

# Identifikasi Kondisi Permukaan Jalan Menggunakan K-Means Clustering

## Road Surface Conditions Identification Using K-Means Clustering

Siti Asmiatun<sup>1\*</sup>, Nur Wakhidah<sup>2</sup>, Astrid Novita Putri<sup>3</sup>

<sup>1</sup>Universitas Semarang

Jl Soekarno Hatta, Tlogosari, Semarang, 0246702757

<sup>2,3</sup>Universitas Semarang

Jl Soekarno Hatta, Tlogosari, Semarang, 0246702757

Siti.asmiatun@usm.ac.id<sup>1\*</sup>, ida@usm.ac.id<sup>2</sup>, astrid@usm.ac.id<sup>3</sup>

**Abstrak** – Kecelakaan lalu lintas merupakan salah satu penyebab utama kematian atau penyakit pada manusia. Peristiwa Kecelakaan lalu lintas menjadi salah satu permasalahan global yang membutuhkan penanganan serius. Hal ini terjadi karena perkembangan infrastruktur tidak seimbang dengan sector lain seperti industry dan pembangunan perumahan. Penelitian ini memfokuskan cara pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer menggunakan metode K-means. Metode K-means sendiri dikenal sebagai salah satu metode yang menghasilkan tingkat akurasi tinggi. 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. Tingkat akurasi yang dihasilkan dari penerapan metode sebanyak 83% yang artinya metode K-Means cukup akurat dalam mengelompokkan kondisi jalan

**Kata Kunci:** k-means, clustering, kecelakaan, kondisi jalan

**Abstract** – Traffic accidents are one of the main causes of death or disease in humans. Traffic accident has become one of the global problems that requires serious handling. This happens because the development of infrastructure is not balanced with other sectors such as industry and housing development [1]. This research focuses on data processing by grouping data obtained from the accelerometer sensor using the K-means method. K-means method itself is known as one method that produces a high degree of accuracy. From the data used a total of 638 data have been successfully grouped into 4 clusters that can indicate the level of frequency of roads in good condition, moderate, lightly damaged and heavily damaged. The level of accuracy resulting from the application of the method is 83%, which means the K-Means method is quite accurate in classifying road conditions

**Keywords:** k-means, clustering, accident, conditions road

### 1. Pendahuluan

Kecelakaan lalu lintas merupakan salah satu penyebab utama kematian atau penyakit pada manusia. Peristiwa Kecelakaan lalu lintas menjadi salah satu permasalahan global yang membutuhkan penanganan serius [1]. Hal ini terjadi karena beberapa faktor yang muncul seperti faktor manusia, faktor kendaraan, faktor jalan/lingkungan atau kombinasi dari beberapa faktor tersebut [2]. Di Indonesia sendiri pada tahun 2016 Badan Pusat Statistik mencatat bahwa sepanjang tahun 2015 angka kecelakaan kendaraan darat yang terjadi mengalami peningkatan

fluktuasi. Jumlah peningkatan kecelakaan terjadi dalam 10 tahun terakhir. Peningkatan paling tinggi terjadi pada tahun 2012 yakni mencapai 117,9 ribu kasus. Dengan naiknya angka kecelakaan, selain mewacanakan uji KIR bagi kendaraan pribadi, pemerintah juga harus segera merencanakan perbaikan infrastruktur di berbagai daerah. Hal ini dikarenakan faktor penyebab kecelakaan tidak hanya pada kendaraan tetapi juga kondisi jalan.

Dalam menangani infrastruktur jalan tentu saja tidak terlepas dari pemerintahan. Dinas Pekerjaan Umum atau DPU di kota Semarang merupakan institusi pemerintahan yang bertugas menangani sarana prasarana infrastruktur seperti jembatan, trotoar, gorong-gorong, penerangan jalan dan termasuk juga memperbaiki kondisi jalan yang berkaitan dengan kegiatan masyarakat sehari-hari [3]. Sementara itu, dalam menangani kondisi permukaan jalan masih kurang efektif karena laporan yang diterima masih dalam bentuk data excel. Sehingga menghambat petugas dalam menerima informasi kondisi permukaan jalan.

