

Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang

Siti Asmiatun^{*1}, Nur Wakhidah², Astrid Novita Putri³

^{1,2}Universitas Semarang; Jl Soekarno Hatta Tlogosari, 024-6702757 Universitas Semarang
Program studi Teknik Informatika Universitas Semarang
e-mail: ^{*1}siti.asmiatun@usm.ac.id, ²ida@usm.ac.id, ³astrid@usm.ac.id

Abstrak

Kondisi jalan di Indonesia khususnya di kota Semarang sering mengalami kerusakan pada saat musim hujan. Jika kondisi tersebut tidak segera diperbaiki maka dapat merugikan pengguna jalan seperti mengancam keamanan lalu lintas. Solusi dari permasalahan tersebut salah satunya yaitu menyediakan informasi kondisi permukaan jalan yang up to date sehingga dinas setempat dapat melakukan perbaikan jalan dengan cepat. Penelitian ini membahas tentang pengelompokan data kondisi jalan dengan memanfaatkan metode k-medoids. Hasil dari aplikasi roadroid didapatkan data sejumlah 638 data dengan atribut nama jalan, prosentase kondisi jalan baik, sedang, rusak ringan, rusak berat. Data telah berhasil dikelompokkan menjadi 4 cluster yang dapat menunjukkan tingkat frekuensi jalan kondisi baik, sedang, rusak ringan dan berat. Berdasarkan hasil pengujian kualitas cluster dengan silhouette coefficient dapat disimpulkan bahwa jumlah cluster 2 merupakan cluster yang kualitasnya paling baik karena paling mendekati angka 1. Hal ini berarti untuk analisis pengelompokan data kondisi jalan rusak lebih baik menggunakan cluster 2. Sehingga data yang dihasilkan dari penerapan metode k-medoids akan semakin mendekati ketepatan pengelompokan data kondisi jalan.

Kata kunci—K-medoids, Cluster, K-Means, Kondisi Jalan

Abstract

Road conditions in Indonesia especially in the city of Semarang often suffer damage during the rainy season. If these conditions are not immediately improved, it can be detrimental to road users, such as threatening traffic safety. One of the solutions to these problems is to provide information on up to date road surface conditions so that the local office can make road repairs quickly. This study discusses the grouping of road condition data using the k-medoids method. The results of the roadroid application obtained 638 data with the attributes of the road name, the percentage of good road conditions, moderate, slightly damaged, heavily damaged. The data has been successfully grouped into 4 clusters which can show the level of frequency of roads in good, moderate, light and heavy conditions. Based on the results of cluster quality testing with the silhouette coefficient, it can be concluded that the number of clusters 2 is the best quality cluster because it is closest to number 1. This means that for the analysis of grouping data on damaged road conditions it is better to use cluster 2. So that the data generated from the application of the k-medoids method will be closer to the accuracy of the grouping of road condition data.

Keywords— K-medoids, Cluster, K-Means, Road Condition



1. PENDAHULUAN

Indonesia merupakan negara dengan dua musim dimana ada musim hujan dan musim kemarau. Disaat memasuki musim hujan, banyak kondisi jalan di Indonesia khususnya di kota Semarang yang mengalami kerusakan pada permukaan jalan. Kondisi jalan yang kualitasnya rendah ini merugikan pengguna jalan dan dapat menyebabkan kerusakan pada kendaraan seperti meningkatkan penggunaan bahan bakar, meningkatkan biaya pemeliharaan kendaraan bagi pengguna jalan, mengurangi kenyamanan berkendara bahkan dapat mengancam keamanan lalu lintas bagi pengguna jalan [1]. Informasi tentang kondisi permukaan jalan ini sangat bermanfaat bagi pengguna jalan karena dapat menghindari jalan yang kondisinya buruk dan mengambil jalan alternatif.

