E-ISSN : 2988-1986
https://ejournal.warunayama.org/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 9 No 1 Tahun 2025

KLASIFIKASI PENANGANAN PRIORITAS KERUSAKAN JALAN TOL MENGGUNAKAN ALGORITMA C4.5

Yaumi Karimah, Samudi, Fahrizal
Program Studi Informasi, Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika

Email penulis : yaumikrmh@gmail.com, samudi.smx@bsi.ac.id, fahrizal.fzl@bsi.ac.id

ABSTRAKSI

Jalan tol berfungsi sebagai jalur alternatif yang penting karena mampu memberikan layanan yang lebih unggul dibandingkan jalan biasa, meskipun tetap memiliki potensi kerusakan akibat tingginya intensitas lalu lintas serta pengaruh lingkungan. Kerusakan yang terjadi pada jalan tol dapat berdampak pada keselamatan dan kenyamanan para pengguna, sehingga penanganan yang cepat dan tepat sangat diperlukan. Penelitian ini dilakukan untuk mengkaji kerusakan pada jalan tol sekaligus menetapkan prioritas penanganan dengan memanfaatkan algoritma C4.5. Dengan jumlah sampel sebanyak 142 data yang ditentukan menggunakan rumus Slovin. Atribut yang dianalisis meliputi aspek pelayanan, kategori dan jenis kerusakan, tingkat keparahan, serta kebutuhan penanganan segera. Proses klasifikasi melibatkan tahap pra-pemrosesan data, pembuatan model pohon keputusan menggunakan algoritma C4.5 dan validasi melalui aplikasi RapidMiner. Hasil penelitian memperlihatkan bahwa algoritma C4.5 dapat mengelompokkan prioritas penanganan kerusakan jalan tol secara efektif dan akurat, sehingga dapat dijadikan dasar dalam pengambilan keputusan untuk pemeliharaan infrastruktur jalan tol secara efisien dan tepat sasaran.

Kata Kunci : Algoritma C4.5, Klasifikasi, Data Mining

ABSTRACT

Toll roads function as an important alternative route because they are able to provide superior services compared to regular roads, although they still have the potential for damage due to high traffic intensity and environmental influences. Damage that occurs on toll roads can have an impact on the safety and comfort of users, so fast and precise handling is needed. This study was conducted to examine damage to toll roads and at the same time determine handling priorities by utilizing the C4.5 algorithm. With a sample size of 142 data determined using the Slovin formula. The attributes analyzed include service aspects, categories and types of damage, severity, and the need for immediate handling. The classification process involves data pre-processing stages, creating a decision tree model using the C4.5 algorithm and validation through the RapidMiner application. The results of the study show

Article History
Received: Juli 2025
Reviewed: Juli 2025
Published: Juli 2025

Plagiarism Checker No 677
Prefix DOI : Prefix DOI :
10.8734/Kohesi.v1i2.365
Copyright : Author
Publish by : Kohesi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

that the C4.5 algorithm can group the priority of handling toll road damage effectively and accurately, so that it can be used as a basis for decision making for efficient and targeted toll road infrastructure maintenance.

Keywords: C4.5 Algorithm, Classification, Data Mining

PENDAHULUAN

1.1. Latar Belakang

Jalan tol berperan sebagai jalur alternatif bagi para pengguna selain jalan biasa. Perbedaan utama terletak pada tingkat kualitas layanan yang disediakan. Pengguna jalan tol mendapatkan manfaat seperti waktu perjalanan yang lebih singkat, tersedianya area istirahat untuk perjalanan antar kota, serta akses cepat ke unit darurat yang siap memberikan bantuan. Namun, untuk fasilitas dan keuntungan tersebut, pengguna jalan tol dikenakan biaya tarif tol sebagai bentuk kompensasi. (Saadah dan Puspito 2023). Lalu lintas kendaraan di jalan tol diperkirakan akan meningkat cukup besar, sehingga berisiko menimbulkan berbagai macam kerusakan. Kerusakan pada jalan dapat mengganggu keselamatan serta kenyamanan para pengguna. Oleh sebab itu, penanganan kerusakan jalan yang cepat dan tepat sangat diperlukan guna menjaga mutu layanan dan menjamin keamanan.

Kerusakan jalan dipicu oleh berbagai faktor, di antaranya adalah usia jalan yang sudah melampaui batas rencana pemakaian, sistem drainase yang tidak berfungsi dengan baik, serta beban kendaraan dan volume lalu lintas yang melebihi kapasitas. Kondisi-kondisi ini menyebabkan umur jalan menjadi lebih singkat dibandingkan dengan yang telah ditetapkan dalam standar perencanaan (Saadah dan Puspito 2023).

Dalam konteks ini, algoritma C4.5 menawarkan pendekatan yang efektif untuk menganalisis data kerusakan jalan dan menentukan prioritas penanganan. Algoritma C4.5 adalah pengembangan dari Algoritma ID3 yang dikembangkan oleh Ross Quinlan. Algoritma ini digunakan untuk mengklasifikasikan data yang memiliki atribut baik dalam bentuk numerik maupun kategorikal. Hasil klasifikasi berupa aturan-aturan yang dapat dimanfaatkan untuk memprediksi nilai atribut diskrit pada data baru. (Franseda et al. 2020). Dengan mengklasifikasikan jenis kerusakan berdasarkan karakteristiknya, proses pengambilan keputusan dalam penanganan kerusakan jalan sebaiknya dilakukan secara lebih efektif, efisien, dan merata agar hasilnya optimal. (Justam et al. 2024).

Dengan demikian, penelitian ini bertujuan untuk menganalisis kerusakan pada jalan tol dan menentukan skala prioritas penanganan menggunakan algoritma C4.5. Hasil dari penelitian ini diharapkan dapat menjadi acuan bagi pengambilan keputusan dalam pemeliharaan infrastruktur jalan tol di masa mendatang.

1.2. Identifikasi Masalah

Berdasarkan dari latar belakang maka adapun identifikasi masalah yaitu sebagai berikut:
Bagaimana mengetahui seberapa parah tingkat kerusakan temuan pada jalan tol?
Dengan menggunakan algoritma C4.5 apakah dapat mengetahui klasifikasi penentuan penanganan kerusakan jalan tol?

