

Analisis Perbandingan Decision Tree Dan Algoritma C4.5 Untuk Mengklasifikasikan Penerimaan Mahasiswa Elektro

Rizal Mantopani¹, Iqbal Lail Ramadhan², Raya Putra Samsara³, Riza Samsinar⁴

^{1,2,3,4}Teknik Elektro, Fakultas Teknik, Universitas Muhammadiyah Jakarta
Jl. Cempaka Putih Tengah No.27, Cempaka Putih, Kota Jakarta Pusat 10510
Telp. 0214244016

E-mail: 20210410200005@student.umj.ac.id

Abstrak

Penelitian ini mengkaji dan membandingkan metode Decision Tree dan algoritma C4.5 dalam mengklasifikasikan penerimaan mahasiswa pada program studi Teknik Elektro. Decision Tree adalah metode klasifikasi yang membangun model dalam bentuk pohon keputusan berdasarkan pembagian atribut. Algoritma C4.5 merupakan pengembangan dari Decision Tree yang memanfaatkan gain ratio untuk memilih atribut terbaik serta memiliki keunggulan dalam menangani atribut kontinu dan data hilang. Dataset yang digunakan dalam penelitian ini mencakup berbagai atribut, seperti nilai ujian, pengalaman sebelumnya, dan keterampilan relevan lainnya. Kedua algoritma diterapkan untuk mengevaluasi kinerja mereka dalam hal akurasi dan efisiensi komputasi pada berbagai skenario data, terutama pada dataset dengan atribut kontinu dan data hilang. Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan akurasi yang lebih tinggi dibandingkan dengan Decision Tree konvensional, khususnya dalam menangani dataset yang kompleks. Kemampuannya dalam mengelola atribut kontinu dan data hilang membuatnya lebih andal dalam aplikasi dunia nyata. Namun, kelemahannya adalah membutuhkan waktu komputasi yang lebih lama karena C4.5 memerlukan sumber daya pemrosesan yang lebih besar dibandingkan Decision Tree biasa. Kesimpulannya, algoritma C4.5 lebih cocok digunakan untuk tugas klasifikasi yang melibatkan dataset kompleks dengan atribut kontinu dan data hilang, sementara Decision Tree tetap menjadi pilihan yang efisien untuk dataset yang lebih sederhana dan terstruktur. Penelitian ini memberikan wawasan penting bagi peneliti dan praktisi dalam memilih algoritma klasifikasi yang sesuai berdasarkan kompleksitas data dan kebutuhan aplikasi tertentu.

Kata kunci: Decision Tree, C4.5, penerimaan mahasiswa.

Abstract

This study examines and compares the Decision Tree method and the C4.5 algorithm in classifying student admissions for the Electrical Engineering program. Decision Tree is a widely used classification method that builds a decision model in the form of a tree by splitting attributes to determine classifications. The C4.5 algorithm is an improvement of the Decision Tree, utilizing gain ratio to select the best attributes and offering advantages in handling continuous attributes and missing data. The dataset used in this study includes various attributes, such as exam scores, prior experience, and other relevant skills. Both algorithms are applied to evaluate their performance in terms of accuracy and computational efficiency under different data scenarios, particularly focusing on datasets with continuous attributes and missing values. The results indicate that the C4.5 algorithm achieves higher accuracy compared to the conventional Decision Tree, especially in managing complex datasets. Its ability to handle continuous data and missing values makes it more reliable in scenarios involving real-world applications. However, this comes at the cost of increased computational time, as C4.5 requires more processing resources than the simpler Decision Tree. In conclusion, the C4.5 algorithm is better suited for classification tasks involving complex datasets with continuous attributes and missing data, while the Decision Tree remains a viable option for simpler and more structured datasets. This study provides important

insights for researchers and practitioners in choosing the appropriate classification algorithm based on the complexity of data and specific application needs.

