



## Penerapan Algoritma C4.5 dan Random Forest untuk Pemetaan Kerusakan Jalan dengan WebGIS

Justam<sup>1</sup>, Nur Jamilah<sup>2</sup>, Sitti Mawaddah Umar<sup>3</sup>, Erlita<sup>4</sup>, Jousadrah Ramba<sup>5</sup>

Program Studi Informatika, Universitas Mega Buana Palopo<sup>1,5</sup>

Program Studi Sistem Informasi, Universitas Mega Buana Palopo<sup>2</sup>

Program Studi Manajemen Infomasi Kesehatan, Politeknik Kesehatan megarezky<sup>3</sup>,

Program Studi Sistem Informasi, Institut Teknologi Pendidikan dan Bisnis Qana'ah Sidenreng Rappang<sup>4</sup>,

Jl. Andi Ahmad No. 25, Wara Utara, 91913, Kota Palopo, Sulawesi Selatan, Indonesia

justam@umegabuana.ac.id\*<sup>1</sup>, Nurjamila1989@gmail.com<sup>2</sup>, Sittimawaddahumar05@gmail.com<sup>3</sup>,

erlitazaindin3@gmail.com<sup>4</sup>, jousadrah@gmail.com<sup>5</sup>

### Kata Kunci :

C4.5;  
Random Forest;  
Prediksi  
Perbaikan Jalan;  
WebGIS;  
Pemetaan.

### ABSTRAK

Kerusakan jalan di wilayah Luwu Raya menjadi tanggung jawab BBPJN VI Makassar, yang melakukan pemantauan kondisi jalan dan melaporkan hasilnya untuk tindakan perbaikan. Penelitian ini bertujuan untuk membandingkan algoritma C4.5 dan Random Forest dalam memprediksi prioritas perbaikan jalan dan persebaran kerusakan jalan. Data yang digunakan dalam penelitian ini adalah 6100 data kerusakan jalan pada tiga ruas jalan dari tahun 2021 hingga 2023. Hasil penelitian menunjukkan bahwa meskipun 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%. Penelitian ini menghasilkan sistem informasi pemetaan berbasis WebGIS yang digunakan untuk menentukan prioritas perbaikan jalan. Hasil penelitian ini menunjukkan bahwa Random Forest lebih efektif dalam memprediksi dan menentukan prioritas perbaikan jalan di wilayah Luwu Raya.

### Keywords

C4.5;  
Random Forest;  
Road Repair  
Prediction;  
WebGIS;  
Mapping.

### ABSTRACT

Road damage in the Luwu Raya region is the responsibility of BBPJN VI Makassar, which monitors road conditions and reports the results for corrective actions. This study aims to compare the C4.5 and Random Forest algorithms in predicting road repair priorities and the distribution of road damage. The data used in this study consists of 6100 road damage records from three road segments from 2021 to 2023. The results show that although the accuracy of both algorithms is similar, Random Forest provides more consistent and better results than C4.5. The C4.5 algorithm achieved precision of 87.9%, recall of 82.6%, f1-score of 87.8%, and accuracy of 88%. Meanwhile, Random Forest achieved precision of 86.6%, recall of 86.8%, f1-score of 86.6%, and accuracy of 87%. This research resulted in a WebGIS-based mapping information system used for determining road repair priorities. The findings suggest that Random Forest is more effective in predicting and determining road repair priorities in the Luwu Raya region.

---Jurnal JISTI @2024---



## PENDAHULUAN

Jalan raya merupakan infrastruktur transportasi yang sangat penting karena berfungsi sebagai penghubung utama antara berbagai wilayah, khususnya untuk mendukung distribusi barang dan jasa. Di Indonesia, khususnya di wilayah Sulawesi Selatan, jalan raya nasional dikelola oleh Balai Besar Pelaksanaan Jalan Nasional (BBPJN) VI Makassar, yang berada di bawah Kementerian Pekerjaan Umum dan Tata Ruang. BBPJN VI Makassar bertanggung jawab atas perencanaan, pembangunan, pemeliharaan, serta rehabilitasi jalan raya di wilayah tersebut, dengan tujuan utama untuk memastikan kelancaran arus lalu lintas dan mendukung kegiatan ekonomi.

Di bawah Undang-Undang No. 38 Tahun 2004 tentang Jalan, disebutkan bahwa jalan merupakan prasarana transportasi darat yang harus dikelola dengan baik. Meskipun sebagian besar ruas jalan di Sulawesi Selatan, khususnya di wilayah Luwu Raya, telah diaspal dan dalam kondisi baik, namun beberapa ruas jalan utama, seperti Bts. Kota Palopo-Bts. Kab. Luwu, Bts. Kab. Luwu Utara-Wotu-Tarengge-Malili-Bts. Prov. Sultra, dan Tarengge-Kayulangi-Bts. Prov. Sulteng, masih menghadapi masalah kerusakan. Beberapa jalan alternatif yang lebih kecil dan jarang dilalui juga menunjukkan kondisi yang kurang baik.

Kerusakan jalan yang terjadi di wilayah Luwu Raya disebabkan oleh berbagai faktor, antara lain cuaca ekstrem dengan curah hujan yang tinggi, kepadatan lalu lintas, terutama kendaraan berat seperti truk dan bus, serta kurangnya pemeliharaan yang rutin. Data menunjukkan bahwa kerusakan jalan di Sulawesi Selatan mencapai 2.009 kilometer pada tahun 2022, meningkat signifikan dari 1.500 kilometer pada tahun 2019 (Kompas.com, 2023). Hal ini menambah tantangan bagi pemerintah daerah dalam upaya perbaikan dan pemeliharaan infrastruktur jalan, terutama dengan keterbatasan sumber daya yang ada.

