E-ISSN : 2988-1986
https://ejournal.cahayailmubangsa.institute/index.php/kohesi

Kohesi: Jurnal Multidisiplin Saintek
Volume 9 No 12 Tahun 2025

PERBANDINGAN ALGORITMA KLASIFIKASI DALAM DATA MINING: DECISION TREE C4.5 DAN K-NEAREST NEIGHBOR (KNN) BERDASARKAN ANALISIS LITERATUR

Dewi Rabiatul Zahro[1*], Solichun[2], Elkin Rilvani[3]
[1,2,3]Program Studi Teknik Informatika , Fakultas Teknik, Universitas Pelita bangsa Cikarang
E-mail: dewirabiatulzahro@gmail.com, likhun.andeskra@gmail.com, elkin.rilvani@pelitabangsa.ac.id

ABSTRACT

This study is a comparative analysis aimed at evaluating the performance of Decision Tree C4.5 and K-Nearest Neighbor (KNN) based on a review of twelve scientific journals that applied these algorithms across various domains, including disease classification, graduation prediction, scholarship classification, and sentiment analysis. The evaluation was conducted by comparing key performance metrics such as accuracy, AUC, precision, and recall, while also considering aspects of interpretability and computational efficiency. The analysis revealed that Decision Tree C4.5 excels in interpretability and accuracy with categorical and complex datasets. In contrast, KNN performs better with small datasets featuring non-linear distributions, although it is highly dependent on the selection of the k value and data preprocessing. This study concludes that the choice of classification algorithm should be tailored to the data characteristics, the need for result interpretability, and computational efficiency.

Keywords: Data Mining, Classification, Decision Tree C4.5, KNN, Accuracy, Interpretability, Computation

ABSTRAK

Penelitian ini merupakan studi komparatif untuk menganalisis performa dari Decision Tree C4.5 dan K-Nearest Neighbor (KNN) berdasarkan telaah terhadap dua belas jurnal ilmiah yang menerapkan kedua algoritma dalam berbagai domain, seperti klasifikasi penyakit, prediksi kelulusan, klasifikasi beasiswa, hingga analisis sentimen. Evaluasi dilakukan dengan membandingkan metrik kinerja utama seperti akurasi, AUC, presisi, dan recall, serta mempertimbangkan aspek interpretabilitas dan efisiensi komputasi. Hasil analisis menunjukkan bahwa Decision Tree C4.5 unggul dalam interpretabilitas dan akurasi pada dataset kategorik dan kompleks. Sebaliknya, KNN lebih unggul pada dataset kecil dengan distribusi non-linear, walaupun sangat bergantung pada pemilihan nilai *k* dan *preprocessing* data. Penelitian ini menyimpulkan bahwa pemilihan algoritma klasifikasi harus disesuaikan dengan karakteristik data, kebutuhan interpretasi hasil, dan efisiensi komputasi.

Kata Kunci: Data Mining, Klasifikasi, Decision tree C4.5, KNN, Akurasi, Interpretabilitas, Komputasi

Article History
Received : Agustus 2025
Reviewed: Agustus 2025
Published: Agustus 2025

Plagiarism Checker No 235
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

1. PENDAHULUAN

Data mining Data mining memainkan peran penting dalam mengekstraksi pengetahuan dari data besar, dengan teknik klasifikasi seperti Decision Tree C4.5 dan K-Nearest Neighbor (KNN) menjadi fokus perhatian di berbagai domain. C4.5, yang menggunakan gain ratio untuk membentuk pohon keputusan, dikenal karena interpretabilitasnya, seperti terlihat dalam studi Rizqullah et al. (2024) yang mencatat akurasi 77.9% pada klasifikasi pasien gagal jantung. Sebaliknya, KNN mengklasifikasikan data berdasarkan kedekatan jarak, dengan performa yang unggul dalam kasus tertentu, misalnya Purwanto dan Nugroho (2023) melaporkan akurasi 98.3% pada klasifikasi beasiswa.

