
CLUSTERING DATA KEPADATAN PENDUDUK KOTA PEMATANGSIANTAR (PER KM²) MENURUT KECAMATAN MENGGUNAKAN METODE K-MEANS CLUSTERING

¹Natanael Napitupulu, ²Simon Bonar Purba, ³Dhany Tampe Silalahi,
⁴Adven Sinamo, ⁵Indra Gunawan

Program Studi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar

Email: natanaelnapitupulu1@gmail.com / simonpurba211104@gmail.com / dhanytampe05@gmail.com /
advenagape20@gmail.com / indra@amiktunasbangsa.ac.id

Abstrak

Penelitian ini bertujuan untuk mengelompokkan data kepadatan penduduk per km² pada delapan Kecamatan di Kota Pematangsiantar selama periode tahun 2022 hingga 2024 menggunakan metode K-Means Clustering. Metode ini digunakan untuk mengidentifikasi pola kepadatan penduduk dan mengklasifikasikan kecamatan ke dalam tiga cluster, yaitu: tinggi, sedang, dan rendah. Pengolahan data dilakukan dengan menggunakan perangkat lunak RapidMiner serta perhitungan manual untuk validasi hasil. Hasil penelitian menunjukkan bahwa Kecamatan Siantar Barat dan Siantar Utara termasuk dalam cluster kepadatan tinggi, Siantar Selatan dan Siantar Timur dalam cluster sedang, dan empat kecamatan lainnya termasuk dalam cluster rendah. Metode K-Means Clustering terbukti efektif dalam mengelompokkan data kepadatan penduduk dan dapat digunakan sebagai dasar pengambilan keputusan dalam perencanaan wilayah dan tata kota.

Kata Kunci: Kepadatan Penduduk; K-Means Clustering; Kota Pematangsiantar; Data Mining;

CLUSTERING POPULATION DENSITY DATA OF PEMATANGSIANTAR CITY (PER KM²) BY DISTRICT USING THE K-MEANS CLUSTERING METHOD

Abstract

This study aims to classify population density data per square kilometer across eight districts in Pematangsiantar City for the period 2022 to 2024 using the K-Means Clustering method. This method is employed to identify patterns in population density and to categorize the districts into three clusters: high, medium, and low. Data processing was carried out using the RapidMiner software and validated through manual calculations. The results show that Siantar Barat and Siantar Utara fall into the high-density cluster, Siantar Selatan and Siantar Timur into the medium cluster, and the remaining four districts into the low-density cluster. The K-Means Clustering method has proven effective in grouping population density data and can serve as a basis for strategic decision-making in urban planning and regional development.

Keywords: Population Density; K-Means Clustering; Pematangsiantar City; Data Mining;

1. PENDAHULUAN

Kepadatan penduduk merupakan salah satu indikator penting dalam perencanaan pembangunan wilayah, khususnya di daerah perkotaan. Kepadatan yang tinggi dapat mempengaruhi kualitas hidup masyarakat, beban infrastruktur, dan distribusi layanan publik. Di Kota Pematangsiantar, kepadatan penduduk antar kecamatan menunjukkan perbedaan yang cukup mencolok. Oleh karena itu, analisis yang tepat diperlukan untuk memahami distribusi penduduk dan mendukung kebijakan yang berbasis data.

Tingginya kepadatan penduduk di suatu wilayah tidak hanya berdampak pada aspek fisik dan sosial, tetapi juga dapat memicu meningkatnya angka kriminalitas dan permasalahan sosial lainnya. Penelitian oleh Sabiq (2022) mengungkapkan bahwa wilayah padat penduduk lebih rentan terhadap tindakan kriminal akibat tekanan ekonomi, keterbatasan ruang, serta kurangnya akses terhadap fasilitas publik dan keamanan.[2]

Seiring dengan berkembangnya teknologi informasi, metode Data Mining menjadi alat yang efektif untuk menganalisis data dalam jumlah besar. Salah satu metode yang umum digunakan adalah K-Means Clustering, yang dapat mengelompokkan data berdasarkan kesamaan karakteristik. Dengan menggunakan metode ini, distribusi kepadatan penduduk di setiap kecamatan dapat dianalisis secara sistematis, sehingga menghasilkan informasi yang bermanfaat untuk pengambilan keputusan dalam tata kelola kota.