Keywords: *Decision Tree, C4.5, student admissions.*

1. Pendahuluan

Klasifikasi adalah suatu proses pengelompokkan dengan menggunakan data pelatihan atau pembelajaran dapat berupa data kategori atau continue dengan dilakukan pelabelan terhadap atribut yang menjadi luaran atau kelas record. (Supriyadi, 2023)

Dalam konteks penerimaan mahasiswa pada program studi Teknik Elektro, klasifikasi dapat digunakan untuk memprediksi kelulusan atau penerimaan berdasarkan berbagai faktor seperti nilai ujian, pengalaman, dan keterampilan lainnya. Penelitian ini bertujuan untuk membandingkan kinerja antara Decision Tree dan C4.5 dalam mengklasifikasikan penerimaan mahasiswa Teknik Elektro dengan mempertimbangkan keakuratan dan efisiensi waktu komputasi. Hasil dari penelitian ini diharapkan dapat memberikan wawasan mengenai metode klasifikasi yang paling sesuai untuk aplikasi penerimaan mahasiswa yang melibatkan data kompleks.

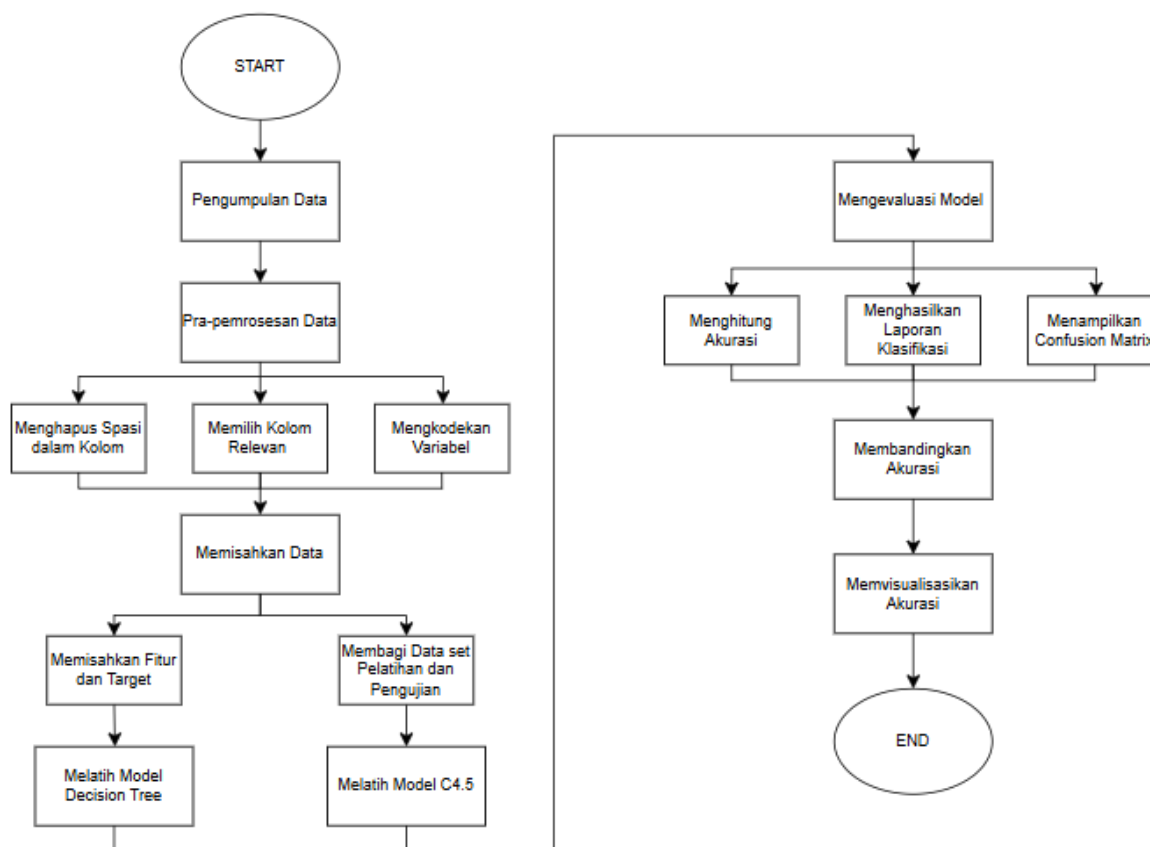
Decision tree merupakan suatu teknik menemukan kumpulan pola atau fungsi yang mendeskripsikan serta memisahkan kelas data yang satu dengan yang lainnya untuk menyatakan objek tersebut masuk pada kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah didefinisikan. Decision Tree (C4.5) merupakan algoritma untuk mengklasifikasikan apa yang akan terjadi di masa depan untuk memperoleh hasil yang maksimal. Decision Tree (C4.5) adalah algoritma pembelajaran yang mengimplementasikannya pencarian heuristik (penemuan) yang cukup sebanding menggunakan pohon keputusan. (Suyadi et al., 2017)

