



A Classification Model for Road Damage Severity Using the Random Forest Algorithm in Kuantan Singingi Regency

Model Klasifikasi Tingkat Kerusakan Jalan Berbasis Algoritma Random Forest di Kabupaten Kuantan Singingi

Riska Mela Sapitri^{1*}, Helda Yenni², Rahmaddeni³, Agustin⁴

^{1,2,3,4}Program Studi Teknik Informatika, Fakultas Teknik dan Informatika,
Universitas Sains dan Teknologi Indonesia, Indonesia

E-Mail: ¹2210031802095@sar.ac.id, ²heldayenni@usti.ac.id,
³rahmaddeni@usti.ac.id, ⁴agustin@usti.ac.id

Received Apr 08th 2026; Revised May 02th 2026; Accepted Jun 12th 2026; Available Online Jun 26th 2026

Corresponding Author: Riska Mela Sapitri

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

Road damage is one of the infrastructure problems that can hinder community mobility and affect transportation effectiveness. Road condition assessment is still commonly performed manually, which requires significant time and may yield inconsistent evaluation results. This study aims to develop a road damage classification model using the Random Forest algorithm based on road infrastructure data from Kuantan Singingi Regency. The dataset was obtained from the Public Works and Spatial Planning Agency (PUPR) of Kuantan Singingi Regency and comprised 1,240 records and 13 variables. The research stages included data preprocessing, data cleaning, data splitting at an 80:20 ratio, feature selection based on feature importance, and the development of a Random Forest classification model. The results showed that the model achieved 83% accuracy, 78% precision, 83% recall, and an F1-score of 80% when classifying road conditions into four categories: good, fair, minor damage, and severe damage. These results indicate that the study's objective was successfully achieved by producing a classification model that automatically assesses road conditions with good performance. The proposed model is expected to assist local governments in conducting road condition assessments more objectively, quickly, and efficiently as a basis for road maintenance and infrastructure development planning

Keywords: Classification, Kuantan Singingi Regency, Machine Learning, Random Forest, Road Damage Severity

Abstrak

Kerusakan jalan merupakan salah satu permasalahan infrastruktur yang dapat menghambat mobilitas masyarakat serta memengaruhi efektivitas transportasi. Penilaian kondisi jalan yang masih dilakukan secara manual membutuhkan waktu yang cukup lama dan berpotensi menghasilkan penilaian yang kurang konsisten. Penelitian ini bertujuan untuk membangun model klasifikasi tingkat kerusakan jalan menggunakan algoritma Random Forest pada data infrastruktur jalan di Kabupaten Kuantan Singingi. Data yang digunakan berasal dari Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Kuantan Singingi sebanyak 1.240 data dengan 13 variabel. Tahapan penelitian meliputi *preprocessing* data, *data cleaning*, pembagian data dengan rasio 80:20, seleksi fitur menggunakan *feature importance*, serta pembangunan model klasifikasi Random Forest. Hasil pengujian menunjukkan bahwa model mampu mengklasifikasikan kondisi jalan ke dalam empat kategori, yaitu baik, sedang, rusak ringan, dan rusak berat dengan nilai *accuracy* sebesar 83%, *precision* sebesar 78%, *recall* sebesar 83%, dan *F1-score* sebesar 80%. Hasil tersebut menunjukkan bahwa tujuan penelitian berhasil dicapai, yaitu menghasilkan model klasifikasi yang mampu melakukan penilaian kondisi jalan secara otomatis dengan performa yang baik. Model yang dihasilkan diharapkan dapat membantu pemerintah daerah dalam melakukan penilaian kondisi jalan secara lebih objektif, cepat, dan efisien sebagai dasar dalam perencanaan pemeliharaan infrastruktur jalan.

Kata Kunci: Kabupaten Kuantan Singingi, Kerusakan Jalan, Klasifikasi, *Machine Learning*, Random Forest

1. PENDAHULUAN

Menurut Undang-Undang Nomor 38 Tahun 2004 tentang jalan. Jalan adalah prasarana transportasi, yang meliputi seluruh bagian jalan, termasuk pembangunan bagian tambahan dan perlengkapan lalu lintas di

darat, air, dan di atas tanah [1]. Di Kabupaten Kuantan Singingi, masalah jalan rusak sering dikeluhkan oleh masyarakat. Banyak ruas jalan yang mengalami kerusakan seperti retak, berlubang, bergelombang, atau bahkan rusak berat. Kondisi ini tentu mengganggu kenyamanan pengguna jalan dan bisa meningkatkan risiko kecelakaan. Karena itu, kondisi jalan perlu dipantau dengan baik dan diperbaiki sesuai kebutuhan masyarakat Kuantan Singingi.

Provinsi Riau masih menghadapi permasalahan kerusakan infrastruktur jalan yang dipengaruhi oleh tingginya aktivitas kendaraan angkutan hasil perkebunan dan pertambangan. Kabupaten Kuantan Singingi sebagai salah satu daerah dengan mobilitas kendaraan logistik yang tinggi juga mengalami kondisi serupa, sehingga berpotensi mempercepat penurunan kualitas perkerasan jalan. Perkembangan pertumbuhan volume kendaraan menjadi semakin meningkat di Kabupaten Kuantan Singingi. sehingga dapat menimbulkan dampak menurunnya tingkat kualitas jalan khususnya perkerasan lentur (aspal). Terdapat hubungan antara volume jenis kendaraan dengan nilai kerusakan jalan. Dengan hasil $R^2 = 0,840$, artinya jenis kendaraan berhubungan terhadap kerusakan jalan sebesar 84% [2]. Volume kendaraan merupakan salah satu faktor penyebab kerusakan jalan, dan tidak adanya drainase memperparah kerusakan, . Jalan Tuanku Tambusai, Rustam S Abrus dan Sudirman, merupakan jalan rigid di Kuantan Singingi. Jalan tersebut banyak di lalui kendaraan berat yang jumlahnya semakin meningkat sehingga sudah mulai rusak. Dari hasil regresi ini dapat disimpulkan bahwa volume kendaraan (X) berpengaruh terhadap kerusakan jalan (Y) sebesar 87,68% dan sisanya dipengaruhi oleh faktor lain [3]. Selain itu, ruas Jalan Taluk Kuantan–Batas Provinsi Sumatera Barat yang merupakan jalur penghubung Provinsi Riau, Jambi, dan Sumatera Barat juga mengalami kerusakan akibat tingginya kendaraan bermuatan berlebih yang berdampak pada penurunan kenyamanan, peningkatan waktu tempuh, serta risiko kecelakaan lalu lintas. Kondisi tersebut menunjukkan bahwa kerusakan jalan masih menjadi permasalahan yang perlu mendapatkan perhatian karena dapat menghambat mobilitas masyarakat dan aktivitas ekonomi daerah.