Selain untuk kebutuhan analisis teknis, pemetaan kepadatan penduduk juga berguna dalam mengidentifikasi dampak sosial ekonomi yang muncul. Dita dan Legowo (2022) menyatakan bahwa tingginya kepadatan bisa memperburuk kemiskinan dan menyebabkan degradasi sosial karena tekanan terhadap sumber daya dan ruang yang terbatas. Oleh sebab itu, penting untuk memiliki gambaran yang akurat tentang konsentrasi penduduk sebagai dasar perencanaan kebijakan sosial dan pembangunan.[3]

Penelitian ini menggunakan data sekunder dari Badan Pusat Statistik Kota Pematangsiantar mengenai jumlah penduduk dan luas wilayah tiap kecamatan di Kota Pematangsiantar selama tahun 2022 hingga 2024. Tujuannya adalah untuk mengelompokkan kecamatan berdasarkan tingkat kepadatan penduduk menggunakan algoritma K-Means. Hasil pengelompokan ini diharapkan dapat memberikan gambaran yang lebih jelas terkait

konsentrasi penduduk, serta mendukung perencanaan pembangunan kota yang lebih merata dan berkelanjutan.

Dengan pendekatan ini, diharapkan Kota Pematangsiantar dapat memiliki dasar yang kuat dalam menata ruang kota, mengatur penyebaran fasilitas umum, serta mendorong pertumbuhan kota yang inklusif. Klasifikasi wilayah berdasarkan kepadatan juga memungkinkan pemerintah daerah mengidentifikasi kebutuhan infrastruktur dan layanan sosial secara lebih presisi, serta meningkatkan efektivitas dalam pengambilan keputusan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Kepadatan Penduduk

Menurut BPS (Badan Pusat Statistik), kepadatan penduduk adalah perbandingan antara jumlah penduduk dengan luas wilayah yang dihuni, dengan kata lain kepadatan penduduk menunjukkan jumlah penduduk yang tinggal di setiap kilometer persegi suatu wilayah. Kepadatan penduduk dihitung dengan membagi jumlah penduduk suatu wilayah dengan luas wilayah tersebut. Kepadatan Penduduk Kasar (Crude Population Density): Merupakan ukuran yang paling umum digunakan, menunjukkan jumlah penduduk per kilometer persegi dari total luas wilayah. Kepadatan penduduk adalah suatu keadaan yang dikatakan semakin padat bila jumlah manusia pada suatu batas ruang tertentu semakin banyak dibandingkan dengan luas ruangnya

kepadatan penduduk merupakan salah satu faktor yang memicu tindakan kriminal karena daerah dengan penduduk yang mengalami padat permasalahan cenderung ekonomi, kesejahteraan, kebutuhan pangan serta kurangnya tingkat keamanan yang berujung pada tindakan kriminal. [2] Kepadatan penduduk sangat berpengaruh terhadap kehidupan sosial ekonomi masyarakat seperti berpengaruh terhadap kemiskinan dan degradasi sosial. Kepadatan penduduk juga dapat mempengaruhi kualitas hidup masyarakat. [3]

Jumlah penduduk yang terus bertambah mengakibatkan pesatnya pembangunan, baik pembangunan tempat tinggal maupun sarana penunjang bagi kehidupan manusia. Pembangunan tersebut seakan bersaing dalam memperebutkan ruang yang ada, sedangkan ruang yang sudah ada ini sudah tidak bisa bertambah lagi. Hal ini tentu akan

menimbulkan masalah sosial, seperti kualitas hidup manusia [4]

Kepadatan Penduduk menurut beberapa ahli : Soekartawi (2004), kepadatan penduduk adalah jumlah penduduk yang tinggal di suatu wilayah tertentu dalam satuan luas area tertentu, biasanya dihitung per kilometer persegi. Kepadatan ini menggambarkan tingkat kepadatan ruang yang dihuni oleh penduduk.[5] Santoso (2013), kepadatan penduduk merupakan ukuran yang menunjukkan jumlah penduduk yang tinggal dalam satu satuan luas wilayah tertentu, yang dapat memberikan gambaran tentang distribusi penduduk dan tekanan terhadap sumber daya wilayah tersebut.[6] Kartodirdjo (1992) mendefinisikan kepadatan penduduk sebagai rasio jumlah penduduk terhadap luas wilayah yang dihuni, yang berpengaruh terhadap berbagai aspek sosial, ekonomi, dan lingkungan.[7]

