



BAB II

TINJAUAN LITERATUR

Dalam penelitian ini, kajian literatur bertujuan untuk memahami serta mengidentifikasi penelitian-penelitian terdahulu yang relevan, khususnya yang berkaitan dengan sistem rekomendasi produk, algoritma C4.5 untuk klasifikasi kebutuhan pengguna, serta implementasi sistem rekomendasi berbasis web. Literatur yang dikaji mencakup beragam pendekatan dalam mengembangkan sistem rekomendasi smartphone, penerapan algoritma C4.5 dalam berbagai klasifikasi. Kajian ini juga memuat referensi tentang perbandingan berbagai varian algoritma, serta metode evaluasi akurasi sistem klasifikasi. Tinjauan literatur ini memberikan gambaran tentang efektivitas algoritma C4.5 dalam menangani data, kelebihan interpretabilitas pohon keputusan dalam konteks retail, serta tantangan dalam mengembangkan sistem rekomendasi untuk toko retail lokal.

2.1 Sistem Rekomendasi

Sistem rekomendasi adalah sistem perangkat lunak yang dirancang untuk membantu pengguna dalam mengambil keputusan dengan menyaring informasi dan menyajikan pilihan yang paling relevan sesuai kebutuhan mereka.

Pada Penelitian Rival Amperawansyah & Dhanny Permatasari Putri (2022) dalam penelitiannya tentang "Pemilihan Jenis Ponsel Pintar Sesuai dengan Kebutuhan", sistem rekomendasi didefinisikan sebagai sistem pakar yang menggunakan basis pengetahuan untuk memberikan saran pemilihan smartphone kepada pengguna dengan akurasi 95%. Dalam penelitian tersebut, sistem rekomendasi diterapkan menggunakan kombinasi Forward Chaining dan Decision Tree untuk

segmen harga 1-3 juta Rupiah, di mana sistem membantu masyarakat menemukan smartphone yang sesuai dengan kebutuhan spesifik mereka[6].

Pada Penelitian Mukhsinin, D.A., dkk. (2024) dalam penelitiannya tentang "Implementation of Decision Tree Algorithm for Movie Recommendation System on Netflix Dataset", sistem rekomendasi didefinisikan sebagai pendekatan klasifikasi rating menggunakan Decision Tree untuk menyaring konten sesuai preferensi pengguna. Penelitian tersebut menerapkan Decision Tree untuk mengklasifikasikan rating film Netflix berdasarkan genre, durasi, dan metadata pengguna, dengan hasil akurasi tinggi yang meningkatkan relevansi rekomendasi. Prosesnya mencakup pengumpulan dataset besar, feature engineering, training model Decision Tree, dan evaluasi untuk mengatasi cold start problem dalam sistem rekomendasi[7].

Pada Penelitian Ivan Zuhdiansyah & Ardytha Luthfiarta (2024) dalam penelitiannya tentang "Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan Singular Value Decomposition", sistem rekomendasi bertujuan mengatasi masalah information overload dalam e-commerce. Penelitian tersebut menerapkan Collaborative Filtering yang menganalisis pola preferensi pengguna untuk memberikan rekomendasi personal[8].

2.2 Data Mining

Data Mining adalah proses ekstraksi informasi dan pengetahuan yang tersembunyi dari basis data berukuran besar dengan menggunakan berbagai teknik komputasi. Proses ini bertujuan untuk menemukan pola, hubungan, dan insight yang dapat digunakan untuk pengambilan keputusan yang lebih baik.

Pada Penelitian Amrin, A., dkk. (2025) dalam penelitian "Analisa Komparasi Model Data Mining Algoritma C4.5, Naïve Bayes, dan Random Forest"





mendefinisikan data mining sebagai proses penggalian informasi bermakna dari data besar dengan menggunakan teknik klasifikasi dan pembelajaran mesin untuk mendukung pengambilan keputusan berbasis data. Dalam penelitian ini, algoritma C4.5 menjadi salah satu metode utama yang dianalisa untuk kasus klasifikasi dataset kompleks. Proses dimulai dengan pengumpulan data mentah, dilanjutkan dengan tahapan preprocessing yang meliputi penanganan data hilang dan normalisasi, pemilihan fitur, hingga pembangunan model decision tree C4.5. Penelitian juga melakukan evaluasi performa model menggunakan cross-validation dan visualisasi ROC curve. Analisa hasil menggunakan confusion matrix menunjukkan performa C4.5 yang kompetitif dan konsisten dibandingkan metode lain dalam memprediksi kelas dengan tingkat akurasi tinggi di berbagai domain aplikasi. Hal ini menegaskan bahwa data mining dengan algoritma C4.5 dapat memberikan hasil yang efektif dan handal pada dataset besar yang bervariasi[9]

