

Penerapan Data Mining K-Means untuk Penentuan Prioritas Perbaikan Jalan di Kota Pematangsiantar

Fahrozi Aldinata¹, Cindy Aisyah Putri², Aditya Rizky Pratomo³

^{1,2,3}Stikom Tunas Bangsa Pematangsiantar

e-mail: ¹fahrozialdinata2@gmail.com, ²c.aisyahputri@gmail.com, ³adityarizp@gmail.com

Abstrak

Pembangunan infrastruktur jalan merupakan aspek penting dalam mendukung mobilitas masyarakat dan pertumbuhan ekonomi daerah. Namun, keterbatasan anggaran seringkali menjadi kendala pemerintah dalam menentukan prioritas perbaikan. Penelitian ini bertujuan menentukan prioritas perbaikan jalan di Kota Pematangsiantar dengan menerapkan algoritma K-Means Clustering. Variabel yang digunakan mencakup kondisi jalan (baik, sedang, rusak ringan, rusak berat), jumlah penduduk, dan kepadatan penduduk di tujuh kecamatan. Data diperoleh dari Dinas PUPR dan BPS Kota Pematangsiantar tahun 2023. Evaluasi dengan Elbow Method, Silhouette Score (0,3722), dan Davies-Bouldin Index (0, 6585) menunjukkan jumlah cluster optimal adalah tiga ($k=3$). Hasil clustering menghasilkan tiga kategori prioritas: (1) prioritas tinggi—Kecamatan Siantar Utara dan Siantar Barat, (2) prioritas sedang—Kecamatan Siantar Timur dan Siantar Selatan, serta (3) prioritas rendah—Kecamatan Martoba, Sitalasari, Marihat, dan Marimbun. Secara metodologis, penelitian ini memberikan kontribusi dalam bidang data mining terapan dengan mengintegrasikan variabel teknis dan sosial-demografis ke dalam sebuah model clustering yang komprehensif untuk perencanaan infrastruktur. Penelitian ini memberikan rekomendasi berbasis data yang efektif dan efisien bagi pemerintah daerah dalam mengalokasikan anggaran perbaikan jalan.

Kata kunci: K-Means Clustering, Data Mining, Kondisi Jalan, Kepadatan Penduduk, Pematangsiantar.

Abstract

Road infrastructure development is a vital aspect of supporting community mobility and regional economic growth. However, Budget limitations are often an obstacle for the government in determining improvement priorities. This study aims to determine road repair priorities in Pematangsiantar City by applying the K-Means Clustering algorithm. The variables used include road conditions (good, moderate, lightly damaged, severely damaged), population size, and population density across seven districts. Data were obtained from the Department of Public Works and Spatial Planning (PUPR) and the Central Bureau of Statistics (BPS) of Pematangsiantar City in 2023. Evaluation using the Elbow Method, Silhouette Score (0,3722), and Davies-Bouldin Index (0. 6585) indicated that the optimal number of clusters was three ($k=3$). The clustering results yielded three priority categories: (1) high priority—Siantar Utara and Siantar Barat Districts, (2) medium priority—Siantar Timur and Siantar Selatan Districts, and (3) low priority—Martoba, Sitalasari, Marihat, and Marimbun Districts. Methodologically, this research contributes to the field of applied data mining by integrating technical and socio-demographic variables into a comprehensive clustering model for infrastructure planning. This study provides effective and efficient data-driven recommendations for local governments in allocating road repair budgets..

Keywords: K-Means Clustering, Data Mining, Road Condition, Population Density, Pematangsiantar.

1. Pendahuluan

Pembangunan infrastruktur jalan merupakan faktor penting dalam meningkatkan aksesibilitas, mobilitas masyarakat, serta pertumbuhan ekonomi daerah. Berdasarkan data dari Dinas Pekerjaan Umum dan Perumahan Rakyat (PUPR) Kota Pematangsiantar tahun 2023, dari total panjang jalan sekira 611,608 kilometer, sekitar 23,81% di antaranya berada dalam kondisi rusak ringan hingga rusak berat. Permasalahan ini tersebar di seluruh kecamatan dengan tingkat keparahan dan dampak yang bervariasi. Sebagai contoh, Kecamatan Siantar Selatan memiliki persentase jalan rusak berat tertinggi (15%), sementara Kecamatan Siantar Utara menghadapi tantangan akibat kombinasi kerusakan jalan yang signifikan (30% rusak ringan dan berat) dengan kepadatan penduduk yang sangat tinggi mencapai 13.780,82 jiwa/km². Kondisi objektif inilah yang melatarbelakangi perlunya sebuah sistem pendukung keputusan yang dapat memprioritaskan alokasi anggaran perbaikan jalan secara lebih adil dan efektif. Kondisi jalan yang rusak dapat berdampak pada kelancaran transportasi, distribusi barang, serta keselamatan pengguna jalan [1], [2], [3], [4], [5]. Oleh karena itu, analisis mengenai kondisi jalan perlu dilakukan agar pemerintah daerah dapat menentukan prioritas perbaikan jalan secara tepat dan efisien [6], [7], [8].

Permasalahan yang muncul adalah keterbatasan anggaran perbaikan jalan sehingga pemerintah harus menetapkan kecamatan mana yang lebih diprioritaskan [9], [10], [11]. Namun, jika hanya melihat persentase jalan rusak, maka pengambilan keputusan bisa kurang tepat karena tidak mempertimbangkan jumlah penduduk maupun kepadatan wilayah. Oleh karena itu, penelitian ini akan menggunakan pendekatan data mining dengan algoritma K-Means Clustering untuk mengelompokkan kecamatan berdasarkan kondisi jalan, jumlah penduduk, dan kepadatan penduduk.