Algoritma C4.5 merupakan salah satu algoritma data mining yang dapat digunakan untuk pengelompokkan atau membuat segmentasi dan klasifikasi yang bersifat prediktif. Pembentukan pohon keputusan (Decision Tree) menjadi dasar dari algoritma ini, pohon keputusan dibentuk melalui beberapa tahapan yaitu dimulai dari pengumpulan data, pengolahan data training dan data kombinasi. Selanjutnya masuk ke inti Algoritma C4.5 yaitu penentuan atribut, hitung entropy & gain, menentukan gain tertinggi untuk menentukan akar dari pohon, partisi data dengan melakukan pengulangan perhitungan gain & entropy, dan terakhir dihasilkan pohon keputusan jika proses partisi menghasilkan semua record mendapat kelas N yang sama, tidak adanya atribut di dalam record yang terpartisi serta tidak adanya record di dalam cabang yang kosong. (Asyhari et al., 2024)

Penulis akan melakukan penelitian dengan judul "Analisis perbandingan Decision tree dan algoritma C4.5 untuk mengklasifikasikan penerimaan mahasiswa elektro" dengan jumlah dataset sebanyak 134 yang diambil dari data penerimaan mahasiswa baru teknik elektro tahun 2022 sampai 2024 universitas Muhammadiyah Jakarta. Pengolahan data dan pembentukan model dilakukan dengan menggunakan bahasa pemrograman Python.

2. Metodologi

Flowchart yang disajikan di bawah ini merepresentasikan metodologi yang diterapkan dalam penelitian ini, yang mencakup serangkaian langkah sistematis mulai dari pengumpulan data hingga evaluasi model. Setiap tahap dirancang dengan cermat untuk menjamin integritas data dan akurasi analisis, sehingga hasil yang diperoleh dapat dianggap valid dan relevan.



Gambar 2.1. Flowchart

a. Pengumpulan Data

Proses awal yang melibatkan pengumpulan data yang relevan untuk penelitian. Data ini bisa berasal dari berbagai sumber, seperti survei, database, atau eksperimen.

b. Pra-pemrosesan Data

Tahap ini meliputi beberapa langkah penting :

1. Menghapus Spasi dalam Kolom : Membersihkan data dengan menghapus spasi yang tidak perlu untuk memastikan konsistensi.
2. Memilih Kolom Relevan : Memilih kolom yang memiliki relevansi terhadap analisis yang akan dilakukan.
3. Mengkodekan Variabel : Mengubah variabel kategorikal menjadi format numerik agar dapat digunakan dalam model analisis

c. Memisahkan Data

Dalam proses yang dilakukan data dibagi menjadi 2 bagian :

1. Memisahkan Fitur dan Target : Menentukan fitur (variabel independen) dan Target (variabel dependen) untuk analisis.
2. Membagi dataset pelatihan dan pengujian : Memisahkan data menjadi set pelatihan untuk melatih model dan set pengujian untuk mengevaluasi kinerja modal.

d. Melatih Model

Pada tahapan ini, model Decision Tree dan C4.5 dilatih menggunakan data pelatihan yang telah dipisahkan sebelumnya. Proses pelatihan ini bertujuan untuk mengidentifikasi pola dan hubungan dalam data, sehingga model dapat membuat prediksi yang akurat terhadap data yang belum pernah dilihat sebelumnya.

e. Evaluasi Model

Setelah model dilatih, langkah selanjutnya adalah mengevaluasi kinerja dengan :

1. Menghitung akurasi : Mengukur seberapa baik model dalam memprediksi data.
2. Menghasilkan Laporan klasifikasi : Menyusun laporan yang menunjukkan hasil klasifikasi dari model.
3. Menampilkan Confusion Matrix : Menyajikan matrix yang menunjukkan jumlah prediksi yang benar dan salah.

f. Membandingkan Akurasi

Langkah selanjutnya adalah membandingkan akurasi dengan model yang lainnya . Proses ini penting untuk menilai kinerja model secara realtif.

g. Memvisualisasikan Akurasi

Terakhir, akurasi yang diperoleh akan di visualisasikan menggunakan tabel, sehingga hasilnya dapat dipahami dengan lebih jelas dan memberikan wawasan yang lebih mendalam tentang efektivitas model.