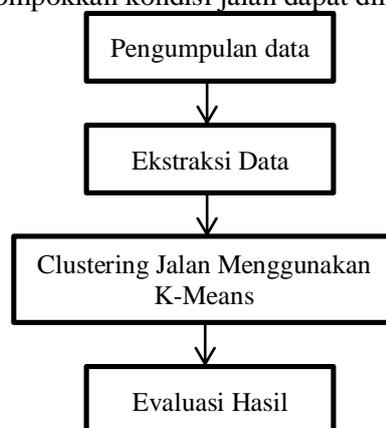
Penelitian tentang mendeteksi kondisi jalan sudah banyak dilakukan. Diantaranya adalah penelitian yang dilakukan oleh Umang Bhatt dkk mengenai *Intelligent Pothole Detection and Road Condition Assessment* telah menghasilkan sebuah sistem yang menampilkan kondisi jalan baik dan tidak baik dengan mengklasifikasikan kondisi menggunakan algoritma SVM. Klasifikasi tersebut menghasilkan akurasi 93% [4]. Sementara itu penelitian yang dilakukan oleh aniket kulkani dkk mengenai *Pothole Detection System using Machine Learning on Android* telah menghasilkan sebuah aplikasi yang dapat mendeteksi lubang jalan menggunakan accelerometer. Accelerometer merupakan sensor yang dapat merekam jalan berlubang dengan memanfaatkan data sumbu x dan z. Sedangkan untuk menghitung nilai ambang batas menggunakan metode neural network. Kemudian data tersebut akan dikirimkan melalui email dalam bentuk csv. Penelitian ini menghasilkan akurasi sebesar 90% [5].

Penelitian ini memfokuskan cara pengolahan data dengan mengelompokkan data-data yang didapat dari sensor accelerometer. Metode K-means sendiri dikenal sebagai salah satu metode yang menghasilkan tingkat akurasi tinggi. Seperti dalam penelitian yang telah dilakukan sachin dan durga dkk yang berjudul *Data mining approach to characterize road accident location* telah berhasil mengelompokkan lokasi kecelakaan di jalan menjadi tiga kategori yaitu frekuensi tinggi, sedang dan rendah [6]. Selain itu penelitian dari nanda dkk terkait clustering kecelakaan daerah rawan kecelakaan telah menghasilkan akurasi sebesar 67,44% dengan menggunakan metode clustering [1]. Dari beberapa penelitian terkait, penelitian ini memperbaharui tentang pengelompokkan data kondisi jalan dengan menggunakan metode K-Means. Penggunaan metode K-Means diharapkan dapat menghasilkan akurasi yang lebih baik, sehingga dapat menyajikan informasi yang lebih tepat, akurat dan efisien bagi petugas dinas yang membutuhkan.

## 2. Metode Penelitian

### 2.1. Tahapan Penelitian

Diagram alir dari pengelompokkan kondisi jalan dapat dilihat dari gambar 1 dibawah ini :



Gambar 1 Diagram Alir Penelitian

Dapat dilihat pada gambar 1 menjelaskan tentang alur dari penelitian ini. Dimulai dari tahap pengumpulan data menggunakan sensor *accelerometer* yang dapat merekam data kondisi jalan berupa ukuran kerusakan, titik koordinat dan sebagainya. Kemudian dari hasil pengumpulan data tersebut, data harus diekstrak terlebih dahulu untuk mengkonversi titik coordinate UTM ke dalam titik koordinat decimal degree. Karena didalam google earth hanya dapat membaca titik koordinat decimal degree. Kemudian data tersebut diolah untuk mengelompokkan ke dalam kondisi jalan baik, sedang, dan buruk menggunakan K-Means Clustering. Setelah proses clustering selesai, evaluasi hasil clustering untuk mengetahui keakuratan dari metode tersebut.

## 2.2. Dataset

Dataset yang digunakan dalam penelitian ini bersumber dari hasil pengumpulan data kondisi jalan rusak menggunakan aplikasi DPU yang memanfaatkan sensor *accelerometer*. Untuk atribut yang digunakan dalam mengelompokkan kondisi jalan dapat dilihat dari tabel 1 dibawah ini :

Tabel 1. Daftar Atribut

No	Nama Atribut	Type
1.	Kondisi Baik	Real
2.	Kondisi Sedang	Real
3.	Kondisi Rusak Ringan	Real
4.	Kondisi Rusak Berat	Real
5.	Jalan	Label
6.	Kondisi	Prediksi