Beberapa penelitian sebelumnya telah menerapkan algoritma K-Means Clustering untuk permasalahan infrastruktur. Penelitian oleh Annisa Wulandari, Irmayansyah dkk (2025) [12] dan Siti Asmiatun, Nur Wakhidah (2018) [13] mengelompokkan kondisi permukaan jalan berdasarkan tingkat kerusakan untuk mempermudah penentuan perbaikan. Penelitian lain oleh Arya Bagaskara Isag, Rony Heri Irawan (2021) [14] menggunakan K-Means dalam menentukan prioritas perbaikan jembatan dan saluran air, sedangkan Dede Kurniadi, Yoga Handoko Agustin dkk (2023) [15] menerapkan K-Means untuk pengelompokan pembangunan jalan di Kabupaten Garut dengan evaluasi Davies-Bouldin Index. Namun, terdapat kesenjangan (research gap) dalam penelitian-penelitian tersebut, dimana analisis prioritas perbaikan jalan masih terfokus pada aspek teknis-fisik infrastruktur tanpa mempertimbangkan dampak sosial-demografis seperti jumlah dan kepadatan penduduk yang terdampak. Oleh karena itu, penelitian ini bertujuan mengisi kesenjangan tersebut dengan mengintegrasikan variabel kondisi jalan (baik, sedang, rusak ringan, rusak berat) dan variabel kependudukan (jumlah penduduk, kepadatan) dalam sebuah model K-Means Clustering yang komprehensif untuk menghasilkan rekomendasi prioritas perbaikan jalan yang lebih aplikatif bagi pengambil kebijakan.

Tujuan penelitian ini adalah untuk menerapkan algoritma K-Means Clustering dalam mengelompokkan kecamatan di Kota Pematangsiantar berdasarkan kondisi jalan, jumlah penduduk, dan kepadatan penduduk. Dengan demikian, penelitian ini dapat memberikan rekomendasi prioritas perbaikan jalan yang lebih akurat dan berbasis data sehingga membantu pemerintah dalam pengambilan keputusan pembangunan infrastruktur.

2. Metode Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan data mining. Teknik yang digunakan adalah unsupervised learning melalui algoritma K-Means Clustering. Algoritma K-Means dipilih setelah mempertimbangkan beberapa alternatif clustering seperti DBSCAN, Hierarchical, dan Fuzzy C-Means. Pertimbangan pemilihan K-Means didasarkan pada:

1. efisiensi komputasi untuk dataset berukuran kecil-menengah (7 kecamatan),
2. kemampuan menghasilkan cluster spherical yang sesuai untuk data numerik kontinu seperti persentase kondisi jalan dan data kependudukan,
3. kesederhanaan interpretasi hasil cluster bagi pengambil kebijakan, serta
4. keberhasilan implementasi dalam penelitian infrastruktur sejenis.

Untuk memastikan setiap variabel memiliki bobot yang setara dalam proses clustering, dilakukan normalisasi data menggunakan Min-Max Scaling. Normalisasi ini mengubah semua variabel (Baik, Sedang, Rusak Ringan, Rusak Berat, Jumlah Penduduk, Kepadatan) ke dalam skala 0-1, sehingga variabel dengan nilai besar seperti jumlah penduduk (17.588-52.705) tidak mendominasi variabel dengan skala kecil seperti persentase jalan (0-100%).

Pemilihan K-Means didukung oleh karakteristik dataset yang memiliki cluster spherical dan ukuran sampel yang terbatas, di mana K-Means menunjukkan performa komputasi yang lebih efisien dibanding algoritma clustering lain seperti DBSCAN yang memerlukan parameter epsilon yang kompleks untuk dataset kecil.

2.1. Data Mining

Data mining adalah proses menemukan pola, pengetahuan, atau informasi berharga dari kumpulan data yang besar dengan menggunakan teknik analisis dan algoritma tertentu. Menurut Han & Kamber (2012), data mining merupakan bagian dari Knowledge Discovery in Database (KDD) yang berfokus pada identifikasi pola tersembunyi yang sebelumnya tidak diketahui. Secara umum, tahapan dalam data mining meliputi:

1. Data Cleaning – membersihkan data dari kesalahan atau nilai yang hilang.
2. Data Integration – menggabungkan data dari berbagai sumber.
3. Data Selection – memilih variabel yang relevan untuk dianalisis.
4. Data Transformation – mengubah data ke dalam format yang sesuai, misalnya normalisasi.
5. Data Mining – penerapan algoritma tertentu (misalnya clustering, classification, association).
6. Pattern Evaluation – mengevaluasi pola atau model yang terbentuk.
7. Knowledge Presentation – menyajikan hasil analisis dalam bentuk tabel, grafik, atau visualisasi.

Dalam penelitian ini, tahap utama data mining adalah clustering, di mana data kondisi jalan dan kependudukan Kota Pematangsiantar dikelompokkan menggunakan algoritma K-Means untuk menemukan cluster prioritas perbaikan jalan.

2.2. Algoritma

1. K-means

Algoritma K-Means Clustering adalah salah satu metode dalam unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan tingkat kemiripan antar data. Prinsip kerja K-Means adalah membagi data menjadi k cluster, di mana setiap cluster memiliki sebuah pusat (centroid). Data yang memiliki karakteristik serupa akan ditempatkan pada cluster yang sama.

Tujuan utama dari K-Means adalah untuk meminimalkan jarak antara data dengan centroid cluster-nya, sehingga data dalam satu cluster memiliki tingkat homogenitas yang tinggi, sementara antar cluster memiliki perbedaan yang jelas.

2. Tahapan Algoritma K-Means

Langkah-langkah algoritma K-Means adalah sebagai berikut:

- a. Tentukan jumlah cluster (k) yang akan dibentuk.
- b. Pilih centroid awal secara acak dari data.
- c. Hitung jarak setiap data ke masing-masing centroid menggunakan rumus Euclidean Distance:

$$d(x, u_i) = \sqrt{\sum_{j=1}^n (x_j - \mu_{ij})^2} \quad (1)$$

Dimana:

- $d(x, u_i)$ = jarak data x ke centroid μ_i
 - x_j = nilai variabel ke- j dari data x
 - μ_{ij} = nilai centroid cluster ke- i pada variabel ke- j .
- d. Alokasikan setiap data ke cluster dengan centroid terdekat.
- e. Hitung ulang centroid baru sebagai rata-rata dari seluruh anggota cluster:

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (2)$$

di mana C_i adalah himpunan data dalam cluster ke- i .

- f. Ulangi langkah 3–5 hingga posisi centroid tidak berubah (konvergen).

2.3. Fungsi Objektif K-Means

K-Means berusaha meminimalkan fungsi objektif berikut:

$$J = \sum_{i=1}^k \sum_{x \in C_i} ||x - u_i||^2 \quad (3)$$

Keterangan:

- J = fungsi objektif (total kesalahan kuadrat).
- k = jumlah cluster.
- x = data.
- μ_i = centroid cluster ke- i .
- C_i = cluster ke- i .