Pada Penelitian Repository BSI (2024) dalam penelitiannya yang berjudul "Penerapan Algoritma C4.5 Untuk Prediksi Kecanduan Smartphone pada Pelajar SMA" menjelaskan bahwa data mining merupakan teknik analisis data yang memungkinkan identifikasi pola perilaku dari data perilaku pengguna, dalam hal ini kecanduan penggunaan smartphone. Penelitian ini mengumpulkan data kuantitatif berupa durasi penggunaan aplikasi, frekuensi akses, dan kebiasaan lain para pelajar SMA. Algoritma C4.5 digunakan untuk mengklasifikasikan tingkat kecanduan smartphone berdasarkan variabel-variabel tersebut. Dalam prosesnya dilakukan tahapan preprocessing data yang menyangkut penyesuaian nilai atribut, pembagian data training dan testing, pembangunan model pohon keputusan, serta evaluasi output model menggunakan metrik akurasi dan confusion matrix. Hasil

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repository akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

menunjukkan model dengan akurasi 91,47% mampu mendeteksi tingkat kecurduan secara efektif dan membantu penentuan intervensi yang tepat.[10]

Pada Penelitian Iddrus, M., & Dewi Wulan Sari (2023) dalam penelitiannya tentang "Penerapan Metode Decision Tree dalam Klasifikasi Data Pendidikan", data mining didefinisikan sebagai proses penemuan pola bermakna dari data pendidikan menggunakan algoritma klasifikasi. Penelitian tersebut menerapkan Decision Tree untuk klasifikasi data siswa dengan hasil akurasi, precision, recall, dan F1-score sekitar 94%. Proses data mining meliputi pengumpulan data akademik, preprocessing missing values, pembangunan model, dan evaluasi komprehensif dengan confusion matrix[11].

2.3 Klasifikasi

Klasifikasi adalah salah satu teknik data mining yang bertujuan untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas-kelas data, sehingga model tersebut dapat digunakan untuk memprediksi kelas dari objek yang labelnya belum diketahui. Klasifikasi merupakan proses pembelajaran terawasi (supervised learning) di mana model dilatih menggunakan data training yang telah memiliki label kelas.

Pada Penelitian Kristina, S. (2025) dalam penelitiannya tentang "Penerapan Algoritma C4.5 untuk Klasifikasi Kelayakan Penerima Bantuan", klasifikasi didefinisikan sebagai metode supervised learning yang membangun model prediksi menggunakan pohon keputusan dari data berlabel. Penelitian tersebut menerapkan C4.5 untuk mengklasifikasikan penerima bantuan sosial dengan akurasi tinggi. Proses klasifikasi mencakup preprocessing data, training model, dan validasi dengan metrik evaluasi standar[12].





Pada Penelitian Abdul Fath Ramadhan, dkk. (2024) dalam penelitiannya tentang "Penentuan Kelayakan Karyawan Baru Menggunakan Data Mining Klasifikasi C4.5", klasifikasi didefinisikan sebagai proses pembelajaran supervised dengan decision tree yang membagi data menjadi subset homogen menggunakan splitting berbasis informasi gain ratio. Penelitian tersebut menerapkan klasifikasi C4.5 pada 960 data pelamar kerja berdasarkan atribut kualifikasi, dengan hasil akurasi 98,44%. Proses klasifikasi meliputi perhitungan entropy dan gain ratio, pembentukan pohon keputusan, serta pruning untuk mengatasi overfitting[13].

Pada Penelitian Pratiwi, N., & Nugroho, A. (2024) dalam penelitiannya tentang "Klasifikasi Pendapatan Menggunakan Algoritma Random Forest dan Decision Tree", klasifikasi didefinisikan sebagai teknik pembelajaran mesin untuk memprediksi kategori pendapatan berdasarkan atribut demografi dan pekerjaan. Penelitian tersebut menerapkan Decision Tree untuk klasifikasi dengan hasil akurasi kompetitif dibanding ensemble methods melalui cross-validation. Prosesnya melibatkan pembagian data training-testing, pembentukan model, dan evaluasi multi-metrik[14].

2.4 Decision Tree

Decision Tree atau pohon keputusan adalah salah satu metode klasifikasi yang menggunakan struktur pohon atau struktur hierarkis untuk merepresentasikan serangkaian keputusan yang mengarah pada suatu kesimpulan atau prediksi. Metode ini populer dalam data mining karena kemampuannya menghasilkan model yang mudah dipahami dan diinterpretasikan oleh manusia.

Pada Penelitian Desmita, R., dkk. (2025) dalam penelitiannya tentang "Comparative Analysis of Decision Tree and Random Forest", Decision Tree



didefinisikan sebagai struktur hierarkis yang merepresentasikan keputusan rekursif berdasarkan atribut pembagi terbaik menggunakan metrik impurity seperti Gini atau entropy. Penelitian tersebut membandingkan Decision Tree dengan ensemble methods untuk berbagai dataset, menyoroti keunggulan interpretabilitas pohon tunggal. Prosesnya meliputi training, pruning post-pruning, dan evaluasi multi-metrik termasuk F1-score dan ROC-AUC[15].

