PENERAPAN METODE K-MEANS UNTUK KLASTERISASI DAERAH DENGAN RISIKO OVERPOPULASI DI INDONESIA

Ilham Maulana Prasetyo, Mira Orisa, Febriana Santi Wahyuni
Teknik Informatika, Institut Teknologi Nasional Malang
Jalan Raya Karanglo km 2 Malang, Indonesia
2118064@scholar.itn.ac.id

ABSTRAK

Pertumbuhan penduduk Indonesia yang terus meningkat hingga mencapai 281 juta jiwa pada 2024 menimbulkan masalah kepadatan tidak merata di sejumlah wilayah, terutama kota besar seperti Jakarta, Bandung, dan Surabaya. Fenomena overpopulasi ini menimbulkan dampak serius seperti tekanan pada infrastruktur, penurunan kualitas lingkungan, dan kesenjangan sosial. Tujuan penelitian ini adalah mengembangkan aplikasi *web* yang mengimplementasikan algoritma pengelompokan *K-means* untuk mengelompokkan wilayah di Indonesia berdasarkan risiko overpopulasi. Data yang digunakan dalam penelitian ini adalah data sekunder dari Badan Pusat Statistik (BPS) periode 2021–2024, dengan indikator yang dianalisis meliputi luas wilayah (X1), jumlah penduduk (X2), kepadatan penduduk (X3), serta laju pertumbuhan penduduk (X4). Jumlah klaster ditentukan secara manual sebanyak tiga klaster, yaitu rendah, sedang, dan tinggi. Evaluasi hasil klasterisasi menggunakan *silhouette index* menunjukkan nilai berkisar antara 0,721–0,724, yang mengindikasikan bahwa pemisahan antar klaster tergolong sangat baik. Hasil pengujian konsistensi terhadap data lapangan menunjukkan tingkat kecocokan rata-rata 77–78%, sedangkan pengujian pengguna menghasilkan tingkat kepuasan tinggi dengan lebih dari 85% responden menyatakan sistem mudah digunakan dan informatif. Secara keseluruhan, sistem yang dikembangkan dinilai efektif dalam mengelompokkan wilayah berisiko overpopulasi dan dapat dijadikan referensi untuk mendukung kebijakan pemerataan penduduk di Indonesia.

Kata kunci : *Clustering, K-Means, Kepadatan Penduduk, Overpopulasi*

1. PENDAHULUAN

Indonesia merupakan salah satu negara dengan jumlah penduduk terpadat di dunia, dengan angka pertumbuhan yang terus meningkat setiap tahunnya. Berdasarkan data Badan Pusat Statistik (BPS), jumlah penduduk Indonesia mencapai 281 juta jiwa pada tahun 2024, dengan distribusi populasi yang tidak merata. Beberapa daerah, terutama kota-kota besar seperti Jakarta, Surabaya, dan Bandung, mengalami masalah overpopulasi yang memicu berbagai permasalahan sosial, ekonomi, dan lingkungan. Overpopulasi dapat menyebabkan penurunan kualitas lingkungan, peningkatan polusi, serta tekanan terhadap sumber daya alam dan layanan publik [1]. Selain itu, kepadatan populasi yang tidak terkendali di kota-kota besar Indonesia meningkatkan kemungkinan risiko kesehatan dan memperburuk kondisi hidup masyarakat akibat sanitasi yang buruk serta akses terhadap layanan kesehatan yang terbatas [2]. Pertumbuhan populasi yang tidak terkendali juga dapat meningkatkan angka pengangguran, kemiskinan, dan kesenjangan sosial [3].

Overpopulasi di suatu daerah biasanya disebabkan oleh perencanaan urban yang kurang baik, angka kelahiran yang tinggi tanpa adanya migrasi keluar, serta daya tarik ekonomi suatu daerah yang menyebabkan arus migrasi masuk tinggi. Selain itu, ketidakseimbangan dalam pengembangan wilayah mengakibatkan populasi yang lebih padat di kota-kota besar. Apabila kepadatan penduduk di suatu daerah melebihi kapasitas infrastruktur dan kemampuan lingkungan untuk mendukungnya, maka daerah tersebut dianggap mengalami overpopulasi. Hal ini dapat dilihat dari indikator-indikator seperti kepadatan per kilometer persegi yang melebihi batas nasional, tekanan pada fasilitas publik, dan kerusakan lingkungan.

Untuk mengatasi permasalahan overpopulasi, penelitian menerapkan algoritma *K-Means Clustering* untuk mengelompokkan daerah di Indonesia yang memiliki risiko overpopulasi berdasarkan kemiripan atribut seperti jumlah penduduk, kepadatan, laju pertumbuhan, dan luas wilayah. Metode ini memiliki keunggulan dalam efisiensi pengolahan data berskala besar dan fleksibilitas dalam mengelompokkan wilayah tanpa bergantung pada ambang batas angka tetap. Selain itu, visualisasi hasil klasterisasi berdasarkan *centroid* juga memudahkan pengambilan keputusan, khususnya dalam merancang strategi pengendalian penduduk secara lebih terarah. Namun, kekurangan metode ini adalah sensitivitas terhadap pemilihan jumlah klaster (*k*) dan inisialisasi awal pusat klaster (*centroid*), yang dapat mempengaruhi hasil klasterisasi.