Meskipun pemerintah dan otoritas setempat telah berupaya memperbaiki jalan yang rusak melalui program perbaikan dan pemeliharaan, kendala sumber daya dan prioritas penanganan sering kali menghambat kelancaran proses tersebut. Oleh karena itu, diperlukan suatu sistem yang dapat membantu pihak terkait dalam pengambilan keputusan mengenai prioritas perbaikan jalan berdasarkan kondisi kerusakan yang ada.

Salah satu pendekatan yang dapat digunakan adalah dengan memanfaatkan teknologi data mining untuk melakukan klasifikasi terhadap parameter-parameter yang mempengaruhi tingkat kerusakan jalan. Penelitian sebelumnya oleh Nahot Frastian, Senna Hendrian, dan V.H. Valentino (2018) telah menggunakan metode C4.5, Naïve Bayes, dan Random Forest untuk membandingkan akurasi dalam prediksi kelulusan mata kuliah. Berdasarkan penelitian tersebut, metode C4.5 memberikan tingkat akurasi tertinggi sebesar 98,89%, diikuti oleh Naïve Bayes dan Random Forest dengan akurasi 96,67% dan 95,56%, masing-masing.

Penelitian ini bertujuan untuk mengembangkan penelitian sebelumnya dengan mengaplikasikan algoritma C4.5 dan Random Forest untuk melakukan klasifikasi data kerusakan jalan di wilayah Luwu Raya. Hasil dari klasifikasi ini akan divisualisasikan dalam bentuk Website Geographic Information System (WebGIS), yang diharapkan dapat menjadi solusi untuk membantu pihak terkait dalam memantau kondisi jalan dan menentukan prioritas perbaikan jalan raya. Dengan adanya sistem WebGIS ini, pengambilan keputusan dalam penanganan kerusakan jalan diharapkan dapat dilakukan dengan lebih efektif, efisien, dan merata.



---

## KAJIAN PUSTAKA

### 1. Balai Besar Pelaksanaan Jalan Nasional (BBPJN) VI Makassar

Jalan nasional merupakan salah satu infrastruktur transportasi darat yang pengelolaannya berada di bawah kewenangan pemerintah. Jalan ini memiliki beberapa bagian yang tidak hanya berfungsi untuk menghubungkan satu daerah dengan daerah lainnya, tetapi juga dapat dimanfaatkan oleh pihak-pihak lain seperti individu, perusahaan, atau instansi pemerintah dengan izin dari pengelola jalan nasional, yaitu Balai Besar Pelaksanaan Jalan Nasional (BBPJN) (Bagian-Bagian Jalan Nasional et al., 2020). BBPJN adalah lembaga yang dibentuk oleh Kementerian Pekerjaan Umum dan Perumahan Rakyat untuk mengelola infrastruktur jalan. BBPJN berfungsi untuk melaksanakan dan mengendalikan penggunaan dana APBN dalam pembangunan jalan nasional. Berdasarkan Pasal 103 Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat, BBPJN bertanggung jawab kepada Direktur Jenderal Bina Marga (BBPJN, 2024).

Menurut Peraturan Menteri Pekerjaan Umum No. 11/PRT/M/2010 tentang Tata Cara dan Persyaratan Layak Fungsi Jalan, jalan dikatakan layak fungsi jika memenuhi persyaratan teknis yang dapat memberikan keselamatan bagi penggunanya, serta persyaratan administratif untuk memberikan kepastian hukum bagi penyelenggara jalan dan pengguna jalan (Adwang, 2020). Berdasarkan Pasal 30 Undang-Undang RI No. 38 Tahun 2004 tentang Jalan, jalan umum dapat dioperasikan setelah memenuhi persyaratan layak fungsi jalan secara teknis administratif. Jalan umum terdiri dari beberapa kategori berdasarkan statusnya, seperti jalan nasional, jalan provinsi, jalan kabupaten, jalan kota, dan jalan desa (Adwang, 2020).

### 2. Kerusakan Jalan

Jalan raya umumnya terbuat dari aspal beton, yaitu lapisan konstruksi jalan yang terdiri dari campuran aspal keras dan agregat, yang dicampur dan dipadatkan pada suhu tertentu (Sukirman, 1992). Aspal beton ini dirancang untuk memiliki kepadatan tinggi dan nilai struktural yang baik, namun di sisi lain, ketahanan dan keawetan jalan dapat rendah jika konstruksinya kurang baik (Saragi et al., 2021). Kerusakan pada jalan dapat dikelompokkan menjadi dua jenis, yaitu kerusakan struktural dan fungsional. Kerusakan struktural terjadi pada bagian struktur perkerasan jalan yang disebabkan oleh ketidakstabilan tanah dasar, beban lalu lintas, dan kondisi lingkungan. Sementara kerusakan fungsional terjadi pada permukaan jalan yang dapat mengganggu kenyamanan dan keamanan pengguna jalan meskipun perkerasan jalan masih dapat menahan beban (Andi Tenrisukki, 2002).

Pada wilayah Luwu Raya, pengamatan dan pelaporan kondisi jalan dilakukan oleh Penilik Jalan yang bertugas untuk memantau, melaporkan kondisi jalan, dan mengusulkan penanganan gangguan jalan. Hasil pengamatan kemudian diteruskan ke Pejabat Pembuat Komitmen (PPK) untuk pengambilan keputusan (Pekerjaan et al., 2021).