Pada Penelitian Musfikar, R., Sari, D.P., dan Hartono, B. (2023) dalam penelitiannya tentang "Aplikasi Prediksi Prestasi pada Siswa Menggunakan Algoritma C4.5", Decision Tree didefinisikan sebagai model klasifikasi yang membangun pohon keputusan dengan node akar hingga leaf berdasarkan gain ratio. Penelitian tersebut menerapkan Decision Tree untuk prediksi prestasi siswa dengan akurasi 94,12%. Prosesnya mencakup visualisasi pohon keputusan dan interpretasi aturan klasifikasi yang dihasilkan[16].

Pada penelitian Adiat Pariddudin & Fajri Syabani Warsa (2023) dalam penelitiannya tentang "Penerapan Algoritma C4.5 Untuk Rekomendasi Mentor Santri Baru", Decision Tree C4.5 digunakan untuk mengembangkan sistem rekomendasi yang dapat mencocokkan santri baru dengan mentor yang sesuai berdasarkan karakteristik kepribadian, kemampuan akademik, dan preferensi pembelajaran. Penelitian ini berhasil mengatasi masalah keterbatasan staf pengajar dengan menyediakan solusi sistematis berbasis data mining. Decision Tree dalam penelitian ini berperan sebagai model yang dapat menyaring kandidat mentor secara otomatis berdasarkan kriteria tertentu, sehingga proses penempatan mentor menjadi lebih efisien dan objektif. Meskipun nilai akurasi tidak disebutkan secara spesifik,

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

sistem ini terbukti efektif dalam mengurangi beban kerja administratif dan meningkatkan kualitas pencocokan mentor-santri[4].

2.5 Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma Decision Tree yang paling populer dan banyak digunakan dalam klasifikasi data. Algoritma ini dikembangkan oleh Ross Quinlan sebagai perbaikan dari algoritma pendahulunya, ID3 (Iterative Dichotomiser 3), dengan menambahkan berbagai fitur dan perbaikan untuk mengatasi keterbatasan ID3.

Pada Penelitian Musfikar, R., Sari, D.P., dan Hartono, B. (2023) dalam penelitiannya tentang "Aplikasi Prediksi Prestasi pada Siswa Menggunakan Algoritma C4.5", algoritma C4.5 didefinisikan sebagai metode klasifikasi berbasis pohon keputusan yang membangun struktur hierarkis dengan memilih atribut pembagi terbaik berdasarkan perhitungan gain ratio. Penelitian tersebut menerapkan algoritma C4.5 untuk memprediksi prestasi siswa berdasarkan nilai rapor dan tingkat absensi, dengan hasil akurasi model mencapai 94,12%. Proses penelitian meliputi pengumpulan data nilai dan absensi siswa, tahap preprocessing untuk menangani nilai yang hilang, pembangunan model Decision Tree C4.5, serta evaluasi performa model menggunakan confusion matrix yang menunjukkan nilai precision dan recall yang tinggi untuk semua kelas prestasi[16].

Pada Penelitian Nasution, D., dkk. (2024) dalam penelitiannya tentang "Pengujian Algoritma C4.5 Untuk Mengevaluasi Kinerja Pegawai", algoritma C4.5 didefinisikan sebagai model decision tree yang handal untuk klasifikasi kinerja dengan perhitungan gain ratio dan pruning otomatis. Penelitian tersebut menerapkan C4.5 pada data pegawai dengan hasil akurasi 100% melalui confusion matrix



sempurna. Prosesnya mencakup perhitungan accuracy, precision, recall, serta visualisasi matriks untuk validasi ketepatan prediksi[17].

Pada Penelitian AD Kuswanto, dkk. (2024) dalam penelitiannya tentang "Penerapan Algoritma C4.5 dalam Klasifikasi Prestasi Atlet", algoritma C4.5 didefinisikan sebagai metode pembelajaran mesin yang membangun pohon keputusan melalui perhitungan entropy, gain, split info, dan gain ratio untuk memilih atribut terbaik. Penelitian tersebut menerapkan C4.5 untuk mengklasifikasikan prestasi atlet berdasarkan data POHMN XVIII dengan hasil model yang interpretatif. Prosesnya meliputi preprocessing atribut campuran, pembangunan Decision Tree, dan evaluasi performance vector[18].

2.6 Confusion Matrix

Confusion Matrix adalah salah satu metode evaluasi performa model klasifikasi yang paling banyak digunakan dalam machine learning dan data mining. Metode ini menggunakan tabel matriks untuk menampilkan hasil prediksi model dibandingkan dengan nilai aktual, sehingga memudahkan analisis kesalahan klasifikasi yang terjadi.