3. Hasil dan pembahasan

3.1. Proses data

a. Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari informasi penerimaan mahasiswa baru untuk program studi Elektro, dengan tujuan untuk menganalisis dan membandingkan kinerja algoritma Decision Tree dan C4.5.

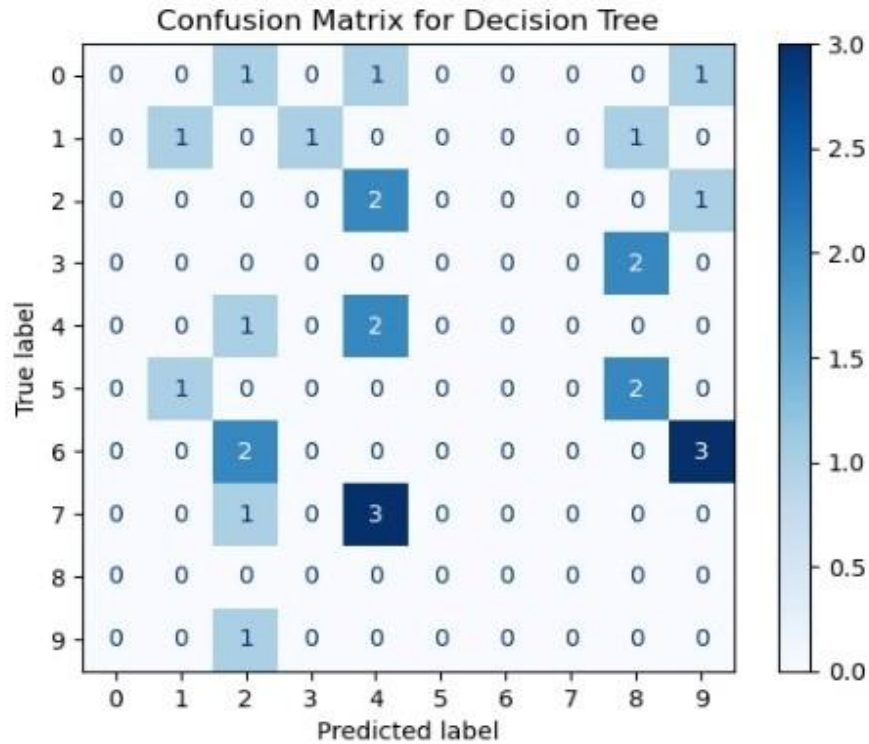
Tabel 3.1 Dataset Mahasiswa Elektro UMJ.

