

Analisis Perbandingan *Decision Tree* Algoritma C4.5 dengan algoritma lainnya: *Systematic Literature Review*

Naurah Nazifah¹, Cahyo Prianto²

^{1,2}Teknik Informatika, Vokasi, Universitas Logistik dan Bisnis Internasional

¹nnaurah988@gmail.com, ²cahyo@ulbi.ac.id*

Abstract

Decision tree is one of the popular methods in data analysis and machine learning. The C4.5 algorithm is one of the most widely used decision tree algorithms because of its ability to produce decision rules that can be understood easily. However, various variations and developments of other decision tree algorithms have emerged, offering improved performance and new features. This study aims to carry out a comparative analysis between the C4.5 decision tree algorithm and several other decision tree algorithms that have been developed. The method used in this research is a systematic literature review, in which the researcher conducts a structured search and evaluation of relevant scientific articles. Researchers will compare the performance of the C4.5 algorithm with other algorithms based on several criteria, including predictive accuracy, computational complexity, interpretability of decision rules, and ability to handle unbalanced data. The results of the analysis show that the selection of a decision tree algorithm must be based on the specific purpose of the analysis and the characteristics of the data used. If the interpretability of decision rules is a major factor, the C4.5 algorithm remains a good choice. However, if predictive accuracy and handling of unbalanced data is a priority, algorithms such as Random Forest, Naive Bayes, or KNN may be a better choice.

Keywords: Data Mining, Prediction, Accuracy, C4.5

Abstrak

*Decision tree merupakan salah satu metode populer dalam analisis data dan pembelajaran mesin. Algoritma C4.5 adalah salah satu algoritma *decision tree* yang banyak digunakan karena kemampuannya dalam menghasilkan aturan keputusan yang dapat dipahami dengan mudah. Namun, berbagai variasi dan pengembangan algoritma *decision tree* lainnya telah muncul, menawarkan peningkatan kinerja dan fitur-fitur baru. Penelitian ini bertujuan untuk melakukan analisis perbandingan antara algoritma *decision tree* C4.5 dengan beberapa algoritma *decision tree* lainnya yang telah dikembangkan. Metode yang digunakan dalam penelitian ini yaitu *systematic literature review*, di mana peneliti melakukan pencarian terstruktur dan evaluasi terhadap artikel-artikel ilmiah yang relevan. Peneliti akan membandingkan kinerja algoritma C4.5 dengan algoritma lainnya berdasarkan beberapa kriteria, termasuk akurasi prediksi, kompleksitas komputasional, interpretabilitas aturan keputusan, dan kemampuan menangani data yang tidak seimbang. Hasil analisis memperlihatkan bahwa pemilihan algoritma *decision tree* harus didasarkan pada tujuan spesifik analisis dan karakteristik data yang digunakan. Jika interpretabilitas aturan keputusan menjadi faktor utama, algoritma C4.5 tetap menjadi pilihan yang baik. Namun, jika akurasi prediksi dan penanganan data yang tidak seimbang menjadi prioritas, algoritma-algoritma seperti *Random Forest*, *Naive Bayes*, atau *KNN* dapat menjadi pilihan yang lebih baik.*

Kata kunci: Data Mining, Prediksi, Akurasi, C4.5

1. Pendahuluan

1.1 Latar Belakang

Analisis perbandingan algoritma *Decision Tree* merupakan topik yang penting dalam studi ilmu data dan kecerdasan buatan. Algoritma *Decision Tree* telah digunakan secara luas untuk memodelkan dan memprediksi data dalam berbagai bidang, termasuk pengambilan keputusan, klasifikasi, dan prediksi. Salah satu algoritma *Decision Tree* yang populer adalah algoritma C4.5. Algoritma ini merupakan pengembangan dari algoritma C4.5 adalah algoritma yang dikembangkan dari algoritma ID3 [1]. C4.5 memiliki beberapa perubahan dan peningkatan yang membuatnya lebih efektif dalam menangani masalah yang lebih kompleks. Dalam penelitian ini, peneliti bertujuan untuk melakukan analisis perbandingan antara algoritma *Decision Tree* C4.5 dengan beberapa algoritma lainnya yang tersedia dalam literatur terkait. Metode yang kami gunakan untuk melakukan analisis adalah sistematis literature review. Dengan *Decision Tree* dibentuk sebuah pohon keputusan dengan atribut data yang akan menghasilkan aturan-aturan dalam mengklasifikasi [2].