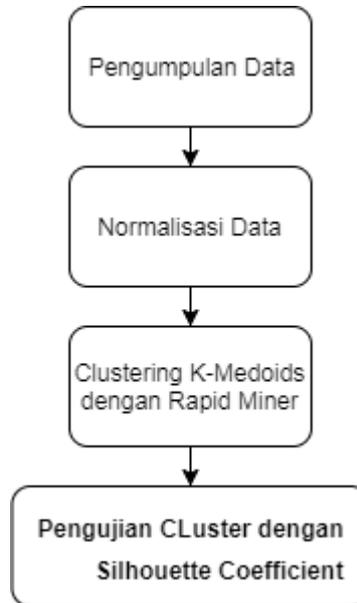
Peningkatan aktifitas di lalu lintas dikarenakan bertambahnya pengguna jalan. Hal tersebut sangat memungkinkan besar manfaatnya jika pengguna jalan dapat mengetahui informasi secara real time tentang kondisi jalan yang akan dilalui. Dalam peraturan daerah, anggaran kota untuk memperbaiki jalan akan dibatasi, jadi informasi yang berkaitan dengan penentuan jalan yang harus segera diperbaiki menjadi sangat penting [2]. Penelitian tentang mendeteksi jalan berlubang dan pemantauan lalu lintas adalah masalah yang banyak diteliti baru baru ini. Banyak penelitian yang menggunakan smartphone sebagai alat pengumpul data karena biaya yang digunakan relatif lebih murah dan fitur yang digunakan lebih mudah [1]. Untuk pengumpulan data biasanya menggunakan sensor khusus seperti GPS, Accelerometer, dan kamera lalu lintas [3].

Penelitian terdahulu sudah sering membahas tentang pengelolaan informasi kondisi jalan. Salah satunya penelitian dari agusta tentang *Web-Based Mapping Untuk Pemetaan Lokasi Kerusakan Jalan Raya Menggunakan Cluster Marker* telah menghasilkan aplikasi *web-based mapping* untuk pemetaan lokasi jalan rusak. Aplikasi ini memanfaatkan Maps API untuk menampilkan data koordinat ke dalam peta google maps [4]. Berbeda dengan lars dan hans yang telah membuat aplikasi Roadroid untuk memonitoring kondisi jalan. Kemjudian dat yang dihasilkan akan diexport ke dalam Sistem Informasi Geografis. Sehingga diharapkan dengan adanya layanan ini pihak terkait dapat memantau kondisi jalan secara rutin [5]. Sedangkan penulis Amr s, Jin lin, Muhammed Rahman dkk menulis tentang *Monitoring Road Surface Anomalies Towards Dinamic Road Mapping For Future Smart City*. Dalam penelitiannya menghasilkan sistem pemantauan kondisi jalan dengan menggunakan *accelerometer* dan GPS. Data yang dihasilkan akan diolah menggunakan *decision tree* untuk membedakan kondisi jalan yang diterima [6].

Melihat beberapa penelitan terkait, penelitian tersebut belum ada yang menerapkan metode tertentu untuk pengolahan data. Penelitian ini memfokuskan penerapan metode untuk pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer menggunakan metode K-medoids. Metode K-medoids sendiri dikenal sebagai salah satu metode yang menghasilkan tingkat akurasi tinggi. Seperti dalam penelitian yang telah dilakukan Ni Nyoman yang berjudul *Otomatisasi Klasifikasi Buku Perpustakaan dengan Menggabungkan Metode K-NN dengan K-Medoids* telah menghasilkan tingkat akurasi sebesar 83% pada metode K-Medoids [7]. Sehingga penelitian ini memutuskan menggunakan metode K-Medoids karena sangat efektif untuk mengelompokkan data. Penerapan metode K-Medoids untuk mengelompokkan kondisi jalan diharapkan dapat memvisualisasikan data kondisi jalan kedalam kelompok status jalannya masing-masing, sehingga dengan adanya informasi pengelompokkan kondisi jalan dapat membantu petugas Dinas Pekerjaan Umum Kota Semarang untuk mendapatkan informasi kondisi jalan dengan cepat dan akurat.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

Gambar 1 merupakan ilustrasi dari tahapan penelitian. Penelitian ini dimulai dari pengumpulan data yang didapat dari Dinas Pekerjaan Umum (DPU) Kota Semarang. DPU kotasemarangtelahmenerapkanaplikasi*roadroid*untukmengumpulkan data kondisijalanrusak. Cara kerjaaplikasitersebutmemanfaatkan sensor accelerometer yang ada di smartphone untukmerekam data kondisijalan. Kemudian dari data mentah yang bebentuk excel akan dinormalisasikan untuk menetralkan atribut. Setelah normalisasi data selesai, data dikelompokkan menggunakan algoritma K-Medoids dengan *tools rapid miner* . Dan terakhir setelah mendapatkan hasil cluster diuji menggunakan pengujian *silhouette coefficient* untuk mendapatkan cluster yang berkualitas.

