



Sistem Klasifikasi Kerusakan Jalan Metode Machine Learning dengan Algoritma K-Means dan Random Forest

<u>INFO PENULIS</u>	<u>INFO ARTIKEL</u>
Fahrim Irhamna Rahman Universitas Muhammadiyah Makassar Fahrim141020@unismuh.ac.id	ISSN: 3026-3603 Vol.3, No. 1 April 2025 http://jurnal.ardenjaya.com/index.php/ajst
Lukman Universitas Muhammadiyah Makassar lukman@unismuh.ac.id	
Hildayati Universitas Muhammadiyah Makassar waodehilda62@gmail.com	

© 2025 Arden Jaya Publisher All rights reserved

Saran Penulisan Referensi:

Rahman, F. I., Lukman., & Hildayati. (2025). Sistem Klasifikasi Kerusakan Jalan Metode Machine Learning dengan Algoritma K-Means dan Random Forest. *Arus Jurnal Sains dan Teknologi*, 3(1), 116-126.

Abstrak

Kerusakan jalan merupakan masalah yang signifikan dalam infrastruktur transportasi, yang dapat mempengaruhi kenyamanan dan keselamatan pengguna jalan. Penelitian ini bertujuan untuk mengklasifikasi kerusakan jalan berdasarkan data yang diambil dari survey lapangan dengan menggunakan algoritma k-means dan random forest. Metode *machine learning* dengan menggunakan algoritma k-means dan random forest dipilih karena kemampuannya dalam mengklasifikasi dan mengelompokkan data berdasarkan kedekatan jarak antar sampel, dan dapat dapat mempercepat proses pemeliharaan jalan. Data yang digunakan mencakup pengukuran kondisi jalan, antara lain panjang jalan dalam kondisi baik, sedang, rusak, dan rusak berat. Fitur-fitur ini dipilih untuk mempresentasikan tingkat kerusakan jalan yang sesuai dengan algoritma k-means dan random forest. Penelitian ini dimulai dengan pra-pemrosesan data dan pengelompokkan data menggunakan algoritma k-means, termasuk penanganan nilai yang hilang dan standarisasi fitur, dan diikuti dengan implementasi algoritma random forest untuk klasifikasi. Kinerja model dievaluasi menggunakan confusion matriks dan classification report, yang menunjukkan bahwa random forest mampu mengklasifikasi kerusakan jalan dengan akurasi yang benar. Kemudian hasil prediksi divalidasi dan disimpan untuk analisis lebih lanjut. Penelitian ini merupakan pengembangan model klasifikasi yang mendukung pengambilan keputusan yang terkait pemeliharaan dan perbaikan jalan secara efektif dan efisien. Penelitian ini berhasil membangun system klasifikasi kerusakan jalan dengan menggabungkan algoritma K-Means untuk pengelompokkan dan Random Forest untuk klasifikasi. Model tersebut dapat memberikan akurasi yang tinggi dalam mengidentifikasi tingkat kerusakan jalan, yang dapat mendukung upaya pemeliharaan jalan yang lebih efektif.

Kata Kunci: Kerusakan Jalan, Klasifikasi, Machine Learning, K-Means, Random Forest

Abstract

Road damage is a significant problem in transportation infrastructure, which can affect the comfort and safety of road users. This study aims to classify road damage based on data taken from field surveys using the k-means and random forest algorithms. The machine learning method using the k-means and random forest algorithms was chosen because of its ability to classify and group data based on the proximity of the distance between samples, and can accelerate the road maintenance process. The data used includes measurements of road conditions, including the length of roads in good, moderate, damaged, and severely damaged conditions. These features are selected to present the level of road damage that is in accordance with the k-means and random forest algorithms. This study began with data pre-processing and data grouping using the k-means algorithm, including handling missing values and feature standardization, and was followed by the implementation of the random forest algorithm for classification. Model performance was evaluated using a confusion matrix and classification report, which showed that random forest was able to classify road damage with correct accuracy. Then the prediction results were validated and stored for further analysis. This study is the development of a classification model that supports decision making related to road maintenance and repair effectively and efficiently. This study successfully built a road damage classification system by combining the K-Means algorithm for clustering and Random Forest for classification. The model can provide high accuracy in identifying the level of road damage, which can support more effective road maintenance efforts.

Keywords: Road Damage, Classification, Machine Learning, K-Means, Random Forest

A. Pendahuluan

Perkembangan pada teknologi dalam mengelola infrastruktur jalan terus mengalami kemajuan, terutama dalam pemanfaatan machine learning untuk mengklasifikasi dan menganalisis kondisi jalan. Kondisi jalan yang tidak terjaga sangat berdampak negative bagi kenyamanan pengguna jalan. Oleh karena itu mengidentifikasi dan mengklasifikasi kondisi jalan yang akurat sangat penting dalam program pemeliharaan yang lebih efisien. Pada penelitian ini, menggunakan algoritma K-Means untuk melakukan pengelompokan data berdasarkan kesamaan pola, sedangkan Random Forest digunakan untuk memprediksi tingkat kerusakan berdasarkan atribut yang relevan. Kombinasi dari kedua metode ini memberikan pendekatan yang komprehensif dalam menganalisis data kondisi jalan.

