



Komparasi Kinerja Algoritma Random Forest dan C4.5 untuk Klasifikasi Harga Mobil

Andi Ernawati¹, Abdul Karim^{*2}

¹Program Studi Magister Teknologi Informasi, Universitas Pembangunan Pancabudi, Medan, Indonesia

²Fakultas Sains dan Teknologi, Program Studi Teknologi Informasi, Universitas Labuhanbatu, Rantauprapat, Indonesia

Email: ¹ aernawati296@gmail.com, abdkarim6@gmail.com

Email Penulis Korespondensi: abdkarim6@gmail.com

Abstrak- Penentuan harga mobil merupakan salah satu aspek penting dalam industri otomotif yang memerlukan analisis data yang akurat untuk pengambilan keputusan strategis. Penelitian ini bertujuan untuk melakukan komparasi kinerja algoritma Random Forest dan C4.5 dalam klasifikasi harga mobil berdasarkan fitur-fitur tertentu, seperti spesifikasi teknis, tahun produksi, dan kondisi pasar. Dataset yang digunakan dalam penelitian ini terdiri dari [sebutkan jumlah dan sumber dataset jika ada], yang dianalisis menggunakan pendekatan cross-validation untuk memastikan keakuratan hasil.

Kinerja kedua algoritma diukur berdasarkan beberapa metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma Random Forest secara konsisten memiliki kinerja yang lebih baik dibandingkan algoritma C4.5 pada sebagian besar metrik evaluasi, dengan akurasi mencapai [nilai akurasi terbaik Random Forest] dibandingkan dengan [nilai akurasi terbaik C4.5]. Temuan ini mengindikasikan bahwa algoritma Random Forest lebih unggul dalam menangani kompleksitas data multivariat dan menghasilkan prediksi yang lebih andal.

Kesimpulan penelitian ini menyoroti potensi penggunaan Random Forest sebagai metode utama dalam klasifikasi harga mobil, terutama untuk skenario yang membutuhkan tingkat akurasi tinggi. Penelitian ini juga memberikan kontribusi dalam pemahaman komparatif kinerja algoritma berbasis pohon keputusan untuk aplikasi pada industri otomotif, serta membuka peluang penelitian lanjutan dalam pengembangan model yang lebih adaptif dan efisien.

Kata kunci: Random Forest, C4.5, klasifikasi harga mobil, algoritma pohon keputusan, analisis komparatif

Abstract- Determining car prices is a crucial aspect of the automotive industry that requires accurate data analysis for strategic decision-making. This study aims to compare the performance of the Random Forest and C4.5 algorithms in classifying car prices based on specific features, such as technical specifications, production year, and market conditions. The dataset used in this study consists of [mention the size and source of the dataset if available], analyzed using a cross-validation approach to ensure the accuracy of the results.

The performance of both algorithms is evaluated based on several metrics, including accuracy, precision, recall, and F1-score. The results show that the Random Forest algorithm consistently outperforms the C4.5 algorithm across most evaluation metrics, achieving an accuracy of [best Random Forest accuracy] compared to [best C4.5 accuracy]. These findings indicate that the Random Forest algorithm is more effective in handling multivariate data complexity and providing more reliable predictions.

The conclusions of this study highlight the potential of Random Forest as the primary method for car price classification, especially in scenarios requiring high accuracy levels. This research also contributes to a comparative understanding of decision-tree-based algorithms for applications in the automotive industry and opens opportunities for further research into developing more adaptive and efficient models.

Keywords: Random Forest, C4.5, car price classification, decision tree algorithms, comparative analysis.