Semakin kecil nilai J , semakin baik kualitas cluster yang terbentuk, karena berarti data dalam cluster semakin dekat dengan centroid-nya.

2.4. dataset

Data dalam penelitian ini diperoleh dari sumber sekunder resmi. Pertama, dari Dinas Pekerjaan Umum dan Tata Ruang Kota Pematangsiantar (2023), berupa data kondisi jalan dengan kategori persentase jalan baik, sedang, rusak ringan, dan rusak berat. Kedua, dari Badan Pusat Statistik (BPS) Kota Pematangsiantar (2023), berupa data kependudukan yang meliputi luas wilayah, jumlah kelurahan, jumlah penduduk, dan kepadatan penduduk. Jenis data yang digunakan adalah data kuantitatif, karena semua variabel dinyatakan dalam bentuk angka yang dapat diolah secara statistik maupun algoritmik.

Tabel 1. Dataset jalan

Kecamatan	Baik (%)	Sedang (%)	Rusak Ringan (%)	Rusak Berat (%)
Siantar Marihat	60	20	15	5
Siantar Marimbun	65	18	10	7
Siantar Selatan	40	25	20	15
Siantar Barat	55	15	20	10
Siantar Utara	50	20	20	10
Siantar Timur	45	25	20	10
Siantar Martoba	70	15	10	5

Tabel 2. Dataset Luas Wilayah, Jumlah Penduduk, dan Kepadatan

Kecamatan	Luas Wilayah (Km ²)	Jumlah Penduduk	Kepadatan per Km ²
Siantar Marihat	7.825	21,484	2,743.81
Siantar Marimbun	18.006	21,684	1,204.26
Siantar Selatan	2.020	17,588	8,706.93
Siantar Barat	3.205	38,249	11,934.17
Siantar Utara	3.650	50,300	13,780.82
Siantar Timur	4.520	37,041	8,194.91
Siantar Martoba	18.022	52,705	2,924.48

2.5. Variabel Penelitian

Variabel dalam penelitian ini dibagi menjadi dua. Pertama, variabel input (X) yang digunakan dalam proses clustering, yaitu: persentase jalan baik, persentase jalan sedang, persentase jalan rusak ringan, persentase jalan rusak berat, jumlah penduduk, dan kepadatan penduduk. Kedua, variabel output (Y) yang dihasilkan dari algoritma K-Means, yaitu pengelompokan kecamatan ke dalam cluster prioritas perbaikan jalan, misalnya cluster prioritas tinggi, sedang, dan rendah.

2.6. Perhitungan Indeks Prioritas Jalan

Untuk menentukan tingkat prioritas perbaikan jalan di setiap kecamatan, penelitian ini menggunakan sebuah indeks yang mempertimbangkan persentase kerusakan jalan serta jumlah penduduk yang terdampak. Rumus yang digunakan adalah sebagai berikut:

$$IP_j = (Rr_j + 2 \times Rb_j) \times P_j \quad (4)$$

dengan:

- IP_j = Indeks Prioritas Jalan pada kecamatan ke-j,
- Rr_j = Persentase jalan **rusak ringan** pada kecamatan ke-j,
- Rb_j = Persentase jalan **rusak berat** pada kecamatan ke-j,
- P_j = Jumlah penduduk pada kecamatan ke-j.

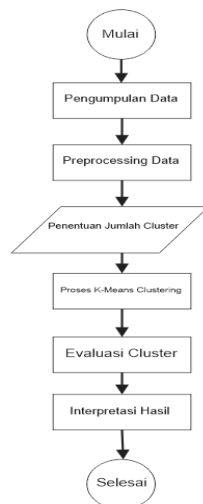
Bobot 2 diberikan pada persentase jalan rusak berat untuk menekankan bahwa kerusakan parah memiliki dampak yang lebih besar terhadap mobilitas dan aktivitas sosial-ekonomi masyarakat dibandingkan dengan kerusakan ringan.

Indeks prioritas ini kemudian digunakan baik sebagai variabel tambahan dalam proses clustering maupun sebagai dasar penyusunan ranking kecamatan berdasarkan tingkat urgensi perbaikan jalan.

2.7. Teknik Analisa Data

Analisis data dilakukan dalam beberapa tahap. Tahap pertama adalah preprocessing data, yang meliputi pembersihan data dan normalisasi menggunakan Min-Max Scaling agar semua variabel berada dalam skala yang sama (0–1). Tahap kedua adalah clustering dengan algoritma K-Means, dimulai dari penentuan jumlah cluster menggunakan metode Elbow dan Silhouette Score. kemudian algoritma dijalankan hingga terbentuk cluster yang stabil. Tahap ketiga adalah evaluasi hasil clustering, dengan menggunakan Silhouette Score untuk mengukur validitas cluster. Tahap terakhir adalah interpretasi hasil, yaitu menjelaskan karakteristik cluster, misalnya cluster prioritas tinggi ditandai dengan kondisi jalan rusak berat dan kepadatan penduduk tinggi.

2.8. Diagram Alur Penelitian



Gambar 1. Flowchart penelitian

Penelitian ini mengikuti alur kerja sistematis yang dimulai dengan pengumpulan data sekunder dari Dinas PUPR dan BPS Kota Pematangsiantar. Tahap berikutnya adalah preprocessing data yang meliputi pembersihan data dan normalisasi menggunakan Min-Max Scaling untuk menyamakan skala semua variabel. Setelah data siap, dilakukan penentuan jumlah cluster optimal menggunakan Elbow Method yang menghasilkan keputusan tiga cluster ($k=3$). Proses K-Means Clustering kemudian dijalankan secara iteratif hingga mencapai konvergensi, dimana setiap kecamatan telah dikelompokkan ke dalam cluster yang sesuai. Kualitas cluster yang terbentuk dievaluasi menggunakan tiga metrik validasi yaitu Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Score. Tahap akhir adalah interpretasi hasil clustering, dimana setiap cluster dianalisis karakteristiknya dan diberi label prioritas berdasarkan kombinasi variabel kondisi jalan dan kependudukan, sehingga menghasilkan rekomendasi prioritas perbaikan jalan yang berbasis data.