Kerusakan jalan merupakan salah satu infrastruktur yang sering ditemukan disetiap Negara, termasuk di Indonesia. Jalan memiliki peran sebagai fasilitas yang diperuntukan untuk lalu lintas umum, dan dapat menunjang pertumbuhan ekonomi masyarakat dikarenakan mobilitas yang tinggi. Namun, faktanya kerusakan jalan masih menjadi salah satu masalah yang belum dapat diselesaikan serta menjadi masalah yang serius. Kerusakan jalan dapat disebabkan oleh beberapa hal, diantaranya beban volume lalu lintas yang tinggi dan berulang, cuaca yang ekstrim, serta konstruksi jalan yang kurang layak.

Kerusakan jalan juga dapat mengganggu pertumbuhan ekonomi dan mobilitas masyarakat, serta menyebabkan masalah yang serius terhadap pengguna jalan, salah satunya adalah kecelakaan kendaraan. Langkah awal dari pemeliharaan jalan adalah dengan mengklasifikasi kerusakan pada suatu jalan, sehingga dapat menentukan tindakan apa yang perlu dilakukan. Metode yang digunakan untuk mengklasifikasi/mengidentifikasi kondisi kerusakan jalan yaitu menggunakan metode machine learning dengan algoritma k-means dan random forest.

algoritma K-Means berhasil mengelompokkan kondisi jalan berdasarkan Panjang jalan yang berada dalam kondisi baik, sedang, rusak, dan rusak berat. Algoritma K-Means ini memungkinkan pengelompokan data jalan berdasarkan kemiripan fitur, sehingga jalan dengan karakteristik yang serupa akan dikelompokkan. Model Random Forest dilakukan menggunakan metrik evaluasi seperti *confusion matrix* dan *classification report*, yang mencakup *precision*, *recall*, *F1-score*, dan *akurasi*. Model secara keseluruhan memiliki kinerja yang sangat baik dalam mengklasifikasi data ke kategori atau prediksi yang benar.

1. Tinjauan Pustaka

A. Machine Learning

1. Pengertian machine learning

Machine learning adalah serangkaian algoritma komputer yang digunakan untuk mengoptimalkan kinerja komputer atau sistem berdasarkan data sampel yang ada. Kemampuan utama dari machine learning adalah modifikasi dan adaptasi keputusan dalam merespon perubahan. Kegunaan dari machine learning antara lain sebagai berikut:

- a. klasifikasi adalah metode *machine learning* yang digunakan untuk memprediksi suatu nilai/kelas sebuah individu dalam sebuah populasi.
 - b. pencocokan kemiripan adalah metode *machine learning* yang digunakan untuk mengidentifikasi kemiripan antar individu berdasarkan data yang ada.
 - c. pengklasteran adalah metode *machine learning* yang digunakan untuk mengelompokkan individu dalam grup yang sama berdasarkan kesamaan yang dimiliki.
2. kelemahan dan kekurangan *machine learning*

Adapun kelebihan dan kekurangan *machine learning*:

Kelebihan *machine learning*, Pola dan tren data dapat lebih mudah untuk diidentifikasi. Teknologi *machine learning* menjadi salah satu teknologi yang digunakan untuk mengolah data yang berukuran besar, termasuk untuk menemukan pola dan tren big data tersebut dalam waktu yang relatif singkat.

Kelemahan *machine learning*, Rentan terjadi kesalahan. Tentunya ada banyak hal yang mungkin tidak sesuai dengan harapan dari developer ketika mesin dibiarkan bekerja sendirian untuk menggantikan manusia. Selain itu, untuk proses diagnosis kesalahan yang terjadi, ternyata cukup memakan waktu.

B. Algoritma K-Means

Algoritma *K-Means* merupakan salah satu algoritma dengan partitional, karena *K-Means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Algoritma *K-Means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan titik *centroid* akhir sebagai output. Metode *K-Means* akan memilih pola *k* sebagai titik awal *centroid* secara acak atau random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan *centroid cluster* yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi.

Dalam penyelesaiannya, algoritma *K-Means* akan menghasilkan titik *centroid* yang dijadikan tujuan dari algoritma *K-Means*. Setelah iterasi *K-Means* berhenti, setiap objek dalam dataset menjadi anggota dari suatu *cluster*. Nilai *cluster* ditentukan dengan mencari seluruh objek untuk menemukan *cluster* dengan jarak terdekat ke objek. Algoritma *K-Means* akan mengelompokkan item data dalam suatu dataset ke suatu *cluster* berdasarkan jarak terdekat.

C. Random forest

Random forest merupakan sebuah metode ensemble, yang aman metode ensemble merupakan cara untuk meningkatkan akurasi metode klasifikasi dengan mengkombinasikan metode klasifikasi dari sebuah pemilah tunggal yang tidak stabil melalui banyak kombinasi penyaringan dari suatu metode yang sama dengan proses keputusan (*voting*) untuk memperoleh prediksi klasifikasi akhir.

Random forest diawali dengan teknik dasar dari data mining yaitu *decision tree*. Pada proses *decision tree*, Dimana input berupa data akan dimasukkan pada bagian atas proses *tree* berupa akar pohon (*root*) kemudian akan dibawa turun kebagian bawah akan berupa daun pohon (*leaf*) pada proses, untuk menentukan data *input*-an tersebut termasuk kedalam kelas apa pada proses.