Namun, hasil penelitian menunjukkan variasi signifikan. Rahmadhani et al. (2023) mencatat akurasi C4.5 98.52% pada klasifikasi jamur, lebih tinggi dari KNN (96.62%), sementara Jasmine Putri et al. (2022) menemukan C4.5 hanya 54.0% dalam jadwal kerja, menunjukkan ketergantungan pada jenis data. Nazifah dan Prianto (2023) dalam Systematic Literature Review mereka mencatat akurasi C4.5 berkisar 73.24%-98.18% dan KNN 45.21%-93.33%, tergantung dataset. Meskipun banyak studi membandingkan keduanya, seperti Adenta Sari et al. (2023) dengan akurasi C4.5 99.75% pada anemia, belum ada sintesis lintas domain yang mengintegrasikan akurasi, interpretabilitas, dan efisiensi. Penelitian ini bertujuan mengisi celah tersebut dengan telaah 12 jurnal untuk memberikan panduan pemilihan algoritma yang lebih terstruktur

2. METODE PENELITIAN

Penelitian ini merupakan studi komparatif dengan pendekatan kualitatif-deskriptif, yang bertujuan untuk menganalisis dan membandingkan kinerja algoritma klasifikasi Decision Tree C4.5 dan K-Nearest Neighbor (KNN) berdasarkan tinjauan literatur dari berbagai jurnal ilmiah. Penelitian ini menggunakan metode Systematic Literature Review (SLR) dengan menganalisis 12 jurnal relevan yang membahas penerapan algoritma Decision Tree C4.5 dan K-Nearest Neighbor (KNN) di berbagai domain.

Data penelitian ini diperoleh dari artikel jurnal ilmiah yang diterbitkan antara tahun 2019 dan 2023, bersumber dari portal jurnal nasional terpercaya seperti Garuda, Google Scholar, dan DOAJ. Proses pengumpulan data dilakukan melalui pencarian dengan kata kunci seperti :

“Decision Tree C4.5”, ‘K-Nearest Neighbor”, “Klasifikasi”, dan Perbandingan algoritma klasifikasi”.

Kriteria inklusi jurnal meliputi :

a. Menjelaskan penerapan algoritma C4.5 dan/atau KNN,

b. Menyajikan hasil evaluasi klasifikasi (akurasi, presisi, recall, AUC),
c. Berasal dari jurnal peer-reviewed

Dari hasil pencarian, terpilih 12 jurnal yang dianalisis lebih lanjut.

Teknik Analisis

Data dianalisis menggunakan metode studi literatur riview dengan fokus pada:

a. Algoritma yang digunakan
b. Domain permasalahan (misalnya:Kesehatan, Pendidikan, sosial)
c. Keunggulan masing-masing algoritma,
d. Kelemahan masing-masing algoritma,
e. Tingkat interpretabilitas dan efisiensi waktu komputasi (jika tersedia).

Setiap jurnal dikaji dan hasilnya direkap dalam bentuk tabel komparatif untuk kemudian dianalisis secara deskriptif.

Instrumen Penelitian

Alat utama yang digunakan dalam penelitian ini adalah tabel pengkodean jurnal yang merekam informasi berikut:

a. Nama penulis & tahun publikasi,
b. Judul jurnal,
c. Algoritma yang digunakan,
d. Domain kasus,
e. Metrik evaluasi & hasilnya
f. Temuan dan simpulan jurnal

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan dan Akurasi

Untuk mendapatkan hasil perbandingan dari akurasi dilakukanlah studi literatur terhadap 12 jurnal dengan berbagai domain klasifikasi. Berdasarkan hasil telaah, diperoleh nilai akurasi dari masing-masing algoritma yang disajikan pada tabel 1.