Sistematis literature review adalah proses pengumpulan, evaluasi, dan sintesis dari studi sebelumnya yang relevan dengan topik penelitian yang sedang dipelajari. Dengan menggunakan metode ini, kami dapat mengidentifikasi, menganalisis, dan membandingkan studi sebelumnya yang telah dilakukan oleh peneliti lain dalam bidang analisis perbandingan algoritma klasifikasi lainnya.

Tujuan dari penelitian ini adalah untuk memperoleh pemahaman yang mendalam tentang kelebihan dan kelemahan algoritma *Decision Tree* C4.5 dibandingkan dengan algoritma lainnya. Dengan melakukan analisis perbandingan ini, diharapkan dapat menyediakan wawasan yang berharga bagi peneliti dan praktisi di bidang ilmu data dan kecerdasan buatan untuk memilih algoritma *Decision Tree* yang paling sesuai dengan kebutuhan mereka. Penelitian ini menggunakan beberapa pengujian seperti akurasi, presisi, *recall*, dan F-measure dimana hal tersebut dapat menjadi nilai tambah [3].

Dalam penelitian ini, peneliti akan mengumpulkan artikel-artikel ilmiah terkait yang telah dipublikasikan dalam jurnal-jurnal dan konferensi terkemuka. Selanjutnya akan melakukan evaluasi terhadap studi-studi tersebut berdasarkan kriteria-kriteria yang telah ditentukan, seperti metode yang digunakan, dataset yang digunakan, dan metrik evaluasi yang digunakan.

Selanjutnya, peneliti akan menganalisis dan membandingkan hasil-hasil yang ditemukan dalam

studi-studi tersebut. Mengeksplorasi kelebihan dan kelemahan algoritma *Decision Tree* C4.5 dibandingkan dengan algoritma lainnya dalam hal akurasi prediksi, waktu komputasi, kompleksitas model, dan kemampuan mengatasi masalah dengan atribut yang hilang atau kontinu. Algoritma *machine learning* bersifat konstruktif dalam membuat prediksi yang akurat dan memberikan analisis yang akurat [4].

Kesimpulan dari analisis perbandingan ini akan memberikan gambaran yang jelas tentang keunggulan dan batasan dari algoritma *Decision Tree* C4.5 dibandingkan dengan algoritma-algoritma *Decision Tree* lainnya. Hasil penelitian ini diharapkan dapat memberikan panduan yang bermanfaat bagi peneliti dan praktisi dalam memilih algoritma *Decision Tree* yang sesuai dengan kasus yang sedang mereka hadapi.

1.2 Tinjauan Literatur

Decision Tree(C4.5)

Decision Tree merupakan algoritma klasifikasi yang dinyatakan sebagai partisi rekursif dari ruang contoh. *Decision Tree* terdiri dari node yang membentuk pohon berakar, yang berarti pohon diarahkan dengan simpul yang disebut akar. Sebuah node dengan tepi keluar disebut internal atau tes node. Semua node yang lain disebut daun. Dalam pohon keputusan, setiap simpul internal membagi ruang misalnya menjadi dua atau lebih sub ruang sesuai dengan fungsi diskrit tertentu dari atribut nilai [5].

Sebuah pohon keputusan terdiri dari internal node yang menentukan tes pada variabel masukan individu atau atribut yang membagi data menjadi himpunan bagian yang lebih kecil, dan serangkaian node daun menetapkan kelas untuk masing-masing pengamatan di segmen yang dihasilkan. Pada penelitian ini, C4.5 membangun pohon keputusan dengan menggunakan konsep entropi informasi. Entropi sampel S dari pengamatan yang diklasifikasikan diberikan oleh

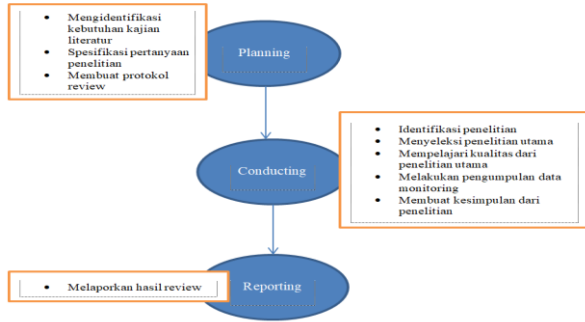
$$\text{Entropy}(S) = -p_1 \log_2(p_1) - p_0 \log_2(p_0),$$

Dimana p_1 adalah proporsi dari kelas yang nilainya 1 dan p_0 proporsi dari kelas yang nilainya 0 dalam sample S . C4.5 memeriksa informasi Gain normalisasi (perbedaan entropi) yang dihasilkan dari memilih atribut untuk membagi data. Atribut dengan informasi Gain normalisasi tertinggi adalah yang digunakan untuk membuat keputusan. Algoritma kemudian terbagi lagi menjadi subset lebih kecil [5].