1.3. Rumusan Masalah

Adapun rumusan masalah dari penulisan skripsi ini adalah:
Untuk pengambilan keputusan dengan cepat dalam pemeliharaan jalan tol bagi badan pemilik usaha jalan tol;

Untuk memperoleh hasil klasifikasi penanganan kerusakan yang akurat dengan metode algoritma C4.5.

1.4. Tujuan Penelitian

Adapun tujuan penelitian dari penulisan skripsi ini adalah:
Untuk mengetahui penanganan kerusakan pada jalan tol lebih cepat menggunakan algoritma C4.5;
Sebagai salah satu syarat kelulusan Program Sarjana Prodi Sistem Informasi Universitas Bina Sarana Informatika.

1.5. Ruang Lingkup

Dalam hal ini terdapat ruang lingkup permasalahan, maka penulis membuat batasan sebagai berikut:
Penerapan data mining dengan metode C4.5 untuk klasifikasi penentuan penanganan prioritas kerusakan pada jalan tol;
Penelitian ini dilakukan dengan mengambil data dari observasi di lapangan;
Penelitian ini menggunakan *tools RapidMiner*. |

METODOLOGI PENELITIAN

2.1. Data Mining

Data mining yang juga dikenal sebagai *knowledge discovery in database* (KDD) merupakan proses untuk mengumpulkan dan memanfaatkan data historis untuk menemukan pola, keteraturan atau hubungan tertentu dalam kumpulan data besar. Secara sederhana, KDD adalah proses untuk menemukan informasi yang bermanfaat dari data yang tersedia. Dalam praktiknya, data mining memanfaatkan berbagai teknik seperti statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengidentifikasi informasi dan pengetahuan yang relevan dari database berukuran besar. Dengan kata lain, data mining adalah rangkaian langkah untuk menggali nilai tambah berupa pengetahuan baru dari data yang sebelumnya tidak dapat ditemukan secara manual (Husna, Rahman, dan Juhari 2022).

Menurut Vulandari, 2017 dalam (Yunus et al. 2021) “Data mining adalah serangkaian proses yang bertujuan menggali informasi berharga yang sebelumnya sulit atau bahkan tidak mungkin ditemukan secara manual dari sebuah basis data. Dari berbagai definisi tersebut, dapat disimpulkan bahwa data mining merupakan metode pengolahan data dalam jumlah besar bahkan mencapai ribuan dengan menggunakan berbagai teknik dan perhitungan khusus untuk mengorganisir data secara terstruktur sehingga menghasilkan informasi yang berkualitas.”

2.2. Pohon Keputusan (*Decision Tree*)

Menurut Z. Sitorus And A. Widarma, 2022 dalam (Maulana Sidiq, Faqih, dan Dwilestari 2024) “*Decision Tree* merupakan salah satu metode klasifikasi dalam *machine learning* dan data mining yang menggunakan bentuk struktur pohon atau model hirarkis untuk membantu proses pengambilan keputusan. Dalam struktur pohon ini, terdapat beberapa elemen penting, yaitu node, cabang dan daun. Node merepresentasikan fitur atau atribut data, cabang menunjukkan aturan atau kondisi keputusan, sedangkan daun menggambarkan hasil akhir dari keputusan tersebut. Cara kerja *Decision Tree* adalah dengan memecah dataset menjadi bagian-bagian yang lebih kecil berdasarkan atribut tertentu, sehingga memudahkan dalam proses klasifikasi data secara sistematis.”

2.3. Algoritma C4.5 (*Decision Tree*)

Algoritma *Decision Tree* (C4.5) adalah salah satu metode dalam data mining yang digunakan untuk mengklasifikasikan data, sehingga memudahkan dapat membantu menyelesaikan berbagai masalah klasifikasi (Khotimah 2022).

Menurut Novianti, 2016 dalam (Justam et al. 2024) Data mining adalah serangkaian proses yang bertujuan menggali informasi berharga yang sebelumnya sulit atau bahkan tidak mungkin ditemukan secara manual dari sebuah basis data. Dari berbagai definisi tersebut, dapat disimpulkan bahwa data mining merupakan metode pengolahan data dalam jumlah besar, bahkan mencapai ribuan, dengan menggunakan berbagai teknik dan perhitungan khusus untuk mengorganisir data secara terstruktur sehingga menghasilkan informasi yang berkualitas.
Secara garis besar, algoritma C4.5 membangun pohon keputusan melalui beberapa tahapan utama, yaitu:
Pilihlah atribut yang akan dijadikan sebagai akar pohon keputusan
buatlah cabang untuk setiap kemungkinan nilai dari atribut tersebut
Bagi data ke dalam masing-masing cabang sesuai dengan nilainya
Lakukan proses ini secara berulang pada setiap cabang hingga seluruh data dalam satu cabang memiliki kelas yang sama. Pemilihan atribut sebagai akar didasarkan pada atribut yang memiliki nilai gain tertinggi di antara atribut-atribut yang tersedia.
Untuk menghitung gain digunakan rumus berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} \; x \; Entropy(S_i)$$

Sumber: (Nurdiyani Sari dan Azharina 2024)
Gambar II.1
Rumus untuk menghitung Gain

Keterangan:
S = himpunan kasus
A = atribut
n = jumlah partisi atribut A
|Si| = jumlah kasus pada partisi ke-i
|S| = jumlah kasus dalam S

Sementara itu, perhitungan nilai entropy digunakan rumus:

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

Sumber: (Nurdiyani Sari dan Azharina 2024)
Gambar II.2
Rumus untuk menghitung entropy

Keterangan:
S = himpunan kasus
A = fitur
n = jumlah partisi S
pi = proporsi dari Si terhadap S
Metode yang dipakai pada penelitian merupakan teknik data mining dengan algoritma C4.5 yang diterapkan pada software *RapidMiner*.