Berdasarkan pertimbangan tersebut, studi ini berupaya menerapkan algoritma pengelompokan *K-means* untuk mengelompokkan wilayah-wilayah di Indonesia berdasarkan risiko. Melalui hasil klasterisasi ini, diharapkan dapat diperoleh pemetaan yang lebih jelas tentang daerah yang berpotensi mengalami kepadatan penduduk ekstrem, sehingga dapat menjadi acuan bagi pemerintah dalam membuat strategi mitigasi yang lebih efektif. Klasterisasi ini juga dapat digunakan sebagai alat bantu mencegah dampak

negatif dari kepadatan penduduk yang berlebihan terhadap lingkungan dan kesejahteraan masyarakat.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian berjudul Penerapan Algoritma *K-Means* untuk Pengelompokan Kepadatan Penduduk di Provinsi DKI Jakarta oleh Handayanna dan Sunarti bertujuan untuk mengklasifikasikan wilayah kabupaten/kota di Provinsi DKI Jakarta berdasarkan tingkat kepadatan penduduk selama periode 2019 hingga 2022. Melalui penerapan metode *K-Means* dengan bantuan perangkat lunak RapidMiner, penelitian ini berhasil mengidentifikasi tiga wilayah dengan tingkat kepadatan tertinggi, yaitu Jakarta Selatan, Jakarta Timur, dan Jakarta Barat, yang menunjukkan kecenderungan pertumbuhan penduduk yang stabil. Hasil pengelompokan ini memberikan informasi yang berguna untuk merumuskan kebijakan pengendalian penduduk bagi pemerintah daerah, serta menjadi pertimbangan bagi masyarakat dalam menentukan lokasi tempat tinggal yang tepat [4].

Dalam penelitiannya yang berjudul Implementasi Algoritma *K-Means Clustering* dengan Jarak *Euclidean* dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor, Iqbal dkk. menyelidiki penggunaan metode *K-Means* untuk mengelompokkan wilayah berdasarkan tingkat penyebaran COVID-19. Jumlah klaster yang optimal ditentukan menggunakan metode *Elbow* dan indeks *Silhouette*. Hasil pengelompokan mengungkapkan dua klaster, yaitu klaster yang terdiri atas 36 kecamatan dengan tingkat risiko sedang, serta klaster C2 yang mencakup 4 kecamatan dengan tingkat risiko tinggi. Validitas hasil pengelompokan diperoleh melalui perhitungan *silhouette index* sebesar 0,71676, yang menunjukkan bahwa struktur klaster yang terbentuk cukup kuat dan konsisten [5].

Penelitian Nabiilah dkk. yang berjudul Penerapan Algoritma *K-Means* dan *K-Medoids* Dalam Pengelompokan Kepadatan Penduduk Provinsi Riau menganalisis permasalahan distribusi penduduk yang tidak merata di Provinsi Riau dan mengevaluasi efektivitas metode pengelompokan yang digunakan. Data diperoleh dari Badan Pusat Statistik (BPS) periode 2010–2021 dan mencakup 12 kabupaten/kota di wilayah tersebut. Hasil analisis menunjukkan bahwa algoritma *K-Means* memberikan hasil pengelompokan yang lebih baik, dengan Indeks *Davies-Bouldin* (DBI) sebesar 0,001 ketika jumlah klaster (k) ditentukan sebanyak dua. Hasil ini menunjukkan bahwa *K-Means* memiliki kinerja yang lebih baik dibandingkan metode *K-Medoids* dalam pengelompokan data berdasarkan tingkat kepadatan penduduk [6].

Sujak dkk. dalam penelitiannya yang berjudul Implementasi *K-Means Clustering* untuk Optimalisasi Anggaran Penyakit Tidak Menular bertujuan untuk mengelompokkan alokasi anggaran kesehatan yang ditujukan untuk penanganan penyakit tidak menular, seperti diabetes, hipertensi, obesitas, dan gangguan mental. Dengan menerapkan algoritma *K-Means clustering*, diperoleh empat klaster dengan nilai *silhouette score* sebesar 0,6156, yang memberikan gambaran mengenai pola distribusi anggaran kesehatan oleh pemerintah daerah. Hasil klasterisasi menunjukkan bahwa alokasi dana dapat dioptimalkan, baik untuk mendukung program kesehatan lainnya maupun sebagai dana cadangan dalam menghadapi potensi pandemi di masa mendatang [7].

Dalam penelitiannya yang berjudul Analisis Jumlah Penduduk Menggunakan Algoritma K-Means Berdasarkan Kabupaten/Kota Di Indonesia, Nurhayah dkk. bertujuan untuk mengelompokkan kabupaten/kota di Indonesia berdasarkan jumlah populasinya menggunakan algoritma *K-Means*. Data yang digunakan diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) untuk periode 2020–2022. Hasil pengelompokkan mengungkapkan empat kategori kepadatan penduduk: sangat padat, padat, sedang, dan rendah. Menurut evaluasi menggunakan Indeks *Davies-Bouldin* (DBI), nilai terbaik diperoleh ketika jumlah klaster (k) = 4, dengan skor DBI sebesar 0,127. Klaster dengan tingkat kepadatan sangat padat hanya mencakup satu kabupaten, sedangkan klaster dengan tingkat kepadatan rendah mencakup 360 kabupaten/kota [8].

2.2. Data Mining

Penambangan data adalah fase pemrosesan data yang bertujuan untuk mengekstrak informasi berharga dari kumpulan data yang besar dan kompleks. Proses ini memungkinkan penemuan pola, hubungan, dan pengetahuan tersembunyi di dalam kumpulan data tersebut. Manfaat dari *data mining* antara lain meningkatkan kualitas pengambilan keputusan, mengidentifikasi pola dan tren, memperkirakan kondisi di masa mendatang, melakukan segmentasi pelanggan, meningkatkan efisiensi operasional, mengelola risiko secara lebih baik, mengoptimalkan strategi pemasaran, serta memperbaiki kualitas layanan kepada pelanggan [9]. Dalam bidang *data mining*, terdapat sejumlah metode utama yang digunakan untuk mengekstraksi pola serta pengetahuan dari data. Metode-metode tersebut meliputi klasifikasi (*classification*), klasterisasi (*clustering*), penambangan asosiasi (*association rule mining*), regresi (*regression*), deteksi anomali/*outlier* (*anomaly/outlier detection*), reduksi dimensi (*dimensionality reduction*), serta peringkasan data (*summarization*).