Tabel 1. Perbandingan Akurasi Algoritma C4.5 dan KNN berdasarkan 12 jurnal

No	Topik	Penulis	Tahun	Algoritma Dibandingkan	Akurasi C4.5	Akurasi K-NN	Ket.

1	Evaluasi performa pelajar SMA dengan berbagai metode klasifikasi	Eri mardiani et al.	2023	k-NN, Naive Bayes, Decision Tree, Ensemble, Linear Regression	-	82.3%	Analisis performa pelajar SMA berdasarkan kursus persiapan ujian dan nilai.
2	Penilaian sentimen pengguna layanan BPJS menggunaka n pendekatan klasifikas	Rani Puspita, Agus Widodo	2020	C4.5, KNN, NB	96.13%	96.01%	C4.5 unggul. Precision positif = 0%. Netral dominan. Dataset: tweet layanan BPJS.
3	Penilaian algoritma untuk memprediksi kelangsungan pasien gagal jantung	Muhammad Fakhri Rizqullah et al.	2024	C4.5, k-NN, Naive Bayes	77.9%	65.5%	Klasifikasi pasien gagal jantung
4	Analisis klasifikasi jamur dengan C4.5 dan MKNN	R. Rahmadhan i et al.	2023		98.52%	96.62%	Klasifikasi jamur, data Kaggle
5	Penentuan jadwal kerja karyawan menggunaka	Jasmine Putri, et al.	2022	C4.5, k-NN	54.0%	16.0%	Jadwal kerja karyawan, data private 501 entri, akurasi

	n C4.5 dan KNN						C4.5 54%, KNN 16%
6	Evaluasi algoritma untuk klasifikasi penerima beasiswa	Agung Purwanto, Handoyo Widi Nugroho	2023	C4.5, k-NN	97.23%	98.3%	Evaluasi menggunakan Confusion Matrix dan ROC/AUC, KNN: precision 98.08%, recall 98.00%, AUC 1.000; C4.5: precision 94.43%, recall 100.00%, AUC 0.956
7	Perbandingan metode untuk klasifikasi serangan jaringan IoT	Zishwa Muhammad Jauhar Nafis, et al.	2024	Decision Tree, k-NN	98.0%	90.0%	Menggunakan Edge-IIoTset dataset, fokus klasifikasi biner, Decision Tree lebih baik (selisih akurasi 0.08)
8	Klasifikasi status kesehatan dengan Decision Tree dan perbandinga n metode lain	Arell S. Biyantoro, Budi Prasetiyo	2024	Decision Tree, k-NN, Naive Bayes	95.45%	90.91%	Menggunakan dataset dummy kesehatan (110 data setelah preprocessing), Decision Tree paling akurat (95.45%)

9	Prediksi potensi pendonor darah menggunakan berbagai algoritma klasifikas	Hermanto Wahono, Dwiza Riana	2020	Decision Tree C4.5, k-NN, Naïve Bayes	93.83%	84.10%	Menggunakan dataset 3710 pendonor dari PMI Kota Bekasi, Decision Tree C4.5 paling akurat (93.83%, AUC 0.978)
10	Prediksi kelulusan mahasiswa dengan pendekatan klasifikasi	Embun Fajar Wati, Biktra Rudianto.	2022	KNN, Naive Bayes, C4.5	91.00%	56.33%	Prediksi kelulusan mahasiswa tepat waktu berdasarkan faktor internal, akademik, tempat tinggal, keuangan, dan perilaku; Naive Bayes paling akurat (100.00%, AUC 1.000), C4.5 (91.00%, AUC 0.978), KNN terendah (56.33%, AUC 0.258); Precision NB 100.00%, Recall NB 100.00%