### 3. Data Mining

Data Mining (DM) adalah proses untuk menggali pengetahuan yang tersembunyi dalam kumpulan data besar yang sulit untuk dianalisis secara manual. Teknik yang sering digunakan dalam data mining meliputi clustering, classification, association rule mining, neural network, dan genetic algorithm (Arif, 2019). Data mining memungkinkan untuk menemukan pola atau hubungan dalam data yang dapat digunakan untuk membuat keputusan yang lebih baik. Proses data mining termasuk seleksi data, preprocessing, transformasi, serta penerapan algoritma untuk mengekstraksi pengetahuan dari data yang telah diproses (Arif, 2019).



#### 4. Metode Comparing

Metode komparasi digunakan untuk mengetahui perbedaan antara dua kelompok atau lebih, serta untuk membandingkan variabel atau objek penelitian yang berbeda. Tujuan utama dari metode komparasi adalah untuk menemukan hubungan sebab-akibat dan membuat kesimpulan berdasarkan perbandingan data yang ada (Meikalyan, 2017).

#### 5. Decision Tree

Decision Tree (pohon keputusan) adalah model yang sering digunakan dalam klasifikasi data. Pohon keputusan membagi kumpulan data menjadi sub-kumpulan berdasarkan aturan keputusan yang dibuat pada simpul-simpul internal, sedangkan simpul daun mewakili label kelas. Algoritma C4.5 dan Random Forest adalah dua metode yang menggunakan pohon keputusan untuk klasifikasi dan regresi (Supriyadi et al., 2020).

#### 6. Classification and Regression Trees (C4.5)

C4.5 adalah algoritma pohon keputusan yang dikembangkan dari ID3. Algoritma ini menggunakan teknik "Gain Ratio" untuk memilih atribut yang akan dijadikan sebagai akar pohon keputusan. C4.5 menghitung Entropy, Gain, dan Gain Ratio untuk menentukan atribut terbaik yang membagi dataset (Novianti et al., 2016). Algoritma ini sering digunakan karena kemampuannya dalam menghasilkan model yang akurat dan mudah dipahami (Yuli Mardi, 2016).

#### 7. Random Forest

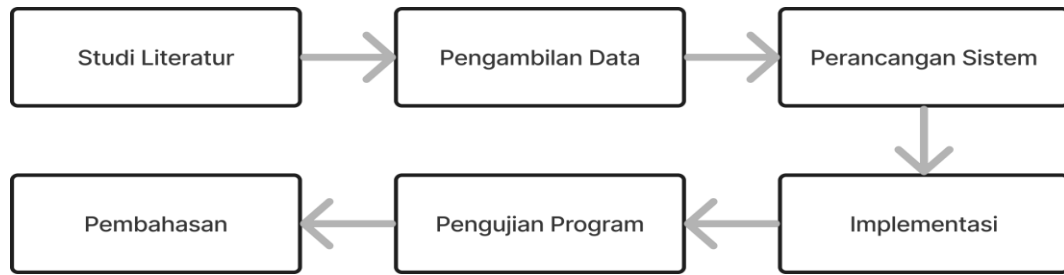
Random Forest (RF) adalah teknik ensemble learning yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi prediksi. RF bekerja dengan membangun banyak pohon keputusan berdasarkan subset data yang dipilih secara acak. Keunggulan RF termasuk kemampuannya untuk menangani data besar dengan efisien, mengatasi missing values, dan mengurangi overfitting (Pamuji & Ramadhan, 2021). Teknik ini pertama kali diperkenalkan oleh Leo Breiman pada tahun 2001 dan telah menjadi salah satu metode yang paling populer dalam machine learning (Pahrul, 2023).

### METODE PENELITIAN

Berisi tentang tahapan penelitian, rancangan/model, tempat dan waktu, teknik pengumpulan data, dan teknik/metode analisis data. Metode penelitian disusun dengan membentuk sub bab (**bold**) seperti berikut (contoh):

#### 1. Tahapan Penelitian

Penelitian ini membandingkan metode C4.5 dan Random Forest untuk mengklasifikasikan data kerusakan jalan di wilayah Luwu Raya. Hasil klasifikasi tersebut kemudian divisualisasikan dalam bentuk Web Geographic Information System (WebGIS). Data yang digunakan dalam penelitian ini mencakup informasi seperti Pejabat Pembuat Komitmen (PPK), provinsi, satuan kerja (satker), nomor ruas jalan, nama ruas jalan, tahun, nomor identitas, lokasi/km, subkategori (kategori kerusakan), kondisi kerusakan ruas kiri dan kanan, waktu pengamatan, nama wilayah, nama penilik jalan, panjang (m), lebar (m), tebal/tinggi (m), volume, balai, dan keterangan. WebGIS yang mengintegrasikan data mining menghasilkan output berupa klasifikasi data kerusakan jalan dalam kategori rutin biasa, rutin kondisi, dan holding. Tampilan tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1



Gambar 1. Tahapan Penelitian

## 2. Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini bersifat sekunder dan diperoleh langsung dari sumber data terkait. Sumber data tersebut mencakup tiga ruas jalan nasional yang dikelola oleh Balai Besar Pelaksanaan Jalan Nasional VI Makassar pada periode 2021–2023. Penelitian ini melibatkan tiga Pejabat Pembuat Komitmen (PPK) dengan total 6.100 data kerusakan jalan. Adapun data yang diperoleh terdiri dari Dataset Kerusakan Jalan dari BBPJJN Dataset ini mencakup 20 variabel sebagai berikut:

Tabel 1. Tabel Variabel Kerusakan jalan

No	Variable	No	Variable
1	PPK	11	Rusak Ruas Kanan Jalan
2	Provinsi	12	Waktu Pengamatan
3	Satker	13	Nama Wilayah
4	No Ruas Jalan	14	Nama Penilik Jalan
5	Nama Ruas Jalan	15	Panjang (m)
6	Tahun	16	Lebar (m)
7	No Identitas	17	Tebal/Tinggi (m)
8	Lokasi/KM	18	Volume
9	Sub Kategori	19	Balai
10	Rusak Ruas Kiri Jalan	20	Keterangan

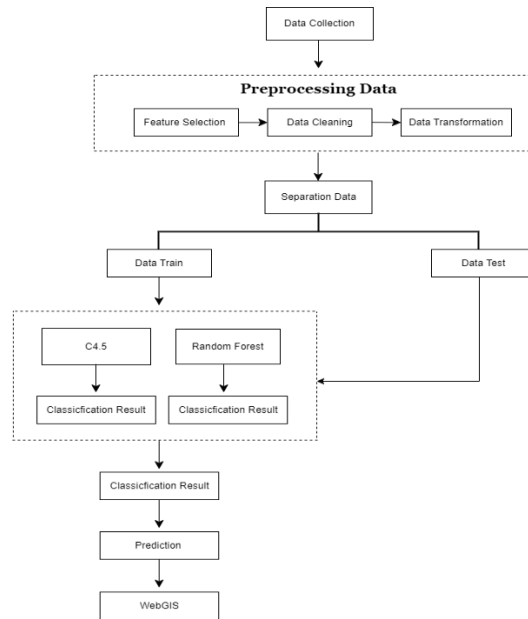
Tampilan dataset kerusakan jalan dapat dilihat pada Gambar 2.

Nama Ruas Jalan	Tahun	No Identitas	Lokasi / KM	Sub Kategori
Bts Kota Palopo - Bts Kab.Luwu Utara	2023	008	406 + 360	120 (Terkelupas)
Bts Kota Palopo - Bts Kab.Luwu Utara	2023	002	406 + 600	114 (Ambblas)
Bts Kota Palopo - Bts Kab.Luwu Utara	2023	001	409 + 450	252 (Kehilangan Permukaan)
Bts Kab.Luwu - Masamba	2023	001	411 + 650	840 (Lain-Lain)
Bts Kab.Luwu - Masamba	2023	001	412 + 100	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	001 / 001	412 + 700	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	002	412 + 800	840 (Lain-Lain)
Bts Kab.Luwu - Masamba	2023	001	413 + 300	114 (Ambblas)
Bts Kab.Luwu - Masamba	2023	002	415 + 515	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	001	415 + 700	120 (Terkelupas)
Bts Kab.Luwu - Masamba	2023	003 / 001	417 + 350	117/111 (Retak Buaya/Lubang)
Bts Kab.Luwu - Masamba	2023	004	418 + 475	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	004	418 + 650	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	005 / 002	419 + 185	117/120 (Retak Buaya/Terkelupas)
Bts Kab.Luwu - Masamba	2023	002	419 + 450	840 (Lain-Lain)
Bts Kab.Luwu - Masamba	2023	005	419 + 750	117/120 (Retak Buaya/Terkelupas)
Bts Kab.Luwu - Masamba	2023	005	420 + 700	117 (Retak Buaya)
Bts Kab.Luwu - Masamba	2023	001	424 + 000	511 (Patok KM Rusak)
Masamba - Bts.Kab Luwu Timur	2023	001 / 001	456 + 070	117/120 (Retak Buaya/Terkelupas)
Masamba - Bts.Kab Luwu Timur	2023	001	456 + 600	115 (Jembul)
Masamba - Bts.Kab Luwu Timur	2023	002	457 + 550	115 (Jembul)

Gambar 2. Dataset Kerusakan Jalan

## 3. Perancangan Sistem

Pada tahap ini dirancang alur sistem seperti pada Gambar 6 berikut.

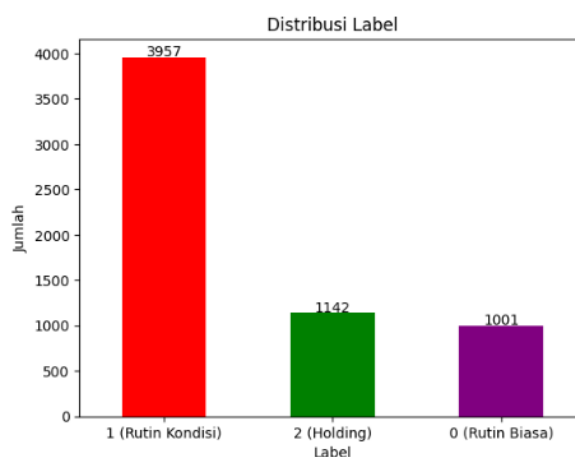


Gambar 1 *Flowchart* sistem prediksi kerusakan jalan

## HASIL DAN PEMBAHASAN

### 1. Pengumpulan Data dan Pelabelan Data

Dataset yang berhasil terkumpulkan berasal dari tiga ruas jalan. Dataset ini terdiri dari beberapa fitur tetapi hanya menggunakan 7 fitur yang dipilih untuk diklasifikasikan kategori kerusakan jalan dan disimpan dalam file dengan format CSV. Setelah dataset dikumpulkan dan dilakukan *feature selection*, setiap kerusakan diberi label secara manual untuk membagi kelas menjadi kelas ‘rutin biasa’, kelas ‘rutin kondisi’ dan kelas ‘holding’. Kelas rutin biasa diberi label 0, kelas rutin kondisi diberi label 1, kelas holding diberi label 2. Tampilan grafik pembagian kelas pada dataset setelah pengumpulan data ditunjukkan pada Gambar 7 berikut.