2.2 Pengumpulan Data

Data yang kami gunakan adlaah dataset yang telah dihasilkan dari aplikasi *roadroid* yang memanfaatkan sensor *accelerometer* berfungsi sebagai perekam kondisi jalan . Data ini diambil pada kisaran bulan januari tahun 2019 oleh ppihak Dinas Pekerjaan Umum Kota Semarang. Data kondisi jalan dapat dilihat pada tabel 1 berikut:

Tabel 1. Data Kondisi Jalan

Jalan	Baik	Sedang	Ringan	Berat	Kondisi
Jl. Barito	80.72	19.28	0	0	Baik
Jl. Bugangan	50.68	49.32	0	0	Baik
Jl. BuganganRegol	73.4	26.6	0	0	Baik

Jl. Cilosari	55.4	44.6	0	0	Baik
Jl. Citandui	0	100	0	0	Sedang
Jl. Citandui 3	0	100	0	0	Sedang
Jl. Citandui 4	0	100	0	0	Sedang
Jl. Citandui Selatan	0	100	0	0	Sedang
Jl. Citarum	46.65	53.35	0	0	Sedang
Jl. Dargo	49.02	50.98	0	0	Sedang
Jl. Halmahera	82.82	17.18	0	0	Baik
Jl. Hawa	0	100	0	0	Sedang
Jl. Indragiri	0	100	0	0	Sedang
Jl. Inspeksi Barat	86.98	13.02	0	0	Baik
Jl. Inspeksi Timur	0	100	0	0	Sedang
Jl. Karimata	100	0	0	0	Baik
Jl. Kartini	83.9	16.1	0	0	Baik
Jl. Krakatau	100	0	0	0	Baik
Jl. Labuan	0	100	0	0	Sedang
Jl. LiguTengah	100	0	0	0	Baik
Jl. M.T. Haryono	91.56	8.44	0	0	Baik
Jl. Maluku	0	100	0	0	Sedang
Jl. Manisharjo	0	100	0	0	Sedang

2.3 Normalisasi Data

Normalisasi merupakan transformasi yang dilakukan untuk merubah data. Normalisasi digunakan untuk menetralkan skala atribut data kedalam range yang spesifik dan kecil seperti -1 sampai 1 atau 0 sampai 1. Maka dibutuhkan dan Min-Max Normalization. Min-Max Normalization sendiri adalah teknik normalisasi dengan melakukan transformasi linier pada atribut data asli untuk menghasilkan range nilai yang sama [8]. Min-Max Normalization memetakan sebuah value v dari atribut A menjadi v' kedalam range $[new_min_A, new_Max_A]$ dengan Persamaan 1.

$$v' = \frac{v - min_A}{max_A - min_A} (new_{max_A} - new_{min_A}) + new_{min_A} \quad (1)$$

Pada tahapan normalisasi, penulis menggunakan teknik Min-Max, dimana menggunakan nilai $min = 0$, nilai $max = 100$, nilai $newmin = 0$ dan nilai $newmax = 1$. Dengan menerapkan rumus 1 diatas menghasilkan normalisasi data sebagai berikut :

Tabel 2. Data Sample Hasil Normalisasi Min Max

Jalan	Baik	Sedang	Ringan	Buruk
Jl. Barito	0,8072	0,1928	0	0
Jl. Bugangan	0,5068	0,4932	0	0
Jl. BuganganRegol	0,7305	0,2606	0	0
Jl. Cilosari	0,4406	0,4406	0	0
Jl. Citandui	0	1	0	0
Jl. Citandui 3	0	1	0	0
Jl. Citandui 4	0	1	0	0
Jl. Citandui Selatan	0	1	0	0
Jl. Citarum	0,465	0	0	0
Jl. Dargo	0	0,5098	0	0