11	injauan literatur sistematis perbandingan C4.5 dengan algoritma lain	Naurah Nazifah, Cahyo Prianto	2023	C4.5, KNN, Naive Bayes, Random Forest, Fuzzy, ID3	73.24% - 98.18%	45.21% - 93.33%	Systematic Literature Review membandingkan C4.5 dengan algoritma lain; akurasi C4.5 berkisar 73.24%-98.18%, KNN 45.21%-93.33%, tergantung dataset; C4.5 unggul dalam interpretabilitas , sedangkan Random Forest dan KNN lebih baik untuk akurasi dan data tidak seimbang
12	Evaluasi klasifikasi penyakit anemia dengan C4.5 dan KNN	Nur Avia Adenta Sari, et al	2023	C4.5, K-NN	99.75%	99.20%	Menggunakan dataset anemia dari Kaggle (1.421 entri, 5 atribut); C4.5 unggul dengan akurasi 99.75%, precision 100%, recall 99.50%; K-NN akurasi 99.20%, precision

							98.51%, recall 98.51%

Berdasarkan tabel 1, algoritma Decision Tree C4.5 cenderung menghasilkan akurasi yang lebih unggul dibandingkan K-Nearest Neighbor (KNN) pada sebagian besar penelitian yang dikaji. C4.5 memiliki keunggulan dalam mengolah data kategorik serta menghasilkan model klasifikasi yang stabil. Di sisi lain, KNN menunjukkan performa yang cukup baik pada data berukuran kecil atau dengan distribusi non-linear, meskipun performanya sangat bergantung pada pemilihan parameter k dan preprocessing data. Rata-rata akurasi C4.5 dan KNN dari 12 jurnal (kecuali No. 1 di mana C4.5 tidak disebutkan):

- Rata-rata akurasi C4.5: (96.13 + 77.9 + 98.52 + 54.0 + 97.23 + 98.0 + 95.45 + 93.83 + 91.00 + 73.24 + 99.75) / 11 = 88.68%
- Rata-rata akurasi KNN: (82.3 + 96.01 + 65.5 + 96.62 + 16.0 + 98.3 + 90.0 + 90.91 + 84.10 + 56.33 + 93.33) / 11 = 79.27%

  Tambahkan kalimat seperti: "Secara keseluruhan, rata-rata akurasi C4.5 (88.68%) lebih tinggi dibandingkan KNN (79.27%), menegaskan keunggulan C4.5 dalam sebagian besar kasus."

3.2 Interpretabilitas dan Kompleksitas

Selain akurasi, interpretabilitas dan kompleksitas merupakan faktor penting dalam mengevaluasi algoritma klasifikasi. Interpretabilitas mengacu pada kemudahan pemahaman model oleh pengguna, seperti yang ditunjukkan oleh Decision Tree C4.5 yang menghasilkan struktur pohon keputusan, sebagaimana terlihat dalam studi Nazifah dan Prianto (2023, No. 11) yang menyoroti keunggulan C4.5 dalam interpretabilitas lintas domain. Sebaliknya, K-Nearest Neighbor (KNN) cenderung kurang interpretabel karena hanya mengandalkan perhitungan jarak, tanpa menghasilkan aturan eksplisit, seperti terlihat dalam analisis Purwanto dan Nugroho (2023, No. 6) yang fokus pada metrik akurasi tanpa struktur jelas. Dari sisi kompleksitas, C4.5 menunjukkan efisiensi pada dataset besar, misalnya dalam klasifikasi anemia oleh Adenta Sari et al. (2023, No. 12) dengan akurasi 99.75%, sementara KNN memerlukan waktu lebih lama perhitungan jarak, terutama pada data besar seperti yang digunakan oleh Rizqullah et al. (2024, No. 3).

Algoritma Decision Tree C4.5 memiliki keunggulan interpretabilitas karena menghasilkan pohon keputusan, seperti terlihat dalam penelitian Biyantoro dan Prasetiyo (2024, No. 8) yang mencapai akurasi 95.45% pada klasifikasi kesehatan dengan aturan yang mudah dipahami. Struktur ini memungkinkan pengguna, termasuk non-teknis, memahami logika klasifikasi

melalui jalur dari akar ke daun. Hal ini menjadikan C4.5 populer di bidang seperti pendidikan dan kesehatan, sebagaimana ditunjukkan oleh Fajar Wati dan Rudianto (2022, No. 10) dengan akurasi 91.00%.

