

ANALISIS JUMLAH PENDUDUK MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN KABUPATEN/KOTA DI INDONESIA

Nurhayah, Nana Suarna, Willy Prihartono

Teknik Informatika, STMIK IKMI Cirebon

Jalan Perjuangan no.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat 45135

nhayah048@gmail.com

ABSTRAK

Jumlah penduduk di Indonesia secara konsisten, terjadi peningkatan dari satu tahun ke tahun berikutnya. Populasi terpadat di Indonesia terdapat pada Kabupaten Bogor. Kepadatan penduduk yang tinggi biasanya menunjukkan bahwa suatu wilayah memiliki populasi yang besar dalam luas wilayah yang terbatas. Di era sekarang, salah satu masalah yang terkait dengan aspek kependudukan adalah disparitas distribusi populasi yang tidak merata. Tujuan penelitian ini adalah untuk mengelompokkan jumlah penduduk menggunakan algoritma K-Means. Pada konteks penelitian ini, sumber data yang dipakai diperoleh dari Badan Pusat Statistika mencakup informasi jumlah penduduk Kabupaten/Kota di Indonesia Tahun 2020 sd. 2022. **Metode** penelitian ini menerapkan Algoritma K-Means dengan memanfaatkan pendekatan Knowledge Discovery and Data Mining (KDD) dan tahapan yang digunakan yaitu Selection data, Preprocessing data, Transformation, Data Mining dan Evaluation. Melalui analisis Data Mining, pengelompokan jumlah penduduk menghasilkan 4 cluster yaitu cluster jumlah penduduk sangat padat, padat, sedang, dan rendah. **Hasil** dari penelitian ini, Dari pengujian yang telah dilakukan dengan Davies Boulding Index didapatkan cluster yang optimal yaitu $k=4$ dengan nilai DBI 0.127. Cluster 0 dengan jumlah penduduk rendah sebanyak 360 Kabupaten/Kota, cluster 1 dengan jumlah penduduk sedang sebanyak 93 Kabupaten/Kota, Cluster 2 dengan jumlah penduduk sangat padat sebanyak 1 Kabupaten, dan cluster 3 dengan jumlah penduduk padat sebanyak 28 Kabupaten/Kota.

Kata kunci : Jumlah Penduduk, Clustering, Data Mining, K-Means

1. PENDAHULUAN

Indonesia, sebagai negara yang masih dalam tahap pembangunan, menghadapi sejumlah tantangan terkait kesejahteraan masyarakat. Kepadatan penduduk merupakan salah satu indikator penting dalam memahami bagaimana manusia terdistribusi disuatu wilayah atau daerah tertentu. Tingkat kepadatan penduduk memiliki implikasi yang signifikan dalam perencanaan wilayah, pembangunan sosial dan ekonomi, serta pengelolaan sumber daya alam dan lingkungan. Jika pertumbuhan populasi suatu kota tidak ditangani dengan baik dan terorganisir, dampak yang merugikan dapat muncul, seperti peningkatan tingkat pengangguran, peningkatan pencemaran lingkungan, pengurangan lahan hijau akibat perkuasan pemukiman penduduk, serta dampak-dampak ekonomi dan sosial lainnya [1]. Oleh karena itu, analisis jumlah penduduk menjadi sangat relevan dalam konteks perencanaan wilayah yang berkelanjutan.

Masalah yang terkait dengan kependudukan di era sekarang yaitu ketimpangan jumlah penduduk yang tidak merata. Ketidapkahaman masyarakat terkait lokasi padat penduduk disetiap kota dapat disebabkan oleh kurangnya pengetahuan atau informasi mengenai pengelompokan kepadatan penduduk disetiap kota. Hal ini dapat mengakibatkan masyarakat memilih lokasi tempat tinggal tanpa pertimbangan matang, sehingga rentan menghadapi permasalahan akibat dampak padat penduduk disetiap kota. Wilayah dengan padat penduduk tinggi mungkin menghadapi tantangan seperti kemacetan,

keterbatasan sumber daya alam, dan tekanan pada infrastruktur. Oleh karena itu, pemerintah disetiap daerah seharusnya dapat memastikan bahwa setiap penduduknya memperoleh akses menyeluruh terhadap layanan yang berkaitan dengan kebutuhan dasar. Keberhasilan ini memiliki signifikansi yang besar dalam pembangunan sumber daya manusia yang penting dalam konteks pembangunan saat ini. Dengan memahami tingkat jumlah penduduk, dapat diidentifikasi Kabupaten/Kota yang memerlukan upaya khusus dalam meningkatkan layanan dasar, pendidikan, dan kesehatan. Berdasarkan permasalahan dan dampak yang mungkin dihadapi oleh masyarakat akibat tingginya jumlah penduduk di beberapa Kabupaten/Kota di Indonesia, maka dalam penelitian ini akan dijalankan proses clustering Kabupaten/Kota yang ada di Indonesia ke dalam beberapa cluster, yaitu cluster jumlah penduduk sangat padat, padat, sedang dan rendah. cluster rendah (C3) mencakup tujuh belas kecamatan.