2. Metode Penelitian

Metode Penelitian ini menggunakan metode tinjauan literatur sistematis. Tinjauan literatur sistematis atau

yang dikenal luas sebagai *Systematic Literature Review* adalah tinjauan literatur yang mengikuti serangkaian aturan baku digunakan untuk mengidentifikasi dan mensintesis semua studi yang relevan serta memberikan penilaian terhadap apa yang diketahui dari topik studietodologi penelitian yang digunakan pada penelitian ini menggunakan metode studi literatur yang didapat melalui informasi dari jurnal, buku, atau sumber lain yang telah dibuat sebelumnya [7]. Adapun alur dari penelitian yang dilakukan dalam penelitian ini digambarkan pada gambar 1.



Gambar 1. Diagram Alur Metodologi Penelitian

RQ1	Bagaimana perbandingan kinerja antara decision tree algoritma C4.5 dengan metode-metode lainnya dalam hal akurasi klasifikasi dan kompleksitas model?	Mengidentifikasi beberapa kinerja algoritma klasifikasi dengan algoritma decision tree C4.5 dalam hal akurasi klasifikasi model pada pembelajaran.
RQ2	Metode apa yang sering digunakan untuk melakukan perbandingan penelitian?	Mengidentifikasi metode yang digunakan pada penelitian sebelumnya untuk melakukan perbandingan.
RQ3	Bagaimana pengaruh variasi parameter pada algoritma C4.5 terhadap kinerja yang dihasilkan?	Mengidentifikasi hasil variasi parameter pada kinerja algoritma C4.5

2.1. Planning

Dalam tahap ini penting untuk membuat Research Question. Tahapan ini sangat penting karena didasari dari ketertarikan atau pusat tujuan dari penulisan *Sistematic Literature Review*[8]. Planning merupakan tahap awal dalam melakukan tinjauan literatur secara sistematis. SLR adalah pendekatan sistematis untuk mengumpulkan, mengevaluasi, dan menyintesis studi literatur yang relevan dengan topik penelitian tertentu. Tahap perencanaan dalam SLR penting untuk mengarahkan langkah-langkah selanjutnya dan memastikan bahwa tinjauan literatur dilakukan secara terorganisir dan komprehensif.

Perencanaan yang cermat dan terorganisir pada tahap awal SLR sangat penting untuk memastikan validitas dan kualitas tinjauan literatur yang dilakukan. Hal ini juga membantu menghindari bias dan meminimalkan kesalahan dalam proses pencarian dan evaluasi literatur yang relevan.

Research Question

Pertanyaan penelitian dibuat untuk menentukan metode yang akan digunakan dalam penelitian yang akan dilakukan oleh penulis nantinya. Berikut beberapa pertanyaan penelitiannya :

ID	Pertanyaan Penelitian	Motivasi
----	-----------------------	----------

2.2 Conducting

Conducting sebagai Tahapan pelaksanaan SLR terdiri dari tahapan-tahapan yang sudah ditetapkan. Menetapkan “search string” merupakan hal utama serta mengacu kepada PICOC (Population, Intervention, Comparison, Outcomes, Context) [9]. Pada tahap pelaksanaan atau penerapan metode yang telah direncanakan sebelumnya untuk melakukan tinjauan literatur sistematis. Tahap ini melibatkan proses praktis dalam mengumpulkan, mengevaluasi, dan menganalisis studi literatur yang relevan.

Population(P)	Apa populasi target yang akan digunakan dalam penelitian ini?
Intervention(I)	Algoritma decision tree c4.5 yang digunakan untuk mengklasifikasikan apa dalam penelitian ini?
Comparison(C)	Dengan apa algoritma decision tree lainnya akan dibandingkan algoritma c4.5?
Outcomes(O)	Apa hasil yang akan diukur atau dinilai dalam penelitian ini?

Context(C)	Apakah ada konteks khusus yang perlu diperhatikan dalam penelitian ini?
------------	---

- Melakukan analisis pada bukti empiris dari paper yang terpilih Dalam proses pencarian sumber penelitian didapatkan sebanyak 25 paper yang sesuai dan dengan kriteria.