Dalam menggunakan metode k-means, tahapan yang pertama harus melakukan preprosesing data yaitu membersihkan data sebelum di proses lebih lanjut yang bertujuan untuk memperbaiki data yang tidak valid. Adapun teknik yang harus dilakukan untuk membersihkan data yaitu mengisi missing value dan membuang outlier. Missing values merupakan informasi yang tidak tersedia, sulit dicari atau informasi yang dicari tidak tersedia. Sedangkan untuk outlier merupakan data yang berbeda dari data lainnya

## 2.3. K-Means Clustering

Clustering merupakan metode untuk membagi satu set data menjadi sejumlah kelompok tertentu. Salah satu metode paling populer yaitu K-Means. Dalam pengelompokkan K-Means memecah kumpulan data menjadi beberapa kelompok. Dibutuhkan  $K_y$  sebagai input parameter dan partisi satu set objek  $n$  dari  $K_y$  cluster. Nilai rata-rata objek diambil sebagai kesamaan parameter untuk membentuk cluster. *Mean* cluster atau pusat terbentuk oleh pemilihan acak objek  $K_y$  dengan membandingkan sebagian besar kemiripan objek lain yang ditugaskan ke cluster [7]. Untuk setiap vektor data, algoritma ini menghitung jarak antara vektor data dan setiap cluster centroid menggunakan persamaan (1)

$$D(Z_p, M_j) = \sqrt{\sum (Z_p - M_j)^2} \quad (1)$$

$Z_p$  adalah  $p^{\text{th}}$  data poin  $M_j$

$M_j$  adalah  $j^{\text{th}}$  centroid dari cluster.

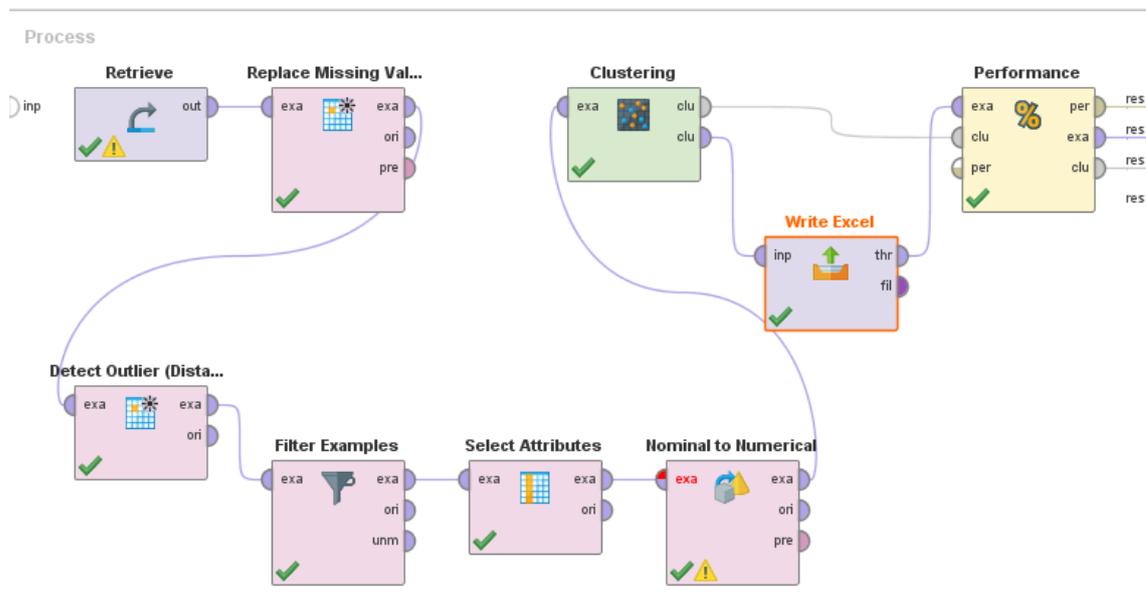
Sedangkan untuk menghitung centroid dapat dihitung ulang setiap kali penambahan titik data dalam cluster  $j$ . Dihitung menggunakan persamaan (2)