Gambar 2. Grafik pembagian kelas dataset

Gambar 7 menampilkan grafik pembagian kelas dataset yang digunakan untuk membangun model. Rincian jumlah kerusakan pada setiap kelas yang ada pada dataset ditunjukkan pada Tabel 2. Penjelasan setiap pelabelan pada Gambar 7 :



- **Label 0** : Dikatakan Rutin Biasa, jika kerusakan tidak mempengaruhi badan jalan, terletak di pinggir jalan atau bahu jalan, dengan jumlah volume sebesar 0 – 0.099 m<sup>3</sup> dan tidak ada kerusakan pada sisi kanan atau kiri jalan.
- **Label 1** : Dikatakan Rutin Kondisi, jika kerusakan mengenai badan jalan tetapi tidak mengenai lapisan bawah aspal, dengan jumlah volume sebesar 1.000 – 99.000 m<sup>3</sup>, dan ada kerusakan pada sisi kanan atau kiri jalan, sama halnya jika tidak ada kerusakan tetapi sisi rusak keduanya.
- **Label 2** : Dikatakan Holding, jika kerusakan mengenai badan jalan dan mengenai lapisan paling bawah aspal, dengan jumlah volume sebesar 100.00 - 675 m<sup>3</sup>, dan ada kerusakan pada sisi kanan dan kiri jalan.

Tabel 2. Rincian jumlah kerusakan jalan pada setiap kelas

Kelas	Jumlah Kerusakan Jalan
Rutin Biasa	1001
Rutin Kondisi	3957
Holding	1142
Jumlah	6100

## 2. Splitting Dataset

Dalam dataset yaitu data kerusakan jalan, dilakukan pemrosesan splitting data untuk membedakan antara fitur dan label. Setelah pemisahan ini, dataset tersebut selanjutnya dibagi menjadi dua bagian yaitu set pelatihan atau data *train* dan set pengujian atau data *test*, seperti ditunjukkan pada Tabel 3.

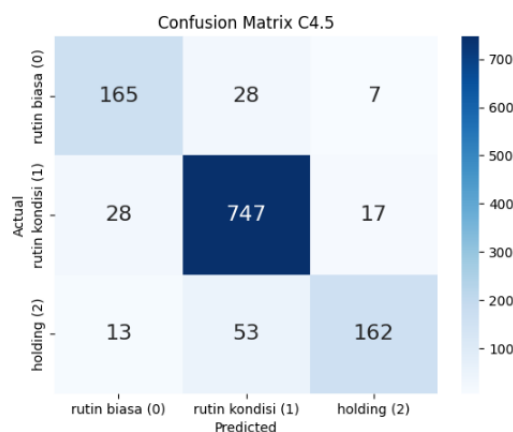
Tabel 3. Distribusi Data Train dan Data Test

Sub-dataset	Proporsi	Total Kerusakan Jalan
<i>Data train</i>	80%	4880
<i>Data test</i>	20%	1220
Jumlah	100%	6100

Tabel 3 menampilkan distribusi dataset untuk kerusakan jalan, di mana 80% dari total kerusakan yang jumlahnya 4,880 dijadikan sebagai data *train*. Sedangkan 20% yang jumlahnya 1,220 kerusakan dijadikan sebagai data *test*. Secara keseluruhan, dataset ini terdiri dari 6,100 kerusakan yang digunakan untuk pelatihan dan pengujian model.

## 3. Analisis Klasifikasi dengan C4.5

Untuk melakukan evaluasi kinerja model, penelitian ini menggunakan *confusion matrix*. *Matrix* ini digunakan untuk mengukur akurasi, presisi, *recall* dan *F1-Score* dari setiap kelas. Tampilan *confusion matrix* dari model C4.5 ditunjukkan pada Gambar 8.



Gambar 3 *Confusion matrix* model C4.5

Data yang salah prediksi dalam *confusion matrix* pada algoritma C4.5 yang menunjukkan ketidakcocokan antara label test dan label prediksi. Serta hasil perhitungan nilai akurasi, *recall*, presisi dan *F1-Score* menggunakan model C4.5 ditunjukkan pada Tabel 4.

Tabel 4 Evaluasi model C4.5

Kelas	Presisi	Recall	F1-Score
Rutin Biasa	0,80	0,82	0,81
Rutin Kondisi	0,90	0,94	0,92
Holding	0,87	0,71	0,78
Akurasi			0,88

Tabel 4 menampilkan evaluasi model C4.5 yang menghasilkan presisi sebesar 0,80 untuk kelas rutin biasa, yang menyatakan bahwa 80% dari semua sampel yang dikategorikan sebagai rutin biasa memang benar-benar berada dalam kelas tersebut. Model menghasilkan *recall* sebesar 0,82, yang menandakan bahwa dari semua sampel yang sebenarnya menjadi kelas rutin biasa, model berhasil mengklasifikasikan 82% dari data tersebut ke dalam kelas rutin biasa. *F1-Score* yang dihasilkan model adalah 0,81. Untuk kelas rutin kondisi, model menghasilkan presisi sebesar 0,90, yang menyatakan bahwa 90% dari semua sampel yang dikategorikan sebagai rutin kondisi memang benar-benar berada dalam kelas tersebut. Model menghasilkan *recall* sebesar 0,94, yang menandakan bahwa dari semua sampel yang sebenarnya menjadi kelas baik, model berhasil mengklasifikasikan 94% dari data tersebut ke dalam kelas rutin kondisi. *F1-Score* yang dihasilkan model adalah 0,92. Untuk kelas holding, model menghasilkan presisi sebesar 0,87, yang menyatakan bahwa 87% dari semua sampel yang dikategorikan sebagai holding memang benar-benar berada dalam kelas tersebut. Model menghasilkan *recall* sebesar 0,71, yang menandakan bahwa dari semua sampel yang sebenarnya menjadi kelas holding, model berhasil mengklasifikasikan 71% dari data tersebut ke dalam kelas holding. *F1-Score* yang dihasilkan model adalah 0,78. Secara umum, model mencapai akurasi sebesar 0,88, yang artinya model mampu memprediksi dengan akurat 88% dari seluruh kasus.