Menurut Manual Pemeliharaan Jalan Bina Marga No. 03/MN/B/1983, kerusakan jalan diklasifikasikan atas: retak (*cracking*), distorsi, cacat permukaan (*disintegration*), pengausan (*polish aggregate*), kegemukan (*bleeding* atau *flushing*), serta penurunan bekas galian atau penanaman utilitas. Dalam penelitian ini, fokus diarahkan pada klasifikasi dengan jenis kerusakan cacat permukaan (*disintegration*). Pemilihan kategori ini didasarkan pada kondisi lapangan, di mana kerusakan berupa lubang (*potholes*), butiran lepas (*ravelling*), dan permukaan terkelupas merupakan jenis kerusakan yang paling sering dijumpai pada ruas-ruas jalan di Kabupaten Kuantan Singingi [4].

Permasalahan utama dalam penelitian ini adalah penilaian kondisi jalan masih menggunakan cara manual. Cara ini membutuhkan waktu yang lama dan hasilnya sering tidak konsisten. Untuk mengatasi permasalahan tersebut, penelitian ini menawarkan solusi dengan menggunakan algoritma Random Forest untuk mengklasifikasi tingkat kerusakan jalan secara otomatis. Random Forest bekerja dengan membentuk banyak pohon keputusan untuk menghasilkan klasifikasi yang lebih stabil dan akurat. Metode ini mampu mengolah banyak variabel sekaligus serta memiliki ketahanan terhadap kesalahan data [5]. Selain itu, Random Forest juga mampu mengurangi kesalahan model dan menghasilkan klasifikasi yang lebih baik [6]. Kondisi tersebut menunjukkan pentingnya sistem identifikasi kerusakan jalan yang mampu memberikan hasil secara cepat dan akurat. Informasi mengenai tingkat kerusakan jalan sangat dibutuhkan oleh pemerintah daerah sebagai dasar dalam menentukan prioritas pemeliharaan dan perbaikan jalan. Apabila proses penilaian kondisi jalan masih dilakukan secara manual, maka pengambilan keputusan dapat menjadi lebih lambat dan berpotensi menimbulkan ketidakkonsistenan hasil penilaian. Oleh karena itu, diperlukan pendekatan berbasis *machine learning* yang mampu membantu proses klasifikasi kondisi jalan secara otomatis sehingga perencanaan pemeliharaan infrastruktur dapat dilakukan secara lebih efektif, objektif, dan efisien.

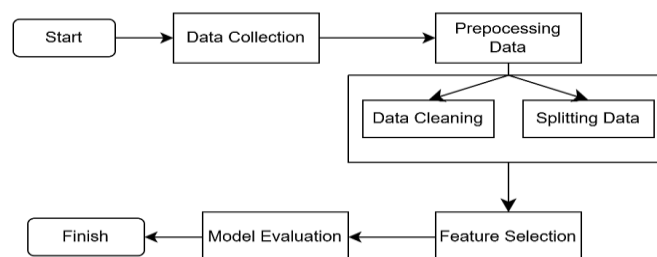
Pemilihan algoritma Random Forest didasarkan pada beberapa keunggulan yang telah dibuktikan dalam berbagai penelitian sebelumnya. Random Forest mampu menangani data berdimensi banyak, memiliki kemampuan mengurangi *overfitting* melalui mekanisme *ensemble learning*, serta menghasilkan tingkat akurasi yang relatif stabil dibandingkan algoritma pohon keputusan tunggal. Beberapa penelitian sebelumnya telah mencoba menganalisis kerusakan jalan menggunakan berbagai metode seperti algoritma C4.5 dan Random Forest dalam sistem WebGIS, tetapi penelitian ini lebih berfokus pada pemetaan dibandingkan klasifikasi. Hasil penelitian menunjukkan bahwa akurasi antara kedua algoritma hampir sama, algoritma Random Forest memberikan hasil yang lebih konsisten dan lebih baik dibandingkan C4.5. Dengan menggunakan algoritma C4.5, didapatkan nilai presisi sebesar 87,9%, *recall* 82,6%, *F1-score* 87,8%, dan akurasi 88%. Sementara itu, Random Forest menghasilkan presisi 86,6%, *recall* 86,8%, *F1-score* 86,6%, dan akurasi 87% [7]. Ada penelitian sebelumnya tentang kerusakan jalan di Kabupaten Kuantan Singingi menggunakan metode SDI yang umum digunakan juga masih mengandalkan pengamatan langsung di lapangan dan belum mampu melakukan prediksi secara otomatis. Hasil perhitungan dari penelitian menggunakan metode SDI adalah persentase tingkat kerusakan yang terdapat pada jalan tersebut : retak 82,5%, berlubang 59%, dan bekas roda 17,4%. Dengan rata-rata nilai SDI antara 100– 150 sehingga tingkat kerusakannya termasuk dalam kondisi rusak ringan [8]. Berdasarkan temuan tersebut, algoritma Random

Forest dipilih karena dinilai sesuai untuk menangani karakteristik data kerusakan jalan yang terdiri atas berbagai atribut numerik dan kategorikal.