$$M_j = 1/N_j \sum_{p \in C_j} \forall Z_p \in C_j \quad (2)$$

Dimana  $N_j$  adalah nomor data poin di cluster  $j$

### 3. Hasil dan Pembahasan

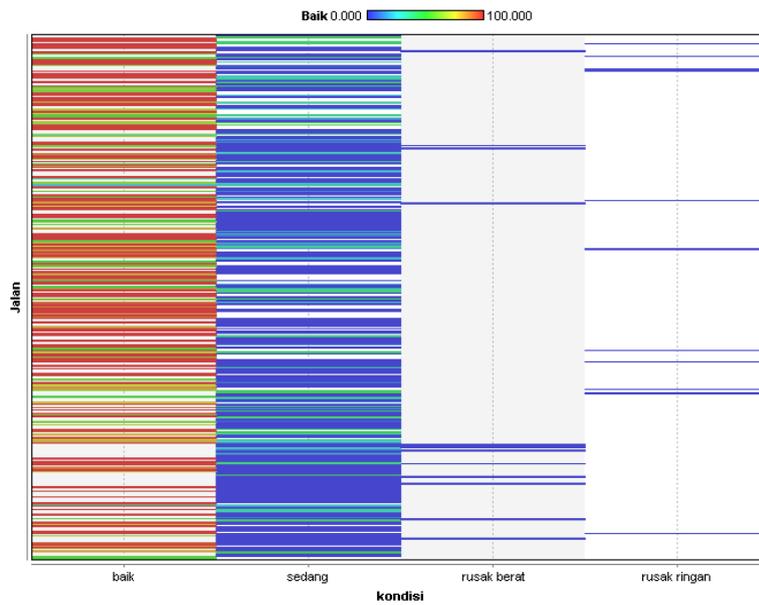
Dalam penelitian ini, penerapan metode K-Means menggunakan operator rapid miner. Karena pada operator rapid miner mampu menghitung data dengan menghubungkan node-node tertentu dalam jumlah yg sangat besar. Hasil yang ditmpilkan pada rapid miner dapat berupa gambar dan grafik. Pada kasus penelitian ini menggunakan data kondisi jalan sekota semarang dengan jumlah 638 data. Konfigurasi rapid miner yang akan digunakan dapat dilihat pada gambar 1 dibawah ini :



Gambar 1. Konfigurasi Rapid Miner

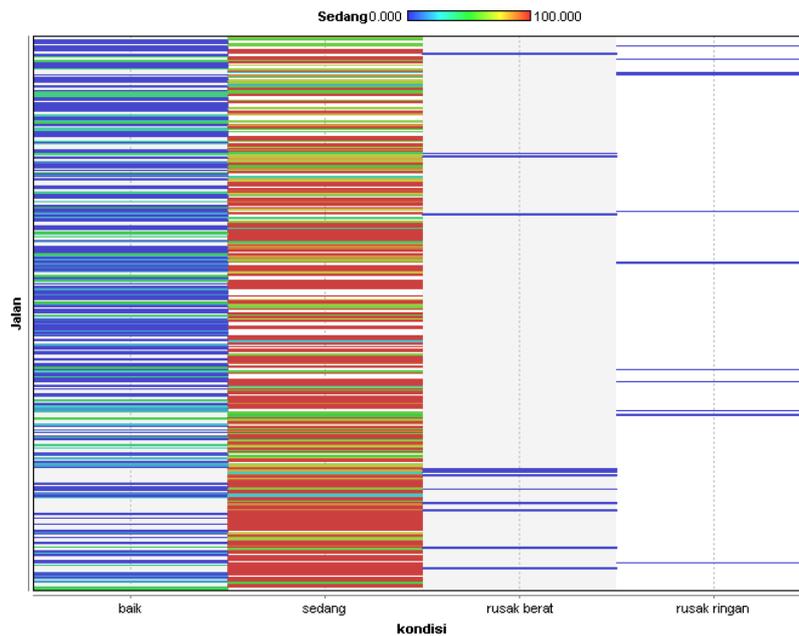
Pada node pertama terdapat replace missing value yang berfungsi untuk preprosesing dengan cara mengecek apakah ada data yang kosong atau tidak jika ada akan dilakukan penggantian dengan mengambil nilai rata-rata. Kemudian dilakukan proses outlier untuk menghapus data yang berbeda. Data yang bukan outlier akan diseleksi untuk dimasukkan ke cluster tertentu. Sebelum melakukan proses clustering, data didistraksi untuk mengconvert dari data nominal menjadi data numeric. Karena proses clustering hanya menerima data numeric.