Pada Penelitian Firnanda, P.A., dkk. (2025) dalam penelitiannya tentang "Analisis Perbandingan Decision Tree dan Random Forest untuk Pengenalan Pola Dasar Angka berdasarkan Gerakan Tangan", Confusion Matrix didefinisikan sebagai metode evaluasi komprehensif yang merangkum jumlah prediksi benar dan salah dalam bentuk tabel silang, dikelompokkan berdasarkan masing-masing kelas untuk menghitung metrik akurasi, precision, recall, dan F1-score secara granular. Penelitian tersebut menerapkan Confusion Matrix untuk membandingkan performa Decision Tree dan Random Forest pada dataset gerakan tangan, dengan analisis





menunjukkan perbedaan karakteristik kesalahan klasifikasi antar algoritma. Proses evaluasi meliputi pembuatan matriks 2x2 untuk klasifikasi biner, visualisasi heatmap kesalahan, perhitungan True Positive, False Positive, True Negative, dan False Negative, serta interpretasi implikasi kesalahan tipe I dan tipe II dalam konteks pengenalan pola visual. Hasil menunjukkan Confusion Matrix efektif mengidentifikasi kelemahan model yang tidak terlihat dari akurasi keseluruhan saja[19].

Pada Penelitian Ningsih, L.N. (2025) dalam penelitiannya tentang "Implementasi Algoritma C4.5 Untuk Klasifikasi Kepuasan Pelanggan", Confusion Matrix didefinisikan sebagai tabel evaluasi standar yang membandingkan hasil prediksi model dengan label aktual untuk mengukur tingkat kesalahan prediksi secara detail melalui komponen TP, FP, TN, dan FN. Penelitian tersebut menerapkan Confusion Matrix pada model C4.5 untuk klasifikasi tingkat kepuasan pelanggan berdasarkan data survei layanan, menghasilkan tingkat kesalahan prediksi yang sangat kecil dengan akurasi tinggi. Proses analisis meliputi pembuatan matriks klasifikasi multi-kelas, perhitungan metrik precision-recall per kelas, visualisasi distribusi kesalahan, serta interpretasi dampak false positive terhadap keputusan bisnis seperti retensi pelanggan. Temuan penelitian menegaskan bahwa Confusion Matrix krusial untuk memahami sensitivitas model terhadap variasi data pelanggan dan memvalidasi keandalan prediksi kepuasan[20].

Pada Penelitian Tim Peneliti Teknois UNBIN (2024) dalam penelitiannya tentang "Penerapan Algoritma C4.5 Untuk Prediksi Kelayakan Pengajuan Kartu Kredit", Confusion Matrix didefinisikan sebagai alat pengukuran akurasi yang membandingkan hasil prediksi dengan data aktual berdasarkan variabel seperti status rumah, penghasilan, status pernikahan, dan umur untuk menentukan kelayakan

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.

3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

nasabah secara akurat. Penelitian tersebut menerapkan Confusion Matrix pada model C4.5 dengan hasil akurasi 83,33%, mengidentifikasi pola kesalahan prediksi pada aplikasi kredit melalui rumus akurasi standar $(TP+TN)/(TP+FP+FN+TN)$. Proses evaluasi mencakup pembuatan Tabel Confusion Matrix lengkap, analisis gain per atribut, perbandingan prediksi vs aktual, serta pembahasan implikasi finansial dari false negative (penolakan nasabah layak) dan false positive (penerimaan nasabah berisiko). Confusion Matrix dalam penelitian ini berperan penting dalam mengukur keakuratan model untuk pengambilan keputusan kredit yang objektif[21].



1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



2.7 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

No.	Nama Penulis	Judul	Metode	Hasil Penelitian	Keterbatasan yang Ditemukan
(1)	(2)	(3)	(4)	(5)	(6)
1.	Adiat Pariddudin & Fajri Syabani Warsa (2023)	Penerapan Algoritma C4.5 Untuk Rekomendasi Mentor Santri Baru	Algoritma C4.5	Model C4.5 berhasil dikembangkan untuk merekomendasikan mentor bagi santri baru dan mengatasi keterbatasan staf pengajar.	Tidak disebutkan spesifik dalam cuplikan. Saran: Kemungkinan keterbatasan data atau belum diuji komparasi dengan algoritma lain untuk memvalidasi efisiensi C4.5.
2.	Deyana Fitri Sri Septiana, dkk. (2025)	Penerapan Metode Decision Tree Algoritma C4.5 Dalam Sistem Rekomendasi Jurusan Bagi Calon Mahasiswa Baru	Decision Tree C4.5	Sistem rekomendasi menghasilkan akurasi model sebesar 80%, memberikan rekomendasi yang lebih objektif berdasarkan atribut minat dan bakat.	Akurasi model masih bisa ditingkatkan (masih 80%). Penelitian ini belum menyertakan komparasi dengan algoritma klasifikasi lain untuk perbandingan performa.
3.	Untung Suropati & Deny Saputra (2025)	Implementasi Algoritma Decision Tree C4.5 Untuk Klasifikasi Pengangkatan Karyawan Tetap	Decision Tree C4.5	Akurasi klasifikasi keseluruhan sebesar 89,25%, mengidentifikasi faktor-faktor kunci pengangkatan karyawan tetap.	Recall untuk klasifikasi karyawan Tetap lebih rendah (78,57%) dibandingkan kelas Kontrak (91,14%), menunjukkan model kurang sensitif pada kelas 'Tetap'.
4.	Rival Amperawansyah & Dhanny Permatasari Putri (2022)	Pemilihan Jenis Ponsel Pintar Sesuai dengan Kebutuhan	Forward Chaining dan Decision Tree	Sistem pakar berbasis web memberikan hasil rekomendasi smartphone dengan akurasi 95%.	Penggunaan kombinasi dua metode (Forward Chaining dan Decision Tree) membuat kontribusi spesifik masing-masing metode menjadi kurang jelas.