2.3 Reporting

Hasil dari tinjauan literatur secara sistematis ini berupa tulisan, baik untuk dipublikasikan dalam bentuk paper ke jurnal ilmiah atau untuk menyusun Literature Review dari skripsi/tesis/disertasi. Struktur penulisan dari tinjauan literatur secara sistematis biasanya terdiri dari 3 bagian besar, yaitu: Pendahuluan (Introduction), Utama (Main Body) dan Kesimpulan (Conclusion) [10]. Penyusunan laporan akhir atau makalah ilmiah yang menggambarkan secara terperinci proses, temuan, dan kesimpulan dari tinjauan literatur yang telah dilakukan. Tahap ini penting untuk mengkomunikasikan secara jelas metodologi yang digunakan, hasil tinjauan literatur, serta implikasi atau kontribusi penelitian terhadap pengetahuan yang ada.

Tahap "reporting" pada SLR memastikan bahwa temuan dan kesimpulan dari tinjauan literatur disampaikan secara terperinci dan transparan kepada pembaca atau pihak yang berkepentingan. Hal ini memungkinkan pengulangan dan peneliti lainnya untuk memahami dan memanfaatkan pengetahuan yang diperoleh dari tinjauan literatur yang telah dilakukan.

Strategi Pencarian

Dalam penelitian ini, strategi pencarian sumber yang digunakan adalah sebagai berikut:

- Menyeleksi paper sesuai dengan kriteria yang telah ditentukan
- Melakukan pencarian paper pada berbagai basis data seperti, Google Scholar, Crossref

Seleksi Studi

Setelah tahapan pencarian sumber penelitian dengan strategi yang telah ditentukan pada penelitian, tahap selanjutnya sejumlah 25 paper diseleksi kembali. Tahapan seleksi studi dilakukan dengan membaca dan menganalisis bagian abstrak dari setiap sumber yang berkaitan dengan topik pada tahapan seleksi studi didapatkan paper yang relevan dan sesuai dengan kriteria yang telah ditentukan. Kemudian dilakukan review secara mendetail dengan membaca seluruh bagian dari paper yang sesuai dengan pertanyaan penelitian dengan topik.

Quality Assessment

Pada tahap selanjutnya yaitu melakukan pengecekan dan evaluasi pada data sumber paper dengan beberapa pertanyaan berikut ini :

QA1	Apakah literature tersebut diunggah dalam waktu 5 tahun terakhir?
QA2	Apakah literature tersebut menggunakan metode yang valid dan dapat dipertanggung jawabkan?
QA3	Apakah proses penafsiran hasil yang dilakukan telah mencakup semua aspek relevan, efisien dan bersifat konsisten?

Hasil penelitian

Table 1 Penelitian Terdahulu

No	Nama Peneliti	Tahun	Judul dan Kesimpulan Penelitian
1.	Agus Purwanto, Eko Agus Darmadi2 [11].	2018	Judul : perbandingan minat siswa smu pada metode klasifikasi menggunakan 5 algoritma Kesimpulan : 1. Berdasarkan hasil uji banding dari ke lima algoritma yaitu algoritma K-Nearest Neighbor, Naïve Bayes, Pohon Keputusan (C4.5), Rule Induction dan Deep Learning variable yang berpengaruh terhadap hasil adalah minat siswa.2. Berdasarkan dari nilai precision dan accuracy, metode Decision Tree (C4.5) memiliki nilai lebih tinggi daripada algoritma yang lainnya dengan nilai precision sebesar 89,09 % dan nilai accuracy sebesar 84,00 %.
2.	Ajeng Setianingrum, Ayu Hindayanti, Dita	2021	Judul : Perbandingan metode algoritma k-nn & metode algoritma c4.5 pada analisa kredit macet (studi kasus pt