Dalam proses clustering, penelitian ini menggunakan menentukan 4 titik centroid secara acak. Data yang digunakan sebanyak 46 telah berhasil dikelompokkan ke dalam 3 cluster. Cluster pertama berjumlah 2 item, cluster kedua berjumlah 24 item, cluster ketiga berjumlah 14 item dan cluster keempat berjumlah 6 item. Hal ini dapat disimpulkan bahwa dari 638 data telah berhasil dikelompokkan menjadi 4 cluster dimana grafik dapat dilihat pada gambar-gambar dibawah ini :



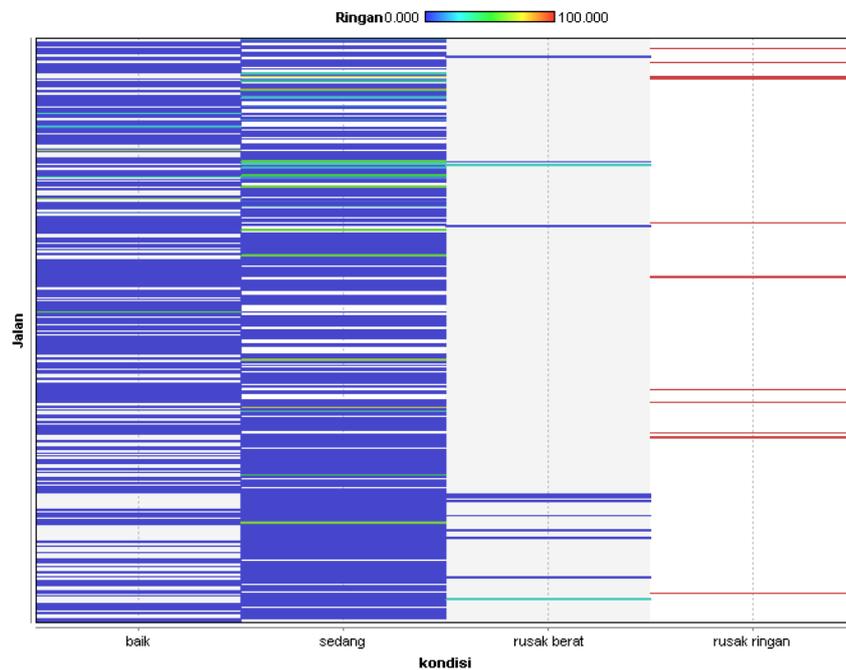
Gambar 2. *Group By* Kondisi jalan Baik

Gambar 2 menjelaskan grafik yang menyatakan bahwa kondisi jalan baik terlihat pada label diatas grafik untuk ukuran baik dari angka 0 berwarna biru sampai dengan angka 100.000 berwarna merah. Dari grafik menunjukkan bahwa blok yang berwarna merah sampai kecoklatan berada pada kondisi Baik.



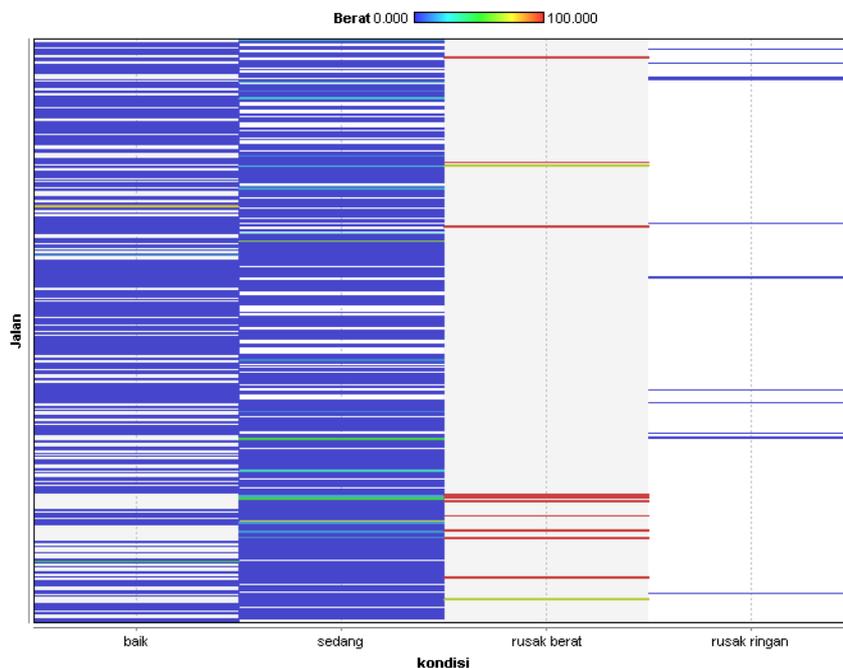
Gambar 3. *Group By* Kondisi jalan Sedang

Sedangkan gambar 3 diatas menjelaskan tentang kondisi jalan sedang. Terlihat ukuran jalan sedang dari warna biru bernilai 0 sampai dengan merah bernilai 100.000 ada di kondisi Sedang.