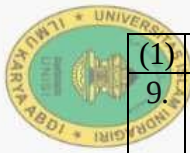
Tabel 2.2 Penelitian Terdahulu Lanjutan

(1)	(2)	(3)	(4)	(5)	(6)
5.	Ahmad Fitriansyah, dkk. (2022)	Sistem Rekomendasi Pemilihan Smartphone Dengan Metode Case Based Reasoning	Case Based Reasoning (CBR)	Sistem rekomendasi berbasis CBR terbukti dapat membantu pelajar dan orang tua dalam mengambil keputusan pembelian smartphone.	Data Kasus Tidak Real-time: Basis data smartphone terbatas pada data April-Juni 2022, sehingga tidak mengikuti perkembangan pasar yang cepat.
6.	Ahmad Taufiq Ramadhan, dkk. (2023)	Penerapan Algoritma Decision Tree Dalam Melakukan Analisis Klasifikasi Harga Handphone	Decision Tree	Algoritma Decision Tree memberikan akurasi rata-rata 81% dalam mengklasifikasi harga smartphone.	Akurasi model relatif rendah (81%) dan penelitian mengakui adanya <i>false positives</i> dan <i>false negatives</i> . Disarankan untuk melakukan komparasi dengan algoritma lain.
7.	Neni Firda Wardani Tan & Kusnawi (2022)	Klasifikasi Menggunakan Algoritma C4.5 dalam Memprediksi Status Keorganisasian Mahasiswa	Algoritma Decision Tree C4.5	Model C4.5 berhasil menganalisis kecenderungan mahasiswa untuk menjadi aktif atau tidak aktif dalam organisasi.	Jumlah dataset yang digunakan relatif kecil, yaitu 167 data, yang dapat mengurangi validitas generalisasi model. Saran: Perlu penambahan fitur/atribut lain.
8.	Edy, Rino, & Desiyanna Lasut (2025)	Sistem Rekomendasi Jurusan Pendidikan Menggunakan Algoritma C4.5 Berbasis Web untuk Siswa Sekolah Menengah Atas	Algoritma C4.5	Penelitian sukses mengembangkan sistem berbasis web menggunakan C4.5 untuk merekomendasikan jurusan SMA sesuai potensi siswa.	Tidak disebutkan spesifik. Saran: Model pohon keputusan perlu diuji komparasi dengan metode klasifikasi lain untuk menguji keunggulan C4.5.



Tabel 2.3 Penelitian Terdahulu Lanjutan

(1)	(2)	(3)	(4)	(5)	(6)
9	Ivan Zuhdiansyah & Ardytha Luthfiarta (2024)	Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan Singular Value Decomposition	K-Means Clustering dan SVD	Sistem ini berhasil mengatasi masalah information overload dalam e-commerce melalui Collaborative Filtering.	Keterbatasan Metode: Tidak menggunakan Decision Tree/C4.5. Metode ini rentan terhadap masalah sparsity (data yang jarang) dan cold start (pengguna/item baru).
10	Anif Wicaksono Gegadang Putro & Tedy Setiadi (2023)	Penerapan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar	Decision Tree (C4.5)	Model prediksi kelulusan mencapai akurasi sangat tinggi, yaitu 96,67%.	Penelitian hanya menggunakan data dari satu kecamatan (Kecamatan Juai), sehingga keterbatasan terletak pada validitas generalisasi model di wilayah lain.
11	Bima Hamdani Mawaridi, Muhammad Faisal (2023)	Rekomendasi Merek Mobil Untuk Calon Pembeli Menggunakan Algoritma Decision Tree	Algoritma Decision Tree CART	Algoritma Decision Tree berhasil mengklasifikasikan merek mobil dengan akurasi tertinggi 85%.	Akurasi 85% menunjukkan masih ada ruang untuk perbaikan. Saran: Perlu penambahan atribut yang lebih detail (misal: fitur keamanan atau biaya perawatan) untuk model yang lebih presisi.
12	Faisal, Harry Dhika, Helmi Veris (2021)	Penerapan Algoritma Decision Tree Dalam Penjualan Handphone	Algoritma Decision Tree	Model prediksi restock barang mencapai akurasi 88,89%.	Recall untuk kelas tidak restok hanya 50%. Model gagal mengidentifikasi setengah dari kasus yang seharusnya tidak perlu restock, berisiko menyebabkan penumpukan stok.