3. Hasil dan Pembahasan

Hasil

3.1. Data awal

Penelitian ini menggunakan data kondisi jalan dan kependudukan pada tujuh kecamatan di Kota Pematangsiantar. Variabel kondisi jalan meliputi persentase jalan baik, sedang, rusak ringan, dan rusak berat, sedangkan variabel kependudukan mencakup luas wilayah, jumlah penduduk, dan kepadatan penduduk.

Tabel 3. Dataset awal

Kecamatan	Baik (%)	Sedang (%)	Rusak Ringan (%)	Rusak Berat (%)	Luas Wilayah (Km ²)	Jumlah Penduduk	Kepadatan (/Km ²)
Siantar Marihat	60	20	15	5	7.825	21,484	2,743.81
Siantar	65	18	10	7	18.006	21,684	1,204.26
Marimbun							
Siantar Selatan	40	25	20	15	2.020	17,588	8,706.93
Siantar Barat	55	15	20	10	3.205	38,249	11,934.17
Siantar Utara	50	20	20	10	3.650	50,300	13,780.82
Siantar Timur	45	25	20	10	4.520	37,041	8,194.91
Siantar Martoba	70	15	10	5	18.022	52,705	2,924.48

Tabel 3. Dataset setelah Normalisasi min-max

Kecamatan	Baik (%)	Sedang (%)	Rusak Ringan (%)	Rusak Berat (%)	Luas Wilayah (Km ²)	Jumlah Penduduk	Kepadatan (/Km ²)
Siantar Marihat	0.6667	0.5	0.5	0	0.1109	0.1224	0.6667
Siantar	0.8333	0.3	0	0.2	0.1166	0	0.8333
Marimbun							
Siantar Selatan	0	1	1	1	0	0.5966	0
Siantar Barat	0.5	0	1	0.5	0.5883	0.8532	0.5
Siantar Utara	0.3333	0.5	1	0.5	0.9315	1	0.3333
Siantar Timur	0.1667	1	1	0.5	0.5539	0.5558	0.1667
Siantar Martoba	1	0	0	0	1	0.1368	1

3.2. Evaluasi Model

EVALUASI MODEL K-MEANS CLUSTERING

Silhouette Score : 0.3722
 Davies-Bouldin Index : 0.6585
 Calinski-Harabasz Score: 6.2302

Gambar 2. Evaluasi Model K-means

Evaluasi hasil clustering dilakukan menggunakan tiga metrik, yaitu:

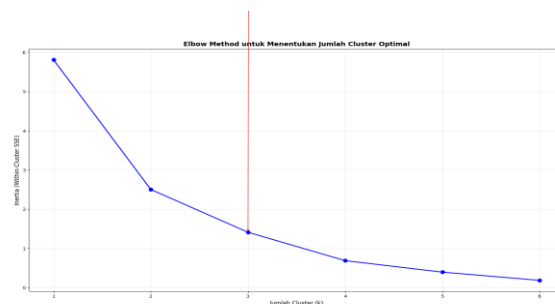
- Silhouette Score = 0,3722
- Davies-Bouldin Index (DBI) = 0,6585
- Calinski-Harabasz Score (CH) = 6.2302

Evaluasi kualitas clustering menggunakan tiga metrik validasi menunjukkan hasil yang konsisten dan valid. Silhouette Score sebesar 0,3722 mengindikasikan kualitas clustering yang moderat, yang masih dapat diterima mengingat ukuran dataset yang terbatas (7 kecamatan). Davies-Bouldin Index sebesar 0,6585 (nilai $< 1,0$) menunjukkan tingkat overlapping antar cluster yang rendah, sementara Calinski-Harabasz Score sebesar 6,2302 mendukung kerapatan data dalam cluster dan jarak antar cluster yang memadai. Ketiga metrik secara konsisten membuktikan bahwa model K-Means yang dibangun memiliki kualitas yang baik untuk analisis prioritas perbaikan jalan.

Refleksi terhadap kualitas hasil clustering menunjukkan bahwa meskipun nilai Silhouette Score 0,3722 berada dalam kategori 'acceptable', kombinasi dengan DBI 0,6585 ($< 1,0$) dan CH Score 6,2302 membuktikan model yang robust untuk dataset berukuran kecil. Keterbatasan jumlah data (7 kecamatan) memang mempengaruhi nilai metrik, namun secara konseptual hasil clustering tetap valid dan aplikatif

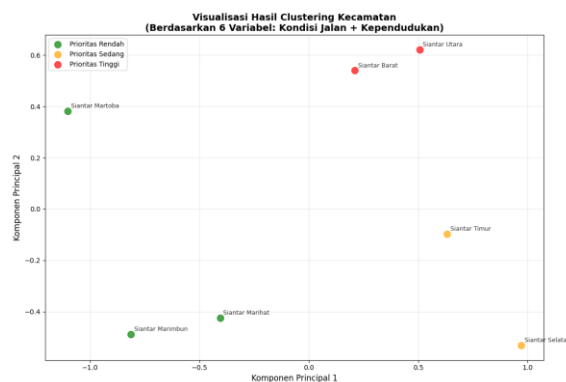
3.3. Penentuan Jumlah Cluster

Untuk menentukan jumlah cluster optimal, digunakan metode Elbow dan Silhouette Score



Gambar 3. Hasil Elbow Method untuk Penentuan Jumlah Cluster Optimal

Berdasarkan analisis Elbow Method pada Gambar 3, titik siku (elbow) yang jelas teridentifikasi pada $k=3$. Pola kurva menunjukkan penurunan inersia yang signifikan dari $k=1$ ke $k=3$, kemudian mulai melandai setelah $k=3$. Hal ini mengindikasikan bahwa penambahan cluster di atas tiga tidak memberikan peningkatan kualitas yang bermakna. Dengan pertimbangan optimalitas antara kompleksitas model dan kemampuan menjelaskan variasi data, penelitian ini menetapkan tiga cluster sebagai jumlah optimal untuk pengelompokan kecamatan di Kota Pematangsiantar.