2.4. Penelitian Terkait

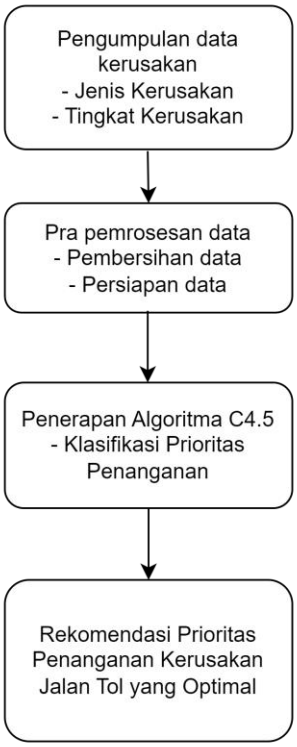
Penelitian mengenai klasifikasi penanganan Klasifikasi Penanganan Prioritas Kerusakan Jalan Tol Menggunakan Algoritma C4.5 adalah sebagai berikut:

Dalam penelitian (Justam et al. 2024) Penelitian ini menggunakan algoritma C4.5 dan Random Forest untuk pemetaan kerusakan jalan berbasis WebGIS, dengan hasil algoritma C4.5 mencapai presisi 87,9%, recall 82,6%, f1-score 87,88%, dan akurasi 88%. Sistem informasi pemetaan yang dikembangkan berfungsi untuk menentukan prioritas perbaikan jalan secara efektif. Hasil penelitian menunjukkan bahwa kedua algoritma memberikan performa yang baik, namun sistem ini khususnya memanfaatkan algoritma C4.5 dalam menghasilkan peta prioritas perbaikan jalan melalui platform WebGIS.

Dalam penelitian (Karyadiputra dan Hijriana 2019) judul penelitian ini adalah Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Penentuan Daftar Prioritas Pengembangan Jembatan yang mencakup seluruh tahapan mulai dari pengujian awal hingga evaluasi akhir hasil penelitian maka dapat disimpulkan bahwa penerapan teknik klasifikasi data mining menggunakan algoritma C4.5 bisa diterapkan sebagai salah satu alternatif untuk mengevaluasi daftar prioritas pengembangan jembatan di Provinsi Kalimantan Selatan dengan akurasi sebesar 82.84% dengan nilai AUC sebesar 0,825 sehingga termasuk dalam kategori *good classification*.

2.5. Kerangka Pemikiran

Penelitian ini penentuan prioritas dalam penanganan kerusakan pada jalan tol dengan memanfaatkan metode klasifikasi menggunakan algoritma pohon keputusan C4.5. Sebagai infrastruktur yang sangat penting, jalan tol sering mengalami berbagai jenis kerusakan dengan tingkat keparahan yang berbeda-beda, yang berpengaruh pada keselamatan pengguna jalan. Karena itu, dibutuhkan adanya sistem yang mampu mengklasifikasikan serta menetapkan prioritas penanganan kerusakan secara efektif dan efisien.



Gambar II.3
Alur Penelitian

Proses penelitian dimulai dengan pengumpulan data kerusakan jalan tol berupa jenis dan tingkat kerusakan. Tahap berikutnya adalah melakukan pra-pemrosesan data guna membersihkan serta mempersiapkan data agar siap untuk dianalisis. Kemudian, algoritma C4.5

diterapkan untuk mengklasifikasikan tingkat prioritas penanganan berdasarkan pola data yang ditemukan, hasil klasifikasi ini dievaluasi menggunakan metrik akurasi. Dengan demikian, hasil penelitian diharapkan dapat memberikan rekomendasi pemeliharaan jalan dapat dilakukan secara efisien dan tepat pada sasaran.

2.6. Hipotesis Penelitian

Hipotesis penelitian merupakan suatu dugaan awal yang disusun berdasarkan teori dan kajian pustaka, yang kemudian akan diuji kebenarannya melalui pengumpulan data dan analisis selama proses penelitian.
H0: Algoritma C4.5 mampu membangun model klasifikasi prioritas yang memiliki tingkat akurasi yang tinggi.
H1: Faktor-faktor seperti jenis kerusakan, tingkat keparahan dan volume lalu lintas berpengaruh signifikan terhadap klasifikasi prioritas penanganan.

HASIL DAN PEMBAHASAN

3.1. Proses Pengumpulan Data

Data yang diterapkan dalam membentuk pohon keputusan adalah data kerusakan jalan tol PT Jasamarga Jalanlayang Cikampek dengan kriteria yang telah ditetapkan dalam mengklasifikasi penanganan segera kerusakan jalan tol. Selanjutnya, dilakukan tahap pra-pemrosesan data dengan tujuan membersihkan serta menyiapkan data agar siap digunakan dalam proses pembuatan pohon keputusan.

3.2. Tahap Pra Pemrosesan Data

Menurut Sugiyono, 2013 dalam (Subhaktiyasa 2024) “Populasi adalah objek atau subjek yang menjadi yang menjadi fokus penelitian, sedangkan sampel yakni sebagian dari populasi yang dipilih untuk mewakili sifat dan karakteristik keseluruhan populasi tersebut”. Adapun populasi dalam penelitian ini adalah data kerusakan bulan Januari, Februari, Maret, April tahun 2025
Menurut Sugiyono, 2013 dalam (Subhaktiyasa 2024) “Sampel merupakan bagian dari populasi yang dipilih untuk menggambarkan atau mewakili ciri-ciri keseluruhan populasi”. Menurut Sugiyono, 2019 dalam (Cahyadi 2022) “Sampel ialah bagian dari populasi yang dipilih untuk mewakili karakteristik yang ada pada seluruh populasi tersebut”.

1. Penentuan Jumlah Sampel Dengan Rumus Slovin

Populasi sasaran berjumlah 220 data, yang kemudian diolah dengan menggunakan metode rumus Slovin, dengan tingkat margin of error yang ditetapkan sebesar 5% atau 0,05 dengan rumus berikut:

$$n = \frac{N}{1 + N(e)^2}$$

Gambar III.4
Rumus Slovin

Keterangan:
n= Ukuran sampel
N = Ukuran populasi
e = Tingkat kesalahan (5%)
adapun perhitungan sampel dengan rumus slovin sebagai berikut:
n = 220 (1 + (220 x 0,0025))
n = 220 / (1+ 0,55)
n = 220 / 1,55
n = 141,95
n= 142

Jumlah sampel yang dilakukan dalam penelitian ini sebanyak 142 data.
Penjelasan dari atribut data yang dipakai adalah sebagai berikut:

1. Substansi Pelayanan

Menjelaskan aspek pelayanan yang menjadi fokus utama, yaitu antara "Kondisi Jalan Tol" dan Keselamatan".