Kebaruan penelitian ini terletak pada penerapan teknik *machine learning* untuk mengklasifikasikan kondisi jalan ke dalam empat kategori secara otomatis, yaitu baik, sedang, rusak ringan, dan rusak berat, sehingga mampu memberikan hasil yang lebih objektif, konsisten, dan efisien dibandingkan metode manual. Selain itu, penelitian ini juga memanfaatkan proses seleksi fitur untuk meningkatkan performa model dalam mengolah data kerusakan jalan yang ada di Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Kuantan Singingi. Berbeda dengan penelitian sebelumnya yang umumnya berfokus pada pemetaan kerusakan jalan menggunakan WebGIS atau analisis kondisi jalan menggunakan metode Surface Distress Index (SDI), penelitian ini berfokus pada pembangunan model klasifikasi berbasis *machine learning* menggunakan data riil Kabupaten Kuantan Singingi. Kebaruan penelitian terletak pada: (1) penggunaan dataset lokal yang berasal langsung dari Dinas PUPR Kabupaten Kuantan Singingi, (2) penerapan *feature selection* berbasis *feature importance* pada Random Forest untuk mengidentifikasi faktor yang paling berpengaruh terhadap tingkat kerusakan jalan, serta (3) klasifikasi kondisi jalan ke dalam empat kategori secara otomatis guna mendukung pengambilan keputusan pemeliharaan jalan yang lebih cepat dan objektif.

2. METODE PENELITIAN

Penelitian ini menggunakan algoritma Random Forest untuk membangun model klasifikasi tingkat kerusakan jalan berdasarkan data infrastruktur jalan Kabupaten Kuantan Singingi. Metodologi penelitian terdiri dari beberapa tahapan, yaitu pengumpulan data (*data collection*), *preprocessing* data, *data cleaning*, pembagian data (*splitting data*), seleksi fitur (*feature selection*), pelatihan model Random Forest, dan evaluasi model. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1. Data collection

Data collection merupakan tahap fundamental dalam proses penelitian yang menentukan kualitas dan keabsahan temuan ilmiah. Dalam dunia akademik, beragam pendekatan metodologis digunakan sesuai dengan karakteristik masalah penelitian dan tujuan yang ingin dicapai [9]. Teknik pengumpulan data merupakan langkah awal dan sangat penting dalam proses penelitian karena melalui teknik inilah peneliti memperoleh informasi yang dibutuhkan untuk menjawab rumusan masalah penelitian [10]. Pengumpulan data ialah organisasi dari mencari, mencatat, dan mengumpulkan semua secara objektif dan apa adanya. Lewat hasil observasi dan wawancara di lapangan yaitu pencatatan data dan berbagai bentuk data yang ada di lapangan [11]. Tahap *data collection* pada penelitian ini dilakukan dengan mengumpulkan dataset kerusakan jalan yang diperoleh dari Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Kuantan Singingi. Dataset yang digunakan berjumlah 1240 data dan terdiri dari 13 variabel, yang mencakup informasi identitas ruas jalan serta atribut kondisi jalan seperti panjang ruas, lebar jalan, dan tingkat kerusakan yang diklasifikasikan ke dalam kategori baik, sedang, rusak ringan, dan rusak berat dalam bentuk panjang (km) dan persentase kondisi.

2.2. Preprocessing Data

Preprocessing data merupakan proses penting dalam analisis data mining yang bertujuan untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat dalam proses analisis [12]. Tahap *Preprocessing* bertujuan untuk menyaring data dari elemen-elemen yang tidak diperlukan tidak memiliki relevansi atau mengganggu, seperti duplikasi, tanda baca, atau kata-kata kosong [13]. Tahap pre-processing dalam penelitian ini dilakukan untuk menyiapkan data agar dapat digunakan dalam proses pemodelan. Proses yang dilakukan meliputi *data cleaning* dan *splitting data*.

1. Data cleaning

Data cleaning merupakan tahap untuk mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) serta menghapus karakter yang tidak relevan seperti angka, simbol, tautan, dan tanda baca [14]. *Data cleaning* bertujuan untuk memfilter data sebelum data diolah dan dianalisis [15]. Proses pembersihan data dilakukan untuk mengatasi missing values dan outliers yang dapat

memengaruhi kualitas analisis serta validitas hasil penelitian. Missing values merupakan salah satu masalah utama dalam pengolahan data karena dapat menyebabkan bias dalam interpretasi serta mengurangi akurasi model prediktif yang digunakan dalam analisis lanjutan [16]. Pada penelitian ini proses *data cleaning* difokuskan pada penanganan nilai yang tidak valid, data yang hilang, serta inkonsistensi format agar data numerik yang digunakan menjadi lebih akurat dan siap untuk tahap pemodelan.

2. Splitting data

Langkah selanjutnya adalah *Splitting data* merupakan tahapan membagi data menjadi data latih dan data uji [17]. Proses pemisahan data membagi dataset menjadi dua bagian yaitu data pelatihan dan data pengujian. Data pelatihan bertujuan untuk mengajarkan algoritma, dan data pengujian sebesar satu pertiga dari jumlah dataset [18]. Pada penelitian ini digunakan rasio pembagian data sebesar 80:20, di mana 80% data digunakan sebagai data latih dan 20% sebagai data uji.