2.3. Clustering

Clustering atau pengelompokan adalah teknik analisis data yang bertujuan mengelompokkan data ke dalam sekumpulan klaster berdasarkan tingkat kesamaan atribut antardata. Tujuan utama teknik ini adalah memastikan bahwa data dalam satu klaster memiliki sifat atau ciri yang lebih serupa satu sama lain dibandingkan dengan data dalam klaster lainnya. Menurut Anjani dan Bahtiar, klasterisasi merupakan

proses mengelompokkan data ke dalam grup-grup yang memiliki kesamaan karakteristik sehingga tiap kelompok bersifat homogen, tanpa memerlukan label kelas yang telah ditetapkan sebelumnya atau dikenal dengan istilah *unsupervised learning* [10]. Klasterisasi memungkinkan peneliti atau analis untuk menemukan struktur tersembunyi dari data besar yang kompleks tanpa perlu mengetahui label sebelumnya. Fungsi utama dari klasterisasi adalah sebagai alat bantu eksplorasi data, deteksi pola tersembunyi, penyederhanaan data, dan segmentasi.

2.4. Algoritma K-Means

Algoritma *K-Means* adalah metode pengelompokan non-hierarkis yang digunakan untuk membagi kumpulan data menjadi beberapa kelompok berbeda. Algortima ini bekerja dengan mengelompokkan data yang memiliki sifat atau ciri serupa ke dalam klaster yang sama, sedangkan data berbeda sifat atau cirinya akan ditempatkan dalam klaster yang berbeda [11]. Tujuan algoritma ini adalah meminimalkan nilai fungsi objektif yang telah ditentukan, dengan cara menekan variasi internal dalam setiap klaster serta memaksimalkan perbedaan antar klaster. Julyantari dan rekan-rekan menyatakan bahwa algoritma *K-Means* terdiri dari dua proses inti, yaitu penentuan titik pusat klaster (*centroid*) dan penetapan keanggotaan data terhadap masing-masing klaster. Secara umum, tahapan pelaksanaan algoritma ini dapat dijabarkan sebagai berikut [12]:

a. Tetapkan jumlah klaster (*k*) yang akan dibentuk.
b. Pilih sejumlah titik pusat awal (*centroid*) secara acak sebanyak *k* buah.
c. Hitung jarak antara setiap objek data dengan setiap *centroid* masing-masing klaster.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Keterangan:
$d(x,y)$ = *Euclidean Distance* antara objek *x* dan *y*
x_i = Nilai variabel ke-*i* dari objek *x*
y_i = Nilai variabel ke-*i* dari objek *y*
n = Jumlah variabel (dimensi) data

d. Tempatkan setiap objek ke dalam klaster yang memiliki jarak paling dekat dengan titik pusat (*centroid*). Pendekatan yang digunakan adalah *hard clustering*, artinya setiap data hanya dapat menjadi anggota dalam satu klaster berdasarkan kedekatan jaraknya dengan *centroid*.
e. Perbarui posisi *centroid* di setiap klaster dengan menghitung rata-rata semua data di klaster. Posisi rata-rata dihitung menggunakan persamaan (2).

$$c_j = \frac{1}{|C_j|}\sum_{x_i \in C_j} x_i \quad (2)$$

Keterangan:
c_j = Titik pusat (*centroid*) dari klaster ke-*j*
C_j = Himpunan objek yang menjadi anggota klaster ke-*j*
x_i = Objek data dalam klaster tersebut
$|C_j|$ = Jumlah objek dalam klaster ke-*j*

f. Ulangi langkah 3 hingga posisi *centroid* tidak berubah lagi dan tidak terjadi perpindahan anggota antar klaster. Kondisi ini menandakan bahwa algoritma telah mencapai konvergensi.

2.5. Silhoutte Index

Silhouette Index adalah salah satu metode evaluasi internal untuk menilai kualitas hasil pengelompokan. Metode ini menilai kecocokan suatu data terhadap klaster tempatnya berada, serta membandingkannya dengan kedekatan data tersebut terhadap klaster lainnya [13]. Nilainya berkisar antara -1 hingga 1, nilai yang mendekati 1 menunjukkan pengelompokan data yang benar, sementara nilai negatif menunjukkan kedekatannya dengan klaster lain dan oleh karena itu terdapat kemungkinan kesalahan pengelompokan [14]. *Silhouette Index* ini tidak hanya mengevaluasi tingkat kekompakan (*cohesion*) suatu klaster, tetapi juga mengukur pemisahan (*separation*) antar klaster, sehingga sangat bermanfaat dalam menentukan jumlah klaster yang optimal. Perhitungan indeks ini dilakukan melalui beberapa tahapan, yaitu [15]:

a. Hitung nilai *a(i)*, yang menunjukkan jarak rata-rata antara titik data *i* dengan semua titik data lain dalam klaster yang sama.

$$a(i) = \frac{1}{|C_i| - 1}\sum_{\substack{j \in C_i \\ j \neq i}} d(i,j) \quad (3)$$

Keterangan:
C_i = Klaster tempat objek *i* berada
$|C_i|$ = Jumlah objek dalam klaster C_i
d(i, j) = Jarak antara objek ke-*i* dan ke-*j*

b. Hitung nilai *b(i)*, yang menunjukkan jarak rata-rata antara titik data *i* dengan semua titik data dalam klaster lainnya. Nilai *b(i)* yang digunakan adalah nilai terkecil di antara semua klaster yang bukan tempat objek tersebut berada.

$$b(i) = \min_{C_j \neq C_i}\left(\frac{1}{|C_j|}\sum_{j \in C_j} d(i,j)\right) \quad (4)$$

Keterangan:
C_i = Klaster tempat objek *i* berada
C_j = Klaster lain yang bukan C_i
$|C_j|$ = Jumlah objek dalam klaster C_j
d(i, j) = Jarak antara objek ke-*i* dan ke-*j*

c. Setelah menentukan nilai *a(i)* dan *b(i)*, langkah selanjutnya adalah menghitung nilai *Silhouette s(i)* untuk setiap titik data *i*. Nilai ini menggambarkan sejauh mana kecocokan suatu data terhadap klaster tempatnya berada dibandingkan dengan kedekatannya terhadap klaster lainnya.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5)$$