2. Kategori Kerusakan

Mengelompokkan jenis kerusakan berdasarkan kategori tertentu, seperti "Perkerasan Jalan Utama", "Penerangan Jalan Umum".

3. Jenis Kerusakan

Menjelaskan jenis kerusakan atau spesifikasi kerusakan yang terjadi, seperti "karet proteksi", "lampu mati", "parapet pecah".

4. Tingkat Kerusakan

Menjelaskan tingkat keparahan kerussakan yaitu "parah" atau "sedang".

5. Penanganan Segera

Merupakan target atau label yang menunjukkan apakah suatu kasus membutuhkan penanganan segera atau tidak.

No	substansi_pelayanan	kategori_kerusakan	jenis_kerusakan	tingkat_kerusakan	penanganan_segera
1	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
2	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
3	Kondisi Jalan Tol	Perkerasan Jalan Utama	Expansion joint	Parah	Ya
4	Kondisi Jalan Tol	Perkerasan Jalan Utama	Expansion joint	Parah	Ya
5	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
6	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Ya
7	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
8	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Ya
9	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Sedang	Tidak
10	Kondisi Jalan Tol	Perkerasan Jalan Utama	Expansion joint	Parah	Ya
11	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
12	Kondisi Jalan Tol	Perkerasan Jalan Utama	Expansion joint	Parah	Ya
13	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
14	Kondisi Jalan Tol	Median Concrete Barier	Parapet pecah	Sedang	Tidak
15	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
16	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
17	Kondisi Jalan Tol	Perkerasan Jalan Utama	beton transisi	Parah	Ya
18	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
19	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Parah	Ya
20	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
21	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
22	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
23	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
24	Kondisi Jalan Tol	Perkerasan Jalan Utama	Karet proteksi	Sedang	Tidak
...	...	...	...	...	...
142	Keselamatan	Penerangan Jalan Umum	lampu mati	Parah	Ya

Sumber: Penelitian (2025)

Gambar III.5
Dataset

3.3. Tahap Pre-Processing

Tahap ini dilakukan untuk memastikan bahwa dataset yang digunakan tidak mengandung data ganda maupun nilai yang kosong atau tidak valid (Missing Value). Langkah ini penting agar proses pengolahan data di *RapidMiner* dapat berjalan dengan lancar. Setelah dilakukan pemeriksaan, tidak ditemukan data duplikat maupun nilai yang hilang, seperti gambar dibawah ini:

Name	Type	Missing	Statistics		Filter (5 / 5 attributes): Search for Attributes
substansi_pelayanan	Binominal	0	Negative: Kondisi Jalan Tol	Positive: Keselamatan	Values: Kondisi Jalan Tol (107), Keselamatan (35)
kategori_kerusakan	Nominal	0	Least: Median C [...] arier (2)	Most: Perkeras [...] ama (105)	Values: Perkerasan Jalan Utama (105), Penerangan Jala
jenis_kerusakan	Nominal	0	Least: Parapet pecah (2)	Most: Karet proteksi (38)	Values: Karet proteksi (38), lampu mati (35), ...[5 more]
tingkat_kerusakan	Binominal	0	Negative: Parah	Positive: Sedang	Values: Parah (98), Sedang (44)
penanganan segera	Binominal	0	Negative: Ya	Positive: Tidak	Values: Ya (100), Tidak (42)

Gambar III.6
Pengecekan Missing Values

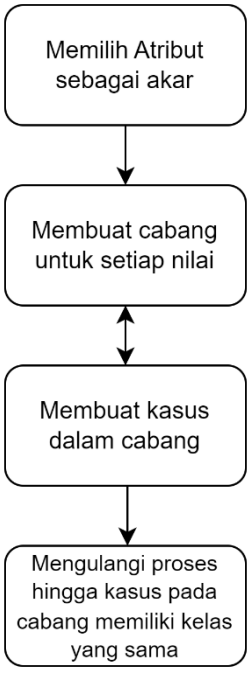
3.4. Pemrosesan Data Menggunakan Metode Algoritma C4.5

3.4.1. *Decision Tree*

Berikut ini adalah tahapan serta proses perhitungan dan pengolahan data dengan menggunakan *Decision Tree*.

Melakukan perhitungan menggunakan *Decision Tree*

Pada umumnya, tahapan dalam membangun *Decision Tree* dapat dilakukan melalui beberapa langkah berikut ini:



Sumber: Penelitian (2025)

Gambar III.7
Proses *Decision Tree*

a. Pilihlah atribut yang akan dijadikan sebagai akar pohon keputusan
b. buatlah cabang untuk setiap kemungkinan nilai dari atribut tersebut
c. Bagi data ke dalam masing-masing cabang sesuai dengan nilainya
d. Lakukan proses ini secara berulang pada setiap cabang hingga seluruh data dalam satu cabang memiliki kelas yang sama. Pemilihan atribut sebagai akar didasarkan pada atribut yang memiliki nilai gain tertinggi di antara atribut-atribut yang tersedia

Untuk menghitung gain digunakan rumus berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{S} \; x \; Entropy(S_i)$$

Keterangan:
S = himpunan kasus
A = atribut
n = jumlah partisi atribut A
|Si| = jumlah kasus pada partisi ke-i
|S| = jumlah kasus dalam S
Sementara itu, perhitungan nilai entropy digunakan rumus:

Keterangan :
S = himpunan kasus
A = fitur
n = jumlah partisi S
pi = proporsi dari Si terhadap S