#	B	C	D	E	F	G	H	I	J	K	L	M
1	NIM	Nama	NIK	JK	Nama_Ibu	Agama	Asal_Sekolah	Tanggal_Lahir	Alamat	HP	LULUSAN	
2	22040200001	FACHRY AL HADZAMI NUF	3,17105E+15	Pria	ARYANTI	Islam	JAKARTA	24-Sep-00	Jl. Mardani raya gang c RT/007 RW	81213388917	SMK - ELEKTRO	
3	22040200002	KENAZ NITIPRANAJA KRIS	3,17405E+15	Pria	ANY KRISWIJAYANTI	Islam	JAKARTA	7 Januari 2005	Jl. Bambu Kuning 1 No.61	83892011009	SMA - IPA	
4	22040200003	RAHMAT PRASTYOBUDE	3,17304E+15	Pria	HARNANI	Islam	JAKARTA	18-Nov-00	Jl. Krendang barat gg.2	8,15692E+11	PESANTREN	
5	22040200004	MUHAMMAD HANIF AL A	3,17405E+15	Pria	DIAN PUSPITA	Islam	JAKARTA	18 Mei 2004	Jalan Juraganan No 13, RT 006 RW	89684761040	SMK - TITL	
6	22040200005	ALDINO RESTU NUGROH	3,17103E+15	Pria	ASIAH	Islam	JAKARTA	21 Maret 1997	Jl Cempaka Sari 3 No 31 RT 016 RW	85695438825	SMA - IPS	
7	22040200006	FERDI HASAN	3,17204E+15	Pria	JAMILAH	Islam	JAKARTA	6 Januari 2001	Jln.cilincing lama 1	81383046771	SMK - MESIN	
8	22040200007	IHSAN RIYADMOKO	3,27503E+15	Pria	SRI LESTARI	Islam	BEKASI	5 Januari 2001	Wisma Seroja Jl. Duwet 9A	8,95359E+11	SMA - IPA	
9	22040200008	MOHAMAD ARIFIN BAHTI	3,17204E+15	Pria	SAADIAH	Islam	JAKARTA	22 Desember 2003	Jln.tipar cakung rt002/005 Kelurahan	89529183502	SMK - TITL	
10	22040200009	MUHAMMAD ERSYAD	3,27504E+15	Pria	YULIA WIDYASTUTI	Islam	BEKASI	2 Desember 2004	GRIYA METROPOLITAN BLOK F.4/8	81296252277	PESANTREN	
11	22040200010	KHAIRIL ANHAR	3,17203E+15	Pria	SURYATI	Islam	DEPOK	1 Oktober 2003	Jl. Pembangunan 3, RT 10/RW 03 No	85707706696	SMA - IPS	
12	22040200011	RANDI ARIF ALDRIAN	3,17204E+15	Pria	SUKAESIH	Islam	JAKARTA	12 Desember 1998	JLN. BAKTI X CLINCING	81295727916	SMK - MESIN	
13	22040200012	MUHAMAD RAFU RAMAD	3,17204E+15	Pria	ELI KOMALA	Islam	JAKARTA	23-Nov-02	JLBELIMBING TERUSAN	83877368915	SMA - IPA	
14	22040200013	CANDRA SUBAGIA	3,21602E+15	Pria	NINING SUHAENI	Islam	TANGERANG	29 Januari 2003	Ujung Harapan RT 005/004 Kelurahan	89507152352	SMK - ELEKTRO	
15	22040200014	MUHAMMAD RIFIN AL M	3,17307E+15	Pria	KUNILAH	Islam	JAKARTA	5 Oktober 2007	Jln. Suacambah Barat VII No. 71	85155707660	PESANTREN	

3.2. Implementasi algoritma

a. Decision Tree

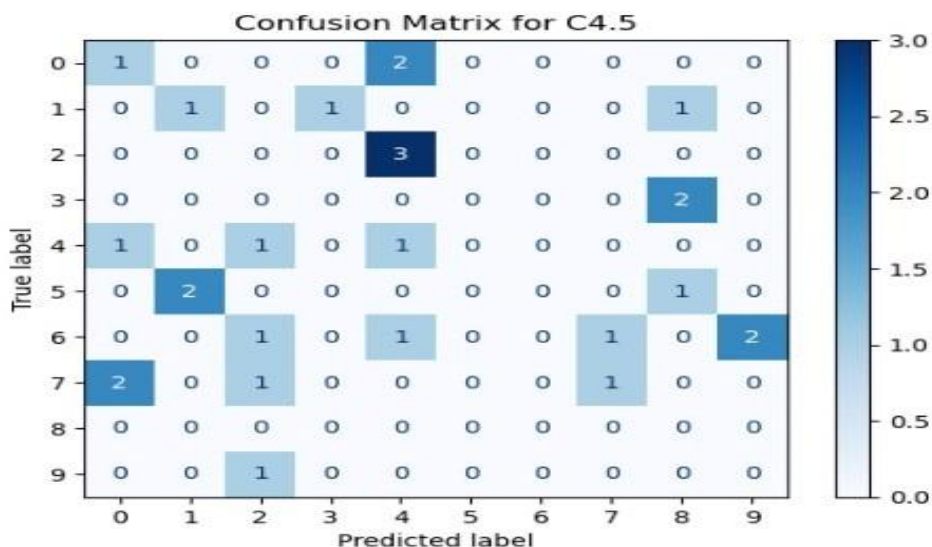
Decision tree adalah metode pemodelan prediktif dalam analisis data yang menggunakan struktur pohon. Tujuan decision tree adalah untuk menggambarkan serta membuat keputusan berdasarkan serangkaian aturan dan kondisi. Dikutip dari pengertian lainnya, decision tree juga disebut sebagai alat yang kuat dalam machine learning, dan dapat digunakan untuk memprediksi hasil berdasarkan serangkaian fitur atau variabel input. (Mukhsinin et al., 2024)