Tujuan penelitian ini adalah untuk memperoleh pemahaman yang lebih menyeluruh mengenai hasil pengelompokan tingkat kepadatan penduduk berdasarkan jumlah penduduk di berbagai Kabupaten/Kota di Indonesia, dengan menerapkan teknik data mining, penelitian ini memiliki tujuan untuk mengidentifikasi pola dan hubungan yang mungkin tersembunyi dalam data jumlah penduduk dari tahun 2020 hingga 2022. Melalui analisis clustering dengan menerapkan Algoritma K-Means, penelitian ini menyajikan wawasan yang lebih mendalam terkait karakteristik padat penduduk di

setiap cluster. Dengan demikian, diharapkan penelitian ini dapat berkontribusi dalam pemahaman regionalisasi tingkat padat penduduk di Indonesia, yang pada gilirannya dapat mendukung perencanaan wilayah, pengembangan kebijakan populasi, dan strategi pembangunan yang lebih terarah.

Metode yang diterapkan dalam penelitian ini yaitu Algoritma K-Means Clustering dengan menggunakan pendekatan Knowledge Discovery and Data Mining (KDD). Analisis kluster merupakan metode dalam data mining yang termasuk dalam kategori analisis tanpa pengawasan (unsupervised analysis). K-Means cluster analysis, disisi lain merupakan salah satu metode analisis kluster non-hierarki yang bertujuan untuk membagi objek-objek dalam dataset ke dalam satu atau lebih kluster berdasarkan karakteristiknya. Dengan pendekatan ini, objek-objek yang memiliki karakteristik serupa dikelompokkan dalam satu kluster, sementara objek-objek dengan karakteristik yang berbeda ditempatkan dalam kluster yang berbeda pula [2].

Hasil dari penelitian ini berpotensi memberikan perspektif baru kepada individu atau kelompok yang terlibat dalam transmigrasi, urbanisasi, atau imigrasi karena alasan pekerjaan, menetap, atau faktor lainnya. Dengan menerapkan metode clustering dalam analisis padat penduduk, penelitian ini dapat meningkatkan pemahaman tentang urgensi mempertimbangkan jumlah penduduk dalam perencanaan wilayah yang berkelanjutan. Penelitian ini mencerminkan bahwa tanpa memperhatikan dengan cermat tingkat padat penduduk, proses transmigrasi, urbanisasi, atau imigrasi berpotensi menimbulkan sejumlah tantangan dan dampak yang signifikan. Sebagai akibatnya, temuan dari penelitian ini mampu memberikan panduan dalam merancang kebijakan terkait transmigrasi, urbanisasi, atau imigrasi serta membantu dalam pengembangan strategi pembangunan yang lebih efektif. Dengan memahami dan mempertimbangkan secara cermat terhadap padat penduduk. Keputusan kebijakan dapat diambil dengan lebih tepat oleh para pengambil kebijakan dalam mengadvokasi distribusi penduduk yang seimbang, mengurangi risiko kemacetan, mengelola sumber daya alam secara berkelanjutan, dan meningkatkan kualitas infrastruktur diberbagai wilayah.

2. TINJAUAN PUSTAKA

2.1. Penduduk

Menurut definisi resmi dari Badan Pusat Statistik, penduduk merujuk kepada semua individu yang tinggal disuatu wilayah geografis di Republik Indonesia selama 6 bulan atau lebih, atau mereka yang tinggal kurang dari 6 bulan tetapi dengan niat untuk menetap. Definisi umum dari penduduk merujuk pada seluruh individu yang menetap di suatu daerah geografis negara selama periode waktu tertentu dan telah memenuhi kriteria yang ditetapkan oleh hukum negara. Di Indonesia, status penduduk

diberikan kepada seseorang atau kelompok yang telah berdomisili di wilayah Indonesia selama lebih dari enam bulan atau tinggal kurang dari enam bulan dengan tujuan untuk menetap [3].

Pertumbuhan penduduk adalah fluktuasi jumlah individu dalam suatu populasi sepanjang periode waktu tertentu, yang diukur sebagai perbedaan dalam jumlah individu per unit waktu. Dalam suatu negara, pertumbuhan penduduk merupakan indikator penting, karena pertumbuhan penduduk dapat dianggap sebagai input potensial yang berperan sebagai faktor produksi untuk meningkatkan produktivitas rumah tangga dan perusahaan. Peningkatan jumlah penduduk umumnya merujuk pada frekuensi penambahan individu manusia [4].

Kepadatan penduduk mengacu pada rata-rata jumlah penduduk di setiap wilayah dengan luas satu kilometer persegi. Nilai tingkat kepadatan penduduk yang bervariasi di setiap wilayah, dan dipengaruhi oleh faktor internal, seperti pertumbuhan penduduk yang cukup tinggi (dengan tingkat kelahiran melebihi tingkat kematian), serta sifat strategis wilayah. Faktor eksternal, seperti mobilitas penduduk dan produktivitas wilayah disekitarnya juga berperan dalam menentukan tingkat kepadatan penduduk [5].