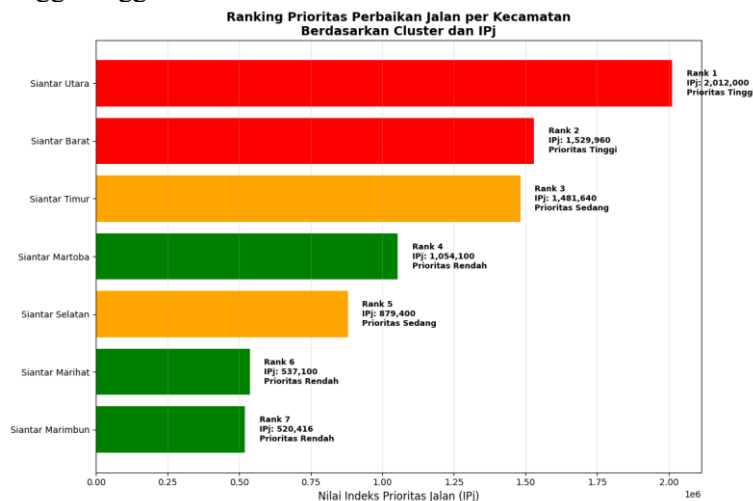


Gambar 4. visualisasi scatter plot

Berdasarkan hasil visualisasi clustering pada gambar, terlihat bahwa tujuh kecamatan di Kota Pematangsiantar terbagi menjadi tiga kelompok prioritas berdasarkan enam variabel yang mencerminkan kondisi jalan dan kependudukan. Kecamatan Siantar Martoba, Siantar Marihat, dan Siantar Marimbun tergolong dalam kategori prioritas rendah (ditandai warna hijau), menunjukkan bahwa wilayah tersebut memiliki kondisi infrastruktur jalan yang relatif baik serta kepadatan penduduk yang stabil. Kecamatan Siantar Timur dan Siantar Selatan termasuk dalam prioritas sedang (warna oranye), yang menandakan bahwa wilayah tersebut memiliki beberapa aspek yang perlu diperhatikan, namun belum memerlukan penanganan mendesak. Sementara itu, Kecamatan Siantar Barat dan Siantar Utara masuk dalam prioritas tinggi (warna merah), yang berarti kedua wilayah ini memiliki kondisi jalan dan kependudukan yang paling memerlukan perhatian dan perbaikan segera. Sebaran titik pada scatter plot memperlihatkan bahwa kecamatan yang berada berdekatan memiliki karakteristik serupa, sedangkan yang berjauhan menunjukkan perbedaan signifikan, sehingga hasil ini dapat menjadi dasar dalam menentukan strategi pembangunan dan alokasi sumber daya secara lebih tepat sasaran.

3.4. Ranking Prioritas Kecamatan

Berdasarkan perhitungan IPJ yaitu Indeks Prioritas jalan menghasilkan urutan kecamatan dari prioritas tertinggi hingga terendah.

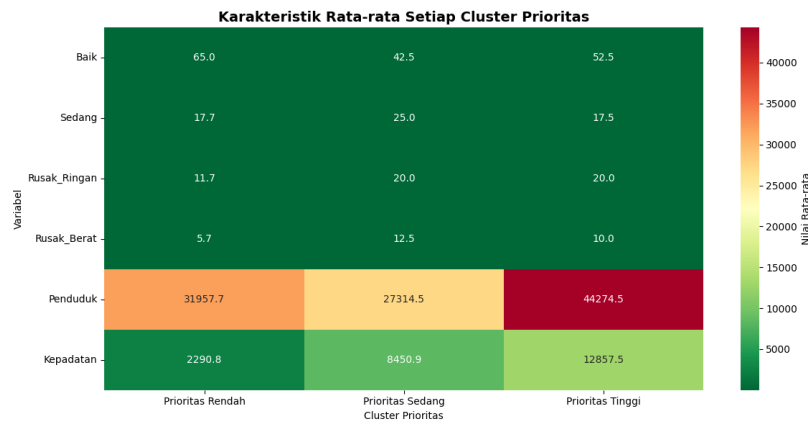


Gambar 5. Ranging prioritas

Berdasarkan chart IPJ diperoleh urutan prioritas dari tertinggi hingga terendah. Hasil menunjukkan Siantar Utara dan Siantar Barat menempati peringkat teratas dengan nilai IPJ tertinggi. Terdapat fenomena menarik dimana Siantar Martoba berada di peringkat 4 namun termasuk Prioritas Rendah, sementara Siantar Selatan peringkat 5 namun Prioritas Sedang. Hal ini disebabkan perbedaan dasar analisis: clustering mempertimbangkan 6 variabel secara simultan, sedangkan ranking hanya berdasarkan formula IPJ.

Fenomena peringkat vs cluster ini justru menguatkan keunggulan pendekatan multivariat. Meskipun Siantar Martoba memiliki IPJ tinggi karena jumlah penduduk besar, karakteristik wilayah yang luas (18,022 km²) dan kondisi jalan terbaik (70% baik) mengurangi urgensi perbaikan. Sebaliknya, Siantar Selatan dengan IPJ lebih rendah namun kombinasi kerusakan parah (15% rusak berat) dan kepadatan tinggi (8.707/km²) memerlukan perhatian lebih segera.

3.5. Analisis Karakteristik Cluster



Gambar 6. Karakteristik Rata rata cluster

Analisis karakteristik cluster mengidentifikasi variabel dominan yang membentuk setiap cluster. Pada Cluster Prioritas Tinggi, variabel kepadatan penduduk (12.857/km²) dan jumlah penduduk (44.274) menjadi faktor penentu utama. Cluster Prioritas Sedang didominasi oleh variabel persentase jalan rusak berat (12,5%), sementara Cluster Prioritas Rendah dikendalikan oleh variabel persentase jalan baik (65%) dan kepadatan rendah (2.291/km²). Pola dominasi ini menunjukkan bahwa model berhasil menangkap kompleksitas faktor yang mempengaruhi prioritas perbaikan jalan.

```
=====
INTERPRETASI HASIL CLUSTERING
=====

Prioritas Rendah:
- Rata-rata Jalan Baik: 65.0%
- Rata-rata Jalan Rusak Berat: 5.7%
- Rata-rata Jumlah Penduduk: 31,958
- Rata-rata Kepadatan: 2,291 per Km²
- Rata-rata IPj: 703,872

Prioritas Sedang:
- Rata-rata Jalan Baik: 42.5%
- Rata-rata Jalan Rusak Berat: 12.5%
- Rata-rata Jumlah Penduduk: 27,314
- Rata-rata Kepadatan: 8,451 per Km²
- Rata-rata IPj: 1,180,520

Prioritas Tinggi:
- Rata-rata Jalan Baik: 52.5%
- Rata-rata Jalan Rusak Berat: 10.0%
- Rata-rata Jumlah Penduduk: 44,274
- Rata-rata Kepadatan: 12,857 per Km²
- Rata-rata IPj: 1,770,980
```