2.3. Feature Selection

Feature selection method merupakan salah satu metode yang banyak digunakan dalam melakukan reduksi fitur [19]. Hasil evaluasi akan digunakan untuk menentukan metode *feature selection* dan algoritma klasifikasi terbaik [20]. Beberapa metode *Feature selection* yang termasuk ke dalam metode pembelajaran *supervised* yang dibagi ke dalam 3 kategori utama: *Wrapper Method (Forward Selection, Backward Selection)*, *Filter Method (Chi Square & Mutual Information)*, dan *Embedded method (Lasso & Term Frequency Document Frequency)* [21]. Pada penelitian ini menggunakan *Feature selection* dengan metode *Emmbedded method (lasso & Term Frequency)*. Ditunjukkan pada persamaan 1.

$$dFI_j = T1 \sum t = 1 T1j(t) \quad (1)$$

2.4. Model Evaluation

Tahap *Model Evaluation* merupakan proses untuk menilai kinerja model *machine learning* yang telah dibangun pada fase *modeling*. Evaluasi dilakukan untuk mengetahui sejauh mana model mampu memprediksi nilai target secara akurat dan konsisten terhadap data uji, serta memastikan bahwa model tidak mengalami *overfitting* atau *underfitting* [22]. Evaluasi model ibarat ujian untuk melihat apakah mereka benar-benar memahami materi – atau hanya menghafal jawaban. Hasil evaluasi performa sebuah model, dapat digunakan sebagai acuan untuk melakukan optimasi/ perbaikan kinerja model [23]. Tahap evaluasi performa model klasifikasi dalam penelitian ini diimplementasikan berdasarkan *confusion matrix*. *Confusion Matrix* merupakan sebuah matrik duadimensi yang menggambarkan perbandingan antara hasil prediksi dengan kelas data sebenarnya, kemudian menghitung jumlah prediksi yang benar dan yang salah untuk setiap kategori kelas [24]. Tabel 1 merupakan tabel yang menggambarkan hasil prediksi model dibandingkan dengan nilai sebenarnya.

Tabel 1. Confusion Matrix

Actual/Predicted	Positive (1)	Negative (0)
Positive (1)	True Positive (TP)	False Negative (FN)
Negative (0)	Flase Positive (FP)	True Negative (TN)

Proses evaluasi ini menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* [25]. Rumus yang digunakan untuk perhitungan metrik evaluasi dapat ditunjukkan pada persamaan 2-5.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN) \quad (2)$$

$$Precision = TP/(TP + FP) \quad (3)$$

$$Recall = TP/(TP + FN) \quad (4)$$

$$F1 - Score = 2 * (Precision * Recall)/(Precision + Recall) \quad (5)$$

Tabel 2. Literature Review

Tahun	Peneliti	Judul	Algoritma	Masalah	Hasil
2025	Fahrim Irhamna Rahman, Lukman, Hildayati	Sistem Klasifikasi Kerusakan Jalan Metode <i>Machine</i>	K-Means & Random Forest	Klasifikasi kerusakan jalan berdasarkan data survei lapangan untuk	Berhasil membangun sistem klasifikasi dengan Random Forest yang mampu mengidentifikasi

Tahun	Peneliti	Judul	Algoritma	Masalah	Hasil
2024	Justam, Nur Jamilah, Sitti Mawaddah Umar, Erlita, Jousadrah Ramba	<i>learning</i> dengan Algoritma K-Means dan Random Forest	Algoritma C4.5 dan Random Forest	mempercepat proses pemeliharaan	tingkat kerusakan jalan dengan akurasi yang tinggi.
		Penerapan Algoritma C4.5 dan Random Forest untuk Pemetaan Kerusakan Jalan dengan WebGIS		Membandingkan algoritma untuk memprediksi prioritas perbaikan jalan di wilayah Luwu Raya	Akurasi C4.5 (88%) sedikit lebih tinggi, tetapi Random Forest (87%) dinilai lebih konsisten dan efektif dalam memprediksi prioritas perbaikan
2022	Ahmad Faisal	Pengaruh Volume Kendaraan Terhadap Tingkat Kerusakan Jalan pada Perkerasan Lentur di Kabupaten Kuantan Singingi	Regresi Linear Berganda	Analisa Pengaruh Volume Kendaraan Terhadap Kerusakan Jalan Rigid di Teluk Kuantan	Tingginya volume kendaraan menyebabkan kerusakan jalan pada perkerasan lentur
2022	Gusmulyani, Irawan Ade, Yolanda	Pengaruh Volume Kendaraan Terhadap Kerusakan Jalan Rigid di Teluk Kuantan	Regresi Linear	jalan rigid akibat peningkatan volume lalu lintas	Diperoleh nilai R ² sebesar 0,840 yang menunjukkan bahwa volume kendaraan berpengaruh sebesar 84% terhadap tingkat kerusakan jalan
					Volume kendaraan memberikan pengaruh sebesar 87,68% terhadap kerusakan jalan rigid.

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penerapan algoritma Random Forest dalam klasifikasi tingkat kerusakan jalan berdasarkan data infrastruktur jalan Kabupaten Kuantan Singingi. Pembahasan dilakukan secara bertahap mulai dari hasil *preprocessing* dan *data cleaning*, pembagian data latih dan data uji, proses seleksi fitur menggunakan *feature importance*, hingga evaluasi performa model. Selain menyajikan hasil pengujian, bagian ini juga menginterpretasikan temuan yang diperoleh, menganalisis faktor-faktor yang memengaruhi kinerja model, serta membandingkan hasil penelitian dengan teori dan penelitian terdahulu yang relevan. Dengan demikian, pembahasan tidak hanya berfokus pada penyajian hasil, tetapi juga menjelaskan makna dan implikasi dari hasil yang diperoleh terhadap proses identifikasi tingkat kerusakan jalan.

3.1. Data collection

Dataset pada penelitian ini bersumber dari Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Kuantan Singingi, dengan nama dataset Kerusakan Jalan. Dataset ini bersifat private, serta memiliki jumlah data sebanyak 1240 dan memiliki 13 variabel. Data yang digunakan mencakup informasi terkait identitas ruas jalan, seperti nomor ruas dan nama ruas jalan, serta atribut teknis lainnya seperti panjang ruas jalan dan lebar jalan. Selain itu, dataset juga memuat informasi kondisi jalan yang diklasifikasikan ke dalam beberapa kategori, yaitu baik, sedang, rusak ringan, dan rusak berat, yang masing-masing direpresentasikan dalam bentuk panjang jalan (km) dan persentase kondisi. Data tersebut digunakan sebagai dasar dalam proses analisis dan pengembangan model klasifikasi untuk menentukan tingkat kerusakan jalan secara lebih akurat dan sistematis. Data mentah dari dinas PUPR Kabupaten Kuantan Singingi dapat di lihat pada Tabel 3.

