

Analisis Kinerja Algoritma C4.5 pada Dataset Titanic yang Tidak Seimbang Menggunakan Gain Ratio

Kuncoro Singgih Prasajo^{1*}, Hasbi Firmansyah², Wahyu Asriyani³, Ali Sofyan⁴

^{1,2,4}Fakultas Teknik dan Ilmu Komputer, Prodi Informatika, Universitas Pancasakti Tegal, Jl. Halmahera KM. 01, Mintaragen, Tegal Tim., Kota Tegal, Jawa Tengah, 52121, Indonesia.

³Fakultas Keguruan dan Ilmu Pendidikan, Prodi Pendidikan Bahasa dan Sastra Indonesia, Universitas Pancasakti Tegal, Jl. Halmahera KM. 01, Mintaragen, Tegal Tim., Kota Tegal, Jawa Tengah, 52121, Indonesia.

E-mail: singgihkuncoro340@gmail.com

* Corresponding Author



<https://doi.org/10.31004/jerkin.v4i2.4402>

ARTICLE INFO

Article history

Received: 06 Dec 2025

Revised: 12 Dec 2025

Accepted: 18 Dec 2025

Kata Kunci:

Algoritma C4.5,
Klasifikasi, Rasio
Keuntungan, Data
Tidak Seimbang,
Kumpulan Data
Titanic.

Keywords:

C4.5 Algorithm,
Classification, Gain
Ratio, Imbalanced
Data, Titanic Dataset.

ABSTRACT

Penelitian ini bertujuan untuk menganalisis kinerja algoritma C4.5 dalam mengklasifikasikan status keselamatan penumpang menggunakan dataset Titanic yang memiliki distribusi kelas tidak seimbang. Metode penelitian dilakukan melalui pendekatan kuantitatif dengan tahapan preprocessing data, perhitungan manual entropy, information gain, split information, dan gain ratio menggunakan Microsoft Excel, serta pemodelan algoritma C4.5 menggunakan perangkat lunak RapidMiner. Dataset terdiri dari 800 data penumpang dengan atribut *survived* sebagai label kelas. Hasil perhitungan manual menunjukkan bahwa atribut *Gender* memiliki nilai information gain tertinggi sebesar 0,955 sehingga dipilih sebagai akar pohon keputusan, sementara atribut lain seperti *Pclass*, *Age Group*, dan *Fare Group* memiliki kontribusi informasi yang sangat rendah. Hasil pengujian model menggunakan RapidMiner menghasilkan nilai akurasi sebesar 62,50%, namun seluruh data uji diprediksi sebagai kelas tidak selamat. Kondisi ini menyebabkan nilai precision dan recall untuk kelas selamat bernilai 0%. Struktur pohon keputusan yang dihasilkan juga sangat sederhana dan tidak memiliki percabangan lanjutan yang signifikan. Hasil penelitian ini menunjukkan bahwa ketidakseimbangan kelas pada dataset Titanic berpengaruh besar terhadap kinerja algoritma C4.5, sehingga diperlukan teknik penanganan data tidak seimbang untuk meningkatkan performa klasifikasi.

This study aims to analyze the performance of the C4.5 algorithm in classifying passenger survival status using the Titanic dataset, which exhibits an imbalanced class distribution. The research employed a quantitative approach consisting of data preprocessing, manual calculation of entropy, information gain, split information, and gain ratio using Microsoft Excel, followed by model implementation using RapidMiner. The dataset contains 800 passenger records with the survived attribute defined as the class label. Manual calculation results indicate that the Gender attribute has the highest information gain value of 0.955, making it the root node of the decision tree, while other attributes such as Pclass, Age Group, and Fare Group contribute very limited information. The experimental results show that the C4.5 model achieves an accuracy of 62.50%; however, all test instances are predicted as non-survived, resulting in 0% precision and recall for the survived class. In addition, the generated decision tree structure is very shallow with no significant branching. These findings demonstrate that class imbalance in the Titanic dataset strongly affects the performance of the C4.5 algorithm, indicating the need for imbalanced data handling techniques to improve classification results.



This is an open access article under the CC-BY-SA license.