Tabel 2.4 Penelitian Terdahulu Lanjutan

(1)	(2)	(3)	(4)	(5)	(6)
13.	Kartika Handayani, dkk. (2021)	Komparasi Algoritma C4.5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah	Komparasi C4.5 dan Naïve Bayes	Naïve Bayes (akurasi 93,26%) lebih baik daripada C4.5 (akurasi 93,22%) dalam menentukan kelayakan donor darah.	Performa C4.5 lebih rendah (akurasi 93,22%) dibandingkan Naïve Bayes dan diklasifikasikan sebagai fair classification.
14.	Nur Diana Saputri, dkk. (2022)	Komparasi Penerapan Metode Bagging dan Adaboost pada Algoritma C4.5 untuk Prediksi Penyakit Stroke	C4.5 + Bagging & Adaboost	Penerapan ensemble learning berhasil meningkatkan akurasi hingga 95.02% (C4.5 + Bagging).	Algoritma C4.5 murni memiliki nilai presisi yang sangat kecil (27%) pada kelas stroke, yang mengindikasikan model murni sangat buruk dalam memprediksi kasus stroke yang sebenarnya.
15.	Andi Ernawati, Sri Wahyuni (2024)	Analisis Data Mining Pola Penggunaan Seluler dan Klasifikasi Perilaku Pengguna di Berbagai Perangkat Menggunakan Metode C4.5	Metode C4.5	Model pohon keputusan C4.5 berhasil dibuat untuk klasifikasi perilaku pengguna seluler dan mengidentifikasi atribut kunci yang memengaruhi perilaku.	Keterbatasan Data: Data yang digunakan adalah data sekunder dari Kaggle.com (data global), sehingga tidak representatif untuk mencerminkan pola dan perilaku pengguna seluler di Indonesia.

2.8 Rangkuman

Penelitian oleh Adiat Pariddudin & Fajri Syabani Warsa (2023) mengaplikasikan Algoritma C4.5 untuk mengembangkan Sistem Rekomendasi Mentor Santri Baru. Penelitian ini dilatarbelakangi oleh jumlah santri baru yang besar dan keterbatasan staf pengajar, yang berpotensi menyebabkan inefektivitas proses





pendidikan. Keunggulan dari penelitian ini terletak pada kemampuannya menyediakan solusi sistematis berbasis data mining untuk menyeleksi mentor secara efisien. Meskipun model pohon keputusan C4.5 berhasil dikembangkan, kelemahannya adalah nilai akurasi tidak disebutkan secara spesifik dalam ringkasan penelitian. Selain itu, keterbatasan penelitian ini kemungkinan berada pada ukuran dan variasi data yang digunakan, serta belum diuji komparasi dengan algoritma recommender system lain untuk memvalidasi efisiensi C4.5[22].

Penelitian oleh Deyana Fitri Sri Septiana, dkk. (2025) menerapkan Decision Tree Algoritma C4.5 dalam Sistem Rekomendasi Jurusan Bagi Calon Mahasiswa Baru. Penelitian ini bertujuan mengatasi kebingungan calon mahasiswa dan memberikan rekomendasi yang lebih objektif dan akurat berdasarkan atribut minat dan bakat mereka. Keunggulan C4.5 di sini adalah kemampuannya menghasilkan pohon keputusan yang mudah dipahami. Hasil penelitian menunjukkan bahwa model ini mampu mencapai Akurasi 80%. Namun, kelemahannya adalah akurasi model masih bisa ditingkatkan (80%) dan penelitian ini belum menyertakan komparasi dengan algoritma klasifikasi lain, sehingga tidak dapat dipastikan apakah C4.5 adalah metode paling optimal untuk kasus rekomendasi jurusan ini[23].

Penelitian oleh Untung Suropati & Deny Saputra (2025) fokus pada implementasi Algoritma Decision Tree C4.5 untuk Klasifikasi Pengangkatan Karyawan Tetap di lingkungan perusahaan. Tujuan utamanya adalah menciptakan dasar pengambilan keputusan oleh HRD yang lebih objektif dan transparan. Keunggulan model ini adalah akurasi klasifikasi keseluruhan yang cukup baik, mencapai 89,25%. Kelemahannya teridentifikasi jelas pada metrik performa kelas minoritas: Recall untuk klasifikasi karyawan Tetap hanya 78,57%, jauh lebih rendah

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



dibandingkan kelas Kontrak (91,14%). Ini berarti model kurang sensitif dalam mengidentifikasi karyawan yang layak diangkat menjadi karyawan tetap, berisiko tinggi menimbulkan false negatives[24].

Penelitian oleh Rival Amperawansyah & Dhanny Permatasari Putri (2022) menggunakan pendekatan ganda, yaitu kombinasi Forward Chaining dan Decision Tree, untuk sistem Pemilihan Jenis Ponsel Pintar. Tujuannya adalah membantu masyarakat segmen harga rendah (1-3 juta Rupiah) mengambil keputusan pembelian yang tepat. Keunggulan utama penelitian ini adalah sistem pakar berbasis web yang dihasilkan memiliki Akurasi tinggi 95% dalam memberikan rekomendasi yang sesuai kebutuhan. Namun, kelemahannya adalah penggunaan kombinasi dua metode (Forward Chaining dan Decision Tree) menimbulkan ambiguitas mengenai kontribusi spesifik dari Decision Tree, dan sejauh mana akurasi 95% tersebut dipengaruhi oleh masing-masing metode[6].