3.2. Preprocessing Data

Tahap *preprocessing data* dilakukan untuk mempersiapkan dataset sebelum proses pemodelan agar data yang digunakan memiliki kualitas yang baik dan sesuai dengan kebutuhan algoritma. Proses ini bertujuan untuk memperbaiki struktur data, mengurangi kesalahan atau data yang tidak relevan, serta memastikan data dapat diproses dengan lebih efektif pada tahap klasifikasi.

1. Data cleaning

Data cleaning dilakukan untuk memastikan bahwa data yang digunakan dalam proses pelatihan model berada dalam kondisi yang baik dan siap digunakan pada tahap klasifikasi. Pada penelitian ini, proses *data cleaning* dilakukan dengan menangani nilai yang tidak valid, mengatasi data yang kosong, serta mengubah format data ke dalam bentuk numerik agar dapat diproses oleh algoritma Random Forest.

Tahapan ini penting untuk meningkatkan kualitas dataset sehingga proses pemodelan dapat berjalan lebih optimal dan menghasilkan klasifikasi yang lebih akurat serta konsisten. Tabel 4 merupakan hasil *data cleaning* pada dataset penelitian.

Tabel 3. Dataset Kerusakan Jalan

No	No Ruas	Nama Ruas Jalan	Panjang Ruas (Km)	ROW Jalan (M)	Panjang Tiap Kondisi							
					Baik		Sedang		Rusak Ringan		Rusak Berat	
					Km	%	Km	%	Km	%	KM	%
1	001	Seb.Teluk Kuantan- Seb.Benai	11,70	4,00	7,50	64,10	2,20	18,80	1,40	11,97	0,60	5,13
2	002	Jake- Geringging Baru	13,30	5,50	8,70	65,41	2,00	15,04	1,60	12,03	1,00	7,52
3	003	Jake – Koto Kombu	20,00	6,00	9,80	49,00	2,80	14,00	3,80	19,00	3,60	18,00
...	...	...	...	...	...	...	...	...	...	...	...	...
1238	020	SKP 8 Trans 1 - Teso	11,243	0,5	0,000	0	0,000	0	0,000	0	11,243	100
1239	021	Sei. Sirih- Simp. 4 PT. Wanasari	9,276	0,8	5,161	55,638	0,600	6,46831	0,000	0	3,515	37,893
1240	022	Pasir Mas – Sako Margasari	14,1200,5	0,5	0,980	6,9405	0,000		0,000	0	13,140	93,059

Tabel 4. Hasil *Data cleaning*

No	No Ruas	Nama Ruas Jalan	Baik		Sedang		Rusak Ringan		Rusak Berat	
			Km	%	Km	%	Km	%	Km	%
1	1.0	Seb. Teluk Kuantan – Seb. Benai	7.5	64.10	2.2	18.80	1.4	11.96	0.6	5.12
2	2.0	Jake – Geringging Baru	8.7	65.41	2	15.03	1.6	12.03	1	7.51
3	3.0	Jake – Koto Kombu	9.8	49.00	2.8	14.00	3.8	19.00	3.6	18.00
4	4.0	Sentajo – Muara Langsat	30.4	89.41	1	2.94	2.6	7.64	0	0.00

2. Data Splitting

Data Splitting dilakukan untuk membagi dataset menjadi data latih dan data uji agar proses pelatihan dan pengujian model dapat dilakukan secara terpisah. Pada penelitian ini, pembagian data menggunakan rasio 80:20, di mana 80% data digunakan sebagai data latih dan 20% data digunakan sebagai data uji. Tahapan ini penting untuk mengevaluasi kemampuan model Random Forest dalam melakukan klasifikasi terhadap data yang belum pernah digunakan sebelumnya sehingga hasil evaluasi yang diperoleh dapat lebih objektif dan akurat. Tabel 5 menunjukkan hasil pembagian data pada dataset kerusakan jalan.

Tabel 5. Hasil *Data Splitting 80:20*

Jenis Data	Jumlah Data	Presentase
Data Training	991	80%
Data Testing	248	20%

3.3. Feature selection

Pada penelitian ini, proses *feature selection* dilakukan menggunakan metode embedded dengan memanfaatkan nilai *feature importance* yang dihasilkan oleh algoritma Random Forest. Metode ini digunakan untuk mengetahui tingkat pengaruh masing-masing fitur terhadap proses klasifikasi tingkat kerusakan jalan. Melalui pendekatan tersebut, fitur yang memiliki kontribusi tinggi dapat dipertahankan, sedangkan fitur yang kurang relevan dapat dikurangi sehingga model yang dihasilkan menjadi lebih optimal, efisien, dan mampu meningkatkan performa klasifikasi. Tabel 6 menunjukkan hasil *feature selection* pada penelitian ini.