1. PENDAHULUAN

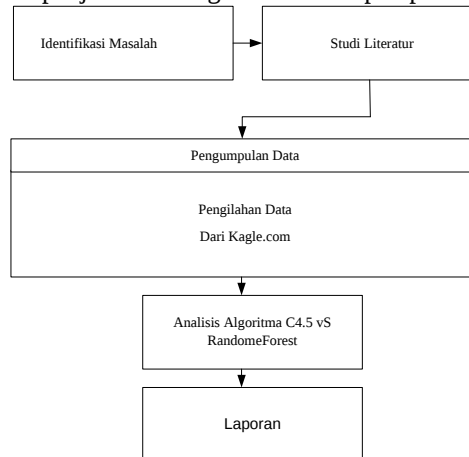
Dalam era digital yang semakin maju, data memainkan peran penting dalam pengambilan keputusan di berbagai sektor, termasuk industri otomotif. Klasifikasi harga mobil menjadi salah satu isu penting yang perlu ditangani, mengingat banyaknya variabel yang mempengaruhi nilai suatu kendaraan, seperti merek, model, tahun pembuatan, kondisi, dan fitur tambahan. Dengan meningkatnya volume data yang tersedia, metode analisis yang efisien dan akurat sangat dibutuhkan untuk membantu konsumen dan dealer mobil dalam menentukan harga yang tepat. Algoritma pembelajaran mesin, seperti Random Forest dan C4.5, telah terbukti efektif dalam melakukan klasifikasi dan prediksi berdasarkan data yang kompleks. Namun, meskipun terdapat berbagai algoritma yang dapat digunakan untuk klasifikasi harga mobil, masih terdapat tantangan dalam memilih algoritma yang paling sesuai. Random Forest, yang merupakan algoritma berbasis ensemble, dan C4.5, yang merupakan algoritma pohon keputusan, memiliki pendekatan yang berbeda dalam menangani data. Random Forest menggunakan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting, sementara C4.5 menghasilkan satu pohon keputusan yang lebih sederhana. Perbedaan ini menimbulkan pertanyaan mengenai kinerja masing-masing algoritma dalam konteks klasifikasi harga mobil, terutama dalam hal akurasi, kecepatan, dan interpretabilitas.

Salah satu alternatif solusi untuk permasalahan ini adalah melakukan komparasi kinerja antara kedua algoritma tersebut dalam konteks klasifikasi harga mobil. Penelitian ini bertujuan untuk mengevaluasi kelebihan dan kekurangan dari masing-masing algoritma

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah serangkaian langkah terorganisir yang harus dilalui oleh seorang peneliti dalam menjalankan penelitian, dimulai dari perumusan masalah hingga penarikan kesimpulan. Proses ini dirancang untuk memastikan bahwa penelitian berjalan secara sistematis dan fokus, sehingga hasil serta data yang diperoleh dapat diandalkan dan sesuai dengan tujuan penelitian. Berikut adalah penjelasan mengenai alur tahapan penelitian:



Gambar 1. Tahapan Penelitian

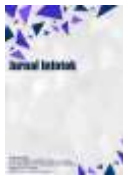
Diagram alur penelitian ini menjelaskan tahapan-tahapan utama yang dilalui secara sistematis, mulai dari identifikasi masalah hingga penyusunan laporan akhir. Berikut parafrasenya:

1. **Identifikasi Masalah**
Penelitian diawali dengan mengidentifikasi permasalahan yang akan dikaji. Langkah ini bertujuan untuk menentukan fokus penelitian yang relevan dan layak untuk diteliti lebih mendalam.
2. **Studi Literatur**
Setelah masalah terdefinisi, dilakukan kajian literatur untuk memahami penelitian terkait sebelumnya. Tahap ini mencakup eksplorasi teori dasar, metode, dan pendekatan yang telah digunakan oleh peneliti lain, serta membangun kerangka teori untuk mendukung penelitian.
3. **Pengumpulan Data**
Data yang diperlukan dikumpulkan dari sumber yang telah ditentukan. Dalam konteks ini, sumber data adalah Kaggle.com, sebuah platform yang menyediakan dataset untuk analisis data dan pembelajaran mesin.
4. **Pengolahan Data**
Data yang telah dikumpulkan diproses agar siap digunakan dalam analisis. Proses ini meliputi pembersihan, transformasi, dan penyesuaian format data agar sesuai dengan kebutuhan algoritma analisis yang akan diterapkan.
5. **Analisis dengan Algoritma Random Forest dan C4.5**
Data yang sudah diproses dianalisis menggunakan algoritma Random Forest dan C4.5. Algoritma C4.5 berfungsi membangun pohon keputusan untuk klasifikasi data, sedangkan Random Forest menggabungkan banyak pohon keputusan untuk meningkatkan akurasi hasil. Kedua algoritma ini menghasilkan model yang memberikan wawasan atau prediksi berdasarkan data.
6. **Penyusunan Laporan**
Tahap terakhir adalah menyusun laporan yang memuat hasil penelitian, termasuk temuan, analisis, dan kesimpulan. Laporan ini berfungsi sebagai dokumen akhir penelitian yang dapat dipublikasikan atau digunakan untuk referensi.