Penelitian oleh Ahmad Fitriansyah, dkk. (2022) mengembangkan Sistem Rekomendasi Pemilihan Smartphone menggunakan Metode Case Based Reasoning (CBR). Penelitian ini bertujuan membantu pelajar dan orang tua dalam menentukan pilihan di tengah banyaknya variasi smartphone yang tersedia. Keunggulan terletak pada sistem rekomendasi berbasis kasus yang intuitif. Namun, kelemahan dan keterbatasannya adalah penelitian ini tidak menggunakan Decision Tree/C4.5. Selain itu, akurasi rekomendasi CBR sangat bergantung pada kualitas dan kelengkapan data kasus yang sudah tersimpan, yang berpotensi gagal memberikan rekomendasi untuk kasus baru (cold start)[25].

Penelitian oleh Ahmad Taufiq Ramadhan, dkk. (2023) menerapkan Algoritma Decision Tree dalam melakukan Analisis Klasifikasi Harga Handphone

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



(rendah/tinggi) berdasarkan 10 fitur. Tujuannya adalah mengidentifikasi fitur yang paling berpengaruh terhadap klasifikasi harga. Hasil penelitian menunjukkan Akurasi rata-rata 81%. Kelemahannya adalah Akurasi model relatif rendah (81%), yang mengindikasikan bahwa model mungkin belum optimal. Penelitian juga mencatat terjadinya false positives dan false negatives, serta menyarankan perlunya komparasi dengan algoritma lain untuk perbaikan performa[26].

Penelitian oleh Neni Firda Wardani Tan & Kusnawi (2022) menggunakan Algoritma Decision Tree C4.5 untuk Memprediksi Status Keorganisasian Mahasiswa (aktif/tidak aktif). Tujuannya adalah menganalisis kecenderungan mahasiswa dalam berorganisasi berdasarkan atribut non-akademik seperti public speaking dan tingkat kepercayaan diri. Keunggulan model ini adalah berhasil mengidentifikasi atribut-atribut non-akademik yang berpengaruh. Kelemahannya adalah jumlah dataset yang digunakan relatif kecil (167 data). Volume data yang kecil ini dapat mengurangi validitas generalisasi dan stabilitas model jika diterapkan pada populasi mahasiswa yang lebih besar[27].

Penelitian oleh Edy, Rino, & Desiyanna Lasut (2025) mengembangkan Sistem Rekomendasi Jurusan Pendidikan menggunakan Algoritma C4.5 Berbasis Web. Tujuannya adalah memberikan rekomendasi jurusan SMA yang sesuai dengan potensi diri siswa berdasarkan minat dan nilai. Keunggulan penelitian ini adalah keberhasilan menghasilkan sistem yang praktis dan mudah diakses (berbasis web). Namun, nilai akurasi tidak disebutkan dalam ringkasan, dan sebagai keterbatasan, disarankan bahwa model pohon keputusan tersebut perlu diuji komparasi dengan metode klasifikasi lain untuk memvalidasi keunggulan C4.5 dalam kasus rekomendasi pendidikan ini[28].

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
 2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
 3. Plagiarisme juga dilarang untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



Penelitian oleh Ivan Zuhdiansyah & Ardytha Luthfiarta (2024) membahas Sistem Rekomendasi Pembelian Smartphone berbasis Algoritma K-Means dan Singular Value Decomposition (SVD). Tujuannya adalah mengatasi masalah information overload di e-commerce dengan memberikan rekomendasi berbasis Collaborative Filtering. Kelemahan dan keterbatasannya adalah penelitian ini tidak menggunakan Decision Tree/C4.5. Lebih lanjut, metode yang digunakan (SVD/Collaborative Filtering) rentan terhadap masalah sparsity (data interaksi yang jarang) dan cold start (kesulitan merekomendasikan produk baru atau pengguna baru)[8].

Penelitian oleh Arif Wicaksono Gegadang Putro & Tedy Setiadi (2023) menerapkan Klasifikasi Decision Tree (C4.5) untuk Memprediksi Kelulusan Siswa Sekolah Dasar. Tujuannya adalah menghasilkan model prediksi yang akurat menggunakan nilai rata-rata mata pelajaran. Keunggulan utama penelitian ini adalah model prediksi kelulusan mencapai Akurasi yang sangat tinggi, 96,67%, dan pohon keputusan yang dihasilkan mudah diinterpretasikan. Kelemahannya terletak pada validitas generalisasi: penelitian hanya menggunakan data dari satu lokasi spesifik (Kecamatan Juai), sehingga stabilitas model perlu diuji dengan data yang lebih beragam dari wilayah lain[29].