2.2. Data Mining

Data mining merupakan suatu proses analisis yang bertujuan untuk mengidentifikasi ciri-ciri yang signifikan pada data yang saat ini diolah, serta untuk mendapatkan pola suatu model yang dapat mendeskripsikan kelas atau konsep dari data [6]. Sedangkan, istilah lain bahwa Data Mining merupakan proses ekstraksi atau eksplorasi data dan informasi yang signifikan yang sebelumnya tidak diketahui, meskipun dapat dimengerti dan memiliki nilai dari dalam basis data yang besar, dan informasi yang ditemukan melalui proses ini dapat digunakan untuk pengambilan keputusan bisnis yang krusial [7].

2.3. Clustering

Clustering adalah suatu metode dalam data mining yang beroperasi dengan mencari dan mengelompokkan data berdasarkan kemiripan karakteristik antar data satu dengan data lainnya yang telah dikumpulkan [8]. Clustering merujuk pada proses pengelompokan elemen-elemen seperti catatan, pemeriksaan, atau ulasan dalam ranah kelas-kelas entitas yang menunjukkan kesamaan. Tujuan dari pengelompokan adalah untuk mengorganisir seluruh set data dalam suatu cluster yang mempunyai tingkat kemiripan yang signifikan, di mana elemen-elemen dalam suatu kluster memiliki tingkat kemiripan yang tinggi, sementara kesamaan di antara berbagai kluster lainnya [5].

2.4. Algoritma K-Means

k-Means merupakan metode pengelompokan data non-hierarki yang memiliki kemampuan membagi data menjadi dua kelompok atau lebih.

Pendekatan ini akan mengorganisir data ke dalam kelompok sehingga data dengan karakteristik serupa akan ditempatkan dalam satu kelompok, sementara data yang mempunyai karakteristik yang berbeda akan disusun di dalam kelompok lainnya [9]. Algoritma K-Means dipilih untuk menganalisis parameter karena kemudahannya dan kinerjanya yang efektif pada dataset yang besar, bila dibandingkan dengan pengelompokan hierarkis [10].

Mengidentifikasi lokasi kluster dilakukan dengan melakukan perbandingan diantara 3 cluster, dan nilai terendah (minimum) dijadikan pilihan, apabila nilai terendah (minimum) teridentifikasi, maka objek dapat dimasukkan ke dalam cluster tersebut [11].

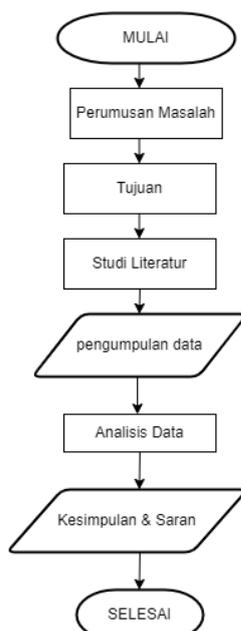
2.5. Davies Boulding Index

David L. Davies dan Donald W. Boulding mengenalkan teknik ini untuk tujuan mengevaluasi cluster. Evaluasi menggunakan metode Davies Boulding Index dilakukan secara internal terhadap cluster, dimana penilaian hasil kluster dilakukan berdasarkan jumlah dan jarak antara kluster. Davies Boulding Index menyesuaikan nilai tengah tingkat kemiripan di antara tiap kelompok. DBI dirancang untuk mengoptimalkan keseimbangan di antara interval didalam kluster sambil mengurangi jarak antara pusat kluster dan objek data yang lain [12]. Dalam penentuan nilai K yang optimal dilakukan dengan mencari nilai Davies Boulding Index (DBI) yang mendekati nol [13].

3. METODE PENELITIAN

3.1. Metode Penelitian

Pendekatan penelitian yang diterapkan dalam kajian ini melalui pendekatan kuantitatif. Penelitian ini mengacu pada tahapan-tahapan penelitian seperti yang tertuang pada gambar 1.



Gambar 1 Tahapan Penelitian

3.2. Sumber Data

Sumber data yang digunakan adalah data sekunder. Penelitian ini menggunakan data yang bersumber dari Badan Pusat Statistika (BPS) Indonesia, yang bisa diakses melalui website <https://www.bps.go.id/>. Data ini mencakup informasi mengenai jumlah penduduk Kabupaten/Kota pada Tahun 2020 sd. 2022. Pada data yang diambil terdapat 514 Record dan 6 Atribut. Data ini akan digunakan untuk mengelompokan tingkat kepadatan penduduk berdasarkan jumlah penduduk menurut Kabupaten/Kota di Indonesia.

3.3. Populasi

Populasi dalam penelitian ini merujuk pada keseluruhan yang menjadi fokus penelitian ini. Dalam dataset jumlah penduduk Kabupaten/Kota di Indonesia yang bersumber dari Badan Pusat Statistik (BPS) Indonesia, yang dapat diakses melalui <https://www.bps.go.id/>. Populasi nya adalah jumlah penduduk dari 514 Kabupaten/Kota pada tahun 2020 sebanyak 271.558.932 jiwa, 2021 sebanyak 272.44.559 jiwa dan 2022 sebanyak 275.928.315 jiwa. Terdapat 6 atribut dan 514 record pada data yang akan diteliti. Dengan mengambil seluruh populasi, tujuan dari penelitian ini adalah untuk menyediakan gambaran menyeluruh tentang distribusi jumlah penduduk di tingkat regional selama tiga tahun terakhir.

