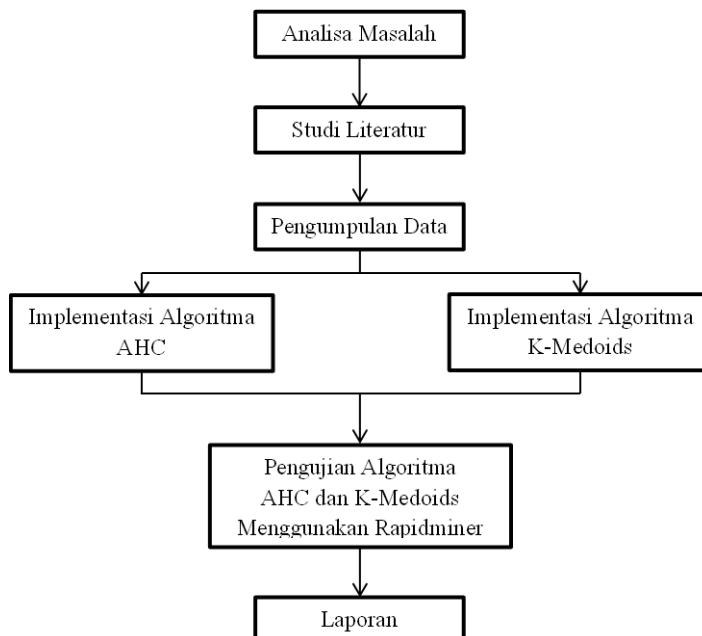
Pada tahun 2021, sebuah penelitian dilakukan oleh Aceng Supriyadi dan timnya. Dalam penelitian ini, dilakukan perbandingan antara Algoritma K-Means dan K-Medoids, serta dilakukan uji validitas terhadap klaster yang terbentuk. Untuk analisis klaster, digunakan Davies Bouldin Index (DBI), yang menghasilkan nilai validitas sebesar 0,67 untuk klaster yang menggunakan Algoritma K-Means, dan 1,78 untuk klaster yang menggunakan Algoritma K-Medoids. Berdasarkan nilai validitas tersebut, Algoritma K-Means dipilih untuk diimplementasikan dalam pembuatan aplikasi klaster armada kendaraan berbasis web, karena memiliki nilai DBI yang lebih rendah dibandingkan dengan K-Medoids. Pengujian yang dilakukan terhadap hasil klaster pada aplikasi web menunjukkan tingkat kesesuaian sebesar 97%, baik menggunakan tool Rapidminer maupun dengan perhitungan manual. Dengan demikian, penggunaan Algoritma K-Means dalam aplikasi ini terbukti efektif dalam menghasilkan klaster yang akurat, sesuai dengan tujuan penelitian tersebut[4].

Tahun 2020 dilakukan penelitian oleh Irmanita Nasution beserta teman timnya, Penelitian yang mereka lakukan menggunakan 34 provinsi sebagai sampel data, yang dibagi menjadi 2 kelompok, yaitu kelompok provinsi dengan tingkat kemiskinan tertinggi dan kelompok provinsi dengan tingkat kemiskinan terendah. Mereka menggunakan algoritma K-Means dalam proses pengelompokan. Melalui pengelompokan tersebut, diperoleh 8 kelompok dengan tingkat kemiskinan tinggi dan 26 kelompok dengan tingkat kemiskinan rendah. Tujuan dari penelitian ini adalah memberikan masukan kepada pemerintah agar dapat memberikan perhatian lebih kepada provinsi-provinsi yang termasuk dalam kategori tingkat kemiskinan tinggi. Diharapkan penelitian ini dapat memberikan wawasan yang berguna dalam upaya penanggulangan kemiskinan[7].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian memiliki peranan yang signifikan dalam mengarahkan serta mengatur proses penelitian secara terstruktur. Dalam melakukan penelitian, tahapan-tahapan tersebut memiliki manfaat dalam merencanakan, mengumpulkan data, menganalisis, dan menginterpretasikan hasil penelitian baik berupa perhitungan manual yang dilakukan hingga pengujian aplikasi yang memudahkan dalam pengecekan kesesuaian dengan hasil perhitungan manual yang telah dilakukan sebelumnya. Di bawah ini disajikan beberapa point penting yang dilakukan dalam penyelesaian penelitian:



Gambar 1. Tahapan Penelitian

Berdasarkan gambar 1, tahapan penelitian yang dilakukan dapat dilihat penjelasan yang lebih jelas dari bagan tersebut sehingga pembaca dapat memahami proses yang dilakukan saat penyelesaian penelitian, berikut penjelasannya:

a. Tahap Analisis Masalah

Pada tahap awal analisis masalah, langkah dilakukan untuk memperoleh pemahaman dan mengidentifikasi akar permasalahan yang sedang dihadapi. Tujuannya adalah untuk menemukan penyebab utama masalah guna menemukan solusi yang tepat dan efektif. Proses ini melibatkan pengumpulan, analisis, dan interpretasi data guna mengidentifikasi faktor-faktor penyebab masalah dan dampaknya pada sistem terkait. Dengan menganalisis masalah secara menyeluruh, kita dapat memahami permasalahan yang sedang dihadapi serta alasan di balik munculnya masalah tersebut. Dengan pemahaman yang mendalam tentang akar masalah, langkah-langkah yang relevan dapat diambil untuk menyelesaikan atau mengatasi masalah tersebut. Analisis masalah berperan penting dalam mengarahkan upaya penyelesaian masalah yang tepat dan efektif dengan fokus pada penanganan sumber permasalahan, bukan hanya gejala yang tampak.

b. Tahap Studi Literatur

Langkah selanjutnya adalah melakukan studi literatur atau tinjauan pustaka yang memiliki peran penting dalam menjalankan penelitian berkualitas. Proses ini melibatkan pengumpulan, analisis, evaluasi, dan penyusunan berbagai sumber informasi atau literatur yang relevan seperti buku, jurnal, artikel, laporan, dan dokumen terkait dengan topik atau permasalahan yang diteliti. Tujuan tahap ini adalah untuk mendapatkan pemahaman yang lebih mendalam tentang topik atau permasalahan yang sedang diteliti serta mengidentifikasi area pengetahuan yang masih kurang atau membutuhkan penelitian lebih lanjut.

c. Tahap Pengumpulan Data

Tahap pengumpulan data melibatkan akuisisi data sesuai dengan metode yang telah ditentukan sebelumnya. Metode yang digunakan termasuk wawancara, survei, observasi, eksperimen, atau analisis dokumen, tergantung pada jenis penelitian yang dilakukan. Selain itu, data juga dapat diperoleh dari badan pemerintahan yang menyediakan dataset untuk digunakan sebagai sampel data pada penelitian. Pengumpulan data yang terstruktur dan sesuai dengan metode yang telah ditetapkan membantu memastikan keandalan dan validitas hasil penelitian.

d. Tahap Implementasi Algoritma AHC dan K-Medoids

Setelah melakukan analisis masalah dan pengumpulan data, langkah berikutnya adalah menerapkan algoritma AHC dan K-Medoids pada data yang telah terkumpul. Algoritma ini digunakan untuk memproses data dengan menggunakan metode yang telah ditentukan sebelumnya. Dalam penelitian ini, algoritma AHC dan K-Medoids clustering digunakan untuk mengatasi masalah yang ada.