	Meilani Cahya, Dini Silvi Purnia [12].		tungmung textile bintang) Kesimpulan : Berdasarkan hasil penelitian dengan judul “Perbandingan Metode Algoritma KNN dan Metode Algoritma C4.5 pada Analisa Kredit Macet(Studi Kasus Koperasi PT. Tungmung Textile Bintang) dengan menggunakan 73 dataset dan 5 atribut. Algoritma KNN mendapatkan nilai akurasi sebanyak 45,21% sedangkan Algoritma C4.5 mendapatkan nilai akurasi sebanyak 61,64%. Penulis dapat menyimpulkan bahwa Algoritma C4.5 lebih akurat untuk menentukan kredit macet daripada Algoritma KNN.
3.	Andy Supriyadi [13].	2023	Judul : Perbandingan Algoritma Naive Bayes Dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi Kesimpulan : Setelah penelitian ini dilakukan, diperoleh hasil penelitian dengan kesimpulan bahwa hasil perbandingan klasifikasi dosen berprestasi menggunakan metode Naive Bayes mendapatkan nilai akurasi sebesar 94,80 %, dengan nilai standar deviasi sebesar 3,79%, nilai presisi kelas berprestasi sebesar 95,25% dan nilai presisi kelas tidak berprestasi sebesar 94,15%.
4.	Jonson Manurung, Bosker Sinaga ,PaskaMarto Hasugian, Logaraj, Sethu Ramen [14].	2022	Judul : analisis algoritma c4.5 dan fuzzy sugeno untuk optimasi rule base fuzzy Kesimpulan : Hasil perbandingan metode fuzzy sugeno yang diintegrasikan dengan algoritma C4.5 mendapatkan hasil akurasi sebesar 88,57 %.
5.	M Fariz Januarsyah, Ermatita Zuhairi, Reza Firsandaya Malik[15] .	2019	Judul : Perbandingan Algoritma Random Forest, Decision Stump, Naive Bayes, Bayesian Network dan Algoritma C4.5 Untuk Prediksi Pola Kartu Poker Kesimpulan : .Random Forest menempati peringkat pertama yang disusul dengan metode C4.5, Decision Stump, Naive Bayes dan Bayesian Network. Metode tersebut membutuhkan waktu yang lama sebesar 2,26 detik, jauh jika dibandingkan dengan metode lain yang hanya menggunakan waktu 0 sampai 0.08 detik.
6.	Tupan Tri Muryono, Ahmad Taufik, Irwansyah [16].	2021	Judul : perbandingan algoritma k-nearest neighbor, decision tree, dan naive bayes untuk menentukan kelayakan pemberian kredit Kesimpulan : Algoritma K-Nearest Neighbor (K-NN) memperoleh hasil akurasi tertinggi sebesar 93.33% pada pengujian ke 5. Algoritma Decision Tree (C4.5) memperoleh hasil akurasi tertinggi sebesar 98.00% pada pengujian ke 3. dan Algoritma Naive Bayes memperoleh hasil akurasi tertinggi sebesar 86.67% pada pengujian ke 3.
7.	Anggita Safitri Febriarini, Erna Zuni Astuti [17].	2019	Judul : Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang Kesimpulan : Dengan pengujian 3 kali yang sudah dilakukan, maka dapat diketahui bahwa pengujian 1 dengan perbandingan data Training dan data testing 60% : 40% adalah nilai akurasi 92,5%, precision 92,42% dan Recall 98,39%, dan pengujian 2 dari perbandingan 70% : 30% adalah nilai akurasi 93,33%, precision 93,88%, dan recall 97,87%, sedangkan pengujian 3 dengan perbandingan 80% : 20% adalah nilai akurasi 95%, precision 96,67%, dan recall 96,67% berhasil dan baik.
8.	Saufika Sukmawati, Sulastri, Hery	2022	Judul : Perbandingan algoritma c 4.5 dan algoritma naive bayes untuk klasifikasi pekerja migran indonesia