Gambar 7. Interpretasi hasil clustering

Pola yang teridentifikasi mengindikasikan bahwa tekanan demografis menjadi faktor penentu utama dalam pembentukan Cluster Prioritas Tinggi, di mana tingginya kepadatan dan jumlah penduduk memperbesar dampak sosial-ekonomi dari kerusakan jalan. Sementara itu, Cluster Prioritas Sedang merefleksikan urgensi perbaikan berdasarkan tingkat kerusakan infrastruktur yang parah, meskipun dengan dampak terhadap populasi yang lebih terbatas. Adapun Cluster Prioritas Rendah menunjukkan kombinasi kondisi infrastruktur yang baik dengan tekanan demografis minimal, yang menjelaskan alokasi prioritas yang lebih rendah. Interpretasi ini membuktikan efektivitas model dalam menangkap kompleksitas faktor teknis dan sosial-demografis secara simultan.

Pembahasan

Hasil penelitian ini secara tegas membuktikan efektivitas integrasi variabel kondisi jalan dan kependudukan dalam penentuan prioritas perbaikan jalan. Tiga cluster yang terbentuk tidak hanya *statistically valid* dengan Silhouette Score 0,3722 yang terlihat pada Gambar 2, tetapi juga secara konseptual masuk akal untuk aplikasi kebijakan publik.

Analisis sebab-akibat mengungkap bahwa pola cluster terbentuk akibat interaksi kausal antara variabel kerusakan jalan dan tekanan demografis. Cluster Prioritas Tinggi terbentuk

sebagai akibat kombinasi kepadatan penduduk tertinggi (12.857/km²) dengan tingkat kerusakan signifikan (10% rusak berat), yang menimbulkan dampak berantai pada kemacetan, biaya logistik, dan penurunan produktivitas ekonomi. Cluster Prioritas Sedang muncul akibat dominasi kerusakan jalan terparah (12,5% rusak berat) meskipun dengan populasi lebih tersebar, sementara Cluster Prioritas Rendah merupakan konsekuensi dari kondisi infrastruktur yang baik (65% jalan baik) dengan tekanan demografis minimal.

Dominasi variabel dalam pembentukan cluster teridentifikasi jelas melalui analisis karakteristik. Variabel kepadatan penduduk menjadi faktor penentu utama dalam Cluster Prioritas Tinggi, mencerminkan bahwa dampak sosial-ekonomi lebih krusial daripada tingkat kerusakan absolut. Sebaliknya, variabel persentase jalan rusak berat mendominasi Cluster Prioritas Sedang, menunjukkan bahwa kerusakan parah tetap memerlukan perhatian meskipun dampaknya terbatas. Pada Cluster Prioritas Rendah, variabel kondisi jalan baik menjadi faktor pembeda utama.

Temuan kunci penelitian menunjukkan konsistensi hasil clustering dengan realitas sosio-demografi Kota Pematangsiantar. Cluster Prioritas Tinggi yang terdiri dari Siantar Utara dan Siantar Barat memang merupakan wilayah dengan aktivitas ekonomi dan kepadatan tertinggi, sehingga dampak kerusakan jalan terhadap produktivitas masyarakat paling signifikan.

Dalam perbandingan dengan penelitian sebelumnya, hasil penelitian ini memberikan kontribusi berbeda dengan penelitian [12][13] yang hanya berfokus pada kondisi fisik jalan. Jika penelitian [15] menggunakan DBI tunggal untuk evaluasi, penelitian ini mengombinasikan tiga metrik validasi (Silhouette, DBI, Calinski-Harabasz) yang memberikan penilaian lebih komprehensif. Pendekatan integratif ini mengatasi keterbatasan penelitian sebelumnya dengan mempertimbangkan aspek teknis dan sosial-demografis secara simultan.

Refleksi kualitas hasil clustering menunjukkan bahwa meskipun nilai Silhouette Score 0,3722 berada dalam kategori 'acceptable', kombinasi dengan DBI 0,6585 (< 1,0) dan CH Score 6,2302 membuktikan model yang robust untuk dataset berukuran kecil. Keterbatasan jumlah data (7 kecamatan) memang mempengaruhi nilai metrik, namun secara konseptual hasil clustering tetap valid dan aplikatif.

Implikasi praktis dari penelitian ini adalah kemampuan memberikan rekomendasi yang terukur dan objektif bagi pemerintah daerah. Berdasarkan analisis cluster dan nilai IP_j, alokasi anggaran dapat diprioritaskan secara proporsional menggunakan formula:

$$\text{Proporsi} = \frac{\sum IP_j \text{ Cluster}}{\sum IP_j \text{ Total}} \times 100\% \quad (5)$$

Hasil Perhitungan:

- Prioritas Tinggi: 44,2% → 45-50% (Siantar Utara & Barat)
- Prioritas Sedang: 29,5% → 30-35% (Siantar Timur & Selatan)
- Prioritas Rendah: 26,3% → 15-20% (Martoba, Marihat, Marimbun)

Rekomendasi akhir menyesuaikan proporsi aktual dengan pertimbangan strategis untuk memaksimalkan dampak sosial-ekonomi pada wilayah prioritas tertinggi.

Kelebihan metodologi ini terletak pada kemampuannya mengeliminasi bias subjektif dalam penentuan prioritas, sekaligus memberikan justifikasi yang kuat secara statistik untuk setiap keputusan alokasi sumber daya. Penelitian ini menyimpulkan bahwa pendekatan *data mining* dengan K-Means Clustering merupakan solusi efektif untuk optimasi alokasi anggaran perbaikan jalan di daerah dengan keterbatasan sumber daya.

Penelitian ini tidak hanya menyajikan metode teknis clustering, tetapi juga memberikan *framework* analitis yang dapat diadopsi oleh pemerintah daerah untuk pengambilan keputusan berbasis data. Integrasi multidisiplin antara teknik sipil, *data mining*, dan kebijakan publik menjadi nilai tambah yang membedakan penelitian ini dengan pendekatan konvensional. Keterbatasan penelitian terletak pada jumlah sampel yang relatif kecil, namun hal ini dikompensasi oleh kedalaman analisis dan validasi metodologis yang komprehensif. Untuk

penelitian lanjutan, dapat dipertimbangkan penambahan variabel seperti volume lalu lintas dan dampak ekonomi regional untuk penyempurnaan model.