Jika *s(i)* mendekati 1, artinya objek sangat sesuai dengan klaster tempatnya berada. Jika mendekati 0, berarti objek tersebut berada di batas

antara dua klaster. Jika bernilai negatif, objek tersebut kemungkinan besar dikelompokkan ke dalam klaster yang tidak sesuai, yang menandakan adanya kesalahan dalam proses klasterisasi.

d. Tahapan akhir evaluasi menggunakan *Silhouette Index* adalah menghitung rata-rata dari seluruh nilai *s(i)*, yang dikenal sebagai *Silhouette Score* atau indeks *Silhouette* keseluruhan. Nilai ini diperoleh dengan menjumlahkan seluruh nilai *s(i)* dan membaginya dengan jumlah total objek. Rumus perhitungan ditunjukkan pada persamaan (6).

$$SI = \frac{1}{N}\sum_{i=1}^{N} s(i) \quad (6)$$

Nilai rata-rata ini menjadi indikator utama kualitas klasterisasi; semakin mendekati 1, semakin baik pemisahan antar klaster dan kekompakan dalam klaster tersebut.

Interpretasi terhadap nilai *Silhouette Indeks* penting untuk menilai sejauh mana hasil klasterisasi telah dilakukan secara optimal. Tabel 2.1 memberikan acuan kualitatif terhadap rentang nilai tersebut dalam mengevaluasi kualitas pengelompokan.

Tabel 1. Interpretasi nilai *silhoutte index*

Nilai Silhouette	Interpretasi Kualitas Klasterisasi
0.71 − 1.00	Klaster memiliki struktur yang sangat kuat
0.51 − 0.70	Klaster terbentuk dengan struktur yang baik
0.26 − 0.50	Klaster menunjukkan struktur yang kurang kokoh
≤ 0.25	Klasterisasi tidak menunjukkan struktur yang jelas

2.6. Overpopulasi

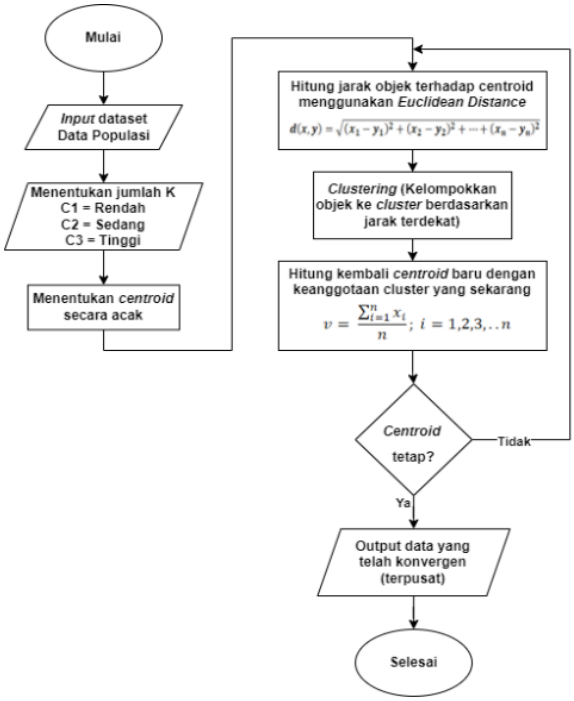
Overpopulasi merupakan kondisi ketika jumlah penduduk suatu wilayah melebihi kapasitas lingkungan untuk menyediakan sumber daya yang cukup guna memenuhi kebutuhan dasar manusia, seperti makanan, air, dan tempat tinggal. Angka kelahiran yang tinggi, tingkat kematian yang rendah, dan migrasi penduduk yang tidak terkendali adalah penyebab overpopulasi terjadi. Menurut Ridwan dkk., overpopulasi dapat menyebabkan kerusakan lingkungan yang serius, terutama di daerah perkotaan dengan tingkat kepadatan penduduk yang tinggi [16]. Di Indonesia, overpopulasi merupakan tantangan utama dalam pembangunan berkelanjutan, di mana pertumbuhan penduduk yang tidak terkendali dapat memperlambat pencapaian kesejahteraan nasional [3]. Oleh karena itu, untuk mengurangi dampak buruk overpopulasi terhadap lingkungan dan kesejahteraan masyarakat secara keseluruhan, diperlukan strategi pengendalian populasi yang efektif.

3. METODE PENELITIAN

3.1. Sumber Data

Studi ini menggunakan data demografi Indonesia dari tahun 2021 hingga 2024, yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS). Data yang dianalisis meliputi informasi luas wilayah, jumlah penduduk, kepadatan penduduk, dan laju pertumbuhan penduduk setiap kabupaten/kota di Indonesia, yang dihitung berdasarkan 514 entri data per tahun.

3.2. Flowchart Algoritma K-Means



Gambar 1. *Flowchart* metode *k-means*

Gambar 1 menyajikan alur kerja dari metode pengelompokan *K-Means*. Proses dimulai dengan memasukkan data populasi serta menetapkan jumlah klaster (*k*), yaitu tiga klaster: rendah, sedang, dan tinggi. Setelah itu, titik pusat awal (*centroid*) ditentukan. Setiap objek kemudian dihitung jaraknya terhadap titik pusat menggunakan metode jarak *Euclidean* untuk menentukan keanggotaan klaster. Proses ini dilakukan secara iteratif hingga titik pusat mencapai kondisi stabil dan tidak mengalami perubahan.

3.3. Implementasi Metode K-Means

Tahapan ini mencakup penerapan awal algoritma *K-Means* guna mengelompokkan wilayah-wilayah berdasarkan tingkat risiko overpopulasi. Analisis dilakukan terhadap data yang mencakup rentang waktu dari tahun 2021 hingga 2024. Informasi terkait data klasterisasi untuk tahun 2021 disajikan dalam Tabel 2.