e. Tahap Pengujian Algoritma AHC dan K-Medoids

Setelah menerapkan algoritma AHC dan K-Medoids, langkah selanjutnya adalah melakukan evaluasi terhadap kinerja algoritma tersebut. Evaluasi dilakukan menggunakan platform RapidMiner. Jika hasil evaluasi konsisten dengan implementasi algoritma sebelumnya, maka evaluasi dianggap berhasil.

f. Tahap Penyusunan Laporan

Pada tahap akhir, peneliti akan menyelesaikan penulisan laporan penelitian yang merangkum semua tahap dan temuan penelitian secara ringkas. Laporan penelitian harus mencakup bagian pendahuluan, metode, hasil, analisis,

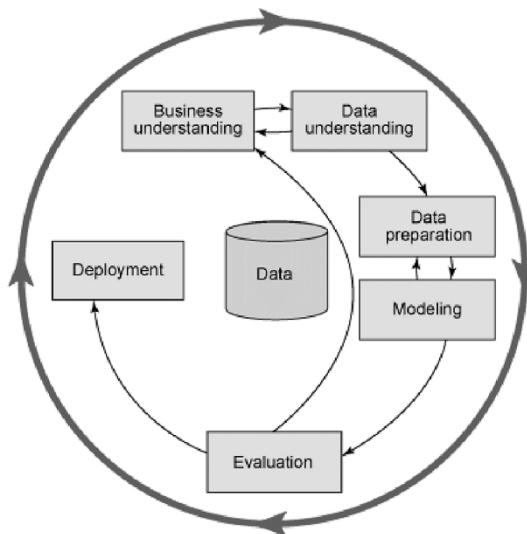
dan kesimpulan. Laporan tersebut juga harus disusun sesuai dengan format dan gaya penulisan yang ditetapkan oleh lembaga atau jurnal penelitian yang relevan.

2.2 Data Mining

Data mining adalah proses eksplorasi dan analisis dataset besar dengan tujuan menemukan pola, hubungan, dan informasi berharga. Tujuan utamanya adalah mengungkap pengetahuan yang tersembunyi dalam data yang kompleks atau besar. Dalam data mining, digunakan teknik dan algoritma statistik, matematika, kecerdasan buatan, dan pemodelan komputasional untuk mengidentifikasi pola yang signifikan dan membuat prediksi yang bermanfaat[8][9][10]. Ada beberapa tujuan penggunaan data mining, di antaranya:

- a. Data mining membantu mengidentifikasi pola tersembunyi dalam data, seperti pola korelasi, urutan, klasifikasi, atau anomali. Dengan mengenali pola ini, kita dapat memahami hubungan antarvariabel, tren, perilaku pelanggan, atau karakteristik data lainnya.
- b. Data mining memungkinkan pembuatan prediksi dan peramalan berdasarkan pola yang ditemukan dalam data historis. Dengan menganalisis pola dan tren masa lalu, kita dapat membuat prediksi tentang peristiwa atau perilaku di masa depan, seperti penjualan produk, permintaan pelanggan, atau tren pasar.
- c. Data mining memungkinkan segmentasi pelanggan berdasarkan karakteristik dan perilaku yang serupa. Dengan mengelompokkan pelanggan ke dalam kelompok yang berbeda, perusahaan dapat mengembangkan strategi pemasaran yang lebih efektif, menyesuaikan produk atau layanan, dan meningkatkan kepuasan pelanggan.
- d. Data mining memberikan wawasan berharga untuk mendukung pengambilan keputusan yang lebih baik. Dengan menganalisis data historis, mengidentifikasi pola, dan memprediksi hasil yang mungkin, keputusan dapat diambil berdasarkan informasi yang akurat dan faktual.
- e. Data mining dapat membantu mengungkap pengetahuan baru melalui analisis data yang mendalam. Hal ini dapat berkontribusi pada pengembangan ilmu pengetahuan, penemuan produk atau teknologi baru, atau pemahaman yang lebih baik tentang fenomena yang kompleks.

Berdasarkan tujuan penggunaan data mining yang telah dijelaskan sehingga data mining digunakan secara luas dalam berbagai bidang, termasuk bisnis, keuangan, ilmu kesehatan, ilmu sosial, pemasaran, dan lainnya. Dengan memanfaatkan kemampuan komputasi dan teknik analisis yang canggih, data mining menjadi alat yang sangat berharga dalam menghadapi tantangan yang terkait dengan volume, kompleksitas, dan keragaman data yang terus meningkat. Berikut gambaran proses data mining sering digunakan dalam memahami ruang lingkup dari data mining[11][12][13].



Gambar 2. Proses siklus iteratif data mining

Gambar 2 adalah proses siklus yang terjadi dalam data mining, mulai dari mengumpulkan data yang relevan dengan tujuan penelitian atau analisis. Data tersebut bisa diperoleh dari berbagai sumber seperti basis data perusahaan, data historis, data transaksi, atau sumber data eksternal. Selain itu, penting untuk memahami konteks bisnis dan tujuan analisis. Tujuan bisnis dan pertanyaan yang ingin dijawab dengan data mining harus ditentukan dengan jelas agar analisis dapat difokuskan pada aspek-aspek yang paling penting bagi bisnis yang disebut sebagai business understanding sehingga melibatkan peninjauan awal terhadap data, pemahaman tentang struktur data, statistik deskriptif, dan identifikasi potensi masalah atau kekurangan dalam data yang dapat mempengaruhi analisis. Data juga perlu dipersiapkan agar siap untuk proses analisis lebih lanjut seperti pembersihan data, transformasi, dan penggabungan data dilakukan (contohnya menghapus data duplikat, mengisi nilai yang hilang, atau melakukan normalisasi data). selanjutnya membangun model atau teknik analisis yang sesuai untuk mencari pola atau hubungan dalam data. Pemilihan algoritma dan metode yang tepat dilakukan berdasarkan jenis data, tujuan analisis, dan pertanyaan bisnis yang ingin dijawab. Contoh model yang sering digunakan meliputi regresi, pengelompokan



(clustering), pohon keputusan, atau jaringan saraf tiruan. Evaluasi juga dilakukan untuk mengukur kinerja dan efektivitas model. Data yang tidak pernah dilihat sebelumnya digunakan untuk menguji kemampuan model dalam membuat prediksi atau menemukan pola yang berguna. Evaluasi ini membantu memahami sejauh mana model dapat diandalkan dan apakah perlu dilakukan penyesuaian atau perbaikan. Implementasi model atau hasil analisis ke dalam lingkungan bisnis yang relevan. Model yang telah diuji dan divalidasi dapat digunakan untuk membuat prediksi, memberikan wawasan, atau mendukung pengambilan keputusan dalam bisnis. Implementasi ini melibatkan integrasi dengan sistem yang ada, pelatihan pengguna, dan pemantauan kinerja model secara berkala. Dalam data mining, ada 5 pengelompokan data mining yang dapat dilakukan oleh peneliti ataupun pihak yang ingin mengelola data agar memperoleh sebuah keluaran (informasi) yang tersembunyi yang dapat digunakan oleh peneliti diantarnya clustering, prediksi, klasifikasi, assosiasi dan estimasi[14][15][16][17][18].