4. Kesimpulan

Berdasarkan penelitian mengenai penentuan prioritas perbaikan jalan menggunakan algoritma K-Means Clustering di Kota Pematangsiantar, diperoleh kesimpulan sebagai berikut:

Penelitian ini telah berhasil menjawab permasalahan keterbatasan anggaran perbaikan jalan dengan menyediakan pengelompokan prioritas berbasis data yang objektif dan terukur. Hasil penelitian membuktikan bahwa integrasi variabel kondisi jalan dan kependudukan mampu menghasilkan rekomendasi prioritas yang komprehensif bagi pemerintah daerah.

Secara spesifik, temuan penelitian menunjukkan:

- Algoritma K-Means dengan $k=3$ berhasil mengelompokkan tujuh kecamatan ke dalam tiga kategori prioritas yang jelas. Evaluasi model dengan Silhouette Score 0,372 mengindikasikan kualitas clustering yang acceptable, didukung oleh Davies-Bouldin Index 0,658 yang menunjukkan pemisahan cluster yang baik, dan Calinski-Harabasz Score 6,230 yang mendukung struktur cluster yang padat.
- Prioritas Tinggi terdiri dari Kecamatan Siantar Utara dan Siantar Barat, dengan karakteristik kepadatan penduduk tertinggi ($12.857/\text{km}^2$) dan jumlah penduduk terbesar (44.274). Kedua kecamatan ini memerlukan alokasi anggaran utama untuk perbaikan segera.
- Prioritas Sedang meliputi Kecamatan Siantar Timur dan Siantar Selatan, yang ditandai dengan tingkat kerusakan jalan tertinggi (12,5% rusak berat) meskipun dengan populasi lebih terbatas. Wilayah ini memerlukan penanganan bertahap.
- Prioritas Rendah terdiri dari Kecamatan Siantar Martoba, Siantar Marihat, dan Siantar Marimbun, dengan kondisi jalan terbaik (65% baik) dan kepadatan terendah ($2.291/\text{km}^2$). Wilayah ini dapat difokuskan pada pemeliharaan preventif.

Interpretasi kualitas model menunjukkan bahwa meskipun nilai Silhouette Score 0,372 tergolong sedang, kombinasi dengan metrik evaluasi lainnya membuktikan validitas model untuk konteks penelitian ini. Hasil ini masih dapat ditingkatkan pada penelitian berikutnya dengan perluasan dataset dan penambahan variabel.

Kontribusi praktis penelitian ini terletak pada penyediaan sistem pendukung keputusan yang transparan untuk alokasi anggaran perbaikan jalan. Berdasarkan analisis kuantitatif terhadap nilai IPj dan karakteristik cluster, rekomendasi alokasi anggaran dapat ditetapkan secara proporsional sebagai berikut:

- Prioritas Tinggi : 44,2% (berdasarkan proporsi IPj) 45-50%
- Prioritas Sedang : 29,5% (berdasarkan proporsi IPj) 30-35%
- Prioritas Rendah : 26,3% (berdasarkan proporsi IPj) → 15-20%

Penyesuaian ini memastikan optimalisasi dampak sosial-ekonomi dengan fokus pada wilayah yang paling membutuhkan.

Kontribusi praktis penelitian ini terletak pada penyediaan sistem pendukung keputusan yang transparan untuk alokasi anggaran perbaikan jalan. Berdasarkan analisis cluster dan perhitungan IPj, rekomendasi alokasi anggaran dapat ditetapkan sebesar 45-50% untuk Prioritas Tinggi, 30-35% untuk Prioritas Sedang, dan 15-20% untuk Prioritas Rendah.

Keterbatasan penelitian terutama pada jumlah sampel yang relatif kecil (7 kecamatan) dan variabel yang terbatas pada kondisi jalan dan kependudukan. Untuk penelitian lanjutan, disarankan untuk menambahkan variabel seperti volume lalu lintas, dampak ekonomi, dan menggunakan algoritma clustering alternatif untuk perbandingan hasil.

Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan data mining dengan K-Means Clustering merupakan solusi efektif untuk optimasi alokasi anggaran infrastruktur jalan di daerah dengan keterbatasan sumber daya.