2.2. Data Mining

Data mining adalah proses pengumpulan dan pengolahan data yang bertujuan untuk mengekstrak informasi penting yang terdapat pada data. Informasi yang didapat bisa berupa angka, maupun informasi yang dapat digunakan untuk berbagai keperluan. Proses pengumpulan dan ekstraksi informasi ini dapat dilakukan menggunakan software dengan bantuan perhitungan statistika, matematika ataupun yang terbaru menggunakan Artificial Intelligence (AI)[8]. Data Mining Data mining adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam database. Data mining merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar[8]

Metode data mining melibatkan proses penggalian data dari berbagai sumber, pembersihan data untuk menghilangkan noise atau informasi yang tidak relevan, transformasi data ke dalam format yang lebih sesuai untuk analisis, serta penerapan algoritma dan teknik analisis untuk mengidentifikasi pola yang bermanfaat. Data mining digunakan dalam berbagai bidang seperti bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya untuk mendapatkan wawasan yang dapat membantu perencanaan strategis dan pengambilan keputusan. Tujuan utama dari data mining adalah untuk mengubah data mentah menjadi wawasan yang dapat diaplikasikan dalam berbagai

bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keuangan, dan lainnya. Dalam konteks bisnis, data mining dapat membantu dalam mengidentifikasi tren konsumen, segmentasi pasar, rekomendasi produk, deteksi penipuan, dan masih banyak lagi[11]

“Data Mining merupakan proses menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi maupun mengidentifikasi data dan informasi yang bermanfaat atau pengetahuan yang terkait dari berbagai database yang besar”. [12]

Menurut beberapa ahli seperti : Han dan Kamber (2012) *Data mining* adalah proses menemukan pola menarik dari data dalam jumlah besar dan menyimpannya dalam basis data, gudang data, atau tempat penyimpanan data lainnya.[13] Witten, Frank, dan Hall (2011) Data mining adalah proses otomatis untuk menemukan pola dan relasi yang berguna dalam kumpulan data yang besar.[14] Turban, Aronson, dan Liang (2007) Data mining adalah proses menemukan informasi penting seperti pola, hubungan, dan tren tersembunyi dari data yang sangat besar yang tersimpan dalam basis data.[15] Fayyad et al. (1996) Data mining merupakan langkah dalam proses Knowledge Discovery in Databases (KDD) yang menggunakan metode statistik, matematis, atau pembelajaran mesin untuk mengekstrak informasi yang berguna dari data. [16]

2.3. Algoritma K-Means Clustering

K-Means adalah metode yang ada pada data mining dengan proses awal mengambil sebagian komponen populasi untuk dijadikan titik pusat di cluster awal. Penetapan kelompok dalam satu cluster dapat dilakukan dengan menghitung jarak setiap objek dengan titik pusat. Algoritma clustering K-Means dapat membagi data berdasarkan jarak antar data pada kelompok yang telah ditetapkan. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi Euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat[9].

Algoritma K-Means Clustering adalah metode pengelompokan data yang menggunakan centroid sebagai pusat kelompok dan berupaya meminimalkan jarak antar data dengan centroid tersebut. Proses ini dilakukan secara iteratif sampai posisi centroid stabil (Nugraha et al., 2022).[8] Metode ini mengelompokkan data berdasarkan kemiripan fitur

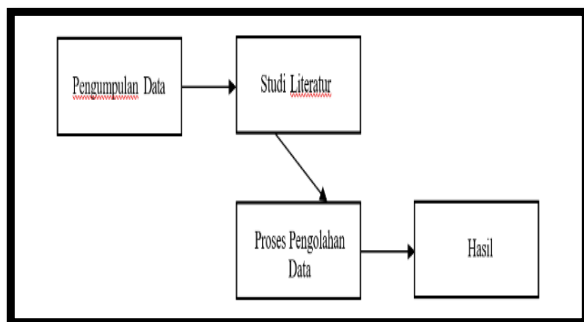
menggunakan jarak Euclidean, dengan tujuan mengoptimalkan variansi dalam setiap cluster dan memaksimalkan perbedaan antar cluster (Arum et al., 2024).[10] Algoritma ini juga efektif untuk mengelompokkan data dalam berbagai bidang, termasuk analisis kesehatan, dengan kemampuan menangani dataset berukuran besar (Prasetyo et al., 2023).[11]