2.3 Clustering

Clustering merupakan metode yang sering digunakan dalam data mining untuk mengelompokkan objek-objek data menjadi kelompok-kelompok yang serupa berdasarkan tingkat kesamaan atau jarak antara mereka. Tujuan utama dari clustering adalah untuk mengungkap struktur tersembunyi dalam data, di mana objek-objek yang memiliki kesamaan yang tinggi dikelompokkan bersama dalam satu kelompok, sedangkan objek-objek yang berbeda ditempatkan dalam kelompok yang berbeda. Dalam proses clustering, digunakan algoritma-algoritma khusus yang melakukan pengelompokan berdasarkan atribut atau fitur yang dimiliki oleh objek data tersebut. Algoritma ini menghitung jarak atau tingkat kesamaan antara objek-objek data dan kemudian menentukan kelompok yang paling sesuai untuk masing-masing objek[19][20][21].

Ada beberapa jenis algoritma clustering yang umum digunakan, seperti K-Means, Hierarchical Clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), dan lainnya. Setiap algoritma memiliki pendekatan dan karakteristik yang berbeda, sehingga pemilihan algoritma clustering yang tepat tergantung pada sifat data dan tujuan analisis yang ingin dicapai. Clustering memiliki berbagai aplikasi dalam berbagai bidang. Misalnya, dalam bidang pemasaran, clustering digunakan untuk mengelompokkan pelanggan berdasarkan perilaku belanja atau preferensi mereka, sehingga strategi pemasaran yang lebih efektif dapat dikembangkan untuk setiap kelompok. Di bidang biomedis, clustering digunakan untuk mengelompokkan pasien berdasarkan karakteristik klinis atau respons terhadap pengobatan, sehingga dapat mendukung pengambilan keputusan medis yang lebih baik. Melalui teknik clustering, kita dapat memperoleh wawasan berharga dari data yang kompleks dan membantu dalam pemahaman serta pengambilan keputusan yang lebih baik[22][4][23].

2.4 Algoritma Agglomerative Hierarchical Clustering (AHC)

AHC (Agglomerative Hierarchical Clustering) merupakan sebuah metode atau algoritma dalam analisis klaster yang bertujuan untuk mengelompokkan data ke dalam struktur klaster hierarkis. Pendekatan ini dimulai dengan menganggap setiap data sebagai sebuah klaster tunggal, lalu secara bertahap menggabungkan klaster-klaster yang memiliki kesamaan tertinggi berdasarkan metrik atau jarak yang telah ditentukan sebelumnya. Proses AHC dimulai dengan mengukur jarak antara setiap pasangan data, dan kemudian menggabungkan dua klaster yang memiliki jarak terdekat, membentuk sebuah klaster baru. Proses ini berlanjut dengan menggabungkan klaster-klaster yang semakin besar hingga semua data tergabung dalam satu klaster utama. Selama proses ini, terbentuklah struktur hirarkis yang menggambarkan hubungan antar klaster, sering kali direpresentasikan dalam bentuk pohon dendrogram atau grafik[21][24][25][26][27].

Salah satu keunggulan dari metode AHC adalah fleksibilitasnya dalam memilih metrik atau jarak yang digunakan untuk mengukur kesamaan antar data. Metrik yang digunakan dapat berupa jarak Euclidean, jarak Manhattan, atau jarak Mahalanobis, tergantung pada karakteristik data dan tujuan analisis yang diinginkan. Selain itu, penggunaan AHC dapat dikontrol dengan memilih jumlah klaster yang diinginkan atau menggunakan kriteria pemotongan (cut-off) untuk menghentikan proses penggabungan klaster. Jarak antara dua klaster dapat dihitung menggunakan jarak Euclidean Distance atau Manhattan dengan menggunakan rumus yang sesuai[28][29].

1. Menghitung matriks jarak

Euclidean Distance:

$$d(x_2, x_1) = \sqrt{\sum_{j=1}^d |x_j - y_j|^2} \quad (1)$$

Manhattan Distance:

$$D_{man}(x, y) = \sum_{j=1}^d |x_{2j} - y_{1j}| \quad (2)$$

2. Pengelompokan AHC

Terdapat tiga mode yang berbeda dalam proses pengelompokan menggunakan algoritma AHC. Mode pertama adalah single linkage, di mana data dikelompokkan berdasarkan jarak terdekat (terkecil) antara dua klaster. Mode kedua adalah complete linkage, yang mengelompokkan data berdasarkan jarak terjauh (maksimum) antara elemen-elemen yang berbeda di dalam klaster. Fokus pada mode ini adalah mencari jarak terbesar antara dua klaster,



sehingga menghasilkan klaster dengan perbedaan maksimum. Mode ketiga adalah average linkage, di mana data dikelompokkan berdasarkan nilai rata-rata dari dua klaster yang dipasangkan. Dalam mode ini, perhatian diberikan pada rata-rata jarak antara dua klaster yang akan digabungkan. Dengan demikian, terdapat beberapa mode pengelompokan AHC yang dapat digunakan[30][31].

- Pengelompokan Single Linkage:

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \quad (3)$$

- Pengelompokan Complate Linkage:

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \quad (4)$$

- Pengelompokan Average Linkage:

$$d_{uv} = \text{average}\{d_{uv}\}, d_{uv} \in D \quad (5)$$

2.5 Algoritma K-Medoids

Algoritma k-medoids adalah metode clustering dalam analisis klaster yang digunakan untuk mengelompokkan data ke dalam kelompok-kelompok berdasarkan kemiripan atau jarak di antara mereka. Tujuan utama dari k-medoids clustering adalah menemukan k titik data yang mewakili masing-masing kelompok atau klaster. Poin data ini disebut sebagai medoid. Proses algoritma k-medoids clustering dimulai dengan memilih secara acak k titik data sebagai medoid awal. Kemudian, setiap data yang tidak menjadi medoid akan diberi label dan ditempatkan dalam kelompok yang memiliki medoid terdekat berdasarkan jarak yang ditentukan. Selanjutnya, medoid dalam setiap kelompok akan diperbarui dengan mencari titik data baru yang memberikan total jarak minimum dengan semua data dalam kelompok tersebut. Langkah ini akan diulang hingga medoid tidak berubah atau iterasi mencapai batas tertentu [4][32].