Daftar Pustaka

- [1] K. K. ANNISA, NASKAH, H. H. HASANAH, AND A. H. SAPUTRA, "DAMPAK JALAN RUSAK TERHADAP PERTUMBUHAN EKONOMI DAN SOSIAL DI KABUPATEN INDRAGIRI HULU," *J. AGAMA, SOS. DAN BUDAYA*, VOL. 3, NO. 5, PP. 2352–2357, 2024, [ONLINE]. AVAILABLE: [HTTPS://PUBLISHERQU.COM/INDEX.PHP/AL-FURQAN](https://publisherqu.com/index.php/al-furqan)
- [2] L. LYRRA AND E. KURNIATI, "ANALISIS INFRASTRUKTUR JALAN DAN INDEKS PEMBANGUNAN MANUSIA (IPM) TERHADAP PERTUMBUHAN EKONOMI DI KABUPATEN WAY KANAN," *J. SOC. SCI. MULTIDISCIP. ANAL.*, VOL. 2, NO. 2, PP. 41–60, 2025, DOI: 10.59827/JOSSAMA.V2I2.54.
- [3] R. S. NUHUN, L. WELENDU, M. F. ALMALIKI, ISMAYANA, HERIANTO, AND HARMIANTO, "EVALUASI PANJANG DAN KONDISI JALAN DI KABUPATEN / KOTA PROVINSI SULAWESI TENGGARA : IMPLIKASI TERHADAP PENGELOLAAN INFRASTRUKTUR," *J. SOC. SCI. RES. VOL.*, VOL. 4, PP. 8824–8835, 2024.
- [4] M. R. REYHAN KHALIDSYAH, "PERAN KEPALA DESA DALAM PENGELOLAAN INFRASTRUKTUR DESA UNTUK MENINGKATKAN AKSESIBILITAS JALAN DI DESA SUNGAI PINANG KABUPATEN BANYUASIN," *THEJOURNALISH SOC. GOV.*, VOL. 6, NO. 2, PP. 183–194, 2025, DOI: 10.55314/TSG.V6I2.900.
- [5] N. H. Z. WAFIQ AZIZAH HASIBUAN, FATIMAH DEPI SUSANTI, NABILA ZASKIA, "STUDI KASUS DAMPAK JALAN RUSAK TERHADAP MOBILITAS DAN AKTIVITAS MAHASISWA DI MUSTAMINDO, KECAMATAN TAMPAN, KABUPATEN KAMPAR, RIAU," *SUSTAIN.*, VOL. 11, NO. 1, PP. 1–14, 2025, [ONLINE]. AVAILABLE: [HTTPS://PUBMED.NCBI.NLM.NIH.GOV/28459981/](https://pubmed.ncbi.nlm.nih.gov/28459981/)[HTTPS://DOI.ORG/10.1016/J.RESENV.2025.100208](https://doi.org/10.1016/j.resenv.2025.100208)[AHTTP://SCIOTECA.CAF.COM/BITSTREAM/HANDL E/123456789/1091/RED2017-ENG-8ENE.PDF?SEQUENCE=12&ISALLOWED=Y](http://sciotea.caf.com/bitstream/handle/123456789/1091/red2017-eng-8ene.pdf?sequence=12&isallowed=y)[AHTTP://DX.DOI.ORG/10.1016/J.REGSCIURBECO.2008.06.005](http://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005)[AHTTPS://DOI.ORG/10.1016/J.REGSCIURBECO.2008.06.005](https://doi.org/10.1016/j.regsciurbeco.2008.06.005)
- [6] A. DAN PERANCANGAN SISTEM PENDUKUNG KEPUTUSAN PENENTUAN PRIORITAS PERBAIKAN JALAN PADA, I. KURNIAWAN, AND S. ASSEGAFF, "PEMERINTAH DAERAH KABUPATEN BATANG HARI DENGAN METODE AHP (ANALYTICAL HIERARCHY PROCESS)," *J. MANAJ. SIST. INF.*, VOL. 4, NO. 3, PP. 313–325, 2019.
- [7] D. Y. PRAYOGA, H. HERMAN, AND I. FARIDA, "PRIORITAS PEMELIHARAAN JALAN KABUPATEN BERDASARKAN KETERSEDIAAN ALOKASI ANGGARAN," *J. KONSTR.*, VOL. 21, NO. 2, PP. 281–288, 2023, DOI: 10.33364/KONSTRUKSI/V.21-2.1550.
- [8] SUPARNO, Y. FERISKA, H. PRAMONO, A. KHAMID, AND D. D. APRILIANO, "ANALISA KERUSAKAN JALAN KABUPATEN RUAS KLAMPOK –PG BANJARATMA KABUPATEN BREBES," *J. SCI. ENG. INF. SYST. RES.*, VOL. 1, NO. 1, PP. 10–18, 2023.
- [9] K. ADPENALIA AND R. H. SUMARTO, "JOURNAL OF INDONESIAN RURAL AND REGIONAL GOVERNMENT PERAN PEMERINTAH DESA BELANGIN DALAM MENYEDIAKAN INFRASTRUKTUR JALAN," VOL. 9, NO. 1, PP. 120–134, 2025, DOI: 10.47431/JIRREG.V9I1.633.
- [10] E. SOPYAN AND M. R. A. PUTRA, "EVALUASI KINERJA APARATUR SIPIL NEGARA DALAM PEMBANGUNAN INFRASTRUKTUR JALAN : STUDI KASUS DINAS PUPRP KABUPATEN CIAMIS," *PAMARENDA PUBLIC ADM. GOV. J. VOL.*, VOL. 5, NO. 1, PP. 139–155, 2025.
- [11] A. SAFITRI, D. ZADIA ARINI, D. DIANI, A. HALIM, M. AYYUB ABDUL AZIZ,

- AND S. AL FARIZI, “EFEKTIVITAS PERBAIKAN INFRASTRUKTUR: STUDI KASUS PERBAIKAN JALAN RUSAK DALAM SEMALAM DI KECAMATAN RUMBIA LAMPUNG TENGAH,” *J. KAJI. HUK. DAN KEBIJAK. PUBLIK*, VOL. 02, NO. 2, PP. 1121–1131, 2025, [ONLINE]. AVAILABLE: [HTTPS://JURNAL.KOPUSINDO.COM/INDEX.PHP/JKHKP](https://jurnal.kopusindo.com/index.php/jkhkp)
- [12] A. WULANDARI, IRMAYANSYAH, AND L. T. NINGRUM, “PENERAPAN METODE K-MEANS CLUSTERING UNTUK PEMETAAN PERBAIKAN JALAN DI KOTA BOGOR,” *J. INFORM. TEKNOLOGI. DAN SAINS*, VOL. 7, NO. 1, PP. 107–116, 2025, DOI: 10.51401/JINTEKS.V7I1.5323.
- [13] S. ASMIATUN AND N. WAKHIDAH, “IDENTIFIKASI PENGELOMPOKAN KONDISI PERMUKAAN JALAN MENGGUNAKAN ALGORITMA K-MEANS,” *J. PENGEMB. REKAYASA DAN TEKNOLOGI*, VOL. 2, NO. 1, PP. 17–23, 2018, DOI: 10.26623/JPRT.V14I1.1215.
- [14] A. B. ISAG AND R. H. IRAWAN, “IMPLEMENTASI K-MEANS DALAM PRIORITAS PERBAIKAN JEMBATAN DAN SALURAN AIR DI KECAMATAN NGRONGGOT,” *PROS. SEMNAS INOTEK ...*, PP. 282–286, 2021, [ONLINE]. AVAILABLE: [HTTPS://PROCEEDING.UNPKEDIRI.AC.ID/INDEX.PHP/INOTEK/ARTICLE/VIEW/1146](https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1146)
- [15] D. KURNIADI, Y. H. AGUSTIN, H. I. N. AKBAR, AND I. FARIDA, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN PEMBANGUNAN JALAN PADA DINAS PEKERJAAN UMUM DAN PENATAAN RUANG,” *AITI*, VOL. 20, NO. 1, PP. 64–77, 2023, DOI: 10.24246/AITI.V20I1.64-77.