Clustering adalah proses membagi objek data menjadi beberapa kelompok yang biasa disebut cluster. Metode ini bermanfaat untuk mengidentifikasi kelompok yang tidak diketahui dalam data, seperti dalam bisnis intelligence yang dimana dapat mengelompokkan data penngangguran terbuka ke berbagai segmen serta mengelompokkan tingkat pengangguran di tiap daerah.[10]

3. METODOLOGI PENELITIAN

3.1. Langkah-langkah Penelitian

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik Kota Pematangsiantar, berupa jumlah penduduk dan luas wilayah delapan kecamatan selama tahun 2022 hingga 2024. Kajian pustaka dilakukan untuk memperoleh dasar teori dan konsep yang relevan, khususnya mengenai kepadatan penduduk dan metode K-Means Clustering dalam analisis data. Setelah data terkumpul, proses pengolahan dilakukan melalui perhitungan manual menggunakan Microsoft Excel serta validasi hasil menggunakan aplikasi RapidMiner. Algoritma K-Means Clustering digunakan untuk mengelompokkan kecamatan ke dalam tiga kategori berdasarkan tingkat kepadatan penduduk, yaitu tinggi, sedang, dan rendah, dengan tujuan menghasilkan informasi yang dapat digunakan dalam perencanaan dan pengambilan keputusan pembangunan wilayah.



Gambar 1 (Langkah-Langkah Penelitian)

Berikut adalah algoritma dari metode K-Means

Clustering :

a. Tentukan jumlah cluster (K)

Pilih jumlah klaster (K) sesuai kebutuhan atau berdasarkan metode seperti Elbow Method.

b. Inisialisasi centroid awal

Pilih secara acak K buah data dari dataset sebagai **centroid awal** (titik pusat klaster).

c. Hitung jarak setiap data ke setiap centroid

Gunakan **Euclidean Distance** untuk menghitung jarak antara data ke- x dan centroid ke- c dengan rumus:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

x = data point

c = centroid dari kluster

n = jumlah fitur

d. Kelompokkan data ke dalam klaster dengan jarak terdekat

Tentukan klaster untuk setiap data berdasarkan jarak terpendek ke centroid menggunakan:

$$Cluster(x) = \arg \min_i d(x, c_i)$$

e. Hitung ulang centroid baru untuk setiap kluster

Setelah semua data dikelompokkan, hitung centroid baru untuk setiap klaster berdasarkan rata-rata semua titik data dalam klaster tersebut:

$$c_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$$

c_j = centroid baru untuk klaster ke - j

S_j = himpunan data dalam klaster ke - j

$|S_j|$ = jumlah anggota dalam klaster ke - j

x_i = data ke- i dalam klaster tersebut

f. Ulangi langkah (c) hingga (e)

Lakukan iterasi langkah (c) sampai (e) **sampai tidak ada lagi data yang berpindah klaster**, atau **centroid tidak berubah** lagi (konvergen), atau hingga batas iterasi tercapai.

4. HASIL DAN PEMBAHASAN

4.1. Analisis dan Pengolahan Data

Penelitian ini menggunakan data dari Badan

Pusat Statistik, yang dapat diakses di <https://siantarkota.bps.go.id/>. Data yang digunakan untuk penelitian ini adalah data Kepadatan Penduduk Kota Pematangsiantar untuk 8 Kecamatan pada tahun 2022, 2023, dan 2024. Data diproses menggunakan metode K-Means Clustering, dan analisisnya bersifat sekunder, yang berarti bahwa data tidak berasal dari sumber langsung, tetapi telah diproses secara terpisah.