Tabel 2. Data populasi tahun 2021

No	Provinsi	Kabupaten/ Kota	X1	X2	X3	X4
1	Aceh	Kab. Simeulue	2051,48	93800	46	0,96
2	Aceh	Kab.Aceh Singkil	2185	128400	59	1,46
3	Aceh	Kab.Aceh Selatan	3841,6	234600	61	0,94
4	Aceh	Kab.Aceh Tenggara	4231,43	224100	53	1,45
5	Aceh	Kab.Aceh Timur	6286,01	427000	68	1,08
6	Aceh	Kab.Aceh Tengah	4318,39	218700	51	1,42
7	Aceh	Kab.Aceh Barat	2927,95	200600	69	0,92
8	Aceh	Kab.Aceh Besar	2969	409500	138	0,97
9	Aceh	Kab.Pidie	3086,95	439400	142	0,94
10	Aceh	Kab.Bireuen	1901,2	439800	231	0,77
....	...	....	....	....	....	....
514	Papua	Kota Jayapura	935,92	404004	432	1,85

Keterangan:
X1: Luas Wilayah
X2: Jumlah Penduduk
X3: Kepadatan Penduduk
X4: Laju Pertumbuhan Penduduk per Tahun

Berikut merupakan implementasi perhitungan metode K-Means dalam mengelompokkan data populasi:

1. Langkah awal menentukan jumlah klaster (k) = 3.
2. Langkah selanjutnya, menentukan nilai *centroid* awal sebanyak *k* buah secara acak dari dataset.

Tabel 3. Menentukan *centroid* awal

Centroid Klaster	X1	X2	X3	X4
C1	678,32	22860	34	1,86
C2	1285,74	287700	224	0,72
C3	2710,62	5489500	2025	1,54

3. Langkah selanjutnya, menghitung jarak antara setiap titik data atau objek dengan setiap *centroid* masing-masing klaster dengan mengunakan rumus *Euclidean Distance* pada persamaan (1).

Tabel 4. Menghitung jarak terdekat

No	X1	X2	X3	X4	C1	C2	C3	Terdekat
1	2051,48	93800	46	0,96	70953,29	193901,59	5395700,40	70953,29
2	2185	128400	59	1,46	105550,76	159302,62	5361100,39	105550,76
3	3841,6	234600	61	0,94	211763,63	53161,72	5254900,49	53161,72
4	4231,43	224100	53	1,45	201271,37	63668,41	5265400,59	63668,41
5	6286,01	427000	68	1,08	404178,90	139389,80	5062501,64	139389,80
....	....	....	....		....	....	....	....
514	935,92	404004	432	1,85	381144,29	116304,71	5085496,56	116304,71

Tabel 4 merupakan hasil iterasi ke-1 algoritma K-Means untuk data tahun 2021. Kolom X1 merupakan Luas Wilayah, X2 adalah Jumlah Penduduk, X3 menunjukkan Kepadatan Penduduk, dan X4 adalah Laju Pertumbuhan Penduduk per Tahun. Kolom C1, C2, dan C3 menunjukkan jarak data ke masing-masing *centroid* klaster, sementara kolom Terdekat menunjukkan *centroid* dengan jarak terkecil yang menentukan data tersebut masuk ke klaster mana.

4. Selanjutnya, menentukan keanggotaan klaster berdasarkan nilai Jarak *Euclidean* terdekat terhadap ketiga *centroid* C1, C2, dan C3, yang ditunjukkan pada kolom Terdekat. Data dikelompokkan ke dalam *cluster* dengan *centroid* yang paling dekat.

Tabel 5. Penentuan klaster

Data	*Cluster*
1	1
2	1
3	2
4	2
5	2
....	....
514	2

5. Langkah selanjutnya, menghitung ulang posisi *centroid* untuk setiap klaster dengan menghitung rata-rata semua data yang menjadi anggota klaster tersebut. Ini dilakukan dengan menghitung rata-rata dari semua data yang menjadi anggota klaster tersebut.

Tabel 6. *Centroid* baru

Centroid Klaster	X1	X2	X3	X4
C1 Baru	4582,34	99925,10	374,76	1,47
C2 Baru	3465,86	631027,55	1291,91	1,19
C3 Baru	1379,60	3732706,60	5332,20	1,56

Setelah *centroid* baru didapatkan maka hitung nilai masing-masing titik data terhadap nilai *centroid* baru pada iterasi ke-2.

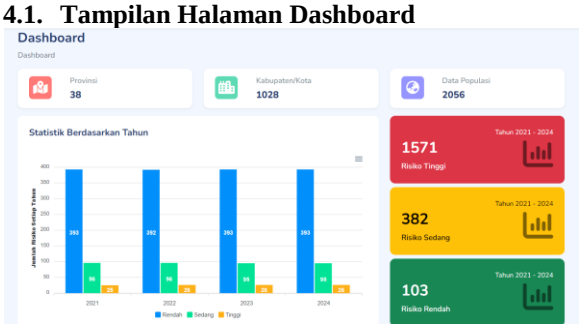
6. Jika posisi *centroid* masih berubah, maka langkah 3 hingga 5 diulang hingga nilai *centroid* stabil dan tidak terdapat perpindahan data antar klaster. Berdasarkan hasil perhitungan hingga iterasi ke-10, dari total 514 data, sebanyak 393 data tergolong dalam klaster C1 (Rendah), 96 data dalam klaster C2 (Sedang), dan 25 data dalam klaster C3 (Tinggi).

Tabel 7. Hasil iterasi ke-10 pada tahun 2021

No	C1	C2	C3	Terdekat	Cluster
1	6635,53	537230,86	3638910,50	157073,43	1
2	28577,38	502630,70	3604310,55	122475,79	1
3	134677,30	396429,64	3498111,44	16272,98	1
4	124175,81	406930,16	3508611,73	26763,20	1
5	327079,48	204050,71	3305714,43	176155,92	1
....	....	....	....	....	....
514	304100,77	227039,28	3328706,24	153184,08	1

4. HASIL DAN PEMBAHASAN

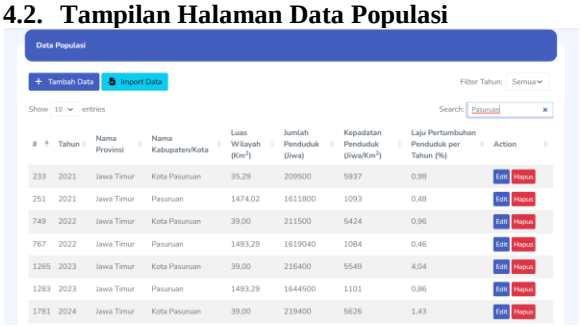
4.1. Tampilan Halaman Dashboard



Gambar 2. Tampilan dashboard

Gambar 2 menampilkan halaman dashboard yang merangkum sejumlah informasi penting, seperti total provinsi, jumlah kabupaten/kota, dan jumlah entri data populasi. Selain itu, tersedia visualisasi grafik yang menunjukkan jumlah klaster berdasarkan tingkat risiko.