Keuntungan dari algoritma k-medoids clustering adalah kemampuannya mengatasi outlier dan data yang tidak terdistribusi secara normal. Selain itu, k-medoids clustering lebih efisien dalam pengelompokan data dengan ukuran yang lebih besar dibandingkan dengan algoritma k-means. Namun, algoritma ini juga memiliki kelemahan yaitu sensitif terhadap pemilihan medoid awal yang acak dan memiliki kompleksitas komputasi yang tinggi. Dengan menggunakan algoritma k-medoids clustering, data dapat dikelompokkan ke dalam kelompok-kelompok yang memiliki kemiripan atau kesamaan, sehingga memudahkan analisis data, penggalian pola, dan pengambilan keputusan. Berikut proses penyelesaian masalah dengan menerapkan metode K-Medoids[33][34]:

- Tentukan jumlah cluster yang akan dibentuk (K)
- Tentukan pusat cluster dengan cara acak (sesuai dengan jumlah k yang akan dibentuk)
- Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance berikut

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

- Tentukan letak cluster berdasarkan nilai minimum dari jarak cost (C1 hingga Cn)
- Hitung total jarak minimum antar medoid (medoid wal)
- Tentukan medoids yang baru (non medoid) diambil dari sampel data secara acak untuk iterasi selanjutnya (pilih data yang bukan medoids)
- Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance seperti langkah 3 dan lakukan juga langkah ke4 dan 5
- Menghitung total simpangan (S)

$$S = \text{Medoid Cost Baru} - \text{Medoid Cost Lama} \quad (2)$$

Jika nilai S < 0 maka ulangi langkah 6 hingga diperoleh nilai simpangan (S) > 0. Jika nilai S > 0 maka proses perhitungan selesai.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, dilakukan pengelompokan terhadap provinsi-provinsi yang memiliki jumlah penduduk kemiskinan yang signifikan. Pengelompokan tersebut didasarkan pada data yang diperoleh dari BPS Profil Kemiskinan di Indonesia pada bulan Maret 2022. Terdapat 34 provinsi yang akan dikelompokkan, dengan jumlah penduduk miskin yang terdiri dari perkotaan dan pedesaan pada bulan September 2021 dan Maret 2022. Pengelompokan akan dilakukan menggunakan algoritma AHC dan K-Medoids. Kedua algoritma tersebut akan menghasilkan pembentukan 3 cluster yang akan dianalisis. Berikut adalah contoh tabel data sampel yang akan digunakan dalam penelitian ini.

Tabel 1. Jumlah Penduduk Miskin Menurut Provinsi (ribu)

Provinsi	September'21	Maret'22
Aceh	850,26	806,82
Sumatera Utara	1273,07	1268,19



Sumatera Barat	339,93	335,21
Riau	496,66	485,03
Jambi	279,86	279,37
Sumatera Selatan	1116,61	1044,69
Bengkulu	291,79	297,23
Lampung	1007,02	1002,41
Kep.Bangka Belitung	69,7	66,78
Kepulauan Riau	137,75	151,68
...
...
Papua Barat	221,29	218,78
Papua	944,49	922,12

3.1 Implementasi Algoritma AHC

Pengelompokan penduduk miskin dengan menerapkan algoritma AHC dimulai dengan menentukan jumlah cluster yang akan dibentuk, menghitung jarak matriks dengan memanfaatkan persamaan euclidian distance kemudian lakukan pengelompokan terhadap data kemiskinan dengan menggunakan mode complete linkage. Lebih jelasnya dapat diikuti langkah penerapan complete linkage AHC berikut:

- Menghitung jarak matriks

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

$$d(Aceh, Sumatera Utara) = \sqrt{|850,26 - 1273,07|^2 + |806,82 - 1268,19|^2} = 625,80$$

$$d(Aceh, Sumatera Barat) = \sqrt{|850,26 - 339,93|^2 + |806,82 - 335,21|^2} = 694,88$$

$$d(Aceh, Riau) = \sqrt{|850,26 - 496,66|^2 + |806,82 - 485,03|^2} = 478,10$$

Jarak matriks untuk matriks “Aceh, Papua” hingga “Papua Barat, Papua” lakukan perhitungan seperti pencarian jarak matriks “Aceh, Sumatera Utara” diatas. Berikut jarak matriks euclidian setelah dilakukan perhitungan terhadap seluruh matriks provinsi:

Tabel 2. Jarak matriks Euclidean

Provinsi	Aceh	Sumatera Utara	Sumatera Barat	Riau	Jambi	Papua
Aceh	0					
Sumatera Utara	625,80	0,00				
Sumatera Barat	694,88	1319,55	0,00			
Riau	478,10	1102,79	216,82	0,00		
Jambi	776,89	1401,51	82,02	298,83	0,00	
Sumatera Selatan	357,11	272,82	1051,95	835,20	1133,96	
Bengkulu	756,02	1380,46	61,32	277,92	21,48	
Lampung	250,66	376,06	943,49	726,74	1025,45	
Kep.Bangka Belitung	1075,61	1700,44	380,89	597,69	298,93	
Kepulauan Riau	967,93	1592,34	273,06	489,84	191,05	
...	
...	
Papua Barat	861,04	1485,77	166,23	383,04	84,27	
Papua	148,91	477,21	842,59	625,78	924,59	0

Berdasarkan tabel 2, telah dihitung jarak matriks keseluruhan sehingga terlihat jarak euclidian terkecil terletak pada matriks provinsi Gorontalo dan Sulawesi Utara dengan nilai 1,97 dan nilai tersebut merupakan nilai terkecil (minimum) diantara seluruh matriks. Oleh karena itu, provinsi-provinsi yang termasuk dalam cluster Gorontalo dan Sulawesi Utara dapat digolongkan sebagai cluster pertama (cluster awal).

- Lakukan pengelompokan dengan mode complete linkage

Langkah berikutnya melibatkan penggabungan cluster yang telah ditentukan menjadi satu cluster tunggal yang disebut sebagai cluster Gorontalo-Sulawesi Utara. Proses pengelompokan dilakukan dengan menggunakan metode complete linkage, di mana jarak antara dua cluster ditentukan berdasarkan jarak terjauh di antara elemen-elemen di dalamnya. Proses ini melibatkan pencarian nilai maksimum antara dua cluster yang akan digabungkan, dan hasilnya akan membentuk sebuah cluster yang lebih besar.

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D$$

$$d_{Gorontalo-Sulawesi Utara, Aceh} = \text{Max}\{910,61; 909,39\} = 910,61$$



$$d_{Gorontalo-Sulawesi Utara, Sumatera Utara} = \text{Max}\{1535,29; 1534,12\} = 1535,29$$

$$d_{Gorontalo-Sulawesi Utara, Sumatera Barat} = \text{Max}\{215,77; 214,58\} = 215,77$$

Lakukanlah perhitungan complete linkage seperti yang dijelaskan sebelumnya terhadap sisa klaster hingga seluruh proses selesai. Cari nilai terjauh dari klaster yang terhubung dengan cluster Gorontalo-Sulawesi Utara. Tidak perlu mencari nilai tertinggi (max) untuk klaster lainnya yang tidak terhubung dengan Gorontalo-Sulawesi Utara. Ambil matriks jarak Euclidean dari Tabel 4 untuk digunakan dalam perhitungan. Setelah selesai mencari nilai tertinggi untuk semua klaster, lanjutkan dengan langkah awal hingga terbentuk 3 klaster untuk provinsi yang ada.