Pada *decision tree* menggunakan *information gain* dan *gini index* untuk perhitungan dalam menentukan *root node* dan *rule*. Sama halnya dengan *random forest* yang akan menggunakan *information gain* dan *gini index* untuk perhitungan dalam membangun *tree*, hanya saja *random forest* akan membangun lebih dari satu *tree*. Masing-masing *tree* dibangun menggunakan dataset dengan atribut atau variable yang diambil secara acak (random) dari data training, yang mana setiap *tree* akan bergantung pada nilai dari sampel *vector* yang bebas (*independent*) dengan pendistribusian yang sama pada setiap *tree* yang dibangun. Selama proses klasifikasi setiap *tree* akan memberikan *voting* kelas yang paling populer.

D. Klasifikasi

Klasifikasi adalah pengelompokan data kedalam kategori atau kelas tertentu berdasarkan fitur atau ciri-ciri yang ada pada data tersebut. *classification* adalah suatu Teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah diklasifikasikan. Penggunaan Teknik ini untuk memberikan klasifikasi pada data baru dengan cara melakukan manipulasi data yang telah diklasifikasikan dan menggunakan hasilnya untuk sejumlah aturan. Tujuan dari klasifikasi adalah untuk memprediksi atau mengidentifikasi kategori atau label yang tepat untuk data yang belum diketahui.

E. Kerusakan jalan

Jalan merupakan infrastruktur yang sangat penting dalam kehidupan sehari-hari. Namun, jalan yang rusak dapat menyebabkan berbagai masalah, seperti kecelakaan, keterlambatan, dan biaya yang tinggi untuk perawatan. Kerusakan jalan dapat disebabkan oleh berbagai factor seperti retak, retak Lelah dan deformasi, lubang-lubang, kegemukkan, striping, dan lain-lain. Dalam beberapa tahun terakhir, kerusakan jalan telah menjadi masalah yang semakin parah di Indonesia, terutama di daerah-daerah yang memiliki intensitas penggunaan jalan yang sangat tinggi.

Kerusakan jalan merupakan salah satu masalah infrastruktur yang sering ditemukan disetiap Negara, termasuk di Indonesia. Kerusakan jalan dapat disebabkan oleh beberapa hal, diantaranya beban folume lalu lintas yang tinggi dan berlubang, cuaca ekstrim, serta kontruksi jalan yang kurang layak

F. Jenis kerusakan jalan

Secara garis besar kerusakan jalan dapat dibedakan menjadi dua bagian, yaitu kerusakan structural dan fungsional. Kerusakan structural yaitu kerusakan yang mencakup kegagalan perkerasan atau kerusakan dari satu atau lebih komponen perkerasan. Kerusakan jenis ini mengakibatkan perkerasan tidak dapat lagi menanggung beban lalu lintas. Sedangkan kerusakan fungsional yaitu kerusakan yang mengakibatkan keamanan dan kenyamanan pengguna jalan terganggu.

B. Metodologi

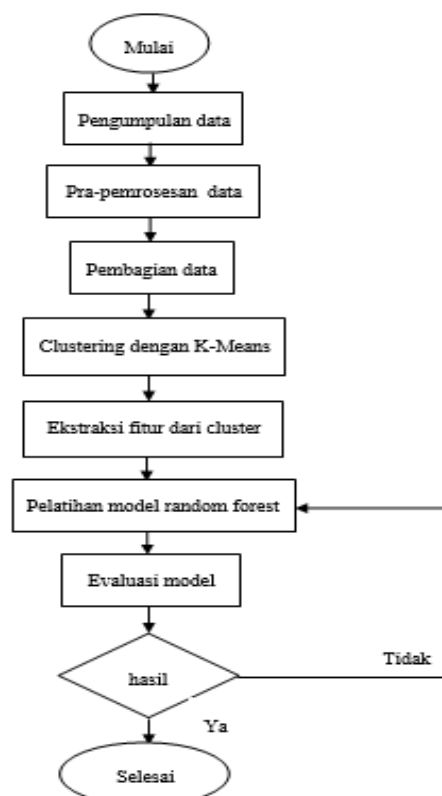
a. Waktu dan Tempat Penelitian

Tempat penelitian merupakan suatu tempat atau objek yang akan dilakukan suatu penelitian. Penentuan lokasi penelitian merupakan langkah penting dalam proses penelitian karena memudahkan peneliti untuk melakukan penelitian. Lokasi pengambilam data pada penelitian ini dilakukan di kecamatan tompobulu, kab. Bantaeng. Waktu kegiatan penelitian dimulai pada bulan agustus 2024 sampai semua proses pengumpulan data selesai.

b. Perancangan Sistem

Dalam membuat suatu sistem diperlukan perancangan sistem, perancangan sistem sangat penting dalam membangun suatu sistem karena menguraikan bagaimana suatu sistem dibangun dari tahap perencanaan hingga tahap pembuatan fungsi-fungsi yang diperlukan untuk pengoperasian sistem. Tujuan dari perancangan sistem yaitu untuk menentukan apakah sistem yang akan dibuat atau dikembangkan dapat menghasilkan hasil yang di inginkan.