Tabel 6. Hasil *Feature selection*

Feature Importance:		
	Fitur	Nilai
3	Label Rusak Berat	0.359416
0	Label Baik	0.339138

Feature Importance:		
	Fitur	Nilai
1	Label Sedang	0.174066
2	Label Rusak Ringan	0.127380

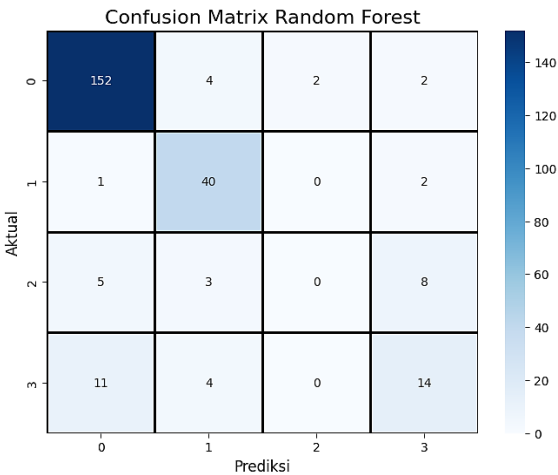
3.4. Model Evaluation

Tahap *model evaluation* dilakukan untuk mengukur performa algoritma Random Forest dalam mengklasifikasikan tingkat kerusakan jalan. Evaluasi dilakukan menggunakan beberapa metrik pengujian, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui tingkat ketepatan model dalam melakukan prediksi terhadap data uji. Hasil evaluasi tersebut digunakan untuk menilai kemampuan model dalam menghasilkan klasifikasi yang akurat dan konsisten pada setiap kategori kerusakan jalan. Tabel 7 menunjukkan hasil *model evaluation* pada penelitian ini.

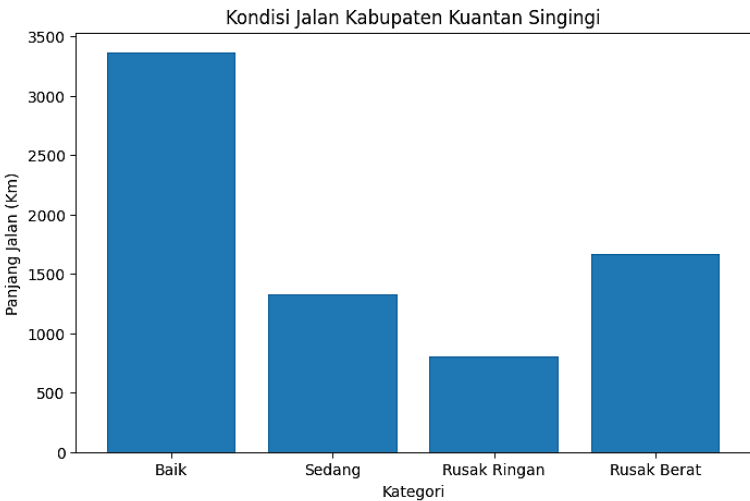
Tabel 7. Hasil Model Evaluation

	Precision	Recall	F1-score
Accuracy	-	0.83	-
Macro Avg	0.78	0.83	0.80
Weighted Avg	0.78	0.83	0.80

Berdasarkan hasil evaluasi model Random Forest, diperoleh nilai *accuracy* sebesar 0.83, *precision* sebesar 0.78, *recall* sebesar 0.83, dan *F1-score* sebesar 0.80. Hasil tersebut menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan tingkat kerusakan jalan pada dataset penelitian. Gambar 4 dan Gambar 5 menampilkan hasil confusion matrix dari model Random Forest yang digunakan untuk mengevaluasi performa klasifikasi tingkat kerusakan jalan di Kabupaten Kuantan Singingi.



Gambar 4. Confusion Matrix



Gambar 5. Diagram Hasil Kondisi Jalan

Berdasarkan Gambar 4, model Random Forest mampu mengklasifikasikan sebagian besar data pada setiap kategori kondisi jalan dengan benar. Kategori "Baik" dan "Sedang" memiliki jumlah prediksi yang tepat lebih tinggi dibandingkan kategori lainnya, yang menunjukkan bahwa karakteristik kedua kelas tersebut lebih mudah dikenali oleh model. Sementara itu, pada kategori "Rusak Ringan" dan "Rusak Berat" masih ditemukan beberapa data yang terklasifikasi ke kelas lain. Kondisi ini mengindikasikan adanya kemiripan karakteristik antar kategori kerusakan sehingga model mengalami kesulitan dalam membedakan batas antar kelas secara tegas. Meskipun demikian, jumlah prediksi yang benar masih lebih dominan dibandingkan prediksi yang salah, sehingga menunjukkan bahwa model memiliki kemampuan yang baik dalam mempelajari pola data kerusakan jalan. Hasil ini juga sejalan dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh, sehingga dapat disimpulkan bahwa model Random Forest cukup efektif digunakan untuk mengklasifikasikan tingkat kerusakan jalan di Kabupaten Kuantan Singingi.

Berdasarkan Gambar 5, kondisi jalan di Kabupaten Kuantan Singingi didominasi oleh kategori "Baik" dengan total panjang lebih dari 3.000 km. Hasil ini menunjukkan bahwa sebagian besar ruas jalan yang menjadi objek penelitian masih berada dalam kondisi layak untuk digunakan dan mampu mendukung aktivitas transportasi masyarakat. Namun demikian, kategori "Rusak Berat" menempati posisi kedua dengan panjang jalan mencapai lebih dari 1.500 km, yang menunjukkan bahwa masih terdapat sejumlah ruas jalan yang memerlukan perhatian dan penanganan secara prioritas. Sementara itu, kategori "Sedang" dan "Rusak Ringan" memiliki proporsi yang lebih kecil dibandingkan dua kategori lainnya. Kondisi ini mengindikasikan bahwa kerusakan jalan di Kabupaten Kuantan Singingi cenderung terpolarisasi antara jalan yang masih dalam kondisi baik dan jalan yang telah mengalami kerusakan cukup signifikan. Tingginya panjang jalan pada kategori rusak berat dapat dipengaruhi oleh berbagai faktor, seperti tingginya volume kendaraan angkutan hasil perkebunan dan pertambangan, beban kendaraan yang melebihi kapasitas jalan, serta faktor lingkungan dan drainase yang kurang optimal. Temuan ini menunjukkan pentingnya upaya pemeliharaan dan rehabilitasi jalan secara terencana agar tingkat kerusakan tidak semakin meningkat dan dapat mengurangi risiko gangguan terhadap mobilitas masyarakat maupun aktivitas ekonomi daerah.