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

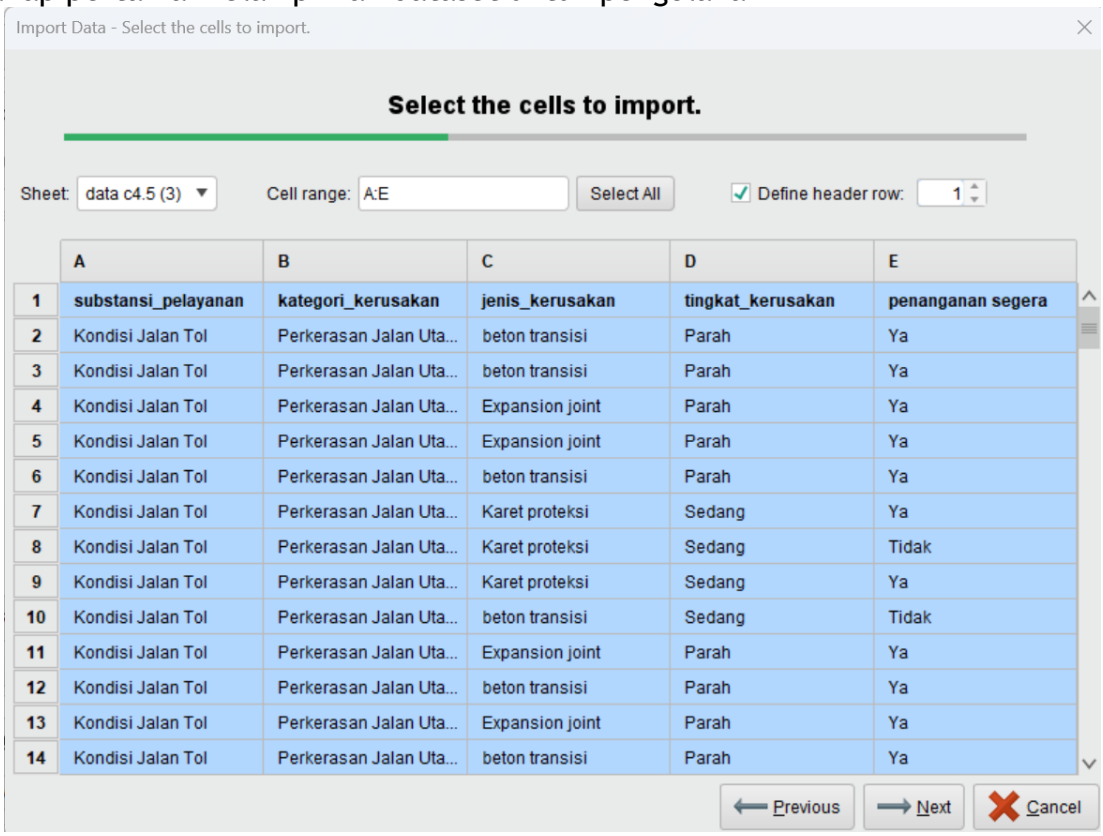
3.5. Pengujian Menggunakan *RapidMiner*

Setelah data diklasifikasikan, langkah berikutnya adalah mengolahnya menggunakan RapidMiner sebagai alat untuk memvalidasi keakuratan data. Jika data terbukti akurat, maka proses pengolahan dilanjutkan untuk memperoleh hasil klasifikasi terkait penanganan segera. Hasil pengolahan tersebut kemudian digunakan sebagai dasar pengambilan keputusan secara tepat melalui *RapidMine*r.

3.5.1. Proses Pengolahan Data Menggunakan *RapidMiner*

Berikut ini adalah model *Decision Tree* yang telah dibuat dengan menganalisis data menggunakan aplikasi *RapidMiner*.

a. Tahap pertama melampirkan dataset untuk pengolahan *Decision Tree*



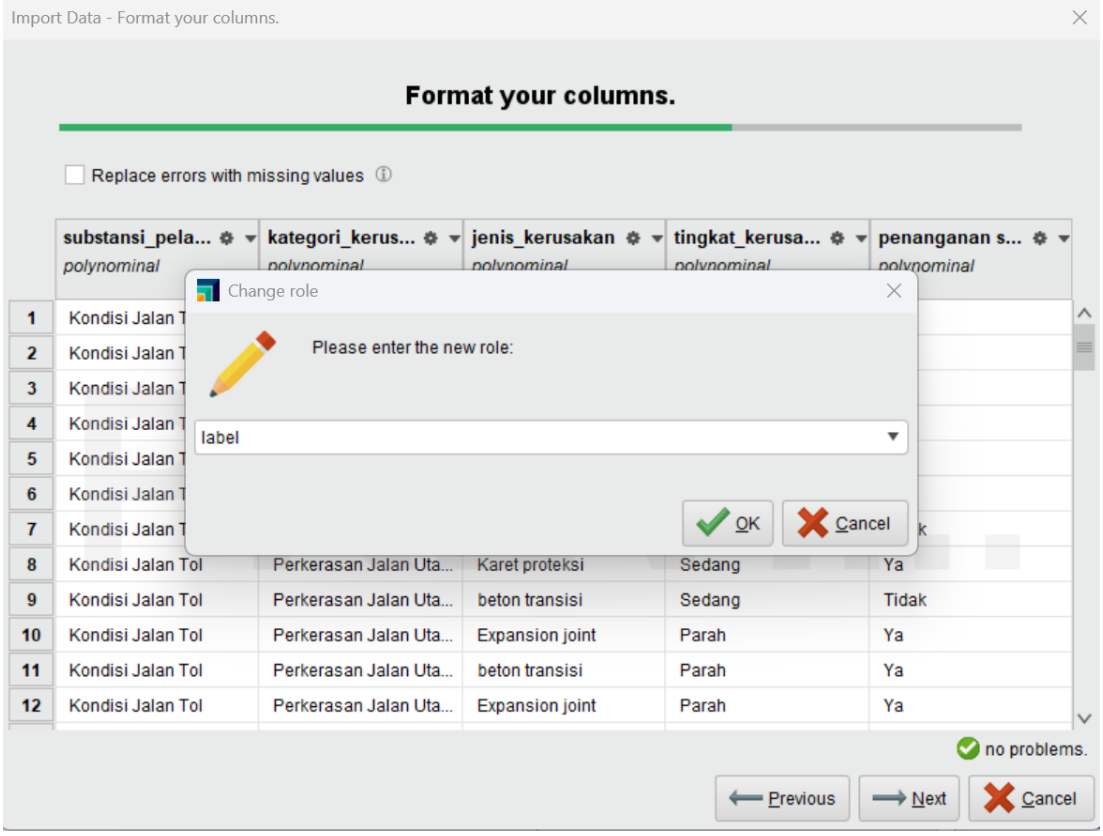
	A	B	C	D	E
1	substansi_pelayanan	kategori_kerusakan	jenis_kerusakan	tingkat_kerusakan	penanganan segera
2	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Parah	Ya
3	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Parah	Ya
4	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Expansion joint	Parah	Ya
5	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Expansion joint	Parah	Ya
6	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Parah	Ya
7	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Karet proteksi	Sedang	Ya
8	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Karet proteksi	Sedang	Tidak
9	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Karet proteksi	Sedang	Ya
10	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Sedang	Tidak
11	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Expansion joint	Parah	Ya
12	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Parah	Ya
13	Kondisi Jalan Tol	Perkerasan Jalan Uta...	Expansion joint	Parah	Ya
14	Kondisi Jalan Tol	Perkerasan Jalan Uta...	beton transisi	Parah	Ya