Sebaliknya, KNN, yang digunakan dalam studi oleh Jasmine Putri et al. (2022, No. 5) dengan akurasi hanya 16.0%, termasuk metode lazy learner yang tidak membentuk model eksplisit, hanya menyimpan data pelatihan untuk perhitungan jarak. Ini membuat KNN kurang interpretatif, terutama pada dataset besar atau berdimensi tinggi, seperti yang tercatat dalam analisis Nafis et al. (2024, No. 7).

Dari sisi kompleksitas komputasi, C4.5 membutuhkan waktu lebih pada pelatihan untuk membangun pohon, seperti pada studi Wahono dan Riana (2020, No. 9) dengan akurasi 93.83%, tetapi klasifikasinya cepat setelah pohon terbentuk. KNN, meskipun tidak memerlukan pelatihan awal, menghadapi beban tinggi saat prediksi dengan perhitungan jarak real-time, seperti pada kasus Rahmadhani et al. (2023, No. 4) dengan akurasi 96.62%. Dengan demikian, C4.5 lebih unggul untuk kebutuhan interpretasi dan efisiensi pada data kompleks, sementara KNN lebih cocok untuk data sederhana dengan kebutuhan akurasi tertentu.

3.3 Rangkuman Analisis

Berdasarkan hasil telaah terhadap 12 jurnal ilmiah yang membandingkan algoritma Decision Tree C4.5 dan K-Nearest Neighbor (KNN), dapat disimpulkan bahwa masing-masing algoritma memiliki keunggulan dan keterbatasan tergantung pada karakteristik data dan tujuan klasifikasi.

Algoritma C4.5 secara umum menunjukkan performa yang lebih konsisten dalam hal akurasi, terutama pada dataset yang bersifat kategorik dan kompleks. Selain itu, interpretabilitasnya yang tinggi menjadikan C4.5 lebih sesuai untuk aplikasi yang membutuhkan transparansi keputusan, seperti prediksi kelulusan mahasiswa atau klasifikasi pasien berdasarkan gejala.

Sementara itu, algoritma KNN unggul dalam menangani dataset kecil dengan distribusi non-linear. Namun, performa KNN sangat bergantung pada pemilihan parameter *k* dan kualitas preprocessing data. Dari sisi komputasi, KNN memiliki kelemahan pada proses klasifikasi karena harus menghitung jarak terhadap seluruh data pelatihan secara langsung.

Secara umum, pemilihan algoritma klasifikasi tidak dapat digeneralisasi, melainkan harus disesuaikan dengan:

a. Tujuan analisis,

b. Ukuran dan jenis data,

c. Kebutuhan interpretasi hasil,

d. Sumber daya komputasi yang tersedia.

Dengan mempertimbangkan aspek-aspek tersebut, algoritma Decision Tree C4.5 lebih direkomendasikan pada kasus yang menuntut kejelasan logika klasifikasi dan struktur data yang kompleks, sementara KNN tetap menjadi alternatif yang efisien untuk data kecil dan sederhana.

KESIMPULAN

Penelitian ini bertujuan untuk membandingkan performa algoritma klasifikasi *Decision Tree C4.5 dan K-Nearest Neighbor (KNN)* melalui telaah sistematis terhadap 12 jurnal ilmiah dari berbagai domain aplikasi. Berdasarkan hasil analisis, dapat disimpulkan bahwa:

a. **Algoritma C4.5** menunjukkan performa akurasi yang lebih baik, khususnya pada dataset dengan atribut kategorik dan struktur data yang kompleks.. Selain itu, interpretabilitas model C4.5 menjadi keunggulan utama karena menghasilkan aturan klasifikasi yang jelas dan mudah dipahami.

b. **Algoritma KNN** menunjukkan performa yang kompetitif, khususnya pada dataset kecil dan distribusi data non-linear. Namun, KNN sangat bergantung pada pemilihan nilai *k*, preprocessing data, dan memiliki kompleksitas komputasi tinggi saat prediksi.

c. Secara umum, **tidak ada algoritma yang selalu unggul dalam semua kasus**. Pemilihan algoritma klasifikasi harus mempertimbangkan karakteristik data, kebutuhan interpretasi hasil, serta efisiensi komputasi.