c. Flowchart Diagram



Gambar 2.1 Flowchart diagram

Perancangan sistem ini dimulai dari login, kemudian pengumpulan data, setelah itu preprocessing data, selanjutnya pembagian data dimana data dibagi menjadi set pelatihan dan set pengujian, sesudah melakukan pembagian data algoritma K-Means akan mengelompokkan data kedalam beberapa cluster berdasarkan fitur, kemudian hasil clustering digunakan untuk menambah label cluster kedata asli sebagai fitur tambahan, dan dilanjutkan dengan pelatihan model random forest menggunakan set pelatihan yang telah diperkaya dengan fitur clustering, kemudian dilanjutkan dengan evaluasi model menggunakan set pengujian untuk mengukur akurasi, presisi, recall, dan lainnya, jika hasil evaluasi model “Ya” proses akan melanjutkan ke hasil untuk menunjukkan bahwa model telah mencapai tujuan yang diinginkan. Sebaliknya jika hasil evaluasi model “Tidak” proses akan kembali ke langkah pelatihan model random forest untuk perbaikan lebih lanjut. selanjutnya hasil klasifikasi digunakan untuk tindakan perbaikan yang sesuai dan kemudian hasil klasifikasi dan tindakan perbaikan di dokumentasikan.

D. Hasil dan Pembahasan

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data angka yang diperoleh dari Dinas PU, Kabupaten Bantaeng. Dataset ini memuat informasi mengenai kondisi jalan berbagai ruas, yang meliputi Panjang tiap kondisi baik, sedang, rusak, dan rusak berat.

Tabel 1. Data Mentah

No Ruas	Nama Ruas Jalan	Kecamatan Yang Dilalui	Panjang Ruas (Km)	Lebar Ruas (M)	Panjang Tiap Kondisi							
					Baik		Sedang		Rusak Ringan		Rusak Berat	
					KM	%	KM	%	KM	%	KM	%
730301	Nipa-Nipa - Banyorang	Pa'jukukang/G. keke/Tompobulu	13,60	5,00	7,200	52,94	5,200	38,24	1,200	8,82	0	0
730302	Banyorang - Bungeng	Tompobulu	7,000	4,00	2,000	28,57	2,500	35,71	1,500	21,43	1,00	14,29
730303	Ereng - Ereng - Labbo	Tompobulu	3,300	4,00	1,900	57,58	0,200	6,06	0,900	27,27	0,300	9,09
730304	Labbo - Panjang	Tompobulu	4,600	4,00	2,800	60,87	1,100	23,91	0,700	15,22	0	0
730305	Bannyorang - Campaga	Timobulu	3,700	4,00	1,700	45,95	1,000	27,03	0,700	18,92	0,300	8,11
730306	Banyorang - Tarico	Tompobulu	2,000	4,00	2,000	100,00	0	0	0	0	0	0
...
730388	Panjang - Benteng Bola'	Tompobulu	2,700	3,50	0,800	29,63	0,300	11,11	1,200	44,44	0,400	14,81
730389	Jambi - Benteng Bola'	Eremerasa/tompobulu	2,300	4,00	1,300	21,74	0,200	8,70	0,500	21,74	1,100	47,83
730390	SP.parangloe - Durian	Eremerasa	2,950	4,00	2,950	69,49	0,400	13,56	0,400	13,56	0,100	3,39
730391	Dalam Kota Lonrong	Eremerasa	1,500	4,00	1,500	100,00	0	0	0	0	0	0
730392	Dalam Kota Kampala	Eremerasa	1,500	4,00	1,500	66,67	0,400	26,67	0	0	0,100	6,67
730393	Pa'bulengan - Batu Langgayya	Sinoa	2,650	3,50	2,650	37,74	0,500	18,87	0,300	11,32	0,850	32,08

2. Ekstraksi dan Pengaturan Data

Program ini memuat dan memproses data dari file excel menggunakan beberapa pustaka seperti *pandas*, *numpy*, *sklearn*, *matplotlib*, dan *seaborn*. Dimana data dimuat dan diperiksa sebelum diterapkan Teknik analisis seperti klastering dengan k-means.

Table 2. Mengolah Data

No Ruas	Nama Ruas Jalan	Kecamatan yang Dilalui	Panjang Ruas (Km)	Lebar Ruas (Km)	Jenis Perkerasan (Km) Hotmix	...	Panjang Tiap Kondisi Sedan (%)	Panjang Tiap Kondisi Rusak (Km)	Panjang Tiap Kondisi Ringan (%)	Panjang Tiap Kondisi Rusak Berat (Km)	Panjang Tiap Kondisi Rusak Berat (%)	LHR	Akses Ke N/P/K
730301	Nipa-Nipa - Banyorang	Pa'jukukang /G/Keke/Tompobulu	13.6	5.0	13.6	...	38.24	1.2	8.82	0.0	0.00	1514	N
730302	Banyorang - Bungeng	Tompobulu	7.0	4.0	7.0	...	35.71	1.5	21.43	1.0	14.29	984	K
730303	Ereng - Ereng - Labbo	Tompobulu	3.3	4.0	3.3	...	6.06	0.9	27.27	0.3	9.09	545	K
730304	Labbo - Panjang	Tompobulu	4.6	4.0	4.6	...	23.91	0.7	15.22	0.0	0.00	181	K
730305	Banyorang - Campaga	Tompobulu	3.7	4.0	3.7	...	27.03	0.7	18.92	0.3	8.11	417	K

K-means mengelompokkan data ke dalam kluster berdasarkan kedekatannya dengan pusat kluster (centroid). Proses terdiri dari beberapa langkah yaitu:

1. Inisialisasi centroid : Tentukan kluster $K=4$, maka memilih 4 pusat kluster secara acak dari data, Pusat kluster ini disebut centroid dan akan diperbarui dalam setiap iterasi.
2. Penugasan data ke kluster : Setiap titik data akan diberikan ke kluster terdekat, yang dihitung menggunakan jarak Euclidean antara titik data dan pusat kluster. Rumus jarak Euclidean antara dua titik $x = (x_1, x_2, \dots, x_n)$ dan (y_1, y_2, \dots, y_n) adalah:

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

Disini menghitung jarak antara data dan setiap centroid (pusat kluster).