Sumber: Penelitian (2025)

Gambar III.8
Dataset Kerusakan Jalan

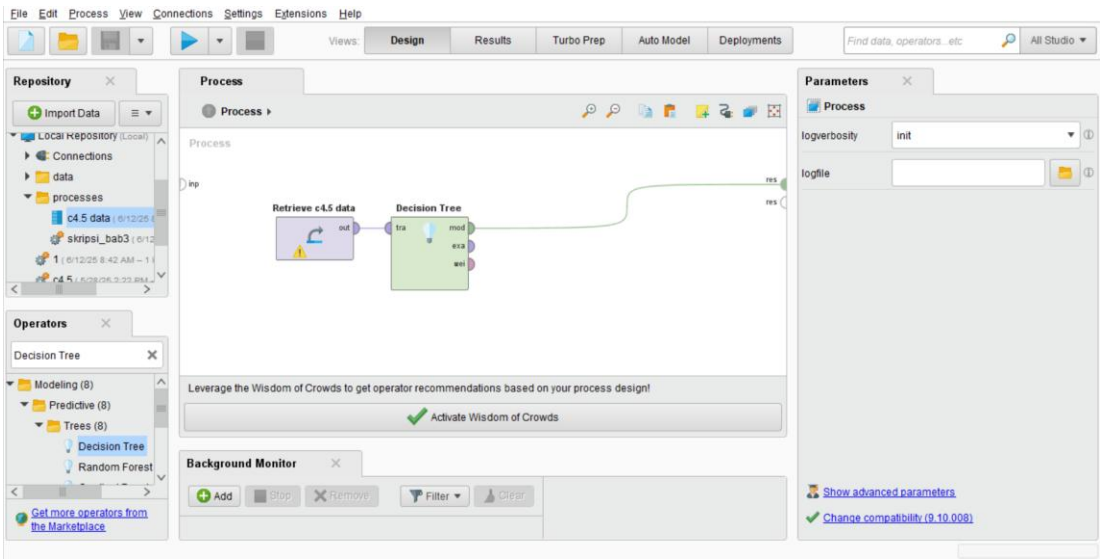
Dataset yang dimasukkan untuk pengolahan data yaitu dari dataset pengujian data kerusakan sebanyak 142 data uji, dengan atribut substansi pelayanan, kategori kerusakan, jenis kerusakan, tingkat kerusakan dan penanganan segera.

b. Selanjutnya, mengubah role menjadi label yang digunakan pada aplikasi *RapidMiner*.

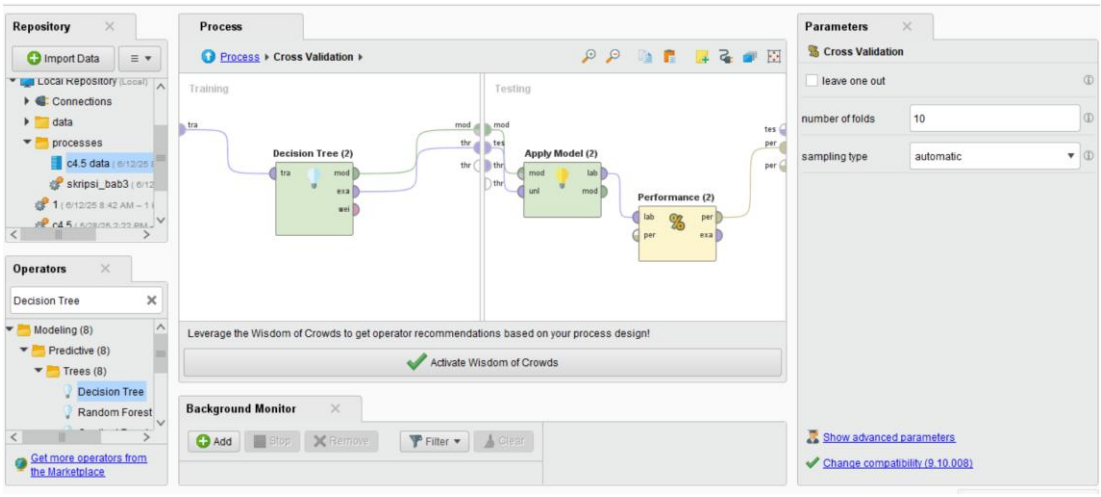


Gambar III.9
Penambahan Label Pada Atribut Label

c. Kemudian, input operator *Decision Tree* dengan mengetik di kolom operator *Decision Tree* untuk mengolah dataset menjadi model *Decision Tree*.

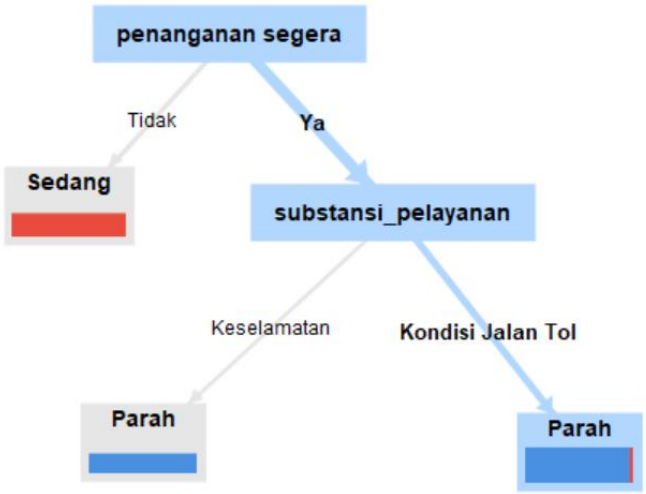


Gambar III.10
Pengolahan Data Menggunakan *RapidMiner*



Gambar III.11
Proses *Apply Model* dan *Performance*

d. Berikut adalah hasil *Decision Tree* yang dibuat dengan menganalisis data menggunakan aplikasi *RapidMiner*.