SARAN

a. Penelitian selanjutnya diharapkan dapat mencakup lebih banyak jurnal atau studi kasus, sehingga hasil analisis yang diperoleh menjadi lebih luas cakupannya dan mencerminkan kondisi yang lebih beragam.
b. Selain menggunakan akurasi sebagai tolok ukur, disarankan untuk mempertimbangkan metrik evaluasi lainnya seperti F1-score, efisiensi waktu pemrosesan, dan ketahanan terhadap data pencilan (outlier) agar analisis yang dilakukan menjadi lebih komprehensif.
c. Perlu dilakukan penelitian lanjutan melalui implementasi langsung algoritma pada dataset aktual, guna memperoleh bukti empiris yang dapat memperkuat temuan dari kajian literatur ini.

DAFTAR PUSTAKA

References

Agung Purwanto, H. W. (2023). ANALISA PERBANDINGAN KINERJA ALGORITMA C4.5 DAN ALGORITMA K-NEAREST NEIGHBORS UNTUK KLASIFIKASI PENERIMA BEASISWA. *Jurnal Teknoinfo*, 236-243.

Arell S. Biyantoro, B. P. (2024). Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan perbandingan KNN dan Naive Bayes . *Institute of Research and Publication Indonesia*, 47-55.

Embun Fajar Wati, B. R. (2022). Penerapan Algoritma KNN, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa. *Jurnal Format*.

Eri Mardiani, N. R. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *INNOVATIVE: Journal Of Social Science Research*, 13880-13892.

Hermanto Wahono, D. R. (2020). Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors dan Decision Tree C4.5 . *JURIKOM (Jurnal Riset Komputer)*, 7-14.

Muhammad Fakhri Rizqullah, N. T. (2024). Komparasi Penerapan Algoritma C4.5, K-Nearest Neighbor, dan Naïve Bayes untuk Keberlangsungan Pasien Gagal Jantung. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2580-2587.

Naurah Nazifah, C. P. (2023). Analisis Perbandingan Decision Tree Algoritma C4.5 dengan algoritma lainnya: Sistematic Literature Review. *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, 57-64.

Nur Avia Adenta Sari, U. M. (2024). KOMPARASI ALGORITMA C4.5 DAN K-NEAREST NEIGHBOR (K-NN) PADA KLASIFIKASI PENYAKIT ANEMIA. *Jurnal Sistem Informasi dan Teknik Komputer*.

R Rahmadhani, A. N. (2023). Analisis Perbandingan Algoritma C4.5 dan Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Jamur . *Junal Sistem Komputer dan Informatika (JSON)*, 226-236.

Rani Puspita, A. W. (2020). Perbandingan Metode KNN, Decision Tree, dan Naïve Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS . *Jurnal Informatika Universitas Pamulang* , 646-654.

Syarifah Jasmine Putri, Q. A. (2022). Klasifikasi Menentukan Jadwal Kerja Data Karyawan Menggunakan Algoritma C4.5 dan K-nearest Neighbor. *Institut Riset dan Publikasi Indonesia (IRPI)* , 2830-3083.

Zishwa Muhammad Jauhar Nafis, R. N. (2024). Perbandingan Algoritma Decision Tree dan K-Nearest Neighbor untuk Klasifikasi Serangan Jaringan IoT. *Komputika:Jurnal Sistem Komputer*, 245-252.