Gambar 3.2.a Confusion matrix decision tree

b. C4.5

Algoritma C4.5 adalah metode pembelajaran mesin berbasis pohon keputusan yang dirancang untuk tugas klasifikasi. Algoritma ini dikembangkan oleh Ross Quinlan sebagai peningkatan dari algoritma ID3. Tujuannya adalah membangun pohon keputusan yang dapat digunakan untuk memprediksi atau menentukan kelas data baru berdasarkan pola dalam data pelatihan. C4.5 bekerja dengan memilih atribut terbaik sebagai pemisah untuk setiap simpul pada pohon keputusan menggunakan konsep entropy dan information gain. Selain itu, algoritma ini dapat menangani atribut kontinu, nilai data yang hilang (missing values), serta melakukan pruning pada pohon untuk mencegah overfitting. (Pada et al., 2025)



Gambar 3.2.b Confusion matrix C4.5

3.3. Evaluasi Hasil Data Decision Tree Dan C4.5

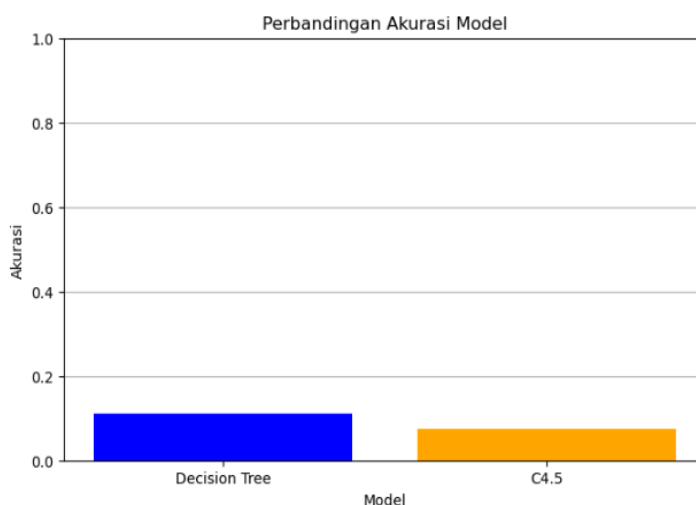
Evaluasi hasil data Decision Tree dan C4.5 di fokuskan pada empat matrix utama, yaitu precision, recall, F1-Score, dan Support. Analisis ini bertujuan untuk mengidentifikasi kekuatan dan kelemahan masing-masing model dalam mengklarifikasikan penerimaan mahasiswa elektro, seperti terlihat pada table 3.3.

Model	Class	Precision	Recall	F1-Score	Support	
Decision Tree	1	0.00	1.00	0.00	0	
	3	0.00	0.00	1.00	3	
	4	1.00	0.33	0.50	3	
	5	0.20	0.33	0.25	3	
	6	0.00	0.00	1.00	2	
	7	0.00	0.00	1.00	3	
	8	1.00	0.00	0.00	3	
	9	1.00	0.20	0.33	5	
	10	0.00	1.00	0.00	0	
	11	0.00	0.00	1.00	4	
	12	0.00	1.00	0.00	0	
	14	0.00	0.00	1.00	1	
		Accuracy			0.11	27
		Macro Avg	0.27	0.32	0.51	27
	Weighted Avg	0.43	0.11	0.63	27	
C4.5	3	0.00	0.00	1.00	3	
	4	0.67	0.67	0.67	3	
	5	0.00	0.00	1.00	3	
	6	1.00	0.00	0.00	2	
	7	0.00	0.00	1.00	3	
	8	1.00	0.00	0.00	3	
	9	0.00	0.00	1.00	5	
	10	0.00	1.00	0.00	0	
	11	0.00	0.00	1.00	4	

	12	0.00	1.00	0.00	0
	13	0.00	1.00	0.00	0
	14	0.00	0.00	1.00	1
	Accuracy			0.07	27
	Macro Avg	0.22	0.31	0.56	27
	Weighted Avg	0.26	0.07	0.78	27

Tabel 3.3 Tabel evaluasi

Perbandingan Akurasi model decision tree dan C4.5 disajikan dalam grafik berikut (Gambar 3.4) memberikan visualisasi yang jelas atas kinerja masing masing model.