How to Cite: Kuncoro Singgih Prasajo, et al (2025). Analisis Kinerja Algoritma C4.5 pada Dataset Titanic yang Tidak Seimbang Menggunakan Gain Ratio, 4(2). <https://doi.org/10.31004/jerkin.v4i2.4402>

PENDAHULUAN

Perkembangan teknologi informasi dalam beberapa dekade terakhir telah mendorong pertumbuhan data dalam jumlah besar di berbagai sektor, seperti transportasi, kesehatan, keuangan, dan pendidikan. Data yang tersimpan tersebut memiliki potensi besar untuk menghasilkan informasi bernilai apabila dianalisis menggunakan pendekatan yang tepat. Salah satu pendekatan yang banyak digunakan untuk menggali pengetahuan dari data berskala besar adalah *data mining*, yang bertujuan menemukan pola, hubungan, dan pengetahuan tersembunyi dari sekumpulan data (Han et al., 2012).

Dalam data mining, teknik klasifikasi berperan penting dalam proses pengambilan keputusan, yaitu dengan mengelompokkan data ke dalam kelas-kelas tertentu berdasarkan atribut yang dimiliki. Salah satu algoritma klasifikasi yang banyak digunakan adalah algoritma C4.5 (Provost & Fawcett, 2013). Algoritma ini merupakan pengembangan dari ID3 dan memiliki keunggulan dalam menangani atribut numerik maupun kategorikal, serta mampu menghasilkan model pohon keputusan yang relatif mudah dipahami dan diinterpretasikan (Quinlan, 1993). Karena sifat interpretabilitasnya tersebut, C4.5 masih sering digunakan dalam penelitian akademik maupun aplikasi praktis.

Namun demikian, performa algoritma C4.5 sangat dipengaruhi oleh karakteristik data yang digunakan, khususnya distribusi kelas. Pada banyak kasus nyata, data yang tersedia tidak selalu memiliki distribusi kelas yang seimbang. Kondisi ini dikenal sebagai *imbalanced dataset*, yaitu ketika jumlah data pada satu kelas jauh lebih besar dibandingkan kelas lainnya (Anggraini, 2018). Beberapa penelitian menunjukkan bahwa algoritma klasifikasi, termasuk decision tree, cenderung mengalami penurunan kinerja pada kondisi data yang tidak seimbang karena model lebih banyak mempelajari pola kelas mayoritas dan mengabaikan kelas minoritas (Syahputra, 2019).

Dataset Titanic merupakan salah satu dataset publik yang sering digunakan sebagai studi kasus dalam penelitian klasifikasi. Dataset ini berisi data historis penumpang kapal Titanic yang tenggelam pada tahun 1912, dengan atribut utama *survived* yang menunjukkan status keselamatan penumpang. Karakteristik utama dataset ini adalah distribusi kelas yang tidak seimbang, di mana jumlah penumpang yang tidak selamat lebih banyak dibandingkan penumpang yang selamat (Ningse, 2022). Kondisi ini menjadikan dataset Titanic relevan untuk mengkaji pengaruh ketidakseimbangan data terhadap kinerja algoritma klasifikasi, khususnya algoritma C4.5.

Beberapa penelitian sebelumnya telah membahas penggunaan algoritma C4.5 pada permasalahan klasifikasi dengan data tidak seimbang. (Dharmawangsa, 2024). menunjukkan bahwa performa C4.5 dapat menurun secara signifikan ketika diterapkan pada dataset dengan distribusi kelas yang timpang. Penelitian lain oleh (BSI, 2020) menekankan bahwa penggunaan *gain ratio* sebagai kriteria pemilihan atribut dapat membantu mengurangi bias pemilihan atribut yang memiliki banyak kategori dibandingkan penggunaan *information gain* semata. Selain itu, (Hofmann & Klinkenberg, 2016) menyatakan bahwa dataset Titanic sering digunakan untuk menguji sensitivitas algoritma klasifikasi terhadap masalah *class imbalance*.

Meskipun demikian, sebagian besar penelitian sebelumnya lebih menitikberatkan pada perbandingan performa antar algoritma atau peningkatan nilai akurasi, sementara analisis mendalam terhadap perilaku algoritma C4.5 dalam membentuk struktur pohon keputusan pada dataset Titanic yang tidak seimbang masih relatif terbatas (Syukriyah, 2023). Oleh karena itu, penelitian ini dilakukan untuk menganalisis kinerja algoritma C4.5 pada dataset Titanic dengan menggunakan pendekatan *gain ratio*, baik melalui perhitungan manual maupun pemodelan menggunakan perangkat lunak RapidMiner. Analisis ini diharapkan dapat memberikan pemahaman yang lebih komprehensif mengenai dampak ketidakseimbangan data terhadap proses pemilihan atribut, struktur pohon keputusan, serta performa model klasifikasi yang dihasilkan (UNNES, 2019).