Tabel 1 (Tabel data)

KECAMATAN	2022	2023	2024	C0	C1	C2
SIANTAR MARIHAT	2673	2701	2736	19234,74	2672,48	10082,26
SIANTAR MARIMBUN	1148	1116	1218	21907,16	0,00	12751,93
SIANTAR SELATAN	8637	9016	8607	8755,21	13156,46	736,29
SIANTAR BARAT	11806	12382	11902	3085,29	18832,51	6112,98
SIANTAR UTARA	13667	13999	13757	0,00	21907,16	9176,13
SIANTAR TIMUR	8129	8553	8071	9625,95	12288,49	800,00
SIANTAR MARTOBA	2794	2819	2923	18989,21	2918,31	9834,62
SIANTAR SITILASARI	1511	1494	1573	21274,24	632,99	12119,74

Tabel 2 (Parameter)

Parameter	Nilai	Keterangan
Data	8	Kecamatan
Atribut	3	Tahun
Cluster	3	Tinggi, Sedang, Rendah

4.2. Pengolahan Data Menggunakan Perhitungan Manual K-Means

Berikut adalah langkah-langkah perhitungan K-Means Clustering:

- Menentukan jumlah cluster
Jumlah cluster yang digunakan sebanyak 3 cluster. Yakni C0, C1, dan C2.
- Penentuan Centroid
Penentuan titik cluster ini dilakukan secara acak. Untuk penentuan awal pusat cluster diambil dengan mengambil nilai maksimum dari masing-masing data pertumbuhan ekonomi 55 tingkat tinggi, nilai median untuk cluster sedang, dan nilai minimum untuk cluster rendah.

Cluster	Centroid		
	2022	2023	2024
C0 (Tinggi)	13667	13999	13757
C1 (Rendah)	1148	1116	1218
C2 (Sedang)	8129	8553	8871

C. Menemukan pusat cluster terdekat.

Setelah nilai k dan pusat cluster diketahui, lalu selanjutnya diukurlah jarak antar pusat cluster menggunakan rumus Euclidian distance, maka akan didapatkanlah jarak C1, C2, dan C3.

$$d_{1,1} = \sqrt{(13667 - 2673)^2 + (13999 - 2701)^2 + (13757 - 2736)^2} = 19234,74$$

$$d_{1,2} = \sqrt{(1148 - 2673)^2 + (1116 - 2701)^2 + (1218 - 2736)^2} = 2672,48$$

$$d_{1,3} = \sqrt{(8129 - 2673)^2 + (8553 - 2701)^2 + (8871 - 2736)^2} = 10082,26$$

$$d_{2,1} = \sqrt{(13667 - 1148)^2 + (13999 - 1116)^2 + (13757 - 1218)^2} = 21907,16$$

$$d_{2,2} = \sqrt{(1148 - 1148)^2 + (1116 - 1116)^2 + (1218 - 1218)^2} = 0,00$$

$$d_{2,3} = \sqrt{(8129 - 1148)^2 + (8553 - 1116)^2 + (8871 - 1218)^2} = 12751,93$$

$$d_{3,1} = \sqrt{(13667 - 8637)^2 + (13999 - 9016)^2 + (13757 - 8607)^2} = 8755,21$$

$$d_{3,2} = \sqrt{(1148 - 8637)^2 + (1116 - 9016)^2 + (1218 - 8607)^2} = 13156,46$$

$$d_{3,3} = \sqrt{(8129 - 8637)^2 + (8553 - 9016)^2 + (8871 - 8607)^2} = 736,29$$

- Menentukan cluster terdekat untuk didapatkan jarak terdekat.
- Setelah didapatkan jarak terdekat data ke cluster, maka didapatkanlah cluster tiap data.

Tabel 3 (Cluster)

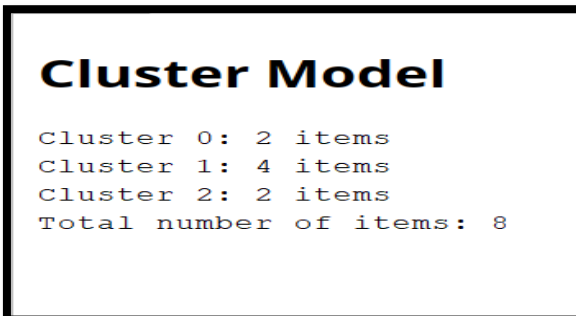
Data	Cluster
Data 1	1
Data 2	1
Data 3	2
...	...
Data 8	1