3.2 Implementasi Algoritma K-Medoids

Pengelompokan penduduk miskin yang ada di indonesia dengan menerapkan algoritma K-Medoids dimulai dengan menentukan jumlah cluster yang akan dibentuk, kemudian memilih medoid dengan jumlah cluster yang ditentukan secara acak, menghitung jarak (cost) menggunakan persamaan euclidian distance dan menghitung total simpangan, lebih jelasnya berikut langkah penyelesaiannya:

- Jumlah cluster yang akan dibentuk sebanyak 3 cluster
- Menentukan pusat 3 cluster setiap medoids dilakukan dengan cara acak (memilih 3 sampel data secara acak)

Tabel 3. Medoid Awal

Cluster	Provinsi	September'21	Maret'22
C1	SulawesiSelatan	765,46	777,44
C2	SulawesiTenggara	323,26	309,79
C3	Gorontalo	184,6	185,44

- Menghitung jarak (cost) dengan menggunakan persamaan euclidian distance sebagai berikut

$$d_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2}$$

Data 1: Aceh

$$C_1 = \sqrt{(850,26 - 765,46)^2 + (806,82 - 777,44)^2} = 89,75$$

$$C_2 = \sqrt{(850,26 - 323,26)^2 + (806,82 - 309,79)^2} = 724,41$$

$$C_2 = \sqrt{(850,26 - 184,6)^2 + (806,82 - 185,44)^2} = 910,61$$

Data 2: Sumatera Utara

$$C_1 = \sqrt{(1273,07 - 765,46)^2 + (1268,19 - 777,44)^2} = 706,05$$

$$C_2 = \sqrt{(1273,07 - 323,26)^2 + (1268,19 - 309,79)^2} = 1349,32$$

$$C_2 = \sqrt{(1273,07 - 184,6)^2 + (1268,19 - 185,44)^2} = 1535,29$$

Data 3: Sumatera Barat

$$C_1 = \sqrt{(339,93 - 765,46)^2 + (335,21 - 777,44)^2} = 613,71$$

$$C_2 = \sqrt{(339,93 - 323,26)^2 + (335,21 - 309,79)^2} = 30,40$$

$$C_2 = \sqrt{(339,93 - 184,6)^2 + (335,21 - 185,44)^2} = 215,77$$

Untuk data ke-4 (Provinsi Riau) hingga data ke-34 (Papua) lakukan pencarian jarak (cost) sesuai dengan perhitungan Data 1 hingga Data 3. Hasil keseluruhan provinsi setelah dihitung jarak (cost) dapat dilihat pada tabel berikut.

Tabel 4. Hasil iterasi 1

Provinsi	C1	C2	C3	Jarak terdekat	Cluster
Aceh	89,75	724,41	910,61	89,75	1
SumateraUtara	706,05	1349,32	1535,29	706,05	1
SumateraBarat	613,71	30,40	215,77	30,40	2
Riau	397,19	246,53	432,59	246,53	2
Jambi	695,62	53,00	133,78	53,00	2
SumateraSelatan	441,28	1081,43	1267,66	441,28	1
Bengkulu	674,51	33,88	154,88	33,88	2
Lampung	330,10	973,27	1159,23	330,10	1
Kep.BangkaBelitung	994,54	351,21	165,17	165,17	3



Provinsi	C1	C2	C3	Jarak terdekat	Cluster
Kepulauan Riau	886,34	243,75	57,75	57,75	3
...
...
Papua Barat	779,89	136,68	49,58	49,58	3
Papua	230,18	872,28	1058,36	230,18	1
Medoid Cost				18201,59	

- d. Menentukan 3 pusat cluster medoids yang baru (non medoid) diambil dari sampel data secara acak untuk iterasi 2 dan dapat dilihat pada tabel berikut.

Tabel 5. Medoid Baru

Cluster	Provinsi	September'21	Maret'22
C1	Lampung	1007,02	1002,41
C2	Kep.BangkaBelitung	69,7	66,78
C3	Kalimantan Timur	233,13	236,25

- e. Lakukan perhitungan jarak (cost) dengan menggunakan persamaan euclidian distance sesuai dengan langkah ke-3. Berikut hasil jarak cost yang diperoleh setelah dilakukan perhitungan terhadap seluruh provinsi.

Tabel 6. Hasil iterasi 2

Kode	C1	C2	C3	Jarak terdekat	Cluster
Aceh	250,66	1075,61	840,48	250,66	1
Sumatera Utara	376,06	1700,44	1465,05	376,06	1
Sumatera Barat	943,49	380,89	145,60	145,60	3
Riau	726,74	597,69	362,41	362,41	3
Jambi	1025,45	298,93	63,58	63,58	3
Sumatera Selatan	117,46	1432,60	1197,54	117,46	1
Bengkulu	1004,41	320,05	84,61	84,61	3
Lampung	0,00	1324,38	1088,99	0,00	1
Kep.BangkaBelitung	1324,38	0,00	235,43	0,00	2
Kepulauan Riau	1216,29	108,81	127,47	108,81	2
...
...
Papua Barat	1109,71	214,67	21,10	21,10	3
Papua	101,77	1223,46	988,16	101,77	1
Medoid Cost Baru				17444,05	

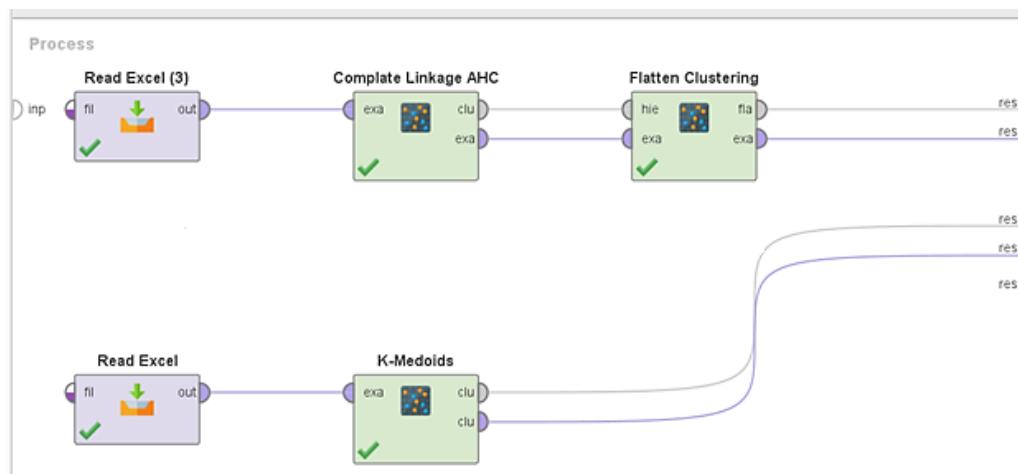
- f. Menghitung total simpangan (S)

$$S = \text{Medoid Cost Baru} - \text{Medoid Cost Lama} = 17444,05 - 18201,59 = -757,54$$

Nilai S < 0 sehingga dihitung lagi proses iterasi selanjutnya dengan mengulangi langkah 2 hingga diperoleh nilai simpangan (S) > 0. Jika nilai S > 0 maka proses perhitungan selesai.