	Februariyanti, Arief Jananto [18].		Kesimpulan : hasil pengujian dengan akurasi tertinggi dihasilkan pada percobaan pertama dengan data training sebanyak 1802 data dan data testing sebanyak 772. Algoritma C 4.5 memiliki tingkat akurasi sebesar 84.84% dan presentasi error rate sebesar 15.16% sedangkan Algoritma Naïve Bayes memiliki nilai akurasi sebesar 58.29% dan error rate sebesar 41.71%.
9.	Titik Faizah, Arief Jananto [19].	2021	Judul : Perbandingan Algoritma C4.5 Dan Id3 Untuk Prediksi Ketepatan Waktu Lulus Mahasiswa Kesimpulan : Hasil pengujian terhadap dataset dengan 3 komposisi berbeda diperoleh nilai akurasi terbaik yaitu pada data testing 30% sebesar 78,75% pada algoritma ID3 dan sebesar 81,8% pada algoritma C4.5. Secara keseluruhan dari hasil implementasi, Algoritma C4.5 memiliki tingkat akurasi yang lebih tinggi dibandingkan algoritma ID3. Dengan demikian dapat dikatakan bahwa algoritma C4.5 adalah algoritma terbaik untuk menentukan prediksi ketepatan lulus mahasiswa.
10.	Victor Saputra Ginting, Kusrini, Emha Taufiq [20].	2020	Judul : Implementasi algoritma c4.5 untuk memprediksi keterlambatan pembayaran sumbangan pembangunan pendidikan sekolah menggunakan python Kesimpulan : Berdasarkan hasil yang didapatkan dari keterangan hasil dan hasil prediksi dengan Algoritma C4.5, maka nilai akurasi, precision, dan recall sebesar 73%, 71%, dan 71%.
11.	Meyer Mega, Jasmir [21].	2023	Judul : Prediksi Masa Studi Mahasiswa Unama Jambi Menggunakan Metode Algoritma C4.5 Kesimpulan : Hasil Klasifikasi Algoritma C4.5 menggunakan data training, Pada tes 5 Fold Cross Validation dengan jumlah kelas Classified Instances 159 dan 249, Incorrectly Classified Instances 92 dan 2, dan persentasi akurasi Correctly Classified Instances sebesar 63.3466% dan 99.2032%
12.	Tedi Wahyudi, Popon Handayani, Rudianto [22].	2022	Judul : perbandingan akurasi c4.5 dan naive bayes untuk Evaluasi kinerja karyawan pt catur sentosa adiprana Kesimpulan : H1 yang menyatakan bahwa algoritma C4.5 memiliki nilai akurasi yang berbeda dengan algoritma Naïve Bayes. Algoritma klasifikasi C4.5 memiliki nilai akurasi sebesar 98.18% lebih unggul 3.03% dibandingkan dengan Naïve Bayes yang memiliki nilai akurasi 95.15%.
13.	Indah Sulihati [23].	2022	Judul : Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi Kesimpulan : Hasil penelitian pada dataset campus recruitment dengan menggunakan metode Decision Tree didapatkan akurasi 82,34%, dengan menggunakan metode Naive Bayes didapatkan akurasi 91,08%.
14.	Chandra Wirawan [24].	2020	Judul: Teknik Data Mining Menggunakan Algoritma Decision Tree C4.5 untuk Memprediksi Tingkat Kelulusan Tepat Waktu Kesimpulan: nilai akurasi 89.82% dengan Precision 52.63% dan Recall 41.67%, dan nilai AUC sebesar 76.6% nilai 50 ini mengindikasikan bahwa performa keakurasian pada percobaan tersebut bernilai cukup baik.
15	Taghfirul Azhima Yoga Siswa, Prihandoko [25].	2018	Judul: perbandingan kinerja algoritma c4.5, naive bayes, k-nearest neighbor, logistic regression, dan support vector

machines untuk mendeteksi penyakit kanker payudara
Kesimpulan: Hasilkinerja terbaik yang diuji menggunakan T-Test didapatkan bahwa algoritma Logistic Regression dan Support Vector Machines memiliki nilai akurasi tertinggi yang sama nilainya yaitu sebesar 0,968.

Pembahasan Hasil

Tahap pembahasan hasil menjawab pertanyaan yang telah ditentukan pada Research Question.

RQ1 Bagaimana perbandingan kinerja antara decision tree algoritma C4.5 dengan metode-metode lainnya dalam hal akurasi klasifikasi dan kompleksitas model?

Dari table 1. Review jurnal yang diperoleh dari beberapa metode yang dipakai mengenai Analisis Perbandingan dari hasil riset 15 jurnal, terdapat beberapa metode dengan akurasi klasifikasi dan kompleksitas model yaitu:

11	C4.5	63,3%	-	-
12	C4.5	98,18%	Naive Bayes	95,15%
13	C4.5	82,34%	Naive Bayes	91,08%
14	C4.5	89,82%	-	-
15	C4.5	-	Naive Bayes, KNN, Logistic Regression, SVM	SVM=96,8%

Tabel 2. Hasil Perbandingan

Jurnal ke-	Algoritma 1	Akurasi	Algoritma 2	Akurasi
1	C4.5	84%	KNN, Naive Bayes	-
2	C4.5	61,64%	KNN	45,21%
3	C4.5	95,25%	Naive Bayes	94,80%
4	C4.5	88,57%	Fuzzy	-
5	C4.5	76%	Random Forest, stump, Naive bayes	-
6	C4.5	98%	KNN, Naive Bayes	93,3% - 86,67%
7	C4.5	96,67%	-	-
8	C4.5	84,84%	Naive Bayes	58,29%
9	C4.5	81,8%	ID3	78,75%
10	C4.5	73%	-	-

RQ 2 Algoritma apa yang sering dibandingkan dengan Algoritma C4.5?