4.3. Hasil Clustering Menggunakan Rapid Miner

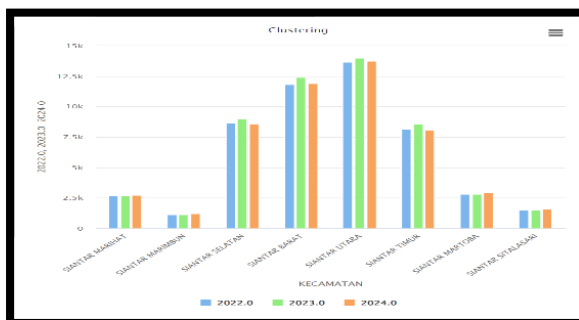
Pada tahapan pengujian algoritma K-

Means Clustering menggunakan tools Rapid Miner digunakan untuk membuktikan benar tidaknya perhitungan manual sebelumnya. Hasil perhitungan ini dapat dilihat dari gambar ini:

	id	KECAMATAN	2022.0	2023.0	2024.0	cluster
1	1	SIANTAR MARIHAT	2673	2701	2736	cluster_1
2	2	SIANTAR MARIMBUN	1148	1116	1218	cluster_1
3	3	SIANTAR SELATAN	8637	9016	8607	cluster_2
4	4	SIANTAR BARAT	11806	12382	11902	cluster_0
5	5	SIANTAR UTARA	13667	13999	13757	cluster_0
6	6	SIANTAR TIMUR	8129	8553	8071	cluster_2
7	7	SIANTAR MARTOBA	2794	2819	2923	cluster_1
8	8	SIANTAR SITALASARI	1511	1494	1573	cluster_1



Gambar 3 (Hasil Clustering Menggunakan Rapid Miner)



Gambar 4 (Hasil Clustering Menggunakan Rapid Miner)

Hasil pengelompokan menggunakan algoritma K-Means menunjukkan bahwa delapan kecamatan di Kota Pematangsiantar dapat dibagi ke dalam tiga kategori tingkat kepadatan penduduk, yaitu tinggi, sedang, dan rendah. Masing-masing kelompok mencerminkan karakteristik wilayah yang berbeda dan memiliki implikasi strategis bagi perencanaan kota.

C0 (Cluster Tinggi)

Terdiri atas Kecamatan Siantar Barat dan Siantar Utara. Kedua wilayah ini merupakan pusat aktivitas ekonomi, pemerintahan, dan perdagangan. Konsentrasi penduduk yang tinggi di kawasan ini

didorong oleh aksesibilitas yang baik terhadap layanan umum serta tingginya mobilitas harian. Namun, kondisi ini juga berpotensi menimbulkan tekanan terhadap ruang, kemacetan, serta beban pada infrastruktur dasar seperti air bersih, sanitasi, dan transportasi.

C2 (Cluster Sedang)

Diisi oleh Kecamatan Siantar Selatan dan Siantar Timur. Kedua kecamatan ini berada di zona transisi antara pusat kota dan wilayah pinggiran. Kepadatan penduduk di sini tergolong moderat dan menunjukkan pertumbuhan yang relatif stabil. Area ini memiliki potensi besar sebagai lokasi pengembangan permukiman baru. Oleh karena itu, kebijakan pembangunan perlu diarahkan agar wilayah ini tidak mengalami lonjakan kepadatan yang tidak terkendali.

C1 (Cluster Rendah)

Mencakup Kecamatan Siantar Marihat, Siantar Marimbun, Siantar Martoba, dan Siantar Sitalasari. Wilayah-wilayah ini umumnya berada di bagian luar kota dengan penggunaan lahan yang masih cukup longgar. Dengan kepadatan yang rendah, area ini memiliki peluang untuk dikembangkan menjadi kawasan hunian baru atau zona penyangga yang mampu meredam kepadatan dari pusat kota. Selain itu, potensi pengembangan fasilitas publik dan infrastruktur masih terbuka luas.

5. KESIMPULAN

Penelitian ini berhasil mengelompokkan delapan kecamatan di Kota Pematangsiantar ke dalam tiga kelompok kepadatan penduduk menggunakan metode K-Means Clustering. Pengelompokan tersebut memberikan gambaran yang jelas mengenai distribusi penduduk selama periode 2022 hingga 2024.

Berdasarkan hasil analisis, Kecamatan Siantar Barat dan Siantar Utara masuk ke dalam kategori kepadatan tinggi, yang menunjukkan konsentrasi aktivitas dan populasi di pusat kota. Sementara itu, Siantar Selatan dan Siantar Timur berada dalam kategori sedang, mencerminkan kawasan peralihan yang potensial untuk pengembangan terkontrol. Empat kecamatan lainnya yaitu Siantar Marihat, Siantar Marimbun, Siantar Martoba, dan Siantar Sitalasari tergolong sebagai wilayah dengan kepadatan rendah, yang berpotensi menjadi zona pengembangan permukiman baru atau kawasan penyangga.