Gambar 3.4 Perbandingan Akurasi Model Decision Tree dan C4.5

4. Kesimpulan dan Saran

4.1. Kesimpulan

Penelitian ini membandingkan algoritma Decision Tree dan C4.5 untuk klasifikasi penerimaan mahasiswa Teknik Elektro Universitas Muhammadiyah Jakarta. Berdasarkan hasil penelitian:

1. Dalam keakurasiannya, Decision Tree menunjukkan akurasi yang lebih tinggi dibandingkan Algoritma C4.5, terutama pada dataset yang kompleks dengan atribut kontinu dan data hilang.
2. Efisiensi Komputasi Decision Tree lebih cepat dan efisien dari segi waktu komputasi, tetapi kurang andal untuk menangani dataset yang kompleks.
3. Keunggulan dari kedua ini, Decision Tree lebih unggul dalam menangani atribut kontinu dan data hilang, membuatnya lebih cocok untuk aplikasi dunia nyata yang kompleks.
4. Keterbatasan pada C4.5 adalah memerlukan sumber daya komputasi yang lebih besar dibandingkan dengan Decision Tree.
5. Dengan demikian, C4.5 lebih direkomendasikan untuk klasifikasi dataset kompleks, sedangkan Decision Tree cocok untuk dataset yang lebih sederhana.

4.2. Saran

1. Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam untuk mengeksplorasi kinerja kedua algoritma dalam konteks yang lebih luas.
2. Studi dapat difokuskan pada optimasi algoritma C4.5 untuk mengurangi waktu komputasi, misalnya dengan parallel computing.
3. Algoritma yang diuji dalam penelitian ini dapat diaplikasikan pada bidang lain, seperti keuangan atau kesehatan, untuk menguji generalisasi kinerja.
4. Penelitian lanjutan disarankan untuk membandingkan algoritma lain seperti Random Forest atau Gradient Boosting untuk memahami efektivitasnya dalam klasifikasi data kompleks.

Daftar Pustaka

- Asyhari, M. R., Alijoyo, F. A., & Vernanda, D. (2024). Penerapan Algoritma C4.5 dan Metode 360 Derajat Pada Sistem Informasi Penilaian Karyawan. *Jurnal Tekno Kompak*, 18(1), 15. <https://doi.org/10.33365/jtk.v18i1.3215>
- Mukhsinin, D. A., Rafliansyah, M., Ibrahim, S. A., Rahmadden, R., & Wulandari, D. (2024). Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 570–579. <https://doi.org/10.57152/malcom.v4i2.1255>
- Pada, C., Pakaian, T., & Busana, R. I. A. (2025). *KLASIFIKASI HASIL PENJUALAN PAKAIAN MENGGUNAKAN ALGORITMA*. 9(1), 1291–1299.
- Supriyadi, A. (2023). Perbandingan Algoritma Naive Bayes dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi. *Generation Journal*, 7(1), 39–49. <https://doi.org/10.29407/gj.v7i1.19797>
- Suyadi, S., Setyanto, A., & Fattah, H. Al. (2017). Analisis Perbandingan Algoritma Decision Tree (C4.5) Dan K-Naive Bayes Untuk Mengklasifikasi Penerimaan Mahasiswa Baru Tingkat Universitas. *Indonesian Journal of Applied Informatics*, 2(1), 59. <https://doi.org/10.20961/ijai.v2i1.13258>
- Azwanti, N., & Putra, N. E. (2024). Analisis Kepuasan Customer pada Sdtechnology Computer dengan Algoritma Decision Tree. *Jurnal Desain Dan Analisis Teknologi*, 3(2), 137–148. <https://doi.org/10.58520/jddat.v3i2.62>
- Tukino, T. (2023). Penerapan Metode Algoritma C4.5 dalam Penilaian Kelayakan Pemberian Kredit Kepada Mitra Usaha PT Arita Prima Sukses. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK)*, 5(September), 306–314. <https://doi.org/10.33884/psnistek.v5i.8098>
- Azwanti, N., & Putra, N. E. (2024). Analisis Kepuasan Customer pada Sdtechnology Computer dengan Algoritma Decision Tree. *Jurnal Desain Dan Analisis Teknologi*, 3(2), 137–148. <https://doi.org/10.58520/jddat.v3i2.62>
- Azwanti, N., & Putra, N. E. (2024). Analisis Kepuasan Customer pada Sdtechnology Computer dengan Algoritma Decision Tree. *Jurnal Desain Dan Analisis Teknologi*, 3(2), 137–148. <https://doi.org/10.58520/jddat.v3i2.62>