Hasil evaluasi menunjukkan bahwa algoritma Random Forest mampu menghasilkan performa klasifikasi yang baik dengan nilai *accuracy* sebesar 83%, *precision* sebesar 78%, *recall* sebesar 83%, dan *F1-score* sebesar 80%. Nilai tersebut menunjukkan bahwa model mampu mengenali sebagian besar pola pada data kerusakan jalan yang digunakan dalam penelitian. Berdasarkan confusion matrix, kategori "Baik" memiliki jumlah prediksi benar yang paling tinggi dibandingkan kategori lainnya. Hal ini menunjukkan bahwa karakteristik data pada kategori tersebut lebih mudah dikenali oleh model. Sebaliknya, kategori "Rusak Ringan" masih menunjukkan beberapa kesalahan klasifikasi ke kategori "Sedang" maupun "Rusak Berat". Kondisi tersebut mengindikasikan adanya kemiripan karakteristik antar kelas sehingga model mengalami kesulitan dalam membedakan batas antar kategori secara tegas.

Hasil *feature importance* menunjukkan bahwa variabel yang berkaitan dengan panjang kerusakan jalan dan persentase kondisi jalan memiliki kontribusi terbesar terhadap proses klasifikasi. Temuan ini menunjukkan bahwa tingkat kerusakan suatu ruas jalan sangat dipengaruhi oleh proporsi panjang kerusakan yang terjadi pada ruas jalan tersebut. Semakin besar tingkat kerusakan yang tercatat pada suatu ruas jalan, semakin besar pula pengaruhnya terhadap keputusan klasifikasi yang dihasilkan oleh model.

Hasil penelitian ini sejalan dengan penelitian Jamilah dan kawan-kawan 2024 yang menyatakan bahwa Random Forest mampu menghasilkan performa klasifikasi yang lebih stabil dibandingkan algoritma C4.5. Selain itu, penelitian Asnur dan kawan-kawan pada tahun 2025 juga menunjukkan bahwa Random Forest memiliki kemampuan yang baik dalam menangani data infrastruktur jalan yang memiliki banyak variabel. Kesamaan hasil tersebut menunjukkan bahwa pendekatan *ensemble learning* yang digunakan Random Forest efektif dalam mengurangi kesalahan klasifikasi dan meningkatkan kemampuan generalisasi model.

Implikasi praktis dari penelitian ini adalah tersedianya model klasifikasi yang dapat membantu Dinas PUPR Kabupaten Kuantan Singingi dalam melakukan evaluasi kondisi jalan secara lebih cepat dan objektif dibandingkan metode manual. Dengan adanya model ini, proses identifikasi prioritas pemeliharaan jalan dapat dilakukan secara lebih efisien sehingga mendukung pengambilan keputusan yang lebih tepat. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari satu wilayah sehingga karakteristik data mungkin belum mewakili kondisi jalan di daerah lain. Selain itu, penelitian hanya menggunakan satu algoritma klasifikasi sehingga belum diketahui perbandingan performanya dengan metode lain seperti XGBoost, LightGBM, Support Vector Machine, maupun Artificial Neural Network.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model klasifikasi tingkat kerusakan jalan menggunakan algoritma Random Forest berdasarkan dataset kerusakan jalan dari Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Kuantan Singingi. Tujuan penelitian untuk menghasilkan model klasifikasi otomatis yang mampu mengidentifikasi tingkat kerusakan jalan telah berhasil dicapai. Hasil pengujian

menunjukkan bahwa model Random Forest mampu menghasilkan nilai *accuracy* sebesar 83%, *precision* sebesar 78%, *recall* sebesar 83%, dan *F1-score* sebesar 80%. Nilai tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam mengklasifikasikan kondisi jalan ke dalam kategori baik, sedang, rusak ringan, dan rusak berat. Penerapan *feature selection* berbasis *feature importance* juga berhasil mengidentifikasi variabel yang paling berpengaruh terhadap proses klasifikasi. Hasil penelitian menunjukkan bahwa informasi panjang kerusakan dan persentase kondisi jalan menjadi faktor dominan dalam menentukan tingkat kerusakan suatu ruas jalan. Penelitian ini memiliki keterbatasan pada penggunaan dataset yang hanya berasal dari satu wilayah serta penggunaan satu algoritma klasifikasi. Oleh karena itu, penelitian selanjutnya dapat mengembangkan model dengan menambahkan dataset yang lebih besar, melakukan perbandingan dengan algoritma *machine learning* lainnya, serta mengintegrasikan teknologi computer vision berbasis citra kerusakan jalan untuk meningkatkan akurasi klasifikasi.