#### 4. Perbandingan Kinerja Model

Tabel 5 menampilkan perbandingan performa dari model - model klasifikasi yaitu C4.5 dan Random Forest berdasarkan perhitungan presisi, *recall*, *F1-Score* dan Akurasi.





Tabel 5 Perbandingan kinerja model

	<b>C4.5</b>	<b>Random Forest</b>
Presisi	87,9%	86,6%
Recall	82,6%	86,8%
<i>F1-Score</i>	87,8%	86,6%
Akurasi	88%	87%

Dalam analisa yang dilakukan berdasarkan hasil, didapatkan nilai presisi, *recall*, *F1-Score* dan akurasi dari algoritma C4.5, dan Random Forest sebagai berikut.

- **C4.5**

a) *Precision*

$$Precision = \frac{[(0,80 \times 200) + (0,90 \times 792) + (0,87 \times 228)]}{1220} = 0,879$$

b) *Recall*

$$Recall = \frac{0,82 + 0,94 + 0,71}{3} = 0,826$$

c) *F1-Score*

$$F1-Score = \frac{[(0,81 \times 200) + (0,92 \times 792) + (0,78 \times 228)]}{1220} = 0,878$$

d) Akurasi

$$Akurasi = \frac{165 + 747 + 162}{1220} = 0,880$$

- **Random Forest**

a) *Precision*

$$Precision = \frac{[(0,85 \times 189) + (0,89 \times 808) + (0,79 \times 223)]}{1220} = 0,866$$

b) *Recall*

$$Recall = \frac{0,75 + 0,93 + 0,74}{3} = 0,868$$

c) *F1-Score*

$$F1-Score = \frac{[(0,79 \times 189) + (0,91 \times 808) + (0,76 \times 223)]}{1220} = 0,866$$

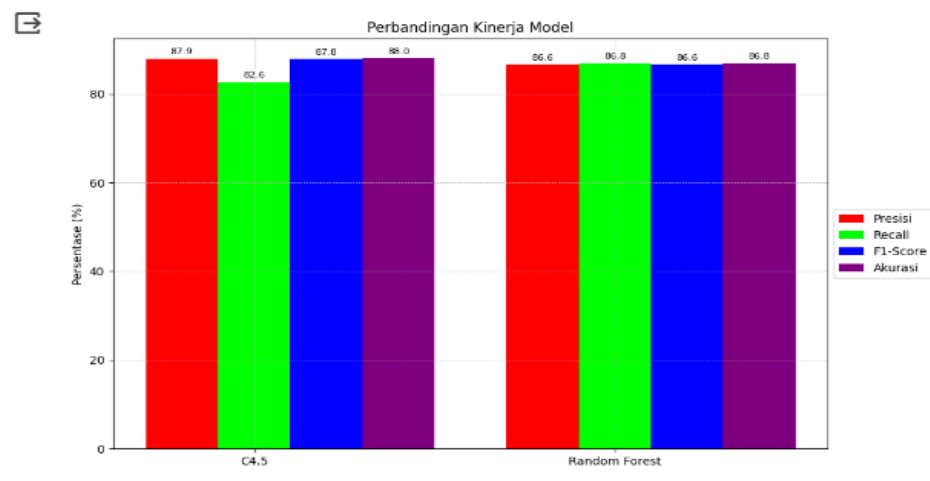
d) Akurasi

$$Akurasi = \frac{136 + 751 + 164}{1220} = 0,868$$

Berdasarkan perhitungan didapat bahwa Random Forest sebagai model terbaik dari model C4.5 yang telah diuji. Model Random Forest mencapai *precision* sebesar 86,6%, *recall* sebesar 86,8%, *F1-Score* sebesar 86,6%, dan akurasi 86,8%. Ini menandakan bahwa model Random Forest mampu memberikan hasil prediksi yang tepat dan konsisten. Di sisi lain, model C4.5 mencatat



precision 87,9%, recall 82,6%, F1-Score 87,8%, dan akurasi 88%. Kesimpulan yang didapat dari hasil evaluasi dan perhitungan model Random Forest yaitu model mampu memberikan prediksi yang lebih akurat dibandingkan dengan model C4.5. Tampilan grafik perbandingan kinerja dari kedua model ditunjukkan pada Gambar 9.



Gambar 94 Grafik Perbandingan Kinerja Model

Berdasarkan Gambar 9, random forest menunjukkan kinerja terbaik dalam klasifikasi kerusakan jalan di Luwu Raya pada tahun 2021-2023 dibanding model C4.5. Meski C4.5 berkinerja baik, random forest tetap baik dalam presisi, recall, F1-Score, dan akurasi.

### 5. Contoh Perhitungan Manual Model

Dalam menguji model menggunakan perhitungan manual dapat dilihat pada Gambar 10 berikut.