Penelitian oleh Bima Hamdani Mawaridi & Muhammad Faisal (2023) menggunakan Algoritma Decision Tree untuk Rekomendasi Merek Mobil (AS, Jepang, Eropa). Tujuannya adalah membantu calon pembeli menemukan merek mobil yang memenuhi spesifikasi. Hasil penelitian menunjukkan Akurasi tertinggi 85% pada kedalaman tree 7. Kelemahannya adalah Akurasi 85% masih memiliki ruang untuk ditingkatkan. Keterbatasan yang dapat ditindaklanjuti adalah perlunya

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.
3. Plagiarisme juga dilarang dan dapat dikenakan sanksi.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.



penambahan atribut yang lebih detail (seperti biaya perawatan atau fitur keamanan) untuk model yang lebih presisi dan relevan bagi calon pembeli[30].

Penelitian oleh Faisal, Harry Dhika, & Helmi Veris (2021) berfokus pada penerapan Algoritma Decision Tree dalam Penjualan Handphone untuk memprediksi produk yang harus direstok atau tidak. Model prediksi mencapai Akurasi 88,89%. Keunggulannya: Presisi prediksi restok tergolong tinggi (87,50%). Kelemahannya teridentifikasi pada Recall untuk kelas 'tidak restok' yang hanya 50%. Ini berarti model gagal mengidentifikasi setengah dari kasus di mana barang seharusnya tidak perlu restock, yang dapat menyebabkan penumpukan stok berlebih dan kerugian[1].

Penelitian oleh Kartika Handayani, dkk. (2021) melakukan Komparasi Algoritma C4.5 dan Naïve Bayes dalam Penentuan Status Kelayakan Donor Darah. Tujuannya adalah mengidentifikasi metode klasifikasi terbaik untuk kasus ini. Hasil penelitian menunjukkan Akurasi C4.5 93,22%. Kelemahannya adalah Performa C4.5 lebih rendah (akurasi 93,22%) dibandingkan Naïve Bayes (93,26%) dan diklasifikasikan sebagai fair classification (AUC 0,758). Ini menyimpulkan bahwa Naïve Bayes lebih optimal daripada C4.5 untuk kasus penentuan kelayakan donor darah ini[31].

Penelitian oleh Nur Diana Saputri, dkk. (2022) mengomparasikan Metode Ensemble Learning (Bagging dan Adaboost) pada Algoritma C4.5 untuk Prediksi Penyakit Stroke. Tujuannya adalah meningkatkan akurasi model prediksi. Keunggulan utama adalah penerapan ensemble learning berhasil meningkatkan akurasi signifikan, dengan C4.5 + Bagging mencapai 95,02% (tertinggi). Kelemahannya adalah Algoritma C4.5 murni memiliki nilai presisi yang sangat kecil (27%) pada kelas stroke sebelum di-boost. Hal ini mengindikasikan bahwa model C4.5 murni sangat

1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.

2. Penggunaan tanpa izin untuk kepentingan akademik, penelitian, dan pendidikan diperbolehkan dengan mencantumkan sumber.

3. Plagiarisme juga dilarang dan dapat dikenakan sanksi. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

peraturan yang berlaku di Indonesia.

buruk dalam mengidentifikasi kasus stroke yang sebenarnya, sehingga wajib diintegrasikan dengan metode boosting atau bagging[32].

Penelitian oleh Andi Ernawati & Sri Wahyuni (2024) menggunakan Metode C4.5 untuk Analisis Data Mining Pola Penggunaan Seluler dan Klasifikasi Perilaku Pengguna. Tujuannya adalah mengklasifikasikan perilaku pengguna di berbagai perangkat. Keunggulan terletak pada keberhasilan mengidentifikasi atribut kunci yang memengaruhi perilaku. Kelemahannya yang paling mendasar adalah Data yang digunakan adalah data sekunder dari Kaggle.com (data global). Karena data tersebut tidak representatif untuk pola pengguna seluler di Indonesia, validitas eksternal temuan untuk konteks lokal menjadi rendah[2].

2.9 Kesimpulan

Berbeda dengan penelitian terdahulu terkait rekomendasi smartphone, seperti kombinasi Forward Chaining dan Decision Tree yang menimbulkan ambiguitas kontribusi masing-masing metode (Amperawansyah & Putri, 2022), atau pendekatan collaborative filtering berbasis K-Means dan SVD yang rentan terhadap masalah cold-start (Zuhdiansyah & Luthfiarta, 2024), penelitian ini membangun dataset sendiri melalui pelabelan manual empat kategori kebutuhan (Gaming, Fotografi/Videografi, Pelajar, Pengguna Umum) dari data Kaggle yang diverifikasi via GSMArena, dilabeli berdasarkan riset artikel dan divalidasi ke Toko Sahabat Ponsel Tembilahan, mengintegrasikan variabel teknis dan budget ke model Decision Tree C4.5 yang dikomparasi terhadap empat algoritma lain (CART, Naive Bayes, KNN, Random Forest) dengan evaluasi lengkap (confusion matrix, precision, recall, F1-score, cross-validation 5-fold), lalu diimplementasikan sebagai platform web SmartPick dengan katalog, perbandingan produk, wishlist, rating, dan dashboard analytics.