REFERENSI

- [1] I. G. Agung, A. Istri, I. G. A. Diputera, I. K. Diartama, and K. Tubuh, "Analisis Penyebab dan Dampaknya Kerusakan Infrastruktur Jalan Terhadap Para Pengguna Jalan dan Masyarakat Sekitar," vol. 11, no. 2, pp. 32–36, 2022, doi: <https://doi.org/10.36733/jikt.v11i2.5427>.
- [2] Gusmulyani, A. Irawan, and Yolanda, "Analisa Pengaruh Volume Kendaraan Terhadap Kerusakan jalan Rigid Di Teluk Kuantan," vol. 4, no. April, pp. 27–37, 2022.
- [3] A. Faisal, "Pengaruh Volume Kendaraan Terhadap Tingkat Kerusakan Jalan Pada Perkerasan Luntur Di Kabupaten Kuantan Singingi," vol. 4, no. 2, pp. 1560–1567, 2021, doi: <https://doi.org/10.36378/jupersatek.v4i2.2349>.
- [4] U. Kerusakan, K. J. Klasifikasi, J. Arteri, J. Kolektor, J. Lokal, and J. Lingkungan, "Bab 3 landasan teori 3.1," 2018.
- [5] H. Asnur, R. Yunita, and A. R. Marsa, "Algoritma Random Forest Untuk Prediksi Perkerasan Tanah," vol. 5, no. 2, pp. 381–387, 2025, doi: <https://doi.org/10.55382/jurnalpustakaai.v5i2.108>.
- [6] M. W. Nugroho, "Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi) Analisis Performa Algoritma Random Forest dalam Mengatasi *Overfitting* pada Model Prediksi," vol. 9, no. December, pp. 1562–1571, 2025, doi: <https://doi.org/10.32520/stmsi.v15i3.6060>.
- [7] N. Jamilah, S. M. Umar, and J. Ramba, "Penerapan Algoritma C4 . 5 dan Random Forest untuk Pemetaan," vol. 7, pp. 326–339, 2024, doi: <https://doi.org/10.57093/jisti.v7i2.270>.
- [8] A. Lubis, GesviFadrizal, and A. Saleh, "s IKLU s," vol. 6, no. 2, pp. 195–203, 2020, doi: <https://10.31849/siklus.v6i2.4647>.
- [9] "Teknik Pengumpulan Data Dalam Penelitian," vol. 10, pp. 315–325, 2025, doi: <https://garuda.kemdiktisaintek.go.id/documents/detail/5668522>.
- [10] P. S. Agasha and R. R. Ulum, "Teknik Pengumpulan Data Dalam Pendekatan Kualitatif," vol. 5, no. 2, pp. 707–713, 2025, doi: <https://Vol. 5, No. 2, Desember 2025>.
- [11] B. A. Habsy, S. Q. A, and I. Widyani, "Manajemen Pengumpulan Data," pp. 34–46, 2024, doi: <https://doi.org/10.59581/jmk-widyakarya.v2i6.4232>.
- [12] A. Agung, A. Daniswara, and I. K. D. Nuryana, "Data *Preprocessing* Pola Pada Penilaian Mahasiswa Program Profesi Guru," vol. 05, pp. 97–100, 2023, doi: <https://doi.org/10.47065/bits.v7i2.8164>.
- [13] B. Warsito and A. Prahutama, "Sentiment Analysis on Tokopedia Product Online Reviews Using Random Forest Method," vol. 16006, pp. 1–10, 2020, doi: <https://doi.org/10.1051/e3sconf/202020216006>.
- [14] S. A. Arian, P. R. Alifia, B. Prihantoro, and M. I. Disriansyah, "Comparison of Supervised *Machine learning* Classification Models in Knowledge Discovery of Pertamina Prabumulih IT Services Perbandingan Model Klasifikasi Supervised *Machine learning* dalam Knowledge Discovery Layanan TI Pertamina Prabumulih," vol. 6, no. April, pp. 705–715, 2026, doi: <https://doi.org/10.57152/malcom.v6i2.2617>.
- [15] W. D. Budimulia and F. Ridho, "Penerapan Komputasi Paralel pada Aplikasi Studi Kasus Survei Sosial Ekonomi Nasional," pp. 7–14, doi: <https://doi.org/10.34123/semnasoffstat.v2019i1.120>.
- [16] J. Elektronika and D. A. N. Komputer, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," vol. 17, no. 2, pp. 657–666, 2024, doi: <https://doi.org/10.51903/elkom.v17i2.2311> <https://journal.stekom.ac.id/index.php/elkom>.
- [17] E. Sahelvi, P. Cikita, and R. M. Sapitri, "Comparison of K-Nearest Neighbors and Random Forest Algorithms for Recommendations for a Healthy Lifestyle in Prevent Heart Disease Perbandingan Algoritma K-Nearest Neighbors dan Random Forest untuk Rekomendasi Gaya Hidup Sehat dalam Mencegah Penyakit Jan," vol. 5, no. July, pp. 830–840, 2025, doi: <https://doi.org/10.57152/malcom.v5i3.1972>.
- [18] I. Ramadhan, R. F. Waleska, S. Elmi, L. Efrizoni, and Rahmaddeni, "Klasifikasi Penjualan Walmart Menggunakan Algoritma C4.5," no. 02, pp. 122–129, doi: <https://doi.org/10.36050/pjbkse24>.
- [19] H. M. Lumbantobing, J. Sisingamangaraja, J. Sisingamangaraja, and J. Sisingamangaraja, "Penerapan

- Metode *Feature selection* pada Algoritma Naïve Bayes dalam Kasus Keyword Extraction,” pp. 6–8, 2020.
- [20] R. Putus, K. Mahasiswa, I. Hadi, A. Ghazali, and A. R. Handoko, “Analisis Efektifitas *Feature selection* dalam Pengkayaan *Machine learning* untuk Deteksi Dini,” vol. 11, no. 1, pp. 149–157, 2025.
- [21] U. Sta, “Feature Evaluation by Filter , Wrapper , and Embedded Approaches,” pp. 29–44, doi: 10.1007/978-3-662-45620-0.
- [22] M. Abror, A. Hilmi, and A. Susanto, “Implementasi dan Evaluasi Model *Machine learning* untuk Optimalisasi Prediksi Penjualan Produk Kue Kering,” vol. 7, no. 3, pp. 1649–1661, 2025, doi: 10.47065/bits.v7i3.8657.
- [23] Unper, *Evaluasi Model Machine learning*. doi: <https://lmsspada.kemdiktisaintek.go.id/>.
- [24] T. Gori *et al.*, “PReprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa,” vol. 11, no. 1, pp. 215–224, 2024, doi: <https://doi.org/10.57152/malcom.v6i2.2620>.
- [25] M. A. K. Neighbors, S. Lestari, and S. Febrianti, “Sentiment Analysis of Shopee Product Reviews on the Instagram Application Using the K-Nearest Neighbors Algorithm Analisis Sentimen Ulasan Produk Shopee di Aplikasi Instagram,” vol. 5, no. October, pp. 1172–1180, 2025, doi: <https://doi.org/10.57152/malcom.v5i4.1595>.