Dari table 1. Review jurnal yang saya dapatkan mengusulkan beberapa metode yang dipakai mengenai Analisis Perbandingan dari hasil riset 15 jurnal, terdapat beberapa metode yang diusulkan oleh peneliti sebelumnya diantara lain :

Naive Bayes	8 Jurnal
Random Forest	1 Jurnal
K-Nearest Neighbor	4 Jurnal
Fuzzy	1 Jurnal

RQ 3 Bagaimana pengaruh variasi parameter pada algoritma terhadap kinerja yang dihasilkan?

Dalam melakukan variasi parameter pada algoritma, perlu diingat bahwa nilai optimal untuk setiap parameter dapat berbeda-beda tergantung pada dataset yang digunakan. Berdasarkan penelitian sebelumnya atribut dataset yang digunakan sebanyak 5-10 parameter, perlu dilakukan eksperimen dan evaluasi untuk menentukan kombinasi parameter yang paling sesuai untuk mencapai kinerja yang diinginkan.

kesimpulan

Berdasarkan hasil *Systematic Literature Review* menunjukkan bahwa algoritma C4.5 memiliki Perbandingan dari hasil riset jurnal, terdapat beberapa metode dengan akurasi klasifikasi dan kompleksitas model dapat digunakan sebagai dasar untuk memilih algoritma perbandingan yang sesuai dengan kebutuhan analisis data yang spesifik, seperti berdasarkan tabel 2, hasil penelitian perbandingan algoritma decision tree C4.5 dan KNN dengan akurasi 98% dan 93%, kemudian decision tree dan *Naive Bayes* dengan akurasi 82% dan 91%. Maka dari itu dalam penelitian ini, algoritma yang sering digunakan sebagai perbandingan dengan hasil akurasi yang tinggi.