2.4 Clustering Metode Non-hierarchical Partitioning atau K-Medoids

K-Medoids atau Partitioning Around Medoids (PAM) merupakan algoritma clustering yang hampir sama dengan algoritma K-Means. Perbedaan dari kedua algoritma ini yaitu algoritma K-Medoids menggunakan objek perwakilan (medoid) sebagai pusat cluster untuk setiap cluster, sedangkan K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster [9]. Kelebihan dari algoritma K-Medoids dapat mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier yang berarti data yang memungkinkan menyimpang dari distribusi data lain. Untuk memahami algoritma ini dapat daat membaca langkah-langkah algoritma K-Medoids sebagai berikut :

- a. Inisialisasi pusat cluster sebanyak k (jumlah cluster) . Tahapan ini menentukan pusat medoid 1 dan medoid 2 dari contoh data tabel 2. Penulis menggunakan medoid 1 = [0,7305,0,2606,0,0] diasumsikan y1 dan medoid 2 = [0,1,0,0] diasumsikan y2
- b. Tahapan ini menghitung jarak setiap data tabel 2 ke dalam cluster terdekat menggunakan persamaan ukuran jarak *Manhattan Distance* dibawah ini:

$$D^{(x,y)} = \sum_{i=1}^n |x_i - y_i| \tag{2}$$

Tabel 3. Contoh Perhitungan Jarak

X1	X2	X3	X4	Jarak x	Jarak y
0,8072	0,1928	0	0	0,1188	0

Hasil dari perhitungan semua data tabel 2 mendapatkan total cost sebanyak 72,2022.

- c. Memilih secara acak data pada masing-masing cluster untuk dijadikan kandidat pada medoid baru. Untuk medoid baru menggunakan medoid 1 = [0,4406,0,4406,0,0] dan medoid 2 = [0,1,0,0]
- d. Menghitung jarak setiap data di dalam *cluster* dengan kandidat medoid baru. Hasil perhitungan jarak pada cluster baru total cost sebanyak 71,6418
- e. Menghitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika $S < 0$, maka data dengan data cluster ditukar untuk membentuk sekumpulan k data baru sebagai medoid. Melihat total cost lama dan baru dapat dihitung

total simpangan (S) = 72,2022 - 71,6418 hasilnya 0,7876 > 0, maka perhitungan dapat dihentikan.

- f. Ulangi langkah c sampai dengan selesai tidak terjadi perubahan medoid, sehingga *cluster* beserta anggota *cluster* siap untuk dipublish.

2.5 Pengujian dengan *silhouette coefficient*

Metode *silhouette coefficient* merupakan gabungan dari dua metode yaitu metode *cohesion* yang berfungsi untuk mengukur seberapa dekat relasi antara objek dalam sebuah *cluster*, dan metode *separation* yang berfungsi untuk mengukur seberapa jauh sebuah *cluster* terpisah dengan *cluster* lain. Tujuan penggunaan metode ini untuk menguji dari kualitas hasil *cluster* yang dihasilkan dari metode K-Medoids. perhitungan *silhouette coefficient* [10]:

- a. Menghitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu *cluster* dengan persamaan:

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Cluster 1: ([0.4406, 0.4406, 0, 0], [0.8072, 0.1928, 0, 0])

$$a(i) = \sqrt{((0,4406 - 0,8072)^2 + (0,4406 - 0,1928)^2)} = 0,1957$$

- b. Menghitung rata-rata jarak objek dengan semua objek lain yang berada pada *cluster* lain, kemudian ambil nilai paling minimum dengan persamaan:

$$d(i, C) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (4)$$

Cluster 2 : ([0, 1, 0, 0], [0, 465, 0, 0])

$$[0,4406, 0,4406, 0,0] \rightarrow [0,1,0,0] = \text{distance} = \sqrt{((0,4406 - 0)^2 + (0,4406 - 1)^2)} = 0,511$$

$$[0,4406, 0,4406, 0,0] \rightarrow [0,465, 0,0,0] = \text{distance} = \sqrt{((0,4406 - 0,465)^2 + (0,4406 - 0)^2)} = 0,1957$$

$$[0,8072, 0,1928, 0,0] \rightarrow [0,1,0,0] = \text{distance} = \sqrt{((0,8072 - 0)^2 + (0,1928 - 1)^2)} = 1,3030$$

$$[0,8072, 0,1928, 0,0] \rightarrow [0,465, 0,0,0] = \text{distance} = \sqrt{((0,8072 - 0,465)^2 + (0,1928 - 0)^2)} = 0,1542$$