3.4. Teknik Pengumpulan Data

Dalam kajian ini, data yang digunakan berasal dari sumber resmi Badan Pusat Statistik (BPS) Indonesia, yang dapat diakses melalui <https://www.bps.go.id/>. Proses pengumpulan data dilakukan dengan secara daring, memanfaatkan kemudahan teknologi dan akses informasi online. Selain itu, perhatian khusus diberikan pada kualitas data untuk memastikan akurasi, kelengkapan, dan keandalan informasi yang diperoleh dari Sumber resmi BPS. Teknik pengumpulan data ini akan memberikan dasar yang kokoh untuk analisis lanjutan terkait kepadatan penduduk berdasarkan jumlah penduduk di berbagai Kabupaten/Kota di Indonesia.

3.5. Teknik Analisis Data

Tahapan data mining Knowledge Discovery and Data Mining seperti yang terlihat pada gambar 2.



Gambar 2 Tahapan KDD

Tahapan-tahapan dalam proses KDD dapat dijelaskan sebagai berikut:

1. Selection

Seleksi data (data target) yang akan dijadikan sampel untuk langkah-langkah berikutnya. Pada tahap ini, peneliti memilih dataset yang sesuai dari Badan Pusat Statistik (BPS) yang berisi informasi jumlah penduduk menurut Kabupaten/Kota di Indonesia dari tahun 2020 hingga 2022. Data ini dipilih dengan cermat berdasarkan relevansi dengan tujuan penelitian.

2. Preprocessing Data

Melaksanakan serangkaian langkah untuk melengkapkan data dan menjaga keseragaman data. Tahap ini peneliti melakukan pra-pemrosesan data yang komprehensif, ini melibatkan pengenalan dan penanganan data yang hilang. Selain itu, peneliti mengidentifikasi outlier dan memutuskan apakah akan menghapus atau menangani outlier tersebut. Data dinormalisasi untuk memastikan skala yang konsisten dan format data diubah jika diperlukan agar siap digunakan dalam analisis.

3. Transformation

Transformasi representasi data untuk memudahkan dan memperbaikinya agar sesuai dengan teknik data mining yang akan digunakan. Pada tahap ini, data yang telah diproses diubah dalam format yang sesuai untuk penggunaan Algoritma K-Means. Peneliti menyeleksi atribut yang akan dimanfaatkan dalam analisis dan mengubahnya ke dalam bentuk yang sesuai dengan ketentuan algoritma.

4. Data Mining

Proses pengembangan model untuk menemukan pola dalam data yang tersedia. Peneliti menerapkan Algoritma K-Means pada data yang telah diubah dalam tahap sebelumnya. Peneliti juga menentukan cluster yang tepat untuk analisis clustering. Penerapan Algoritma K-Means digunakan untuk mengelompokkan wilayah berdasarkan jumlah penduduk.

5. Evaluation

Tahap interpretasi dan penilaian terhadap pola yang dihasilkan untuk menentukan apakah pola tersebut menarik, bermanfaat, atau relevan. Peneliti menganalisis hasil clustering untuk mengidentifikasi pola atau wawasan yang muncul dari cluster yang terbentuk. Menginterpretasikan hasil tersebut dalam konteks masalah ketimpangan jumlah penduduk di Indonesia.

3.6. Algoritma K-Means

Untuk mengelompokkan kepadatan penduduk berdasarkan jumlah penduduk di Indonesia ke dalam beberapa cluster, yaitu cluster padat, Sedang, dan rendah. Metode K-Means clustering diterapkan sebagai suatu contoh untuk clustering data jumlah penduduk, menggunakan langkah-langkah berikut ini[14]:

a. Mengidentifikasi jumlah cluster

Dari 514 Kabupaten/Kota di Indonesia akan dikelompokkan ke dalam 4 cluster yaitu, cluster sangat padat, padat, sedang dan rendah.

b. Menentukan pusat cluster

Dari 514 Kabupaten/Kota akan dipilih 4 pusat cluster "Centroid"

c. Melakukan perhitungan jarak antara objek data dan pusat cluster

Jumlah populasi setiap Kabupaten/Kota akan dihitung menuju pusat cluster yang telah ditentukan dengan menggunakan teorema jarak Euclidean, yang dirumuskan sebagai berikut :

$$D(i, j) = \sqrt{(X1i - x1j)^2 + (X2i - X2j)^2 + \dots + (Xki - Xkj)^2} \quad (1)$$

Keterangan :