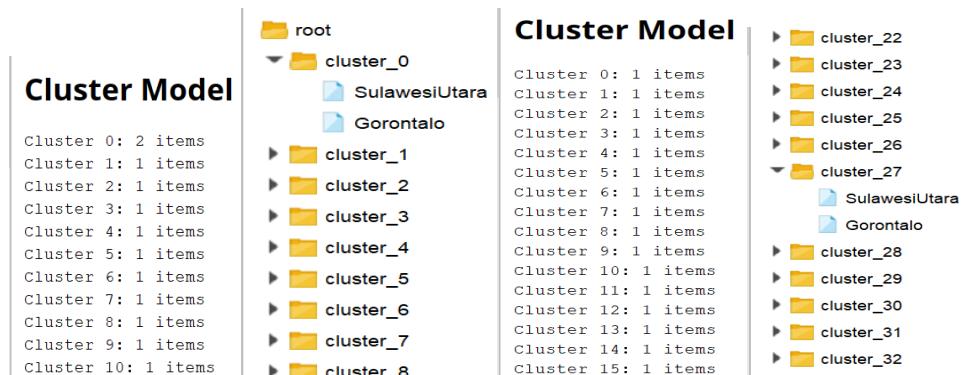
3.3 Pengujian Algoritma AHC dan Algoritma K-Medoids

RapidMiner menyajikan pengalaman pengguna yang ramah dan mudah digunakan melalui antarmuka pengguna dengan fitur drag-and-drop, sehingga pengguna dapat membuat proses analisis data tanpa harus menulis kode secara manual. Dengan antarmuka yang intuitif ini, pengguna bahkan tanpa pengetahuan pemrograman dapat dengan mudah menjalankan analisis data. RapidMiner juga menawarkan berbagai algoritma analisis data yang kuat, seperti regresi, klasifikasi, klustering, analisis asosiasi, dan lainnya. Pengguna memiliki fleksibilitas dalam memilih algoritma yang sesuai dengan kebutuhan mereka, dan mereka dapat membandingkan hasil akhir dari algoritma AHC dan K-Medoids. RapidMiner juga menyediakan berbagai alat preprocessing untuk mempersiapkan data sebelum analisis. Alat-alat ini memungkinkan pengguna untuk membersihkan data, menghilangkan nilai yang hilang, menormalisasi data, menggabungkan dan membagi kolom, serta melakukan transformasi data lainnya. Dengan proses preprocessing yang efisien ini, pengguna dapat dengan cepat menyiapkan data untuk analisis lebih lanjut. Selain itu, RapidMiner menawarkan berbagai fitur visualisasi yang kuat untuk membantu pengguna dalam menganalisis dan memahami data. Grafik, diagram, dan visualisasi lainnya memudahkan pengguna dalam menemukan pola dan hubungan dalam data secara intuitif. Hal ini membantu pengguna untuk memahami data dengan lebih baik dan mendukung pengambilan keputusan yang lebih baik. Dalam konteks pengelompokan provinsi di Indonesia berdasarkan data kemiskinan, gambar inputan operator untuk kedua algoritma AHC dan K-Medoids juga disajikan.



Gambar 3. Proses dataset inputan file K-Means dan AHC

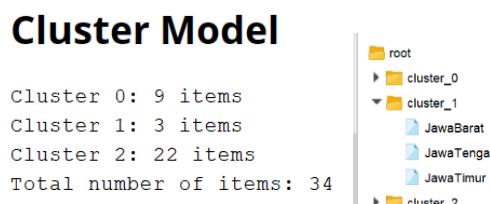
Gambar 3 menjelaskan langkah-langkah penggunaan RapidMiner. Proses dimulai dengan memasukkan file data sebagai input yang akan diproses, seperti file Excel yang telah diduplikat agar dapat dihubungkan ke dua algoritma pengelompokan. Operator pengelompokan AHC dimasukkan dan dihubungkan dengan operator flatten clustering untuk menentukan jumlah cluster yang akan dibentuk (3 cluster) dan metode penghitungan jarak maksimal yang dikenal sebagai metode complete linkage. Metode ini melibatkan perhitungan jarak hierarkis antara setiap pasangan provinsi. Selain itu, operator pengelompokan K-Medoids juga dimasukkan dengan menetapkan jumlah cluster yang akan dibentuk sebanyak 3 cluster. Setelah semua operator terhubung sesuai dengan masing-masing algoritma, proses RapidMiner dapat dijalankan. Untuk memverifikasi kesesuaian antara perhitungan manual sebelumnya dengan penggunaan RapidMiner, peneliti menguji aplikasi RapidMiner dengan membentuk 33 cluster. Cluster pertama yang terbentuk harus konsisten antara perhitungan manual dan hasil penggunaan aplikasi RapidMiner. Tampilan model klaster yang terbentuk dengan total 33 cluster juga disajikan.



Gambar 4. Pembentukan cluster pertama

Dari Gambar 4, dapat dijelaskan bahwa ketika kedua algoritma diterapkan untuk membentuk 33 cluster, terlihat bahwa provinsi Gorontalo dan Sulawesi Utara tergabung dalam cluster pertama (cluster 0). Hal ini konsisten dengan hasil perhitungan manual menggunakan algoritma AHC yang juga mengidentifikasi provinsi Gorontalo dan Sulawesi Utara sebagai cluster pertama (dalam pembuktian ini, hanya algoritma AHC yang digunakan karena algoritma K-Medoids memerlukan iterasi yang lebih banyak sehingga penempatan klaster pada iterasi pertama dapat berbeda dengan iterasi berikutnya). Kesesuaian ini menunjukkan konsistensi antara proses manual dan penggunaan aplikasi RapidMiner. Dengan demikian, dapat disimpulkan bahwa hasil dari kedua metode tersebut cocok, dan penggunaan RapidMiner dapat dilanjutkan. Selanjutnya, fokus akan dialihkan pada pembentukan 3 klaster.

a. Pengujian pada algoritma AHC



Gambar 5. Cluster Model AHC

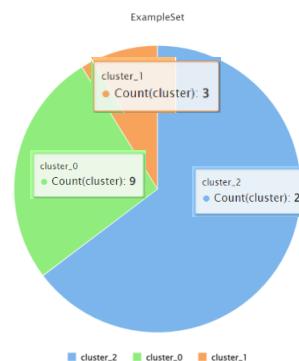


Berdasarkan gambar 5, dapat dijelaskan bahwa dari ketiga cluster yang dibentuk tersebut, cluster 2 memiliki jumlah kelompok provinsi yang lebih banyak dibanding dua cluster lainnya, yaitu 22 items atau lebih jelasnya ada 22 provinsi yang masuk dalam cluster 2. Selain itu ada juga cluster 0 yang memiliki provinsi lebih banyak kedua dengan jumlah 9 provinsi dan cluster dengan jumlah provinsi paling sedikit yaitu cluster 1 dengan jumlah kelompok hanya 3 provinsi. Dari ketiga cluster tersebut, provinsi yang masuk kedalam cluster_1 dinyatakan sebagai provinsi dengan jumlah kemiskinan terbanyak dibanding provinsi yang ada di cluster_0 dan cluster_2. Lebih jelasnya lagi dapat dilihat tabel berikut untuk mengetahui provinsi apa saja yang masuk kedalam 3 cluster tersebut.