Perbarui centroid : Setelah semua titik data diberikan ke kluster terdekat, centroid baru untuk setiap kluster dihitung dengan cara mengambil rata-rata dari semua titik data yang ada dalam kluster tersebut. Rumus untuk menghitung centroid baru:

$$C_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_i$$

Dimana:

C_k Adalah centroid baru untuk kluster k ,

N_k Adalah jumlah data dalam kluster k ,

X_i Adalah data pada kluster k .

Langkah perhitungan dari program:

1. Inisialisasi centroid: program menginisialisasi 4 kluster menggunakan K-Means ($n_clusters=4$) dan memilih centroid secara acak di awal.
2. Penugasan data ke kluster: program menghitung jarak Euclidean antara titik data dan pusat kluster yang telah diinisialisasi. Setiap titik data akan diberikan ke kluster terdekat.
3. Perbarui centroid: setelah penugasan data selesai, program menghitung centroid baru untuk masing-masing kluster berdasarkan rata-rata data yang ada dalam kluster tersebut.
4. Konvergensi: iterasi ini di ulang hingga posisi centroid tidak berubah signifikan (konvergen).
Dalam program ini, K-Means dari *sklearn* menangani perhitungan tersebut secara otomatis menggunakan implementasi optimasi K-Means yang efisien yaitu:
 - a. *Fit predict ()* akan melakukan perhitungan jarak dan penugasan data ke kluster.
 - b. *Cluster_centers_* memberikan posisi centroid setelah konvergensi.
 - c. *Labels_* memberikan informasi kluster yang diperoleh untuk setiap titik data.

3. Pra-Pemrosesan Data

Pra-pemrosesan data merupakan langkah penting untuk memastikan data siap digunakan dalam model klasifikasi, termasuk memuat data, memilih fitur yang relevan, menangani nilai hilang, memisahkan fitur dan label, membagi data menjadi set pelatihan dan pengujian serta menstandarisasi fitur.

4. Klustering dengan K-Means

Algoritma K-means digunakan untuk mengelompokkan data berdasarkan kondisi jalan. Jumlah kluster yang dipilih adalah 4, yang mewakili kategori kondisi jalan: Baik-Baik Saja, Rusak Kecil, Rusak Sedang, dan Rusak Parah.

Table 3. Pemetaan Kluster kedalam Label

NO	RUAS	NAMA RUAS JALAN	Kluster
0	730301	Nipa-Nipa - Bannyorang	Rusak Sedang
1	730302	Bannyorang - Bungeng	Rusak Kecil
2	730303	Ereng-Ereng - Labbo	Rusak Kecil
3	730304	Labbo - Panjang	Rusak Kecil
4	730305	Bannyorang - Campaga	Rusak Kecil

Hasil klustering yang telah diproses disimpan ke dalam file excel baru. Memungkinkan data tersebut untuk digunakan kembali untuk analisis lebih lanjut dilain waktu. Dengan menyimpan hasil klustering ke file excel baru, program ini memastikan bahwa hasil analisis dapat dengan mudah diakses dan dibagikan untuk digunakan sebagai pelaporan, pemantauan kondisi jalan, dan analisis lanjut.

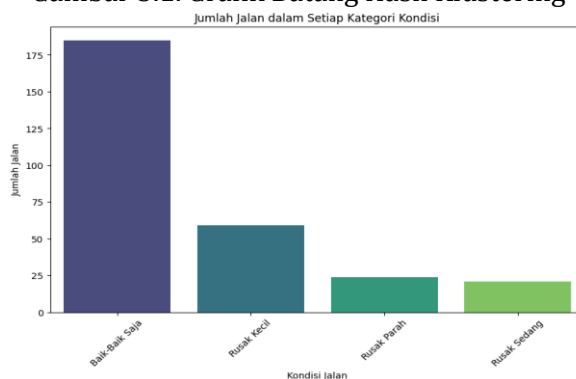
5. Analisis Hasil Klustering

Dalam analisis hasil klustering ini menjelaskan bagaimana menampilkan grafik batang untuk memvisualisasikan distribusi jumlah jalan berdasarkan kategori kondisi (klaster). Dengan menganalisis hasil klustering dan mem visualisasikan jumlah jalan dalam setiap kategori kondisi, program ini memeberikan pemahaman yang lebih jelas tetang distribusi atau pembagian kondisi jalan.