D(i,j) : jarak data ke i ke pusat cluster j

Xki : Data ke I pada atribut data ke k

Xkj : Titik pusat ke j pada atribut data ke k

d. Penempatan data dalam cluster terdekat

Setelah semua populasi seluruh data dari 1 hingga 514 dihitung ke pusat cluster melalui rumus Euclidion Distance, hasil perhitungan akan dialokasikan ke cluster yang paling mendekati.

e. Mengidentifikasi pusat cluster yang baru

Setelah melakukan perhitungan pada seluruh data dan menetapkan hasilnya ke cluster yang terdekat, total populasi di setiap cluster akan dijumlahkan, kemudian dibagi dengan jumlah Kabupaten/Kota yang termasuk dalam cluster tersebut. Hasil dari perhitungan tersebut akan menjadi pusat cluster baru.

f. Iterasi penentuan pusat cluster dilakukan hingga tidak terjadi perubahan pada data

Apabila, nilai pusat cluster ada perubahan, maka lakukan kembali ke langkah 3.

4. HASIL DAN PEMBAHASAN

Temuan dari penelitian yang dibahas dalam konteks ini, yakni khususnya mengenai proses pengelompokan dataset jumlah penduduk menggunakan Algoritma K-Means. Pengelompokan ini dilakukan melalui tahapan pengujian menggunakan metode machine learning, yakni melibatkan perangkat lunak RapidMiner Studio.

4.1. Data Penduduk

Data jumlah penduduk dari 514 Kabupaten/Kota di Indonesia pada periode tahun 2020 hingga 2022 yang telah dikumpulkan untuk di analisis dan dikelompokkan ke dalam tiga cluster, yaitu tingkat penduduk padat, sedang dan rendah. Data yang diterapkan dalam kajian dalam kajian ini diperoleh dari situs resmi Badan Pusat Statistik (BPS) Indonesia, yakni <https://www.bps.go.id/>. Dataset ini mencakup beberapa atribut, antara lain No, Provinsi, Kabupaten/Kota serta Jumlah Penduduk untuk tahun 2020, 2021, dan 2022. Selanjutnya, data yang sudah terkumpul akan diproses menggunakan metode

Algoritma K-Means Clustering. Data jumlah penduduk bisa dilihat pada tabel 1.

Tabel 1 Data Jumlah Penduduk

No	Provinsi	Kabupaten/Kota	2020	2021	2022
1	Aceh	Simeulue	92865	93762	94876
2	Aceh	Aceh Singkil	126514	128384	130787
3	Aceh	Aceh Selatan	232414	234630	237376
4	Aceh	Aceh Tenggara	220860	224119	228308
5	Aceh	Aceh Timur	422401	427032	432849
6	Aceh	Aceh Tengah	215576	218684	222673
7	Aceh	Aceh Barat	198736	200579	202858
8	Aceh	Aceh Besar	405535	409527	414490
9	Aceh	Pidie	435275	439398	444505
10	Aceh	Bireuen	436418	439788	443874
...
514	Bengkulu	Kota Bengkulu	373591	378604	384841

hilang dengan menggunakan parameter Filter Example. Parameter Filter Example ditampilkan pada gambar 5.



Gambar 5 Parameter Filter Examples

Pada tahap ini, dilakukan penghapusan data pada 32 Kabupaten/Kota yang memiliki nilai yang hilang pada atribut jumlah penduduk tahun 2020, 2021, maupun 2022. Proses ini dilakukan untuk memastikan data yang digunakan dalam analisis lebih lanjut telah bersih dari Missing Values. Data yang dihapus ditampilkan pada tabel 2.

Tabel 2 Penghapusan Data

No	Kabupaten/Kota	Provinsi
1	Karimun	Kepulauan Riau
2	Bintan	
3	Natuna	
4	Lingga	
5	Kepulauan Anambas	
6	Kota Batam	
7	Merauke	Papua
8	Jayawijaya	
9	Jayapura	
10	Kepulauan Yapen	
...
32	Deiyai	Papua

4.2. Data Selection

Parameter Select Attributes ditampilkan pada gambar 3.



Gambar 3 Parameter Select Attributes

Dalam melakukan seleksi data untuk penelitian ini, fokus utama jatuh pada dataset jumlah penduduk tahun 2020, 2021, dan 2022 yang diperoleh pada situs resmi Badan Pusat Statistik (BPS). Sebagai hasil dari proses seleksi data, dataset yang diambil terdiri dari 514 record dan 6 atribut, masing-masing mewakili Kabupaten/Kota di Indonesia. Hasil selection dari dataset yang tidak digunakan, hanya 4 atribut yang digunakan. Dengan memasukan atribut-atribut ini, penelitian ini dapat melakukan eksplorasi dan analisis terhadap perubahan kepadatan penduduk di setiap wilayah selama periode tiga tahun terakhir.

4.3. Preprocessing Data

Adanya nilai yang hilang (missing value) terdeteksi pada atribut jumlah penduduk 2021 dan 2022. Jumlah missing value ditampilkan pada gambar 4.