Tabel 7. Cluster Terbentuk AHC

Provinsi	2021	2022	Cluster
Aceh	850.26	806.82	0
SumateraUtara	1273.07	1268.19	0
SumateraBarat	339.93	335.21	2
Riau	496.66	485.03	2
Jambi	279.86	279.37	2
SumateraSelatan	1116.61	1044.69	0
Bengkulu	291.79	297.23	2
Lampung	1007.02	1002.41	0
Kep.BangkaBelitung	69.7	66.78	2
KepulauanRiau	137.75	151.68	2
DKIJakarta	498.29	502.04	2
JawaBarat	4004.86	4070.98	1
...
...
PapuaBarat	221.29	218.78	2
Papua	944.49	922.12	0

Berdasarkan tabel 7, provinsi yang dikelompokkan kedalam cluster pertama (0) berjumlah 9 provinsi diantaranya provinsi Aceh, Banten, Nusa Tenggara Barat, Sulawesi Selatan, Lampung, Papua, Sumatera Utara, Sumatera Selatan dan Nusa Tenggara Timur. Provinsi yang dieklomopkan kedalam cluster kedua (1) berjumlah 3 provinsi diantaranya Jawa Barat, Jawa Tengah dan Jawa Timur. Sedangkan provinsi yang dikelompokkan kedalam cluster ketiga (2) berjumlah 22 provinsi dengan nama provinsi selain cluster pertama dan kedua. Berikut visualisasi dari ketiga cluster yang telah dibentuk.



Gambar 6. Visualisasi Cluster AHC

Gambar 6 terlihat bahwa cluster_2 (biru) menampung lebih banyak provinsi dari cluster yang lain, yaitu dengan jumlah persentase sebesar 64,7% sedangkan cluster_0 (hijau) dengan jumlah provinsi yang ditampung sebanyak 9 provinsi sehingga persentasenya 38,2% lebih kecil dari cluster_0 (biru). Sedangkan cluster_1 (orange) persentasenya hanya 8,8%.

b. Pengujian pada algoritma K-Medoids

Cluster Model

Cluster 0: 9 items
 Cluster 1: 22 items
 Cluster 2: 3 items
 Total number of items: 34

```

root
  ↘ cluster_0
  ↘ cluster_1
  ↘ cluster_2
    ↘ JawaBarat
    ↘ JawaTengah
    ↘ JawaTimur
  
```

Gambar 7. Visualisasi Cluster K-Medoids

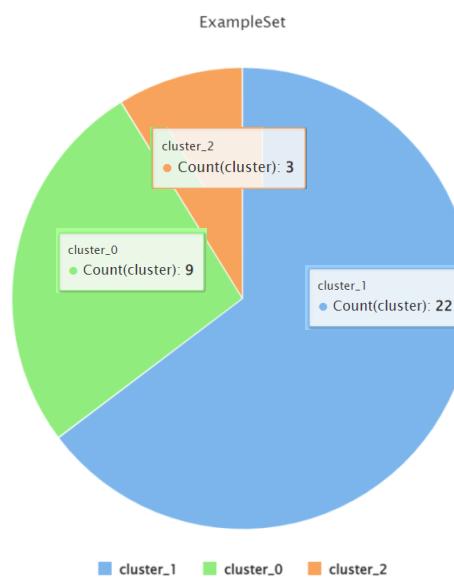


Berdasarkan gambar 6, dapat dijelaskan bahwa dari ketiga cluster yang dibentuk tersebut, cluster 1 memiliki jumlah kelompok provinsi yang lebih banyak dibanding dua cluster lainnya, yaitu 22 items atau lebih jelasnya ada 22 provinsi yang masuk dalam cluster 1. Selain itu ada juga cluster 0 yang memiliki provinsi lebih banyak kedua dengan jumlah 9 provinsi dan cluster dengan jumlah provinsi paling sedikit yaitu cluster 2 dengan jumlah kelompok hanya 3 provinsi. Dari ketiga cluster tersebut, provinsi yang masuk kedalam cluster_2 dinyatakan sebagai provinsi dengan jumlah kemiskinan terbanyak dibanding provinsi yang ada di cluster_0 dan cluster_1. Lebih jelasnya lagi dapat dilihat tabel berikut untuk mengetahui provinsi apa saja yang masuk kedalam 3 cluster tersebut.

Tabel 8. Cluster Terbentuk AHC

Provinsi	2021	2022	Cluster
Aceh	850.26	806.82	0
SumateraUtara	1273.07	1268.19	0
SumateraBarat	339.93	335.21	1
Riau	496.66	485.03	1
Jambi	279.86	279.37	1
SumateraSelatan	1116.61	1044.69	0
Bengkulu	291.79	297.23	1
Lampung	1007.02	1002.41	0
Kep.BangkaBelitung	69.7	66.78	1
KepulauanRiau	137.75	151.68	1
DKIJakarta	498.29	502.04	1
JawaBarat	4004.86	4070.98	2
...
...
PapuaBarat	221.29	218.78	1
Papua	944.49	922.12	0

Berdasarkan tabel 8, provinsi yang dikelompokkan kedalam cluster pertama (0) berjumlah 9 provinsi diantaranya provinsi Aceh, Sumatera Utara, Sumatera Selatan, Lampung, Banten, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Selatan dan Papua. Provinsi yang dieklompokkan kedalam cluster ketiga (2) berjumlah 3 provinsi diantaranya Jawa Barat, Jawa Tengah dan Jawa Timur. Sedangkan provinsi yang dikelompokkan kedalam cluster kedua (1) berjumlah 22 provinsi dengan nama provinsi selain cluster pertama dan ketiga. Berikut visualisasi dari ketiga cluster yang telah dibentuk.



Gambar 8. Visualisasi Cluster K-medoids

Gambar 8 terlihat bahwa cluster_1 (biru) menampung lebih banyak provinsi dari cluster yang lain, yaitu dengan jumlah persentase sebesar 64,7% sedangkan cluster_0 (hijau) dengan jumlah provinsi yang ditampung sebanyak 9 provinsi sehingga persentasenya sebesar 26,5%. Sedangkan cluster_2 (orange) persentasenya hanya 8,8%.

4. KESIMPULAN



Kesimpulan yang diambil pada penelitian ini adalah terdapat peningkatan tingkat kemiskinan di Indonesia dan untuk mengatasi masalah ini, pemerintah perlu melakukan langkah-langkah seperti pemberian bantuan sosial, penciptaan lapangan kerja, dan peningkatan standar pengupahan. Selain itu, metode clustering seperti AHC dan K-Medoids dapat digunakan untuk mengidentifikasi provinsi-provinsi dengan jumlah penduduk miskin terbanyak. Dalam rangka menentukan provinsi-provinsi dengan jumlah penduduk miskin terbanyak dari 34 provinsi yang ada di indonesia, dilakukan pengelompokan menggunakan metode clustering, yaitu metode AHC (Agglomerative Hierarchical Clustering) dan metode K-Medoids. Hasil analisis menunjukkan bahwa terbentuk tiga cluster dengan letak cluster yang berbeda antara metode AHC dan K-Medoids. Namun, jumlah provinsi dalam masing-masing cluster sama. Dalam metode AHC, cluster 2 memiliki jumlah provinsi terbanyak, yaitu 22 provinsi, diikuti oleh cluster 0 dengan 9 provinsi, dan cluster 1 dengan hanya 3 provinsi. Sementara itu, dalam metode K-Medoids, cluster 1 memiliki jumlah provinsi terbanyak, yaitu 22 provinsi, diikuti oleh cluster 0 dengan 9 provinsi, dan cluster 2 dengan hanya 3 provinsi.