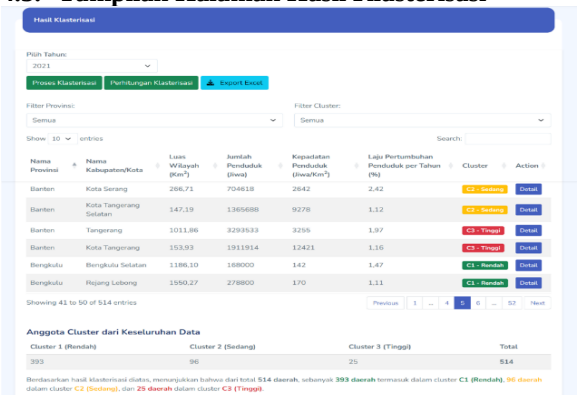
4.2. Tampilan Halaman Data Populasi



Gambar 3. Tampilan halaman data populasi

Gambar 3 berisi tampilan halaman data populasi, yang mencakup informasi seperti tahun, nama provinsi, kabupaten/kota, luas wilayah (km²), jumlah penduduk (jiwa), kepadatan (jiwa/km²), serta tingkat pertumbuhan penduduk tahunan (%).

4.3. Tampilan Halaman Hasil Klasterisasi



Gambar 4. Tampilan halaman hasil klasterisasi

Gambar 4 memperlihatkan halaman untuk menampilkan hasil klasterisasi data populasi berdasarkan tahun yang dipilih. Pengguna dapat menentukan tahun tertentu dan menekan tombol "Proses Klasterisasi" untuk memulai, kemudian sistem akan menampilkan hasil pengelompokan sesuai tahun tersebut.

4.4. Tampilan Halaman Pemetaan

Gambar 5. Tampilan halaman pemetaan

Gambar 5 menampilkan halaman pemetaan wilayah-wilayah dengan potensi risiko overpopulasi di Indonesia berdasarkan hasil klasterisasi yang telah dihasilkan pada menu sebelumnya.

4.5. Pengujian Fungsional

Tabel 8. Pengujian fungsional sistem

Fitur	Skenario	Hasil yang Diharapkan	Ket
Login	Mengisi *username* dan *password* salah	Muncul pesan *error* jika *login* gagal	Berhasil
	Mengisi *username* dan *password*	Menuju ke halaman *dashboard*	Berhasil
Dashboard	Menampilkan informasi terkait penduduk dan klasterisasi	Tampil data sesuai yang tersimpan di *database*	Berhasil
Provinsi	Menampilkan data provinsi	Data provinsi ditampilkan	Berhasil
	Import data provinsi	Sistem berhasil mengimpor data provinsi	Berhasil
	Menambah data provinsi	Sistem berhasil menyimpan data	Berhasil
	Mengedit data provinsi	Sistem berhasil mengubah data	Berhasil
	Menghapus data provinsi	Sistem berhasil menghapus data	Berhasil
Kabupaten/Kota	Menampilkan data kabupaten/kota	Data kabupaten/kota ditampilkan	Berhasil
	Import data kabupaten/kota	Sistem berhasil mengimpor data	Berhasil
	Menambah data kabupaten/kota	Sistem berhasil menyimpan data	Berhasil
	Mengedit data kabupaten/kota	Sistem berhasil mengubah data	Berhasil
	Menghapus data kabupaten/kota	Sistem berhasil menghapus data	Berhasil
Data Populasi	Menampilkan data populasi	Data populasi ditampilkan	Berhasil
	Import data populasi	Sistem berhasil mengimpor data populasi	Berhasil
	Menambah data populasi	Sistem berhasil menyimpan data	Berhasil
	Mengedit data populasi	Sistem berhasil mengubah data	Berhasil
	Menghapus data populasi	Sistem berhasil menghapus data	Berhasil

Hasil Klasterisasi	Memilih tahun untuk menampilkan hasil klaterisasi	Sistem menampilkan hasil klasterisasi dari tahun yang dipilih	Berhasil
	Menekan tombol proses klasterisasi	Sistem melakukan perhitungan k-means dan menyimpan hasil klasterisasi	Berhasil
	Menampilkan perhitungan klasterisasi	Sistem menampilkan hasil iterasi k-means	Berhasil
Pemetaan	Memilih tahun untuk menampilkan pemetaan daerah	Sistem berhasil menampilkan pemetaan daerah berdasarkan hasil klasterisasi dari tahun yang dipilih	Berhasil

Untuk menilai kesesuaian data masukan dan keluaran, dilakukan uji fungsional sistem klasterisasi wilayah risiko overpopulasi di Indonesia. Uji ini dilakukan pada tahap akhir pengembangan sistem untuk memastikan semua fungsi berfungsi sebagaimana mestinya. Metode yang digunakan adalah pengujian kotak hitam, suatu pendekatan yang mengevaluasi fungsionalitas sistem sesuai spesifikasi tanpa mempertimbangkan struktur internal kode program. Berdasarkan hasil pengujian yang disajikan pada Tabel 8, sistem mampu menjalankan ketujuh fitur desain dan kasus uji, serta hasilnya memenuhi kriteria pengujian yang ditetapkan.