Gambar 4 Missing Value

Untuk menangani hal ini, peneliti memutuskan untuk menghapus data yang memiliki nilai yang

4.4. Data Transformasi

Pada tahap berikutnya, dilakukan proses transformasi data dengan menggunakan parameter “Set Role” dan “Normalize”. Transformasi ini bertujuan untuk menetapkan peran atribut dan mengubah skala nilai variabel. Dengan menetapkan peran atribut, data menjadi lebih terstruktur dan siap untuk dianalisis lebih lanjut. Normalisasi, yang mencakup pengubahan skala nilai variabel ke rentang antara 0 dan 1, dilakukan untuk memastikan bahwa semua variabel memberikan dampak yang seimbang pada analisis.

4.4.1. Set Role

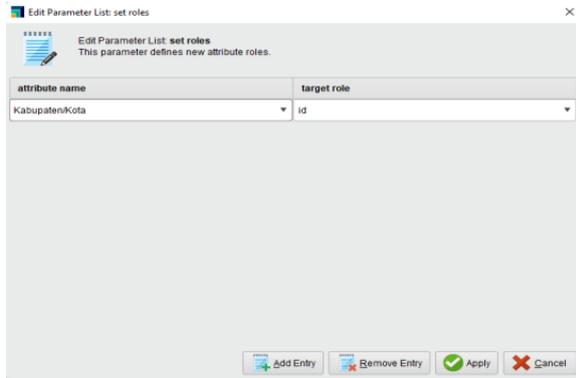
Parameter Set Role yang ditampilkan pada gambar 6.



Gambar 6 Parameter Set Role

Proses “Set Role” merupakan tahap penting dalam persiapan data sebelum menjalani analisis lebih lanjut. Pada tahap ini, peneliti memanfaatkan

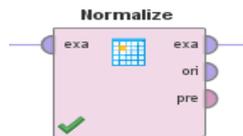
set rol dengan tujuan menetapkan atribut Kabupaten/Kota sebagai ID. Dengan menetapkan atribut Kabupaten/Kota sebagai ID, data menjadi lebih terstruktur dan dapat diakses dengan lebih mudah dalam tahap analisis berikutnya. Proses Set Role ditampilkan pada gambar 7.



Gambar 7 Tampilan Set Role

4.4.2. Normalize

Tahap selanjutnya yaitu dengan menggunakan parameter Normalize. Parameter Normalize ditampilkan pada gambar 8



Gambar 8 Parameter Normalize

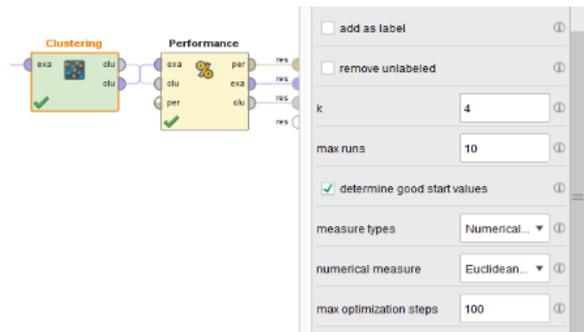
Dengan memanfaatkan parameter normalize, tahap transformasi data dapat dijalankan dengan lebih terstruktur, sehingga data siap untuk dianalisis lebih mendalam. Normalisasi, sebagai metode transformasi data dilakukan untuk mengadaptasi skala nilai variabel, tujuannya untuk menjamin bahwa semua variabel memiliki pengaruh yang seimbang pada proses analisis. Data normalisasi ditampilkan pada gambar 9.

Row No.	Kabupaten	cluster	2018	2019	2020
1	Banyuwangi	cluster_2	0,013	0,013	0,013
2	Asah Tegal	cluster_2	0,079	0,079	0,079
3	Asah Tegal	cluster_2	0,036	0,036	0,036
4	Asah Tegal	cluster_2	0,024	0,027	0,027
5	Asah Tegal	cluster_2	0,087	0,074	0,074
6	Asah Tegal	cluster_2	0,035	0,038	0,038
7	Asah Tegal	cluster_2	0,026	0,023	0,023
8	Asah Tegal	cluster_2	0,038	0,041	0,041
9	Pekalongan	cluster_2	0,028	0,079	0,079
10	Banyuwangi	cluster_2	0,079	0,079	0,079
11	Asah Tegal	cluster_2	0,037	0,107	0,107
12	Asah Tegal	cluster_2	0,023	0,024	0,024
13	Asah Tegal	cluster_2	0,041	0,042	0,042
14	Asah Tegal	cluster_2	0,028	0,023	0,023
15	Asah Tegal	cluster_2	0,023	0,027	0,027

Gambar 9 Normalisasi Data

4.5. Data Mining

Pada tahap data mining, parameter yang digunakan yaitu K-Means Clustering yang akan mengelompokkan data ke dalam 4 cluster, yaitu cluster sangat padat, padat, sedang, dan rendah. Parameter clustering ditampilkan pada gambar 10.



Gambar 10 Parameter Clustering

Hasil dari parameter K-Means Clustering diperoleh 4 cluster yaitu cluster 0 sebanyak 360 items, cluster 1 sebanyak 93 items, cluster 2 sebanyak 1 items, dan cluster 3 sebanyak 28 items dengan total 514 items. Hasil cluster model ditampilkan pada gambar 11.