Gambar III.12
Hasil *Decision Tree* pada *RapidMiner*

Dari pohon keputusan pada gambar III.9 didapatkan Rule untuk klasifikasi penanganan segera yaitu sebagai berikut:

Tree

```
hasil = Tidak: Sedang {Parah=0, Sedang=42}
hasil = Ya
|   substansi_pelayanan = Keselamatan: Parah {Parah=35, Sedang=0}
|   substansi_pelayanan = Kondisi Jalan Tol: Parah {Parah=63, Sedang=2}
```

Gambar III.13
***Rule Model* Pohon Keputusan**

Berdasarkan hasil pemodelan *Decision Tree* pada gambar tersebut, deskripsi dari rule pohon keputusan dapat dijelaskan bahwa:

1. Jika hasilnya "Tidak", maka seluruh data dikategorikan sebagai tingkat kerusakan "Sedang" (dengan jumlah 42 kasus dan tidak ada kasus "Parah").
2. Jika hasilnya "Ya", maka dilakukan pemisahan lebih lanjut berdasarkan atribut "substansi_pelayanan".
3. Jika "substansi_pelayanan" adalah "Keselamatan", maka seluruh kasus masuk ke kategori "Parah" (sebanyak 35 kasus, tanpa ada kasus "Sedang").
4. Jika "substansi_pelayanan" adalah "Kondisi Jalan Tol", maka sebagian besar kasus dikategorikan sebagai "Parah" (63 kasus), dan hanya 2 kasus yang masuk kategori "Sedang".

Berdasarkan hasil yang diperoleh dapat diketahui bahwa atribut "substansi_pelayanan" memiliki dampak terhadap tingkat kerusakan, terutama pada data dengan hasil "Ya". Faktor keselamatan dan kondisi jalan tol yang buruk cenderung menghasilkan tingkat kerusakan yang tinggi.

3.6. Evaluation

1. Nilai *Accuracy*

Nilai *Accuracy* merupakan presentase yang menggambarkan tingkat ketepatan model dalam mengklasifikasikan data. Nilai akurasi dihitung dengan membandingkan jumlah data yang berhasil diklasifikasikan dengan benar terhadap jumlah total data yang diuji. Berikut merupakan dari hasil nilai akurasi algoritma C4.5.

accuracy: 98.59%

	true Ya	true Tidak	class precision
pred. Ya	98	0	100.00%
pred. Tidak	2	42	95.45%
class recall	98.00%	100.00%	

Gambar III.14
Nilai *Accuracy*

2. Nilai *Precision*

Nilai *precision* digunakan untuk mengukur performa sistem dengan menghitung jumlah data yang berhasil diklasifikasikan dengan benar dibandingkan dengan jumlah total data yang diprediksi sebagai positif, termasuk yang salah diklasifikasikan. Nilai presisi diperoleh dengan membagi jumlah prediksi benar dengan total prediksi positif yang dihasilkan oleh model.

accuracy: 98.59%

	true Ya	true Tidak	class precision
pred. Ya	98	0	100.00%
pred. Tidak	2	42	95.45%
class recall	98.00%	100.00%	

Gambar III.15
Nilai *Precision*

3. Nilai *Recall*

Nilai *recall* merupakan rasio antara jumlah data yang benar-benar teridentifikasi sebagai positif dengan total data yang memang berstatus positif yaitu penjumlahan antara *true positive* dan *false negative*.

accuracy: 98.59%

	true Ya	true Tidak	class precision
pred. Ya	98	0	100.00%
pred. Tidak	2	42	95.45%
class recall	98.00%	100.00%	

Gambar III.16
Nilai Recall

3.7. Kesimpulan Pengujian

Berdasarkan hasil pengujian terhadap data kerusakan jalan tol menggunakan algoritma C4.5 di aplikasi RapidMiner, berikut adalah beberapa poin utama yang dapat disimpulkan:

a. Data yang digunakan:

Analisis dilakukan pada 142 sampel data kerusakan jalan tol yang dikumpulkan selama Januari hingga April 2025, dengan 142 data digunakan sebagai data uji setelah melalui pra pemprosesan sehingga data yang digunakan bersih dari duplikasi maupun nilai yang hilang

b. Tahapan Pengujian

Pengujian dilakukan dengan membangun model *Decision Tree* menggunakan algoritma C4.5 di RapidMiner, di mana pemilihan atribut utama didasarkan pada nilai gain tertinggi dan model divalidasi untuk mengetahui performa klasifikasinya

c. Hasil Model *Decision Tree*

Model pohon keputusan yang dihasilkan memunculkan aturan (rule) klasifikasi penanganan segera kerusakan jalan tol, dengan atribut "substansi_pelayanan" sebagai faktor yang sangat berpengaruh, khususnya pada data yang membutuhkan penanganan segera ("Ya")1.
Jika hasil klasifikasi "Tidak", seluruh data masuk kategori tingkat kerusakan "Sedang".
Jika hasilnya "Ya" dan "substansi_pelayanan" adalah "Keselamatan", maka semua kasus dikategorikan sebagai "Parah".
Jika "substansi_pelayanan" adalah "Kondisi Jalan Tol", sebagian besar kasus masuk kategori "Parah" dan hanya sebagian kecil yang "Sedang"

d. Evaluasi Kinerja Model

Kinerja model dinilai berdasarkan nilai accuracy, precision, dan recall yang menunjukkan bahwa model mampu mengklasifikasikan data uji dengan tingkat ketepatan dan performa yang baik.
Model *Decision Tree* berbasis algoritma C4.5 yang dibangun mampu mengklasifikasikan penanganan segera kerusakan jalan tol dengan akurasi yang baik, serta menghasilkan aturan klasifikasi yang jelas dan mudah dipahami. Atribut "substansi_pelayanan" terbukti menjadi faktor utama dalam menentukan tingkat kerusakan dan kebutuhan penanganan segera pada kasus kerusakan jalan tol PT Jasamarga Jalanlayang Cikampek