4.6. Pengujian Pengguna

Tabel 9. Pengujian pengguna

No	Pernyataan	Hasil Skor				
		1	2	3	4	5
1	Saya menemukan bahwa navigasi menu dan fitur yang tersedia pada website mudah digunakan.	0	0	1	6	5
2	Saya tidak mengalami kesulitan saat mencari data atau mengakses informasi dalam website.	0	0	3	5	4
3	Saya merasa bahwa hasil pengelompokan risiko overpopulasi (Rendah, Sedang, Tinggi) mudah dipahami.	0	0	1	6	5
4	Saya pikir label, warna, dan elemen visual lainnya pada tampilan peta dan tabel cukup jelas dan informatif.	0	1	0	6	5
5	Saya merasa terbantu dengan fitur visualisasi data seperti grafik, peta, dan tabel interaktif.	0	0	1	4	7
6	Saya merasa visualisasi data dalam website mempermudah saya memahami kondisi daerah berisiko overpopulasi.	0	0	1	6	5
7	Saya merasa antarmuka pengguna (user interface) website ini menarik secara desain dan mudah dipahami.	0	0	4	5	3
8	Saya pikir website ini secara keseluruhan mudah digunakan, bahkan bagi pengguna yang baru pertama kali mengaksesnya.	0	0	3	6	3
9	Saya pikir informasi yang disajikan dalam website ini relevan dan sesuai dengan kebutuhan saya.	0	1	1	6	4
10	Saya bersedia menggunakan kembali website ini di masa mendatang jika diperlukan.	0	0	1	7	4
Total Jawaban		0	2	16	57	45
Total Seluruh Jawaban		120				

Pengujian pengguna dilakukan untuk mengevaluasi pengalaman pengguna terhadap sistem klasterisasi daerah berisiko overpopulasi berbasis web. Tujuannya adalah menilai kemudahan penggunaan dan tingkat pemahaman pengguna terhadap sistem. Evaluasi dilakukan melalui kuesioner menggunakan skala Likert (1–5) kepada sejumlah responden yang telah mencoba sistem, di mana 1 = Sangat Tidak Setuju, 2 = Tidak Setuju, 3 = Netral, 4 = Setuju, 5 = Sangat Setuju.

Untuk mengetahui distribusi penilaian pengguna terhadap sistem, dilakukan perhitungan persentase setiap kategori jawaban menggunakan rumus:

$$\text{Persentase Total} = \frac{\text{Jumlah Jawaban}}{\text{Total Seluruh Jawaban}} \times 100\% \quad (7)$$

Dari total 120 jawaban yang diperoleh (12 responden × 10 pernyataan), hasil persentase menunjukkan: Sangat Tidak Setuju 0%, Tidak Setuju 1,67%, Netral 13,33%, Setuju 47,5%, dan Sangat Setuju 37,5%. Secara keseluruhan, lebih dari 85% responden memberikan tanggapan positif (Setuju dan Sangat Setuju), yang menandakan bahwa sistem telah diterima dengan baik, mudah digunakan, dan mampu menyajikan informasi dengan jelas.

4.7. Pengujian Konsistensi di Lapangan

Tabel 10. Pengujian konsistensi di lapangan 2024

No	Kabupaten/ Kota	Luas Wilayah	Jumlah Penduduk	Kepadatan Penduduk	Laju Pertumbuhan Penduduk	Klaster Sistem	Temuan Lapangan
1	Kota Jakarta Barat	125	2479600	19837	0,49	3 (Tinggi)	Kota memiliki luas kecil, penduduk tinggi, kepadatan tinggi, dan laju pertumbuhan rendah maka masuk risiko tinggi.

No	Kabupaten/ Kota	Luas Wilayah	Jumlah Penduduk	Kepadatan Penduduk	Laju Pertumbuhan Penduduk	Klaster Sistem	Temuan Lapangan
2	Kota Jakarta Selatan	144,94	2230700	15390	0,05	3 (Tinggi)	Kota memiliki luas kecil, penduduk tinggi, kepadatan tinggi, dan laju pertumbuhan rendah maka masuk risiko tinggi.
3	Kota Bandung	166,59	2528200	15176	0,91	3 (Tinggi)	Kota memiliki luas kecil, penduduk tinggi, kepadatan tinggi, dan laju pertumbuhan rendah maka masuk risiko tinggi.
4	Kota Tangerang	178,35	1964000	11012	0,95	3 (Tinggi)	Kota memiliki luas kecil, penduduk tinggi, kepadatan tinggi, dan laju pertumbuhan rendah maka masuk risiko tinggi.
...	...	...	...	...	...	...	...
514	Merauke	45013,35	239400	5	0,79	1 (Rendah)	Kabupaten memiliki luas besar, penduduk sedang, kepadatan rendah, dan laju pertumbuhan rendah maka masuk risiko rendah.

Berdasarkan hasil validasi terhadap kondisi nyata di lapangan yang disajikan pada Tabel 10, sistem menunjukkan tingkat konsistensi yang cukup tinggi antara hasil klasterisasi dan klasifikasi berdasarkan indikator demografis (luas wilayah, jumlah penduduk, kepadatan, dan laju pertumbuhan). Dari total 514 wilayah, sebanyak 400 wilayah (77,82%) menunjukkan kecocokan hasil pada tahun 2024.

Secara keseluruhan, tingkat konsistensi sistem dari tahun 2021 hingga 2024 berkisar antara 77–78%, sebagaimana ditunjukkan pada Tabel 11 berikut:

Tabel 11. Persentase Pengujian Konsistensi per Tahun

Tahun	Total Data	Jumlah Data Konsisten	Jumlah Data Tidak Konsisten	Konsistensi (%)
2021	514	402	112	78,21%
2022	514	400	114	77,82%
2023	514	399	115	77,63%
2024	514	400	114	77,82%

Hasil ini menunjukkan bahwa algoritma *K-Means* yang diterapkan memiliki stabilitas dan keakuratan yang baik dalam mengelompokkan wilayah sesuai karakteristik demografis tiap tahun.

4.8. Pengujian Kualitas Model

Pengujian kualitas model dilakukan menggunakan *Silhouette Index* untuk mengevaluasi kualitas hasil klasterisasi yang diperoleh dari penerapan algortima K-Means. Pengujian dilakukan dengan jumlah klaster yang telah ditetapkan, yaitu tiga klaster. Nilai *Silhouette Index* dihitung untuk masing-masing tahun 2021-2024 guna menilai seberapa baik struktur klaster yang terbentuk setiap tahun.