Metode K-Means terbukti efektif dalam memetakan pola distribusi penduduk dan memberikan informasi yang berguna untuk perencanaan wilayah berbasis data. Dengan adanya pengelompokan ini, pemerintah daerah dapat lebih mudah mengidentifikasi kebutuhan infrastruktur, merancang tata ruang yang seimbang, serta mengatur distribusi layanan publik secara lebih efisien dan berkelanjutan.

6. DAFTAR PUSTAKA

[1] Badan Pusat Statistik Kota Pematangsiantar, *Data Kepadatan Penduduk Kota Pematangsiantar (per km²)*. [Online]. Available: <https://siantarkota.bps.go.id/id>

[2] R. M. Sabiq, "Pengaruh Kepadatan Penduduk Terhadap Tindakan Kriminal," *Jurnal Kolaborasi Resolusi Konflik*, vol. 3, no. 2, pp. 161–167, 2022. ISSN: 2655-8823 (p), 2656-1786 (e).

[3] C. Y. E. Dita and M. Legowo, "Analisis Kepadatan Penduduk yang Berpengaruh terhadap Kemiskinan dan Degradasi Lingkungan," *Prosiding Seminar Nasional Ilmu-Ilmu Sosial (SNIS)*, Universitas Negeri Surabaya, vol. 1, 2022.

[4] F. L. Yanuar, A. Ahmad, and G. A. Tjahjono, "Pengaruh Kepadatan Penduduk Terhadap Kualitas

Hidup Masyarakat di Kecamatan Pasar Kliwon Kota Surakarta Tahun 2021," *Geadidaktika: Jurnal Pendidikan Geografi*, vol. 4, no. 1, 2024.

[5] Soekartawi, *Ekonomi Sumberdaya dan Lingkungan*, Jakarta: LP3ES, 2004.

[6] Santoso, *Demografi dan Kependudukan*, Yogyakarta: Pustaka Pelajar, 2013.

[7] Kartodirdjo, *Penduduk dan Pembangunan*, Jakarta: Gramedia, 1992.

[8] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Yana Sport Jati," *Jurnal Mahasiswa Teknik Informatika*, vol. 6, no. 2, Sep. 2022.

[9] A. Pangestu and T. Ridwan, "Penerapan Data Mining Menggunakan Algoritma K-Means Pengelompokan Pelanggan Berdasarkan Kubikasi Air Terjual Menggunakan WEKA," *Jurnal Mahasiswa Teknik Informatika*, vol. 11, no. 3, May 2021, pp. 67–71.

[10] S. K. Arum, R. Astuti, and F. M. Basysyar, "Penerapan Algoritma K-Means pada Dataset Pengangguran Terbuka Berdasarkan Pendidikan di Provinsi Jawa Barat," *JATI: Jurnal*

Mahasiswa Teknik Informatika, vol. 8, no. 2, Apr. 2024.

[11] R. B. Prasetyo, Y. A. Pranoto, and R. P. Prasetya, "Implementasi Data Mining Menggunakan Algoritma K-Means Clustering pada Penyakit Pasien Rawat Jalan di Klinik Dr. Atirah Desa Sioyong, Sulteng," *JATI: Jurnal Mahasiswa Teknik Informatika*, vol. 7, no. 4, Aug. 2023.

[12] J. F. Butarbutar, N. B. Nugroho, and W. R. Maya, "Implementasi Data Mining untuk Mengelompokkan Daerah Rawan Tindakan Kriminal di Kota Medan Menggunakan Metode K-Means," *Jurnal CyberTech*, vol. 4, no. 3, Mar. 2021.

[13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, no. 3, pp. 37–54, 1996.

[14] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 3rd ed. Amsterdam: Elsevier, 2012.

[15] E. Turban, J. E. Aronson, and T. P. Liang, *Decision Support Systems and Intelligent Systems*, 7th ed. New Jersey: Pearson Prentice Hall, 2007.

[16] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA: Morgan Kaufmann, 2011.