Tabel 4. Hasil Klustering

Nama Ruas Jalan	Kecamatan yang Dilalui	Panjang Tiap Kondisi Sedang (%)	Panjang Tiap Kondisi Rusak (Km)	Panjang Tiap Kondisi Ringan (%)	Panjang Tiap Kondisi Rusak Berat Km)	Kluster
Nipa-Nipa - Banyoraning	Pa'jukukan / G.keke / Tompobulu	38,24	1,2	8,82	0,0	Rusak Sedang
Banyoraning - Bungeng	Tompobulu	35,71	1,5	21,43	1,0	Rusak Kecil
Ereng-Ereng - Labbo	Tompobulu	6,06	0,9	27,27	0,3	Rusak Kecil
Labbo Panjang	Tompobulu	23,91	0,7	15,22	0,0	Rusak Kecil
Banyoraning - campaga	Tompobulu	23,03	0,7	18,02	0,3	Rusak Kecil

Gambar 3.1. Grafik Batang Hasil Klustering



Output tersebut menunjukkan hasil klustering dari data jalan yang telah disimpan kedalam file Excel *Hasil_Klustering.xlsx*. berikut penjelasan dari masing-masing kluster:

1. Baik-Baik Saja: kluster ini berisi 185 data yang dikategorikan dalam kondisi baik.
2. Rusak Kecil: kluster ini berisi 59 data yang memiliki kerusakan jalan.
3. Rusak Parah: kluster ini berisi 24 data yang menunjukkan kerusakan parah.
4. Rusak Sedang: kluster ini berisi 21 data yang berada dalam kategori kerusakan sedang.

Dari hasil klustering menunjukkan bahwa sebagian besar ruas jalan yang dikategorikan 'Baik-Baik Saja' berisi jumlah jalan paling banyak yang tidak memiliki masalah besar atau hanya memiliki sedikit kerusakan. Sedangkan jumlah ruas jalan dengan kondisi 'Rusak Kecil', 'Rusak Parah', dan 'Rusak Sedang' lebih sedikit dibandingkan dengan kategori 'Baik-Baik Saja'.

6. Klasifikasi dengan Random Forest

Setelah data jalan dikelompokkan, tahap selanjutnya adalah menggunakan algoritma random forest untuk membuat klasifikasi berdasarkan hasil klaster tersebut. Klasifikasi ini bertujuan untuk mempelajari fitur yang ada dan mengklasifikasikan data ke dalam kategori kerusakan. Pra-pemrosesan data merupakan langkah penting yang melibatkan serangkaian proses untuk membersihkan dan mempersiapkan data agar siap digunakan untuk pelatihan model klasifikasi. Tahapan ini termasuk memuat data, memilih fitur yang relevan, menangani nilai hilang, memisahkan fitur dan label, membagi data menjadi set pelatihan dan pengujian.

Algoritma random forest adalah model pembelajaran mesin yang menggunakan *ensemble*, yaitu menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi yang lebih

baik dan stabil. $n_estimators=100$: menentukan jumlah pohon keputusan yang akan dibangun dalam random forest yaitu 100 pohon akan digunakan. Semakin banyak pohon, semakin baik model dalam menangani overfitting dan meningkatkan akurasi, tetapi dengan komputasi yang lebih tinggi. $random_state=42$: menentukan seed untuk menghasilkan hasil yang konsisten setiap kali menjalankan kode ini. Kemudian melatih model random forest menggunakan data pelatihan (x_train untuk fitur dan y_train untuk label). Proses ini akan membangun sejumlah pohon keputusan berdasarkan data yang ada. Setelah model dilatih dengan data pelatihan, kemudian menggunakan data pengujian (X_test) untuk membuat prediksi. *Predict* akan menghasilkan prediksi untuk setiap sampel dalam data pengujian. Kemudian Hasilnya disimpan dalam y_pred , yang merupakan label yang diprediksi oleh model untuk data pengujian.

Langkah-langkah untuk Random Forest:

1. Pembentukan pohon keputusan:
 - a. Random forest terdiri dari banyak pohon keputusan yang dibangun menggunakan subset acak dari data pelatihan.
 - b. Setiap pohon keputusan dibangun berdasarkan Teknik bagging, yang berarti setiap pohon dibangun menggunakan subset data yang dipilih secara acak dengan pengganti (bootstrap sampling).
2. Split pada pohon keputusan:

Setiap pohon keputusan menggunakan kriteria untuk membagi data di setiap simpul (node). Salah satu kriteria yang paling sering digunakan adalah gini impurity atau entropy.

Rumus gini impurity: digunakan untuk mengukur ketidakmurnian suatu simpul. Nilai gini berada antara 0 (sudah murni, semua data pada simpul tersebut berada dalam satu kelas) dan 1 (semua data dalam simpul berasal dari kelas yang berbeda).

Evaluasi model adalah proses untuk mengukur seberapa baik sebuah model pembelajaran mesin dalam memprediksi kelas target pada data yang belum pernah dilihat. Dalam klasifikasi ini, kita menggunakan dua metrik untuk mengevaluasi kinerja model yaitu confusion matrix dan classification report.

7. Penyimpanan Hasil Prediksi

Setelah model random forest digunakan untuk prediksi dan validasi kondisi jalan, hasil prediksi dan validasi disimpan ke dalam file Excel baru untuk analisis lebih lanjut. File *Hasil_Prediksi.xlsx*: berisi data asli (*data*) dengan kolom tambahan prediksi kondisi yang berisi hasil prediksi dari model. File *Hasil_Validasi.xlsx*: berisi dua kolom, yaitu *True_Condition* (label yang benar dari data pengujian) dan *Predicted_Condition* (hasil prediksi oleh model).