METODE

Jenis penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen, yang bertujuan untuk menganalisis kinerja algoritma C4.5 dalam proses klasifikasi data. Pendekatan kuantitatif dipilih karena penelitian ini melibatkan pengolahan data numerik, perhitungan matematis, serta evaluasi performa model klasifikasi berdasarkan metrik tertentu. Metode eksperimen dilakukan dengan

membangun model klasifikasi dan mengujinya menggunakan dataset Titanic yang memiliki distribusi kelas tidak seimbang.

Waktu Dan Tempat Penelitian

Penelitian ini dilaksanakan pada bulan Januari hingga Maret 2025. Proses pengolahan data, perhitungan manual, dan pemodelan dilakukan secara mandiri menggunakan perangkat lunak Microsoft Excel dan RapidMiner Studio pada lingkungan komputer pribadi peneliti.

Data Dan Sumber Data

Data yang digunakan dalam penelitian ini adalah dataset Titanic Survival yang bersifat publik. Dataset ini berisi data penumpang kapal Titanic dengan atribut *survived* sebagai label kelas yang menunjukkan status keselamatan penumpang. Dataset terdiri dari 800 data penumpang dengan distribusi kelas yang tidak seimbang, di mana jumlah penumpang tidak selamat lebih banyak dibandingkan penumpang selamat. Seluruh atribut dalam dataset digunakan sebagai variabel prediktor dalam proses klasifikasi.

Target Dan Subjek Penelitian

Target penelitian ini adalah menganalisis kemampuan algoritma C4.5 dalam mengklasifikasikan data pada kondisi kelas yang tidak seimbang. Subjek penelitian berupa data penumpang Titanic yang tercatat dalam dataset, yang diperlakukan sebagai unit analisis dalam proses pemodelan dan evaluasi klasifikasi.

Prosedur Penelitian

Prosedur penelitian dilakukan secara bertahap dan sistematis. Tahap pertama adalah preprocessing data yang meliputi pembersihan data, penanganan nilai kosong, serta transformasi atribut kategorikal agar dapat diproses oleh algoritma C4.5 (Yunus, 2021). Tahap kedua adalah perhitungan manual menggunakan Microsoft Excel untuk memperoleh nilai entropy, information gain, split information, dan gain ratio pada setiap atribut. Tahap ketiga adalah pembangunan model klasifikasi menggunakan algoritma C4.5 pada perangkat lunak RapidMiner. Dataset dibagi menjadi data latih dan data uji menggunakan operator *Split Data*. Tahap terakhir adalah evaluasi model menggunakan confusion matrix dan metrik performa.

Instrumen Dan Teknik Pengumpulan Data

Instrumen yang digunakan dalam penelitian ini adalah dataset Titanic Survival, perangkat lunak Microsoft Excel untuk perhitungan manual, serta RapidMiner Studio untuk pemodelan dan evaluasi algoritma C4.5. Teknik pengumpulan data dilakukan dengan mengambil dataset dari sumber publik dan mengolahnya sesuai dengan kebutuhan penelitian.

Teknik Analisis Data

Teknik analisis data dilakukan dengan pendekatan matematis dan komputasional. Analisis diawali dengan perhitungan nilai entropy untuk mengukur tingkat ketidakpastian data, yang dirumuskan sebagai berikut (Han et al., 2012).

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

Selanjutnya, nilai information gain dihitung untuk mengetahui pengurangan entropy setelah data dipisahkan berdasarkan suatu atribut (Quinlan, 1993).

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i) \quad (2)$$

Untuk mengurangi bias terhadap atribut dengan banyak kategori, digunakan gain ratio sebagai kriteria pemilihan atribut pada algoritma C4.5, yang dirumuskan sebagai berikut:

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S, A)}{\text{SplitInfo}(S, A)} \quad (3)$$

Hasil perhitungan manual ini digunakan sebagai dasar analisis pemilihan atribut dan dibandingkan dengan hasil pemodelan yang dihasilkan oleh RapidMiner. Evaluasi performa model dilakukan menggunakan confusion matrix serta metrik accuracy, precision, recall, dan F1-score untuk menilai kemampuan klasifikasi algoritma C4.5 terhadap dataset yang tidak seimbang (Halimah, 2022).