REFERENCES

- [1] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan algoritma DBSCAN dan k-means clustering untuk pengelompokan kasus Covid-19 di dunia," *SITEKIN J. Sains, Teknol. Dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.
- [2] U. Rizqi, "Gaussian Mixture Model dengan Algoritme Expectation Maximization untuk Pengelompokan Data Distribusi Air Bersih di Jawa Barat," in *PRISMA, Prosiding Seminar Nasional Matematika*, 2023, vol. 6, pp. 745–750.
- [3] N. K. Zuhal, "Study Comparison K-Means Clustering Dengan Algoritma Hierarchical Clustering: AHC, K-Means Clustering, Study Comparison," in *STAINS (SEMINAR NASIONAL TEKNOLOGI & SAINS)*, 2022, vol. 1, no. 1, pp. 200–205.
- [4] A. Supriyadi, A. Triayudi, and I. D. Sholihat, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 6, no. 2, pp. 229–240, 2021.
- [5] S. Asmiyatun, "Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang," *JATISI (Jurnal Tek. Informat. dan Sist. Informasi)*, vol. 6, no. 2, pp. 171–180, 2019, doi: 10.35957/jatisi.v6i2.193.
- [6] N. G. P. R. TARAM, I. K. G. SUKARSA, and I. G. A. M. SRINADI, "Pengelompokan Tingkat Kriminalitas Dengan Metode Agglomerative Dan K-Means Serta Peubah Pencirinya," *E-Jurnal Mat.*, vol. 8, no. 2, p. 102, 2019, doi: 10.24843/mtk.2019.v08.i02.p241.
- [7] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.
- [8] R. H. Sukarna and Y. Ansori, "Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. Ilm. Sains dan Teknol.*, vol. 6, no. 1, pp. 50–61, 2022, doi: 10.47080/saintek.v6i1.1467.
- [9] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021, [Online]. Available: <https://hostjournals.com/>
- [10] Z. Nabilah, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [11] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth);," *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.
- [12] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, pp. 133–137, 2021, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73>
- [13] S. S. S, A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020, doi: 10.37600/tekinkom.v3i1.131.
- [14] A. S. L. T. T. H. Hafizah, "Data Mining Estimasi Biaya Produksi Ikan Kembung Rebus Dengan Regresi Linier Berganda," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, no. Vol 1, No 6 (2022): EDISI NOVEMBER 2022, pp. 888–897, 2022, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi/article/view/5732/1938>
- [15] Y. L. Nainel, E. Buulolo, and I. Lubis, "Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: PT. Pyridam Farma Tbk)," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, p. 214, 2020, doi: 10.30865/jurikom.v7i2.2097.
- [16] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.
- [17] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [18] H. Maulidiya and A. Jananto, "Asosiasi Data Mining Menggunakan Algoritma Apriori dan FP-Growth sebagai Dasar Pertimbangan Penentuan Paket Sembako," *Proceeding SENDIU 2020*, vol. 6, pp. 36–42, 2020.
- [19] F. Harahap, "Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita," *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.
- [20] M. A. Rofiq, A. Qoiriah, S. Kom, and M. Kom, "Pengelompokan Kategori Buku Berdasarkan Judul Menggunakan Algoritma Agglomerative Hierarchical Clustering Dan K-Medoids," *J. Informatics Comput. Sci.*, vol. 2, no. 03, pp. 220–227, 2021.
- [21] B. Harli Trimulya Suandi As and L. Zahrotun, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN DATA RIWAYAT AKADEMIK SEBELUM KULIAH DAN DATA KELULUSAN MAHASISWA MENGGUNAKAN



- METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING (Implementation Of Data Mining In Grouping Academic History Data Before Students And Stud," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 3, no. 1, pp. 62–71, 2021, [Online]. Available: <http://jtiika.if.unram.ac.id/index.php/JTIKA/>
- [22] E. P. Priambodo and A. Jananto, "Perbandingan Analisis Cluster Algoritma K-Means Dan AHC Dalam Perencanaan Persediaan Barang Pada Perusahaan Manufaktur," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 257, 2022, doi: 10.35889/progresif.v18i2.868.
- [23] H. D. Tampubolon, D. Gultom, L. Y. Hutabarat, F. I. R. H. Zer, and D. Hartama, "Penerapan Algoritma K-Means Untuk Mengetahui Tingkat Tindak Kejahatan Daerah Pematangsiantar," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 146–151, 2020.
- [24] R. A. Setyawan and R. M. Fadilla, "Klasterisasi media pembelajaran daring di era pandemi COVID-19 menggunakan metode Agglomerative," *Inf. Interaktif*, vol. 5, no. 3, 2020, [Online]. Available: [http://www.e-journal.janabadra.ac.id/index.php/informasiinteraktif/article/download/1305/890](http://www.e-journal.janabadra.ac.id/index.php/informasiinteraktif/article/view/1305%0Ahttps://www.e-journal.janabadra.ac.id/index.php/informasiinteraktif/article/download/1305/890)
- [25] Marjiyono, "Penerapan Algoritma Ahc Algorithm Dalam Aplikasi Ppembagian Kelas Siswa Baru," *Semin. Nas. Teknol. Inf. dan Multimed. 2015*, pp. 6–8, 2015.
- [26] T. Li, A. Rezaeipanah, and E. M. T. El Din, "An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement," *J. King Saud Univ. Inf. Sci.*, vol. 34, no. 6, pp. 3828–3842, 2022.
- [27] R. T. Adek, R. K. Dinata, and A. Ditha, "Online Newspaper Clustering in Aceh using the Agglomerative Hierarchical Clustering Method," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 1, pp. 70–75, 2022.
- [28] P. Govender and V. Sivakumar, "Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019)," *Atmos. Pollut. Res.*, vol. 11, no. 1, pp. 40–56, 2020.
- [29] C. Briggs, Z. Fan, and P. Andras, "Federated learning with hierarchical clustering of local updates to improve training on non-IID data," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–9.
- [30] K. Zeng, M. Ning, Y. Wang, and Y. Guo, "Hierarchical clustering with hard-batch triplet loss for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13657–13665.
- [31] N. K. Zuhal, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022, [Online]. Available: <https://jurnal.dharmawangsa.ac.id/index.php/djtechno/article/view/966/867>
- [32] B. Produktivitas, "Perbandingan algoritma k-means dengan k-medoids pada pengelompokan armada kendaraan truk berdasarkan produktivitas," vol. 06, pp. 229–240, 2021.
- [33] D. Wahyuli, I. Parlina, A. P. Windarto, and D. Suhendro, "Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids," no. September, pp. 452–461, 2019.
- [34] S. R. Ningsih, I. S. Damanik, A. P. Windarto, and H. Satria, "Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi," no. September, pp. 721–730, 2019.