Hasil prediksi pada file *hasil_prediksi.xlsx* berisi kluster yang diprediksi oleh model Random Forest untuk setiap ruas jalan dalam dataset. Prediksi kondisi jalan merupakan kolom yang menunjukkan kluster hasil prediksi yang berkaitan dengan tingkat kerusakan jalan.

Tabel 5. Hasil Prediksi

No Ruas	Nama Ruas Jalan	Kecamatan yang Dilalui	...	LHR	Kluster	Prediksi Kondisi
730301	Nipa-Nipa Bannyorang	- Pa'jakukang/ G\nKeke/ Tompobulu	...	1514	Rusak Sedang	Rusak Sedang
730302	Bannyorang Bungeng	- Tompobulu	...	984	Rusak Kecil	Rusak Kecil
730303	Ereng-Ereng Labbo	- Tompobulu	...	545	Rusak Kecil	Rusak Kecil
730304	Labbo - Panjang	Tompobulu	...	181	Rusak Kecil	Rusak Kecil
...
7303191	Panjang - Biringere	Tompobulu	...	96	Rusak Parah	Rusak Parah
7303192	Labbo - Borong Inru	Tompobulu	...	38	Rusak Kecil	Rusak Kecil
7303193	Panjang - Borong Inru	Tompobulu	...	167	Baik-Baik Saja	Baik Saja

True_Condition	Predicted_condition
Baik-Baik Saja	Baik-Baik Saja
Rusak Kecil	Baik-Baik Saja
Baik-Baik Saja	Baik-Baik Saja
...	...
Baik-Baik Saja	Baik-Baik Saja
Rusak-Rusak Kecil	Rusak Kecil
Rusak Kecil	Rusak Kecil

Tabel 4. Hasil Validasi

Hasil validasi pada file Hasil_Validasi.xlsx berisi perbandingan antara label asli (dari data pengujian, yaitu y_{test}) dan hasil prediksi (dari model, yaitu y_{pred}). *True Condition* merupakan label asli dari data pengujian, yaitu nilai sebenarnya yang menunjukkan kluster dan kondisi jalan yang sesungguhnya berdasarkan data yang diketahui. Kolom ini berisi nilai yang diambil dari y_{test} . Sedangkan *Predicted Condition* adalah hasil prediksi dari model untuk data pengujian. Kolom ini berisi nilai prediksi yang dibuat oleh model random Random Forest setelah dilatih menggunakan data pelatihan. Kolom ini diisi dengan nilai dari y_{pred} .

Hasil evaluasi dari model klasifikasi kerusakan jalan yang telah di uji:

Confusion Matrix:

1. Baik-Baik Saja:
 - a. Prediksi Benar: 39
 - b. Prediksi salah: tidak ada.
2. Rusak Kecil:
 - a. Prediksi Benar: 11
 - b. Prediksi salah: (1 prediksi yang salah menjadi 'Baik-Baik Saja').
3. Rusak Parah:
 - a. Prediksi Benar: 3
 - b. Prediksi Salah: (1 prediksi yang salah menjadi 'Rusak Kecil').
4. Rusak Sedang:
 - a. Prediksi Benar: 2
 - b. Prediksi Salah: (1 prediksi yang salah menjadi 'Rusak Kecil').

Classification Report:

Classification report memberikan metrik evaluasi seperti *precision*, *recall* dan *f1-score* untuk masing-masing kelas.

1. *Precision:*
Prediksi positif yang benar:
 - a. Baik-Baik Saja: 97%
 - b. Rusak Kecil: 85%
 - c. Rusak Parah: 100%
 - d. Rusak Sedang: 100%
2. *Recall:*
Kasus positif yang terdeteksi model:
 - a. Baik-Baik Saja: 100%
 - b. Rusak Kecil: 92%
 - c. Rusak Parah: 75%
 - d. Rusak Sedang: 67%
3. *F1-Score:*
Rata-rata harmonic dari precision dan recall:
 - a. Baik-Baik Saja: 0.99
 - b. Rusak Kecil: 0.88
 - c. Rusak Parah: 0,86
 - d. Rusak Sedang: 0.80
4. Akurasi (*Accuracy*):

Model secara keseluruhan memiliki kinerja yang sangat baik dalam mengklasifikasi data ke kategori atau prediksi yang Benar.

- a. *Accuracy* atau akurasi keseluruhan: 0.95 (atau 95).
5. *Macro avg:*
 - a. *Macro avg* menunjukkan bahwa secara keseluruhan, model bekerja dengan baik, tetapi ada

beberapa kelas yang lebih sulit diprediksi (seperti kelas dengan data lebih sedikit), yang terlihat dari recall yang lebih sedikit lebih rendah (0.83).

6. *Weight avg*

Weight avg menunjukkan bahwa model sangat baik dalam hal precision, recall, dan f1-score secara keseluruhan karena mempertimbangkan jumlah data di setiap kelas.

E. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi kerusakan jalan menggunakan algoritma k-means dan random forest, beberapa kesimpulan yang dapat diambil adalah sebagai berikut:

1. Dalam penelitian ini, algoritma K-Means berhasil mengelompokkan kondisi jalan berdasarkan Panjang jalan yang berada dalam kondisi Baik, Sedang, Rusak, Dan Rusak Berat. Algoritma K-Means ini memungkinkan pengelompokkan data jalan berdasarkan kemiripan fitur, sehingga jalan dengan karakteristik yang serupa akan dikelompokkan. Setelah klustering dilakukan, model Random Forest mengklasifikasi kluster-kluster tersebut kedalam kategori yang lebih spesifik, seperti "Baik", "Rusak Kecil", "Rusak Sedang," dan "Rusak Parah". Random forest memberikan prediksi kategori kondisi jalan berdasarkan kluster yang sudah terbentuk sebelumnya.
2. Kinerja model Random Forest dilakukan menggunakan metrik evaluasi seperti Confusion Matrix dan Classification Report, yang mencakup Precision, Recall, F1-score, dan akurasi. Model secara keseluruhan memiliki kinerja yang sangat baik dalam mengklasifikasi data ke kategori atau prediksi yang Benar, akurasi keseluruhan yaitu 0,95 (atau 95). Random Forest menunjukkan kemampuan yang baik dalam mengklasifikasikan kondisi jalan sesuai dengan kluster yang telah dibentuk. Metrik-metrik ini memberikan gambaran yang jelas tentang kualitas prediksi model, serta area yang mungkin memerlukan perbaikan (misalnya, apabila ada ketidak seimbangan dalam jumlah data antar kelas).

F. Saran

Penelitian lebih lanjut dapat mengeksplorasi metode lain yang mungkin lebih efektif untuk meningkatkan kinerja. Dan akan lebih baik jika kedepannya seseorang melakukan pengembangan. Peningkatan dalam pra-pemrosesan data dan fitur engineering juga dapat membantu meningkatkan performa model.

G. Referensi

- (Asmiatun et al., 2019)Aini, N., Arif, M., Agustin, I. T., & Toyibah, Z. B. (2024). Implementasi Algoritma Random Forest untuk Klasifikasi Bidang MSIB di Prodi Pendidikan Informatika. *Jurnal Informatika*, 11(1), 11-16. <https://doi.org/10.31294/inf.v11i1.20637>
- Algorithm, M. C. (2012). *Teori K Means*. 2010, 7-24.
- Asmiatun, S., Wakhidah, N., Putri, A. N., & Kunci, K. (2019). *Identifikasi Kondisi Permukaan Jalan Menggunakan K-Means Clustering Road Surface Conditions Identification Using K-Means Clustering*. November 2019, 23-30.
- Dawkins, R. (2019). Bab Ii Tinjauan Pustaka Dan Dasar Teori 2.1. *Uajy*, 2013, 8-11. <http://e-journal.uajy.ac.id/>
- Fadil Danu Rahman, Mulki, M. I. Z., & Taryana, A. (2024). Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest. *Jurnal SINTA: Sistem Informasi Dan Teknologi Komputasi*, 1(2), 90-97. <https://doi.org/10.61124/sinta.v1i2.15>
- García Reyes, L. E. (2013). Cara Kerja Artificial Neural Network. *Journal of Chemical Information and Modeling*, 53(9), 1689-1699.
- Karyadiputra, E., Setiawan, A., Rizal Abdi, M., & Melanda Febriyanti, R. (2023). Implementasi Data Mining K-Means Clustering Dalam Penentuan Prioritas Perbaikan Jalan Berbasis Web. *Smart Comp: Jurnalnya Orang Pintar Komputer*, 12(4). <https://doi.org/10.30591/smartcomp.v12i4.5685>
- Margolang, K. F., Zarlis, M., & Hartono, H. (2023). Sentiment Classification on Mandalika MotoGP Event Using K-Means Clustering and Random Forest. *International Conference on Information Science and Technology Innovation (ICoSTEC)*, 2(1), 65-69. <https://doi.org/10.35842/icostec.v2i1.39>

- Pramestya, R. H. (2018). Deteksi dan Klasifikasi Kerusakan Jalan Aspal Menggunakan Metode YOLO Berbasis Citra Digital. *Institut Teknologi Sepuluh Nopember*, 91. http://repository.its.ac.id/59044/1/06111650010019-Master_Thesis.pdf
- Rahmanto, A. (2016). Evaluasi Kerusakan Jalan Dan Penanganan Dengan Metode Bina Marga Pada Ruas Jalan Banjarejo - Ngawen. *Simetris*, 10(1), 17–24.
- Ramadhani, D. N., Harjanta, A. T., Tyogi, A., & Ajhara, S. A. (2022). Implementasi Metode K-Means Pada Sistem Informasi Geografis Pemetaan Kerusakan Jalan di Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Jepara. *Science And Engineering National Seminar*, 7(7).
- Risatayn, H., Ekojono, E., & Suryani Hormansyah, D. (2024). Metode Random Forest Untuk Klasifikasi Kerusakan Transformator Daya Berdasarkan Gas Terlarut Pada Duval Triangle Dan Duval Pentagon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3464–3471. <https://doi.org/10.36040/jati.v7i6.7743>
- Suci Amaliah, Nusrang, M., & Aswi, A. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 121–127. <https://doi.org/10.35580/variansiunm31>
- Suprawoto, T. (2016). Klasifikasi Data Mahasiswa Menggunakan Metode K-Means Untuk Menunjang Pemilihan Strategi Pemasaran. *JIKO (Jurnal Informatika Dan Komputer)*, 1(1), 12–18. <https://doi.org/10.26798/jiko.2016.v1i1.9>
- Suryani, T., Faisol, A., & Vendyansyah, N. (2021). Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 380–388. <https://doi.org/10.36040/jati.v5i1.3259>