HASIL DAN PEMBAHASAN

Hasil Penelitian

Bagian ini menyajikan hasil penelitian yang diperoleh dari perhitungan manual algoritma C4.5 menggunakan Microsoft Excel serta hasil pengujian model menggunakan perangkat lunak RapidMiner. Penyajian hasil dilakukan secara berurutan sesuai tahapan penelitian, sehingga keterkaitan antara proses analisis dan temuan yang diperoleh dapat dipahami secara jelas.

Hasil Perhitungan Manual Algoritma C4.5

Tahap awal analisis dilakukan melalui perhitungan manual untuk menentukan atribut yang paling berpengaruh dalam memisahkan kelas *survived* dan *not survived* pada dataset Titanic. Perhitungan mencakup entropy awal dataset, entropy pada setiap subset, serta nilai *information gain* dan *gain ratio* untuk masing-masing atribut yang dianalisis (Azhari, 2022). Tahap ini penting untuk memastikan bahwa pemilihan atribut pada algoritma C4.5 dilakukan secara transparan, terutama karena dataset yang digunakan memiliki distribusi kelas tidak seimbang (Wiraraja, 2018).

Berdasarkan distribusi 501 penumpang tidak selamat dan 299 penumpang selamat, nilai entropy awal dataset diperoleh sebesar 0,9535. Nilai ini menunjukkan tingkat ketidakpastian yang cukup tinggi pada data sebelum dilakukan pemisahan berdasarkan atribut tertentu. Selanjutnya, atribut *Gender*, *Pclass*, *Age Group*, dan *Fare Group* dianalisis untuk melihat sejauh mana masing-masing atribut mampu menurunkan nilai entropy tersebut.

Hasil perhitungan menunjukkan bahwa atribut *Gender* memiliki nilai *information gain* tertinggi, yaitu sebesar 0,955, dengan nilai entropy setelah pemisahan sebesar 0,9531. Nilai ini jauh lebih besar dibandingkan atribut lainnya, sehingga menunjukkan bahwa *Gender* merupakan atribut yang paling informatif dalam memisahkan kelas target. Temuan ini selaras dengan kondisi historis peristiwa Titanic, di mana perempuan memiliki peluang keselamatan yang lebih tinggi dibandingkan laki-laki.

Sebaliknya, atribut *Pclass* hanya menghasilkan nilai *information gain* sebesar 0,000228, sedangkan *Age Group* memiliki nilai *information gain* mendekati nol, yaitu -0,000031. Atribut *Fare Group* memberikan kontribusi informasi yang sedikit lebih besar dibandingkan atribut lain selain *Gender*, namun nilainya tetap rendah, yaitu 0,003004. Rendahnya nilai *information gain* pada atribut-atribut tersebut menunjukkan bahwa atribut tersebut tidak cukup informatif untuk digunakan sebagai pemisah utama pada tahap awal pembentukan pohon keputusan.

Ringkasan hasil perhitungan entropy dan *information gain* untuk setiap atribut disajikan pada Tabel 1, yang memperlihatkan perbedaan kontribusi masing-masing atribut secara kuantitatif.

Tabel 1. Ringkasan Nilai Entropy dan Information Gain Setiap Atribut

Atribut	Entropy Split	Information Gain	Keterangan
Gender	0,95314	0,95500	IG tertinggi, terlih menjadi root
Pclass	0,95357	0,000228	Pengaruh rendah
Age Group	0,95443	-0,000031	Hampir tidak informatif
Fare Group	0,95164	0,003004	Informasi sangat kecil

Untuk memberikan gambaran yang lebih jelas mengenai pengaruh atribut *Gender*, distribusi kelas *survived* berdasarkan jenis kelamin disajikan pada Tabel 2. Tabel ini menunjukkan adanya perbedaan distribusi kelas yang cukup signifikan antara penumpang laki-laki dan Perempuan.

Tabel 2. Distribusi Kelas Survived Berdasarkan Gender

Gender	Survived=0	Survived=1	Total
Male	298	171	469
Female	203	128	331
Total	501	299	800

Berdasarkan hasil tersebut, atribut *Gender* dipilih sebagai node akar dalam pembentukan pohon keputusan karena memiliki kemampuan paling baik dalam menurunkan tingkat ketidakpastian data.

Atribut lain baru berpotensi digunakan pada pemisahan lanjutan setelah data dipisahkan berdasarkan *Gender*. Hasil perhitungan manual ini memberikan dasar yang kuat untuk membandingkan perilaku algoritma C4.5 pada tahap pemodelan otomatis menggunakan RapidMiner.