Cluster Model

```
Cluster 0: 360 items
Cluster 1: 93 items
Cluster 2: 1 items
Cluster 3: 28 items
Total number of items: 482
```

Gambar 11 Cluster Model

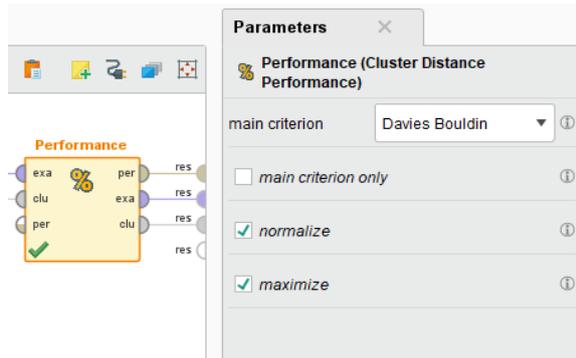
Parameter-parameter yang digunakan dalam RapidMiner ditampilkan pada gambar 12.



Gambar 12 Data Mining

4.6. Evaluation

Operator ini digunakan dalam tahap pengujian kinerja K-Means pada langkah kedua, dengan tujuan memperoleh nilai Davies Boulding Index (DBI) yang optimal dan mendekati nol. Semakin mendekati nol nilai DBI, semakin optimal performa cluster. Oleh karena itu, pada langkah ini dilakukan serangkaian uji coba dengan variasi nilai k untuk menentukan jumlah cluster yang dapat menghasilkan pembagian cluster dengan nilai DBI yang mendekati nilai optimal. Parameter performance ditampilkan pada gambar 13.



Gambar 13 Parameter Performance

Untuk memperoleh nilai k yang optimal, dilakukan percobaan. Kriteria keoptimalan tersebut adalah nilai k yang menghasilkan nilai Davies Boulding Index mendekati 0. Dari percobaan tersebut, didapati bahwa nilai k yang memberikan hasil DBI mendekati 0 adalah k=4, dengan nilai DBI sebesar 0.127. Hasil percobaan ditampilkan pada tabel 3.

Tabel 3 Nilai Davies Boulding Index

No	Cluster	Davies Boulding Index
1	K=2	0.179
2	K=3	0.160
3	K=4	0.127
4	K=5	0.148
5	K=6	0.158
6	K=7	0.157
7	K=8	0.151
8	K=9	0.155
9	K=10	0.152

Hasil dari data mining yang diperoleh dengan menerapkan algoritma K-Means sebagai berikut:

- Dengan menggunakan K-Means, terbentuk 4 cluster, yaitu cluster 0 sebanyak 360 item, cluster 1 sebanyak 93 item, cluster 2 sebanyak 1 item, dan cluster 3 sebanyak 28 item. Data cluster_0 ditampilkan pada tabel 4.

Tabel 4 Data Cluster_0

No	Kabupaten/Kota	Cluster	Keterangan
1	Simeulue	Cluster_0	Rendah
2	Aceh Singkil	Cluster_0	Rendah
3	Aceh Selatan	Cluster_0	Rendah
4	Aceh Tenggara	Cluster_0	Rendah
5	Aceh Timur	Cluster_0	Rendah
6	Aceh Tengah	Cluster_0	Rendah
7	Aceh Barat	Cluster_0	Rendah
8	Aceh Besar	Cluster_0	Rendah
9	Pidie	Cluster_0	Rendah
10	Bireueu	Cluster_0	Rendah
...
391	Kota Bengkulu	Cluster_0	Rendah

Data cluster_1 ditampilkan pada tabel 5.

Tabel 5 Data Cluster_1

No	Kabupaten/Kota	Cluster	Keterangan
1	Asahan	Cluster_1	Sedang
2	Simalungun	Cluster_1	Sedang
3	Langkat	Cluster_1	Sedang
4	Padang	Cluster_1	Sedang
5	Kampar	Cluster_1	Sedang
6	Pekanbaru	Cluster_1	Sedang
7	Kota Batam	Cluster_1	Sedang
8	Ogan Komering Ilir	Cluster_1	Sedang
9	Banyuasin	Cluster_1	Sedang
10	Palembang	Cluster_1	Sedang
...
94	Kabupaten Lombok Timur	Cluster_1	Sedang

Data cluster_2 ditampilkan pada tabel 6.

Tabel 6 Data Cluster_2

No	Kabupaten/Kota	Cluster	Keterangan
1	Bogor	Cluster_2	Sangat Padat

Data cluster_3 ditampilkan pada tabel 7.