KESIMPULAN

4.1. Kesimpulan

Kesimpulan yang diperoleh dari penulisan skripsi ini adalah sebagai berikut:
Berdasarkan penelitian yang memanfaatkan algoritma C4.5 untuk mengklasifikasikan prioritas penanganan kerusakan jalan tol, diperoleh tingkat akurasi sebesar 98,59%. Angka ini diperoleh melalui proses pengujian dan validasi menggunakan aplikasi *RapidMiner* terhadap 142 data kerusakan jalan tol.

Selain itu, penelitian lain yang relevan juga menunjukkan bahwa algoritma C4.5 mampu menghasilkan akurasi yang tinggi pada kasus serupa, yaitu sebesar 82,84% dalam klasifikasi prioritas pengembangan jembatan, dan pada penelitian pemetaan kerusakan jalan, akurasi yang dihasilkan mencapai 88% dengan precision 87,9%, recall 82,6%, dan f1-score 87,88%.

Dengan demikian, hasil penelitian ini membuktikan bahwa algoritma C4.5 sangat efektif dan akurat dalam klasifikasi prioritas penanganan kerusakan jalan tol, dengan tingkat akurasi mencapai 98,59%.

Algoritma C4.5 terbukti efektif dalam mengklasifikasikan prioritas penanganan kerusakan pada jalan tol dengan tingkat akurasi yang cukup tinggi, sehingga mampu mendukung proses pengambilan keputusan pemeliharaan secara lebih tepat dan efisien.

4.2. Saran

Adapun saran dari penulisan skripsi ini yaitu sebagai berikut:
Pengembangan sistem aplikasi yang mengintegrasikan algoritma C4.5 sangat dianjurkan agar proses klasifikasi prioritas penanganan kerusakan jalan tol dapat dilakukan secara otomatis dan real-time.

Pengumpulan data yang lebih lengkap dan berkualitas, termasuk variabel lingkungan dan kondisi lalu lintas yang dinamis, perlu dilakukan agar model klasifikasi lebih akurat dan adaptif terhadap perubahan kondisi jalan.

Implementasi hasil penelitian harus disertai dengan evaluasi dan pemantauan berkala untuk memastikan prioritas pemeliharaan tetap relevan dan efektif.
Perlu juga mempertimbangkan aspek biaya dan manfaat secara menyeluruh dalam model klasifikasi agar keputusan pemeliharaan tidak hanya berdasarkan kondisi teknis tetapi juga aspek ekonomis.

DAFTAR PUSTAKA

Cahyadi, universitas buddhi dharma. 2022. “Pengaruh Kualitas Produk Dan Harga Terhadap Keputusan Pembelian Baja Ringan Di Pt Arthanindo Cemerlang.” *Ekonomi Dan Manajemen Bisnis* 1: 60–73.

Franseda, Afrilio, Wawan Kurniawan, Sita Anggraeni, dan Windu Gata. 2020. “Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas.” *Jurnal Sistem dan Teknologi Informasi (Justin)* 8(3): 282. doi:10.26418/justin.v8i3.40982.

Husna, Faridatul, Hairur Rahman, dan Juhari Juhari. 2022. “Implementasi Data Mining Menggunakan Algoritma C4.5 pada Klasifikasi Penjualan Hijab.” *Jurnal Riset Mahasiswa Matematika* 2(2): 40–46. doi:10.18860/jrmm.v2i2.14891.

Justam, Nur Jamilah, Sitti Mawaddah Umar, dan Jousadrah Ramba. 2024. “Penerapan Algoritma C4 . 5 dan Random Forest untuk Pemetaan.” 7: 326–39.

Karyadiputra, Erfan, dan Nadiya Hijriana. 2019. “Penerapan Algoritma Decision Tree C4.5 Untuk Klasifikasi Penentuan Daftar Prioritas Pengembangan Jembatan.” *Technologia: Jurnal Ilmiah* 10(1): 43. doi:10.31602/tji.v10i1.1757.

Khotimah, Khusnul. 2022. “Teknik Data Mining menggunakan Algoritma Decision Tree (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi.” *Jurnal SIMADA (Sistem Informasi dan Manajemen Basis Data)* 4(2). doi:10.30873/simada.v4i2.3064.

Maulana Sidiq, Cecep, Ahmad Faqih, dan Gifthera Dwilestari. 2024. “Algoritma Decision Tree C4.5 Digunakan Untuk Mengklasifikasikan Data Stroke.” *JATI (Jurnal Mahasiswa Teknik Informatika)* 8(2): 1869–74. doi:10.36040/jati.v8i2.8388.

Nurdiyani Sari, Ratih, dan Syarifah S Azharina. 2024. “Penerapan Algoritma C4.5 Untuk Memprediksi Bencana Gunung Meletus Di Indonesia.” *Jts* 3(2): 1–9.

Saadah, Syarifah Rafizma, dan Hagni Puspito. 2023. “ANALISIS KERUSAKAN JALAN MENGGUNAKAN METODE INTERNATIONAL ROUGHNESS INDEX Studi Kasus pada Ruas Jalan Tol Gempol-Pasuruan Jalur A KM 776+000 sampai dengan 810+000 (Analysis Road Damage

Using the International Roughness Index Method Case Study Case Study o." *Jurnal Artesis* 3(2): 200–206.

Subhaktiyasa, Putu Gede. 2024. "Menentukan Populasi dan Sampel : Pendekatan Metodologi Penelitian Kuantitatif dan Kualitatif." 9: 2721–31.

Yunus, Muhamad, Hanandriya Ramadhan, Dimas Rizki Aji, dan Agus Yulianto. 2021. "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indon[1] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, 'Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP),' Paradig. - J. Komput. dan Inform., vol." *Paradigma - Jurnal Komputer dan Informatika* 23(2).