Daftar Rujukan

- [1] N. H. Inda Arifin and W. J. Shudiq, "Algoritma Decision Tree Dengan Menggunakan Pruning dan Missing Value Untuk Prediksi Kredit Macet," *COREAI J. Kecerdasan Buatan, Komputasi dan Teknol. Inf.*, vol. 3, no. 1, pp. 38–45, 2022, doi: 10.33650/coreai.v3i1.4124.
- [2] P. Issn, A. Fatkhudin, M. Y. Febrianto, F. A. Artanto, M. W. N. Hadinata, and R. Fahlevi, "Mahasiswa Program Studi Manajemen Informatika Ump," vol. 8, no. 2, pp. 83–86, 2022.
- [3] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: 10.29408/edumatic.v5i2.3424.
- [4] K. L. Kohsasih and Z. Situmorang, "Analisis Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022, doi: 10.31294/inf.v9i1.11931.
- [5] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [6] U. Athiyah, A. P. Handayani, M. Y. Aldean, N. P. Putra, and R. Ramadhani, "Sistem Inferensi Fuzzy: Pengertian, Penerapan, dan Manfaatnya," *J. Dinda Data Sci. Inf. Technol. Data Anal.*, vol. 1, no. 2, pp. 73–76, 2021, doi: 10.20895/dinda.v1i2.201.
- [7] A. Sulaiman and S. Azizah, "Problem-Based Learning to Improve Critical Thinking Ability in Indonesia: A Systematic Literature Review," *J. Pedagog.*, vol. 07, no. 01, p. 139, 2020.
- [8] R. A. A. Sugiharto, "Tinjauan Pustaka Sistematis tentang Efektivitas Pembelajaran Group Discussion Online Terhadap Kemampuan Perkembangan Nalar Siswa," *J. IT-EDU*, vol. 6, no. 2, pp. 165–169, 2021, [Online]. Available: <https://ejournal.unesa.ac.id/index.php/it-edu/article/view/46622>
- [9] S. Purworaharjo and G. Firmansyah, "Tinjauan Literatur Secara Sistematis Pada Self-Service Business Intelligence," *Konf. Nas. Sist. Inf.*, pp. 986–990, 2018, [Online]. Available: <http://jurnal.atmaluhur.ac.id/index.php/knsi2018/article/view/482/407>
- [10] S. Wahyu and G. Firmansyah, "Sebuah Tinjauan Literatur Secara Sistematis Pada Enterprise Architecture Framework (EAF)," *Konf. Nas. Sist. Inf.*, pp. 705–710, 2018, [Online]. Available: <http://jurnal.atmaluhur.ac.id/index.php/knsi2018/article/viewFile/437/362>
- [11] A. Purwanto *et al.*, "Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma," vol. 2, no. 1, pp. 43–47, 2018.
- [12] A. Setianingrum, A. Hindayanti, D. M. Cahya, and D. S. Purnia, "Perbandingan Metode Algoritma K-NN & Metode Algoritma C45 Pada Analisa Kredit Macet (Studi Kasus PT Tungmung Textil Bintan)," *EVOLUSI J. Sains dan Manaj.*, vol. 9, no. 2, pp. 78–92, 2021, doi: 10.31294/evolusi.v9i2.11166.
- [13] A. Supriyadi, "Perbandingan Algoritma Naive Bayes dan Decision Tree (C4 . 5) dalam Klasifikasi Dosen Berprestasi," vol. 7, no. 1, pp. 39–49, 2023.
- [14] J. M. Jonson Manurung, B. S. Bosker Sinaga, P. M. H. Paska Marto Hasugian, L. Logaraj, and S. R. Sethu Ramen, "Analisis Algoritma C4.5 Dan Fuzzy Sugeno Untuk Optimasi Rule Base Fuzzy," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 166–171, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2488.
- [15] F. M. Januarsyah, E. Zuhairi, and R. M. Firsandaya, "Perbandingan Algoritma Random Forest, Decision Stump, Naïve Bayes, Bayesian Network dan Algoritma C4.5 Untuk Prediksi Pola Kartu Poker," *Pros. Annu. Res. Semin.*, vol. 5, no. 1, pp. 122–126, 2019, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Poker+Hand>.
- [16] T. T. Muryono, A. Taufik, and I. Irwansyah, "Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit," *Infotech J. Technol. Inf.*, vol. 7, no. 1, pp. 35–40, 2021, doi: 10.37365/jti.v7i1.104.
- [17] A. S. Febriarini and E. Z. Astuti, "Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang," *Eksplora Inform.*, vol. 8, no. 2, pp. 95–103, 2019, doi: 10.30864/eksplora.v8i2.156.
- [18] S. Sukmawati, H. Februariyanti, A. Jananto, U. Stikubank Semarang Jl Tri Lomba Juang No, and S. Kode Pos, "I N F O R M a T I K a Perbandingan Algoritma C 4.5 Dan Algoritma Naive Bayes Untuk Klasifikasi Pekerja Migran Indonesia," *J. Inform. Manaj. dan Komput.*, vol. 14, no. 1, pp. 7–16, 2022.
- [19] T. Faizah and A. Jananto, "Perbandingan Algoritma C4 . 5 Dan Id3 Untuk Prediksi," *Jatisi*, vol. 8, no. 2, pp. 980–990, 2021.
- [20] V. S. Ginting, K. Kusriani, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, pp. 36–44, 2020, doi: 10.35585/inspir.v10i1.2535.
- [21] M. Metode, C. Algoritma, M. S. Informasi, and U. D. Bangsa, "Prediksi Masa Studi Mahasiswa Unama Jambi," vol. 8, no. 1, 2023.
- [22] R. Rudianto, T. Wahyudi, and P. Handayani, "Perbandingan Akurasi C4.5 Dan Naive Bayes Untuk Evaluasi Kinerja Karyawan Pt Catur Sentosa Adiprana," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 61–66, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2350.
- [23] I. Sulihati, "Penerapan Komparasi Algoritma C4.5 dan Naïve Bayes untuk Menentukan Hasil Seleksi Masuk Perguruan Tinggi," *J. Tecnoscienza*, vol. 6, no. 2, pp. 311–320, 2022, doi: 10.51158/tecnoscienza.v6i2.694.
- [24] C. Wirawan, "Teknik Data Mining Menggunakan Algoritma Decision Tree C4.5 untuk Memprediksi Tingkat Kelulusan Tepat Waktu," *Appl. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 47–52, 2020, doi: 10.15408/aism.v3i1.13033.
- [25] P. Prihandoko, "Perbandingan Kinerja Algoritma C4. 5, Naïve Bayes, K-Nearest Neighbor, Logistic Regression, Dan Support Vector Machines Untuk Mendeteksi Penyakit Kanker Payudara," *J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 1–10, 2018.