Rata-rata jarak *cluster* 2 = 0,5409

Langkah yang sama dilakukan pada *cluster* 3 menghasilkan jarak rata-rata sebanyak 0,6345. Kemudian nilai minimal diambil dari rata-rata jarak *cluster* 2 dan 3

$$b(i) = (0,5409 < 0,6345) = 0,5409$$

- c. Hitung nilai *silhouette coefficient* dengan persamaan:

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

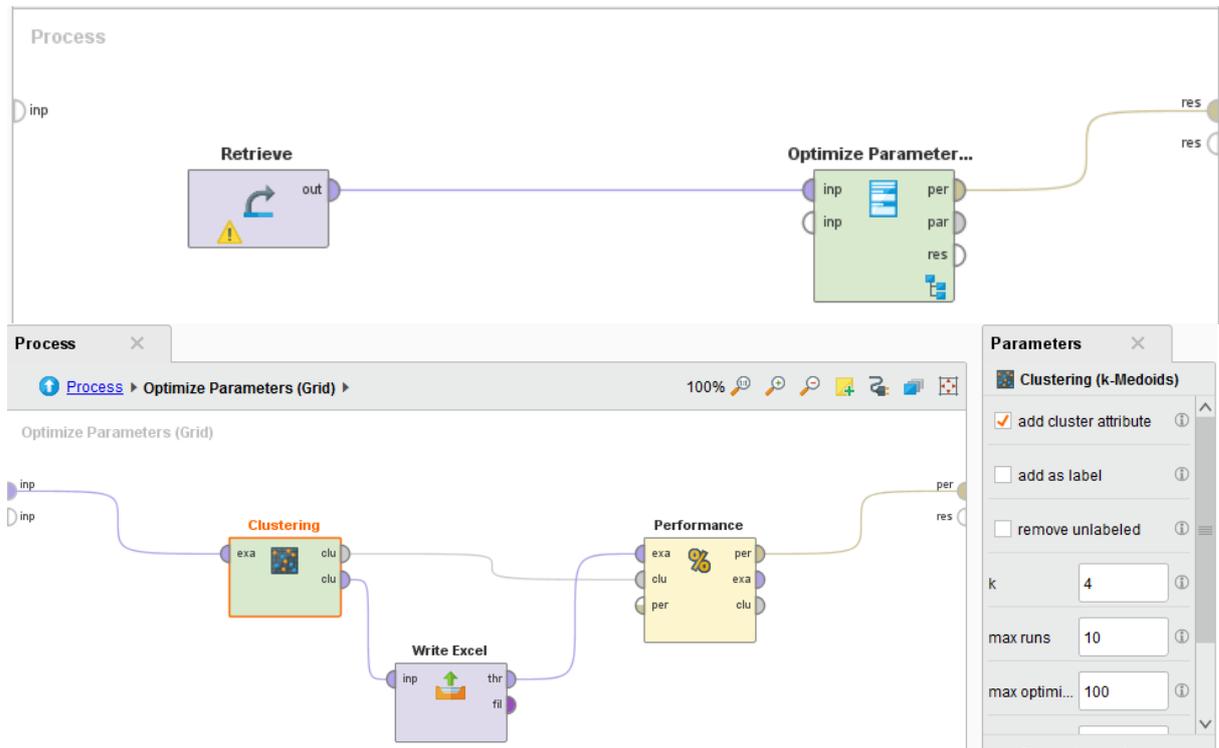
$$s(i) = 1 - (0,1957 / 0,5409) = 1 - 0,3618 = 0,6382$$

- d. Nilai hasil *silhouette coefficient* terletak pada kisaran nilai -1 hingga 1. Semakin nilai *silhouette coefficient* mendekati nilai 1, maka semakin baik pengelompokan data dalam satu *cluster*. Sementara hasil *silhouette coefficient* *cluster* 1 sebanyak 0,6382 sudah cukup baik karena mendekati nilai 1.