Hasil Pengujian Model Menggunakan Rapidminer

Setelah perhitungan manual dilakukan, tahap selanjutnya adalah pengujian model klasifikasi menggunakan perangkat lunak RapidMiner. Pada tahap ini, atribut *Survived* ditetapkan sebagai label kelas, sedangkan atribut lainnya digunakan sebagai variabel prediktor (ITN, 2020). Dataset dibagi menjadi data latih dan data uji menggunakan operator *Split Data* dengan komposisi kelas yang tetap tidak seimbang.

Hasil pengujian menunjukkan bahwa model menghasilkan nilai akurasi sebesar 62,50%. Namun, analisis lebih lanjut menunjukkan bahwa nilai akurasi tersebut tidak mencerminkan performa klasifikasi yang sesungguhnya. Model memprediksi seluruh data uji sebagai kelas *tidak selamat*, sehingga tidak satu pun penumpang selamat berhasil dikenali oleh model.

Distribusi kelas pada dataset Titanic yang digunakan dalam penelitian ini disajikan pada Tabel 3 untuk memperjelas kondisi ketidakseimbangan kelas yang menjadi faktor utama rendahnya performa model.

Tabel 3. Distribusi Kelas Survived pada Dataset Titanic

Kategori Survived	Jumlah
Tidak Selamat (False)	501
Selamat (True)	299
Total	800

Evaluasi performa model menggunakan confusion matrix ditampilkan pada Tabel 4. Tabel tersebut menunjukkan bahwa model berhasil mengklasifikasikan seluruh penumpang tidak selamat dengan benar, tetapi gagal mengidentifikasi seluruh penumpang yang selamat.

Tabel 4. Confusion Matrix Pengujian Model C4.5

Prediksi/Aktual	Tidak selamat	Selamat
Prediksi False	100	60
Prediksi True	0	0

Hasil evaluasi performa model secara keseluruhan disajikan pada Tabel 5, yang mencakup nilai akurasi, *precision*, *recall*, dan metrik evaluasi lainnya. Nilai *recall* dan *precision* untuk kelas *true* bernilai 0%, yang menandakan kegagalan model dalam mengenali kelas minoritas.

Tabel 5. Hasil Evaluasi Performa Model C4.5

Metrik Evaluasi	Nilai
Akurasi	62,50%
Weighted Mean Reccal	50,00%
Weighted Mean Precision	31,25%
Reccal Kelas True	0%
Precision Kelas True	0%
Reccal Kelas False	100%
Precision Kelas False	62,50%

Selain evaluasi numerik, struktur pohon keputusan yang dihasilkan oleh RapidMiner juga dianalisis. Pohon keputusan yang terbentuk terlihat sangat sederhana dan tidak memiliki percabangan lanjutan yang signifikan. Visualisasi struktur pohon keputusan tersebut ditampilkan pada Gambar 2, yang menunjukkan bahwa model berhenti berkembang setelah pemisahan awal.



Gambar 1. Hasil Pohon Keputusan Algoritma C4.5 pada RapidMiner

Gambar 1 menyajikan struktur pohon keputusan yang dihasilkan. Terlihat bahwa pohon sangat dangkal dan tidak mengandung percabangan berarti, menandakan bahwa model gagal menemukan pola berbeda dari atribut yang tersedia.

Pembahasan

Hasil penelitian menunjukkan bahwa algoritma C4.5 tidak mampu memberikan performa klasifikasi yang optimal pada dataset Titanic yang memiliki distribusi kelas tidak seimbang. Ketidakseimbangan antara kelas penumpang selamat dan tidak selamat menyebabkan model lebih banyak mempelajari pola dari kelas mayoritas. Akibatnya, model cenderung menghasilkan prediksi yang mengarah sepenuhnya ke kelas *tidak selamat*.

Temuan ini konsisten dengan hasil perhitungan manual *information gain* yang menunjukkan bahwa hanya atribut *Gender* yang memiliki kontribusi informasi yang signifikan. Atribut lainnya memiliki nilai *information gain* yang sangat rendah, sehingga tidak cukup kuat untuk membentuk percabangan lanjutan pada pohon keputusan. Kondisi ini menjelaskan mengapa struktur pohon keputusan yang dihasilkan oleh RapidMiner bersifat dangkal dan tidak kompleks (Yamantri, 2024).