Tabel 7 Data Cluster_3

No	Kabupaten/Kota	Cluster	Keterangan
1	Deli Serdang	Cluster_3	Padat
2	Medan	Cluster_3	Padat
3	Jakarta Selatan	Cluster_3	Padat
4	Jakarta Timur	Cluster_3	Padat
5	Jakarta Barat	Cluster_3	Padat
6	Jakarta Utara	Cluster_3	Padat
7	Sukabumi	Cluster_3	Padat
8	Cianjur	Cluster_3	Padat
9	Bandung	Cluster_3	Padat
10	Garut	Cluster_3	Padat
...
28	Kota Surabaya	Cluster_3	Padat

- Hasil dari proses analisis K-Means clustering mencakup nilai centroid untuk tiap cluster. Centroid, sebagai titik pusat cluster mempresentasikan rata-rata dari objek-objek yang termasuk dalam cluster tersebut. Nilai centroid yang diperoleh ditampilkan pada gambar 14.

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
2020.0	0.040	0.175	1	0.397
2021.0	0.043	0.195	1	0.435
2022.0	0.043	0.194	1	0.433

Gambar 14 Nilai Centroid

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian clustering jumlah penduduk menggunakan Algoritma K-Means, dapat disimpulkan berikut ini:

Hasil analisis clustering dengan menggunakan metode K-Means pada data jumlah penduduk Kabupaten/Kota berhasil mengelompokan wilayah-wilayah tersebut berdasarkan tingkat kepadatan penduduk. Penelitian ini menghasilkan 4 cluster yaitu

cluster 0 dengan jumlah penduduk rendah sebanyak 360 Kabupaten/Kota, Cluster 1 dengan jumlah penduduk sedang sebanyak 93 Kabupaten/Kota, cluster 3 dengan jumlah penduduk padat sebanyak 28 Kabupaten/Kota, dan cluster 2 dengan jumlah penduduk sangat padat sebanyak 1 Kabupaten. Nilai Davies Boulding Index yang didapatkan dari Algoritma K-Means ini sebesar 0.126 dengan nilai rata-rata jarak centroid yang optimal, ditemukan pada cluster 2 dengan average within centroid distance sebesar 0.000.

DAFTAR PUSTAKA

- [1] S. Sonang, A. T. Purba, dan F. O. I. Pardede, "Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means," *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, hal. 166, 2019, doi: 10.37600/tekinkom.v2i2.115.
- [2] N. I. Febianto dan N. Palasara, "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, hal. 130–140, 2019, doi: 10.32736/sisfokom.v8i2.653.
- [3] A. Apriandi dan M. D. N. Arindi, "Analisis Pengaruh Jumlah Penduduk, Penduduk Usia Sekolah dan Pertumbuhan Ekonomi Terhadap Pengangguran Di Provinsi Sumatera Utara," *J. Simki Econ.*, vol. 6, no. 1, hal. 100–109, 2023, doi: 10.29407/jse.v6i1.51.
- [4] R. Aritonang, L. B. Murbun, R. A. Simatupang, D. M. Rangkuty, U. Pembangunan, dan P. Budi, "Studi Kajian Pertumbuhan Penduduk di Kabupaten Deli Serdang Jumlah Penduduk," vol. 1, no. 4, 2023.
- [5] D. Gultom, I. Gunawan, I. Purnamasari, S. R. Andani, dan Z. A. Siregar, "Penerapan Algoritma K-Means Dalam Pengelompokan Kepadatan Penduduk Menurut Kecamatan di Kabupaten Simalungun," *TIN Terap. Inform. Nusant.*, vol. 2, no. 10, hal. 622–628, 2022, doi: 10.47065/tin.v2i10.1375.
- [6] T. Tendean dan W. Purba, "Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means," *J. Sains dan Teknol.*, vol. 1, no. 2, hal. 5–11, 2020.
- [7] Rayuwati, Husna Gemasih, dan Irma Nizar, "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMREDIKSI TINGKAT PENYEBARAN COVID," *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, hal. 38–46, 2022, doi: 10.55606/jurritek.v1i1.127.
- [8] C. A. Sugianto, A. H. Rahayu, dan A. Gusman, "Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, hal. 39–44, 2020, doi: 10.47292/joint.v2i2.30.
- [9] G. Gustientiedina, M. H. Adiya, dan Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, hal. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [10] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, hal. 606–620, 2022, doi: 10.35957/jatisi.v9i1.1516.
- [11] D. Safira, M. Mustakim, E. D. Lestari, M. Iffa, dan S. Annisa, "Pengelompokan Jumlah Penduduk Sumatera Barat Berdasarkan Angkatan Kerja Menggunakan Algoritma K-Means," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 6, no. 1, hal. 26, 2020, doi: 10.24014/rmsi.v6i1.8682.
- [12] F. J. Fauzan dan N. Nurazizah, "Application of K-Means and K-Medoids Algorithms in Clustering of Densely Populated Areas in Riau Province Penerapan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kepadatan Penduduk Provinsi Riau," hal. 223–229, 2023.
- [13] N. P. Gantara dan I. Ali, "Penerapan Metode K-Means Clustering Pada Penjualan Barang Di Sports Station," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, hal. 28, 2023, doi: 10.30587/e-link.v18i1.5339.
- [14] P. Marpaung dan R. F. Siahaan, "Penerapan Algoritma K-Means Clustering Untuk Pemetaan Kepadatan Penduduk Berdasarkan Jumlah Penduduk Kota Medan," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 1, hal. 503–521, 2021.