No	Sub Kategori	Nama Wilayah	Volume Kerusakan
1	Terkelupas	Palopo	0.084
2	Amblas	Palopo	0.036
3	Retak Buaya	Palopo	0.075

Gambar 105 Contoh Kasus

- Untuk Algoritma C4.5

#### Langkah 1: Menghitung Entropi Total

Entropi dari target variable (Volume Kerusakan):

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2$$

Di mana  $p_i$  adalah proporsi dari setiap nilai Volume Kerusakan. Dalam kasus ini, setiap kategori memiliki 1/3 proporsi (karena ada tiga nilai yang berbeda).



$$\rho_i = \frac{1}{3}, \text{ untuk } i = 1,2,3$$

$$\begin{aligned} \text{Jadi entropy total} &= - \left( \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} + \frac{1}{3} \log_2 \frac{1}{3} \right) \\ &= - 3 \times \frac{1}{3} \log_2 \frac{1}{3} \\ &= - \log_2 \frac{1}{3} \rightarrow = \log_2 3 = 1.585 \end{aligned}$$

### Langkah 2: Menghitung Gain

Karena setiap Sub Kategori unik, setiap subset hanya akan mengandung satu contoh, sehingga entropi subset akan menjadi nol

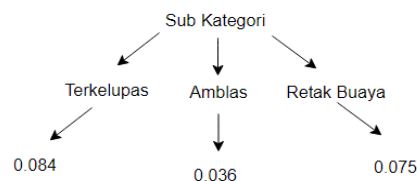
$$\text{Entropi(Terkelupas)} = 0$$

$$\text{Entropi(Ambblas)} = 0$$

$$\text{Entropi(Retak Buaya)} = 0$$

$$\begin{aligned} \text{Gain}(S, A) &= \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \\ &= 1.585 - \left( \frac{1}{3} \times 0 + \frac{1}{3} \times 0 + \frac{1}{3} \times 0 \right) \\ &= 1.585 \rightarrow \text{setiap fitur (0.528)} \end{aligned}$$

Karena setiap Sub Kategori memiliki gain yang sama (0.528) dan entropi yang sama (0), kita bisa memilih salah satu sebagai root node, sehingga, hasil prediksi C4.5 lebih sederhana dan lebih rentan terhadap overfitting karena hanya mengandalkan satu pohon keputusan. Gambar pohon keputusan C4.5 dapat dilihat pada Gambar 12 berikut.



Gambar 6 Pohon Keputusan C4.5

- **Untuk Algoritma Random Forest**

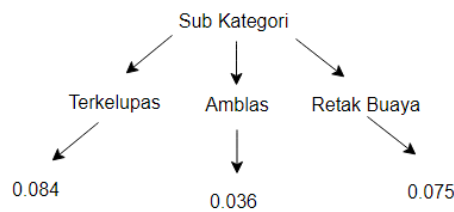
#### Langkah 1: Membuat Subset Data (Bootstrapping):

Kita membuat beberapa subset dari dataset asli dengan metode bootstrap (mengambil sampel dengan penggantian). Misalkan kita membuat 3 subset data:

- Dataset 1: [(Terkelupas, 0.084), (Ambblas, 0.036), (Retak Buaya, 0.075)]
- Dataset 2: [(Ambblas, 0.036), (Retak Buaya, 0.075), (Ambblas, 0.036)]
- Dataset 3: [(Retak Buaya, 0.075), (Terkelupas, 0.084), (Terkelupas, 0.084)]

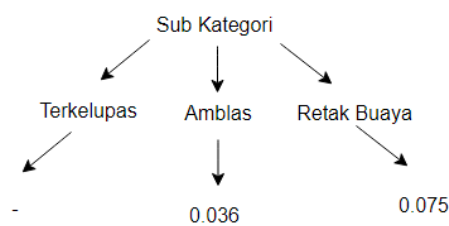
#### Langkah 2: Melatih Pohon Keputusan pada Setiap Subset Data

**Pohon 1** (dari Dataset 1): untuk melatih dataset acak yang telah ditentukan sebelumnya dapat dilihat pada Gambar 13 berikut.



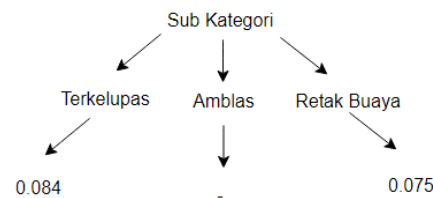
**Gambar 7 Pohon Keputusan 1**

**Pohon 2** (dari Dataset 2): untuk melatih dataset acak yang telah ditentukan sebelumnya dapat dilihat pada Gambar 14 berikut.



**Gambar 8 Pohon Keputusan 2**

**Pohon 3** (dari Dataset 3): untuk melatih dataset acak yang telah ditentukan sebelumnya dapat dilihat pada Gambar 15 berikut.



**Gambar 9 Pohon Keputusan 3**

### Langkah 3: Menggabungkan Hasil

Untuk membuat prediksi akhir, kita menggabungkan hasil dari semua pohon keputusan. Dalam kasus regresi seperti ini, maka mengambil rata-rata dari prediksi semua pohon.