Perbandingan antara hasil perhitungan manual dan hasil pemodelan RapidMiner disajikan pada Tabel 6, yang menunjukkan kesesuaian antara atribut dengan nilai *information gain* tinggi dan perilaku model dalam membentuk struktur pohon keputusan.

Tabel 6. Perbandingan Hasil Perhitungan Manual dan RapidMiner

Atribut	Information Gain (Manual)	Pengaruh Model
Gender	0,955	Menjadi Root Tree
Pclass	0,000228	Tidak signifikan
AgeGroup	-0,000031	Tidak digunakan model
FareGroup	0,003004	Tidak signifikan

Secara keseluruhan, hasil dan pembahasan ini menegaskan bahwa permasalahan utama dalam penelitian ini bukan terletak pada kesalahan implementasi algoritma, melainkan pada karakteristik dataset yang digunakan. Ketidakseimbangan kelas yang cukup besar menyebabkan algoritma C4.5 mengalami kesulitan dalam membangun model yang mampu mengenali kelas minoritas secara efektif. Temuan ini memperkuat hasil penelitian dan memberikan dasar yang jelas untuk perumusan kesimpulan dan rekomendasi penelitian selanjutnya.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma C4.5 belum mampu memberikan performa klasifikasi yang optimal pada dataset Titanic yang memiliki distribusi kelas tidak seimbang. Hasil perhitungan manual menunjukkan bahwa atribut *Gender* merupakan atribut paling informatif dengan nilai *information gain* tertinggi, sedangkan atribut lainnya memiliki kontribusi yang sangat rendah dalam proses pemisahan kelas. Hasil pengujian menggunakan RapidMiner menghasilkan nilai akurasi sebesar 62,50%, namun model gagal mengenali kelas penumpang selamat karena seluruh prediksi mengarah ke kelas mayoritas. Struktur pohon keputusan yang dihasilkan juga bersifat dangkal dan tidak kompleks, yang menandakan keterbatasan model dalam mempelajari pola kelas minoritas. Temuan ini menunjukkan bahwa ketidakseimbangan data menjadi faktor utama yang mempengaruhi kinerja algoritma C4.5, sehingga penerapan teknik penanganan data tidak seimbang atau penggunaan algoritma alternatif perlu dipertimbangkan pada penelitian selanjutnya.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dalam pelaksanaan penelitian ini, baik secara langsung maupun tidak langsung. Ucapan terima kasih juga disampaikan kepada pengelola dataset publik Titanic serta pihak-pihak yang telah menyediakan referensi dan perangkat lunak yang digunakan dalam proses pengolahan data dan penyusunan artikel ini.

REFERENSI

Anggraini, S. (2018). Analisis Data Mining Penjualan Ban Menggunakan Algoritma C4.5. *JITEKI (Jurnal Ilmu Teknik Elektro Komputer Dan Informatika)*.

- Azhari, D. Z. (2022). Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat ... *Jurnal PDSI / e-Journal PDSI*.
- BSI, A. (2020). Penerapan Algoritma C4.5 Dalam Menentukan Prediksi ... *IJEC (Jurnal BSI)*.
- Dharmawangsa, A. (2024). Performance Evaluation Algoritma C4.5 pada Berbagai Kondisi Dataset. *DJTechno*.
- Halimah, D. (2022). Algoritma C4.5 untuk Menentukan Klasifikasi Tingkat ... *Jurnal Nasional (Garuda)*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hofmann, M., & Klinkenberg, R. (2016). *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. CRC Press.
- ITN, A. (2020). Penerapan Pohon Keputusan C4.5 untuk ... *Jurnal ITN*.
- Ningse, W. (2022). Klasifikasi Algoritma C4.5 untuk Penentuan ... *Jurnal Nasional (Garuda)*.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Syahputra, M. (2019). Teknik Klasifikasi C4.5 dalam Menentukan Faktor Utama ... *Jurnal Nasional*.
- Syukriah, R. (2023). Penerapan Algoritma Decision Tree dan SMOTE untuk ... *Repository UII / Prosiding*.
- UNNES, A. (2019). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *Jurnal Universitas Negeri Semarang (UNNES)*.
- Wiraraja, A. (2018). Penerapan Data Mining Menggunakan Algoritma Decision Tree ... *Jurnal Wiraraja*.
- Yamantri, A. B. (2024). Penerapan Algoritma C4.5 untuk Prediksi Faktor Risiko ... *Jurnal Kesehatan / JKA*.
- Yunus, M. (2021). Penerapan Metode Data Mining C4.5 untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP). *Paradigma*.