Gambar 4. *Group By* Kondisi jalan Rusak Ringan

Gambar 4 merupakan grafik yang menggambarkan kondisi jalan rusak ringan. Dimana terlihat pada ukuran ringan warna biru bernilai 0 sampai dengan warna merah berwarna 100.000. Jalan dengan kondisi rusak ringan tercluster di dalam kondisi Rusak Ringan.



Gambar 5. *Group By* Kondisi jalan Rusak Berat

Kondisi jalan rusak berat dapat dilihat pada grafik gambar 5. Grafik yang berwarna merah menunjukkan bahwa semua data dapat tercluster ke dalam kondisi jalan rusak berat.

Untuk mengukur kemampuan system dalam mengidentifikasi permukaan jalan, penelitian ini akan menggunakan perhitungan *presisi*, *recall* dan *F-Measure* [8]. Pada penelitian ini menggunakan percobaan jalan di beberapa kecamatan wilayah Semarang. Dari percobaan penelitian ini menggunakan 638 titik jalan rusak yang tersebar di Kota Semarang. Hasil percobaan rapid miner telah berhasil mengelompokkan data kedalam kondisi baik, sedang, rusak ringan dan rusak berat. Berikut *matrix* dari hasil percobaan tersebut :

**Table 2.** Matrix Hasil Pengumpulan Data

Jenis	True	False
Kondisi Benar	478	160
Kondisi Salah	174	13

$$\text{Recall} = \frac{478}{478+174} = 0,733$$

$$\text{Precision} = \frac{478}{478+13} = 0,973$$

$$\text{F-Measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Precision} + \text{Recall}} = \frac{2 * 0,733 * 0,973}{0,733 + 0,973} = \frac{1,426}{1,706} = 0,836$$

Hasil nilai diatas menunjukkan nilai akurasi recall sebanyak 73%, presisi 97% dan menghasilkan F-Measure sebanyak 83% yang artinya bahwa sistem memiliki akurasi yang cukup tinggi dalam mengelompokkan kondisi jalan di Kota Semarang.

#### 4. Kesimpulan

Pada 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-Means. 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. Tingkat akurasi yang dihasilkan dari penerapan metode sebanyak 83% yang artinya metode K-Means cukup akurat dalam mengelompokkan kondisi jalan. 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

Kami tim penelitian mengucapkan terima kasih kepada Kemenristek Dikti yang telah mensupport penelitian kami melalui program hibah bersaing.

#### Referensi

- [1] N. D. Arumsari, A. L. Nugraha, dan M. Awaluddin, "Pemodelan Daerah Rawan Kecelakaan Dengan Menggunakan Cluster Analysis," *J. Geod. Undip*, vol. 5, hlm. 10, 2016.

- [2] S. Djaja, R. Widyastuti, K. Tobing, D. Lasut, dan J. Irianto, “Gambaran Kecelakaan Lalu Lintas Di Indonesia, Tahun 2010-2014,” vol. 15, no. 1, hlm. 13, 2016.
- [3] Yadi Utama, “Sistem Pendukung Keputusan Untuk Menentukan Prioritas Penanganan Perbaikan Jalan Menggunakan Metode Saw Berbasis Mobile Web,” *J. Sist. Inf.*, vol. 5, no. 1, hlm. 566–584, Apr 2013.
- [4] U. B. Shouvik Mani, Edgar Xi, “Intelligent Pothole Detection and Road Condition Assessment,” *Conference: Data for Good Exchange 2017, At New York, NY*, 2017.
- [5] A. Kulkarni, N. Mhalgi, S. Gurnani, dan D. N. Giri, “Pothole Detection System using Machine Learning on Android,” vol. 4, no. 7, hlm. 5, 2014.
- [6] S. Kumar dan D. Toshniwal, “A data mining approach to characterize road accident locations,” *J. Mod. Transp.*, vol. 24, no. 1, hlm. 62–72, Mar 2016.
- [7] P. Arora, Deepali, dan S. Varshney, “Analysis of K-Means and K-Medoids Algorithm For Big Data,” *Procedia Comput. Sci.*, vol. 78, hlm. 507–512, 2016.
- [8] A. Primajaya dan B. N. Sari, “Random Forest Algorithm for Prediction of Precipitation,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, hlm. 27, Mar 2018.