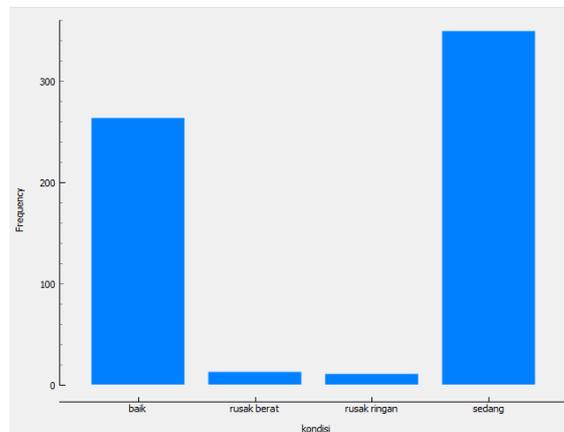
3. HASIL DAN PEMBAHASAN

3.1. Penerapan Algoritma K-Medoids

Pada penerapan algoritma *k-medoids* kali ini menggunakan dataset data kondisi jalan tahun 2018, dimana pada dataset tersebut dipilih 638 sampel data yang dikelompokkan kedalam 4 kluster yaitu kondisi jalan baik, sedang, rusak ringan, dan rusak berat. Untuk iterasi yang digunakan sebanyak 10. Data tersebut dianalisis menggunakan rapid miner seperti gambar 1 dibawah ini:

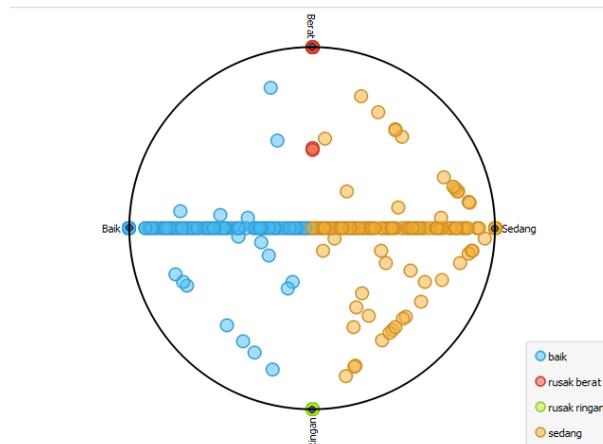


Gambar 2. Implementasi Rapid Miner



Gambar 3. Grafik Kondisi Jalan

Gambar 2 diatas menunjukkan bahwa penerapan algoritma menggunakan K-Medoids telah berhasil dilakukan dengan menghasilkan 4 kluster yaitu jalan kondisi jalan baik, sedang, rusak ringan dan rusak sedang. Tampilan grafik cluster dapat dilihat pada gambar 3.



Gambar 4. Hasil Clustering

3.2. Pengujian

Untuk mengukur kemampuan kualitas cluster, penelitian ini menggunakan pengujian yang berpengaruh terhadap jumlah nilai k . Pengujian pengaruh jumlah cluster dilakukan dengan menggunakan 638 data uji untuk mengetahui jumlah cluster terbaik dengan melihat dari nilai Silhouette Coefficient yang dihasilkan. Nilai cluster (k) yang akan diujikan adalah nilai k 4 sampai 8. Hasil pengujian dapat dilihat pada Tabel 2

Tabel 4. Hasil Uji Jumlah Cluster

Data Uji	Nilai k	Rata-rata Silhouette Coefficient
638	4	0,57432
	5	0,532110
	6	0,501650
	7	0,457142
	8	0,407543

Pada tabel 2 di atas menunjukkan hasil dari pengujian cluster. Pada proses hasil pengujian, kualitas cluster paling baik diperoleh pada cluster ke 4 dengan nilai *Silhouette Coefficient* rata-rata 0,57432.