#### **Prediksi untuk Sub Kategori = Terkelupas:**

- Pohon 1 memprediksi: 0.084
- Pohon 2 tidak memiliki data untuk Terkelupas.
- Pohon 3 memprediksi: 0.084

$$\text{Rata-rata} = \frac{0.084 + 0.084}{2} = \frac{0.168}{2} = 0.084$$

#### **Prediksi untuk Sub Kategori = Amblas:**



- a. Pohon 1 memprediksi: 0.036
- b. Pohon 2 memprediksi: 0.036
- c. Pohon 3 tidak memiliki data untuk Amblas.

$$\text{Rata-rata} = \frac{0.036+0.036}{2} = \frac{0.072}{2} = 0.036$$

**Prediksi untuk Sub Kategori = Retak Buaya:**

- a. Pohon 1 memprediksi: 0.075
- b. Pohon 2 memprediksi: 0.075
- c. Pohon 3 memprediksi: 0.075

$$\text{Rata-rata} = \frac{0.075+0.075+0.075}{3} = \frac{0.225}{3} = 0.075$$

Sehingga, algoritma Random Forest menggabungkan hasil dari beberapa pohon keputusan untuk mengurangi overfitting dan juga setiap pohon dilatih pada subset data yang berbeda, sehingga Random Forest menangkap lebih banyak variasi dalam data

### SIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Implementasi metode Random Forest untuk prediksi prioritas perbaikan pada kerusakan jalan di Balai Besar Pelaksanaan Jalan Nasional VI Makassar, dengan total 6.100 data kerusakan pada tiga ruas jalan antara tahun 2021–2023, menghasilkan prediksi prioritas perbaikan dengan kategori: kelas Rutin Biasa sebesar 15,7%, kelas Rutin Kondisi sebesar 68,3%, dan kelas Holding sebesar 16,0%.
2. Perbandingan antara algoritma C4.5 dan Random Forest menunjukkan bahwa keduanya memiliki akurasi yang hampir serupa. Namun, Random Forest menunjukkan hasil yang lebih konsisten dan sedikit lebih baik. Untuk algoritma C4.5, nilai presisi sebesar 87,9%, recall sebesar 82,6%, f1-score sebesar 87,8%, dan akurasi sebesar 88%. Sedangkan untuk Random Forest, nilai presisi sebesar 86,6%, recall sebesar 86,8%, f1-score sebesar 86,6%, dan akurasi sebesar 87%.
3. Visualisasi hasil prediksi dilakukan dengan memanfaatkan peta digital interaktif. Pengguna dapat mengunggah dan memverifikasi data kerusakan jalan melalui halaman homepage dan halaman data. Halaman hasil klasifikasi menampilkan peta digital yang memungkinkan pengguna untuk mengklik simbol lingkaran berwarna untuk melihat detail prediksi. Selain itu, halaman detail klasifikasi memberikan informasi lebih mendalam tentang hasil klasifikasi, dan halaman penambahan dataset yang dikelola melalui phpMyAdmin memastikan pemeliharaan dataset yang selalu terbaru.

Beberapa saran yang dapat diberikan untuk pengembangan sistem ini di masa depan antara lain:

1. Menggunakan lebih banyak model klasifikasi untuk memperkaya analisis dan prediksi.
2. Menggunakan dataset yang lebih besar dan beragam untuk meningkatkan akurasi dan keandalan model.
3. Menerapkan metode comparing dengan dua atau lebih algoritma untuk mendapatkan perbandingan yang lebih luas.
4. Menggunakan teknik preprocessing data yang lebih canggih dan beragam untuk meningkatkan kualitas data sebelum dilakukan pemodelan.



---

---

## DAFTAR PUSTAKA

- Adwang, M. (2020). *Peraturan Menteri Pekerjaan Umum Nomor 11/PRT/M/2010 tentang Tata Cara dan Persyaratan Layak Fungsi Jalan*.
- Andi Tenrisukki. (2002). *Konstruksi dan Pemeliharaan Jalan*. Jurnal Teknik Sipil.
- Arif, R. (2019). *Pengenalan Data Mining dalam Ilmu Komputer*. Yogyakarta: Penerbit Mitra.
- Bagian-Bagian Jalan Nasional et al. (2020). *Pengelolaan Jalan Nasional*. Jurnal Perencanaan Infrastruktur.
- BBPJN. (2024). *Tugas dan Fungsi Balai Besar Pelaksanaan Jalan Nasional*. Makassar: Kementerian Pekerjaan Umum.
- Frastian, N., Hendrian, S., & Valentino, V.H. (2018). Perbandingan Metode C4.5, Naïve Bayes, dan Random Forest dalam Menentukan Kelulusan Mata Kuliah. *Jurnal Teknologi dan Sistem Komputer*, 6(2), 151-159.
- Kompas.com. (2023). Jalan Rusak di Sulawesi Selatan Capai 2.009 Kilometer.
- Meikalyan, R. (2017). *Metode Komparasi dalam Penelitian*. Jurnal Metodologi Penelitian.
- Zainuddin, Zahir. "Detection and Counting of the Number of Cocoa Fruits on Trees Using UAV." *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, 2023.
- Novianti, A., et al. (2016). *Implementasi Algoritma C4.5 dalam Klasifikasi Data*. Jurnal Informatika.
- Pamuji, T., & Ramadhan, A. (2021). *Penerapan Random Forest untuk Klasifikasi Data*. Jurnal Teknologi dan Sistem.
- Pekerjaan et al. (2021). *Pelaporan Kerusakan Jalan oleh Penilik Jalan*. Jurnal Pekerjaan Umum.
- Pahrul, A. (2023). *Keunggulan Random Forest dalam Data Mining*. Jurnal Teknologi Informasi.
- Saragi, A., et al. (2021). *Kualitas Aspal Beton dan Pengaruhnya Terhadap Ketahanan Jalan*. Jurnal Konstruksi dan Material.
- Sukirman, S. (1992). *Perkerasan Jalan Raya: Teknologi dan Implementasi*. Yogyakarta: Penerbit Teknik.
- Supriyadi, H., et al. (2020). *Decision Tree untuk Klasifikasi Data*. Jurnal Teknik Komputer.
- Wu, X., & Kumar, V. (2016). *The Top Ten Algorithms in Data Mining*. Springer.
- Yuli Mardi. (2016). *Algoritma C4.5 dalam Data Mining*. Jurnal Informatika.