Berdasarkan Tabel 12, nilai *Silhouette Index* yang diperoleh untuk tahun 2021 hingga 2024 berada diatas 0,71. Rentang nilai ini menunjukkan bahwa struktur klaster yang dihasilkan kuat, yang berarti pengelompokan data sudah sangat baik dan batas antar klaster sangat jelas.

Tabel 12. Pengujian *silhouette index*

Tahun	Jumlah Klaster (k)	Nilai Silhouette	Jumlah Risiko Rendah	Jumlah Risiko Sedang	Jumlah Risiko Tinggi
2021	3	0,724046	393	96	25
2022	3	0,721511	392	96	26
2023	3	0,721313	392	96	26
2024	3	0,721605	393	95	26

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, sistem aplikasi yang dikembangkan mampu menjalankan proses klasterisasi wilayah di Indonesia berdasarkan tingkat risiko overpopulasi secara efektif. Aplikasi ini tidak hanya berhasil mengelompokkan data berdasarkan tahun ke dalam beberapa klaster risiko secara otomatis, tetapi juga mampu menyajikan hasil dalam bentuk visualisasi peta yang informatif. Hasil pengujian pengguna menunjukkan tingkat kepuasan yang tinggi, dengan lebih dari 85% responden memberikan penilaian positif bahwa sistem mudah digunakan dan dipahami. Uji konsistensi dengan data lapangan menghasilkan tingkat kecocokan rata-rata 77–78% selama periode 2021–2024, menandakan algoritma *K-Means* cukup akurat dan stabil. Untuk pengembangan penelitian selanjutnya, disarankan menggunakan metode inisialisasi *centroid K-Means++* guna meningkatkan stabilitas hasil klasterisasi, serta mempertimbangkan penambahan metrik evaluasi baru, seperti *Davies-Bouldin Index* untuk memperkuat validasi terhadap kualitas klaster yang dihasilkan.

DAFTAR PUSTAKA

[1] T. Talibi, C. Onyeneke, and C. Eguzouwa, “Natural and Human Impacts of Overpopulation on the Economy,” *International Journal of Science, Technology and Society*, vol. 10, no. 6, pp. 217–223, 2022, doi: 10.11648/j.ijsts.20221006.11.

[2] S. Temalagi, T. B. Sembiring, J. Nasution, and ..., "The Importance of Public Knowledge for a Sustainable Environment," *J Surv Fish Sci*, vol. 10, no. 1, pp. 2869–2875, 2023, doi: 10.53555/sfs.v10i1.1269.

[3] O. Milanda, N. Almira, and G. A. Malik, "Besarnya Pertumbuhan Angka Penduduk Indonesia (Overpopulation) Dalam Perspektif Undang-Undang Nomor 24 Tahun 2013 Tentang Administrasi Kependudukan," *JME: Journal of Multidiscipline & Equality*, vol. 1, no. 2, pp. 43–53, 2024.

[4] F. Handayanna and Sunarti, "Penerapan Algoritma K-Means Untuk Mengelompokkan Kepadatan Penduduk Di Provinsi DKI Jakarta," *Journal of Applied Computer Science and Technology (JACOST)*, vol. 5, no. 1, pp. 50–55, 2024, doi: 10.52158/jacost.v5i1.477.

[5] M. Iqbal, S. Syaripuddin, and M. N. Huda, "Implementasi Algoritma K-Means Clustering dengan Jarak Euclidean dalam Mengelompokkan Daerah Penyebaran COVID-19 di Kabupaten Bogor," *BASIS (Jurnal Ilmiah Matematika)*, vol. 2, no. 1, pp. 47–56, 2023.

[6] Nabiilah, F. J. Fauzan, N. Nurazizah, A. Hamid, and S. F. Octavia, "Penerapan Algoritma K-Means dan K-Medoids Dalam Pengelompokan Kepadatan Penduduk Provinsi Riau," in *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 2023, pp. 223–229.

[7] G. M. M. Sujak, H. N. Rofiq, and F. I. Tawakal, "Implementasi K-Means Clustering untuk Optimalisasi Anggaran Penyakit Tidak Menular," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 67–74, 2025.

[8] Nurhayah, N. Suarna, and W. Prihartono, "Analisis Jumlah Penduduk Menggunakan Algoritma K-Means Berdasarkan Kabupaten/Kota Di Indonesia," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 894–901, 2024, doi: 10.36040/jati.v8i1.8863.

[9] P. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.

[10] I. D. Anjani and A. Bahtiar, "Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Penerima Bantuan Sosial Tunai (BST) Di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 3, pp. 2743–2747, 2024, doi: 10.36040/jati.v8i3.8974.

[11] A. Maulana, R. D. Dana, and N. D. Nuris, "Implementasi Algoritma K-Means Clustering Dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam Di Kabupaten Cirebon," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1417–1424, 2024, doi: 10.36040/jati.v8i2.9024.

[12] N. K. S. Julyantari, I. K. Budiarta, and N. M. D. K. Putri, "Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih)," *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 1, no. 2, pp. 92–101, 2021, doi: 10.25008/janitra.v1i2.134.

[13] M. Orisa and A. Faisol, "Researchers Productivity Level Clustering Based On H-Index and Citation Using The Fuzzy C-Means Algorithm," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 5, no. 1, pp. 18–26, 2023, doi: 10.47709/cnahpc.v5i1.1984.

[14] R. Ananda and A. Z. Yamani, "Penentuan Centroid Awal K-means pada proses Clustering Data Evaluasi Pengajaran Dosen," *JURNAL RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, no. 3, pp. 544–550, 2020.

[15] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.

[16] M. Ridwan, S. Hidayanti, and Nilfatri, "Studi Analisis Tentang Kepadatan Penduduk Sebagai Sumber Kerusakan Lingkungan Hidup," *Jurnal IndraTech*, vol. 2, no. 1, pp. 25–36, 2021