4. KESIMPULAN

Hasil penelitian ini dapat disimpulkan bahwa untuk mengelompokkan data kondisi jalan rusak dapat menggunakan operator rapid miner. Rapid miner mendukung analisis data dengan menggunakan metode K-Medoids. Dari data yang digunakan jumlah 638 data telah berhasil dikelompokkan menjadi 4 cluster yang dapat menunjukkan tingkat frekuensi jalan kondisi baik, sedang, rusak ringan dan rusak berat. Dari hasil pengujian kualitas cluster menggunakan *Silhouette* menunjukkan bahwa cluster 4 menunjukkan kualitas yang paling baik dengan hasil rata-rata nilai *Silhouette Coefficient* sebesar 0,57432. Berdasarkan hasil tersebut dapat disimpulkan bahwa jumlah cluster 2 merupakan cluster yang kualitasnya paling baik karena paling mendekati angka 1. Hal ini berarti untuk analisis pengelompokan data kondisi jalan rusak lebih baik menggunakan cluster 2. Sehingga data yang dihasilkan dari penerapan metode k-medoids akan semakin mendekati ketepatan pengelompokan data kondisi jalan.

5. SARAN

Untuk pengembangan penelitian selanjutnya diharapkan data ini dapat diterapkan pada aplikasi sehingga dapat divisualisasikan dengan baik dan dapat meningkatkan akurasi dengan menggunakan metode yang berbeda.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih kepada Ristekdikti melalui LPPM Universitas Semarang yang telah member dukungan financial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] P. B. Lanjewar, J. Khedkar, R. Sagar, R. Pawar, and K. Gosavi 2015, “*Survey of Road Bump and Intensity Detection algorithms using Smartphone Sensors*,” Vol. 6, No. 6, pp. 5133–5136.
- [2] B. H. Jakob Eriksson, Lewis Girod 2008, “*The Pothole Patrol: Using a Mobile Sensor Network for Road Surface Monitoring*,” Proc. 6th Int. Conf. Mob. Syst. Appl. Serv., pp. 29–39.
- [3] V. Akinwande, O. Bello, K. Adewole, and A. Akintola 2015, “*Automatic and real-time Pothole detection and Traffic monitoring system using Smartphone Technology Automatic and real-time Pothole detection and Traffic monitoring system using Smartphone Technology*,” Int. Conf. Comput. Sci. Res. Innov. (CoSRI 2015), No. August,.
- [4] A. P. R. Pinem 2018, “*Web-Based Mapping untuk Pemetaan Lokasi Kerusakan Jalan Raya Menggunakan Cluster Marker*,” J. Sisfokom (Sistem Inf. dan Komputer), Vol. 7, No. 2, p. 93,
- [5] Lars Forsslöf and Hans Jones 2015, “*Roadroid: Continuous Road Condition Monitoring with Smart Phones*,” J. Civ. Eng. Archit., Vol. 9, No. 4, pp. 485–496,
- [6] A. S. El-Wakeel, J. Li, M. T. Rahman, A. Noureldin, and H. S. Hassanein 2018, “*Monitoring Road Surface Anomalies to Wards Dynamic Road Mapping for Future Smart Cities*,” 2017 IEEE Glob. Conf. Signal Inf. Process. Glob. 2017 - Proc., vol. 2018-Janua, pp. 828–832,.
- [7] N. N. E. Smrti 2013, “*Otomatisasi Klasifikasi Buku Perpustakaan dengan Menggabungkan Metode K-NN dengan K-Medoids*,” Lontar Komput. J. Ilm. Teknol. Inf., Vol. 4, No. 1, pp. 201–214.
- [8] H. Junaedi, H. Budianto, I. Maryati, and Y. Melani 2011, “*Data transformation pada data mining*,” Pros. Konf. Nas. Inov. dalam Desain dan Teknol., Vol. 7, pp. 93–99.

- [9] N. K. Kaur, U. Kaur, and D. Singh 2014, “*K-Medoid Clustering Algorithm- A Review,*” *Int. J. Comput. Appl. Technol., Vol. 1, No. 1, pp. 42–45.*
- [10] M. Fornalski *et al.* 2014, “*Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means pada Pengelompokan Dokumen,*” *Int. J. Comput. Appl. Technol., Vol. 7, No. 2, pp. 93–99.*