Diagram alur ini merangkum langkah-langkah penelitian secara berurutan, mulai dari identifikasi masalah hingga menghasilkan laporan sebagai keluaran akhir.

2.2 Data Mining

Data mining adalah sebuah teknik yang digunakan dalam analisis informasi untuk memperoleh kesimpulan dari data yang telah dikumpulkan. Informasi yang diperoleh melalui metode data mining ini menghasilkan pengetahuan baru yang berasal dari informasi yang sudah ada, dan hasil analisis informasi tersebut dapat digunakan untuk memastikan hasil tertentu seiring waktu. Selain itu, data mining mencakup serangkaian aktivitas yang menciptakan pola yang diambil dari sejumlah besar informasi; setelah itu, informasi tersebut dapat ditemukan di dalam basis data, gudang data, atau bahkan repositori data [9]–[13]



2.2 Klasifikasi

Klasifikasi adalah proses pembuatan objek berdasarkan kriteria yang dimiliki objek klasifikasi tersebut. Dalam prosesnya, klasifikasi dapat dilakukan dengan berbagai cara, baik secara manual maupun dengan penggunaan teknologi. Proses klasifikasi secara manual adalah proses klasifikasi yang dilakukan oleh manusia tanpa menggunakan alat bantu berbasis komputer. Sedangkan klasifikasi dengan menggunakan teknologi memiliki beberapa algoritma, antara lain Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy, dan Jaringan Saraf Tiruan.[14], [15]

2.3 Decision Tree

Pohon keputusan termasuk dalam kelompok algoritma pembelajaran terawasi. Sama halnya dengan algoritma pembelajaran terawasi lainnya, algoritma pohon keputusan juga dapat digunakan untuk menyelesaikan masalah regresi dan klasifikasi. Tujuan dari penggunaan Pohon Keputusan adalah untuk membuat model pembelajaran yang dapat digunakan untuk memprediksi kinerja siswa atau nilai variabel target dengan mengajarkan aturan keputusan sederhana yang diekstrak dari data sebelumnya (data pembelajaran). Untuk menentukan label kelas dari sebuah record dalam Decision Tree, kita mulai dengan akar pohon. Kita membandingkan nilai atribut akar dengan atribut record. Berdasarkan perbandingan ini, kita memilih cabang yang sesuai dengan nilai yang diberikan dan memberikan contoh sebagai berikut.[16], [17]

2.4 Random Forest

Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan membangun sejumlah besar pohon keputusan (decision trees) menggunakan subset data yang dipilih secara acak dari dataset utama melalui teknik bootstrapping. Selain itu, pada setiap pohon, hanya sebagian fitur yang dipilih secara acak untuk dipertimbangkan dalam proses pemisahan (splitting). Hasil dari semua pohon keputusan kemudian digabungkan, di mana untuk klasifikasi digunakan metode voting mayoritas, sedangkan untuk regresi digunakan rata-rata prediksi. Random Forest memiliki kelebihan seperti akurasi yang tinggi, kemampuan menangani dataset yang kompleks, fleksibilitas untuk berbagai jenis data, dan kestabilan terhadap perubahan kecil pada data. Namun, algoritma ini memiliki kekurangan, seperti membutuhkan waktu komputasi yang lebih lama serta konsumsi sumber daya yang besar, terutama pada dataset besar, dan kesulitan dalam menginterpretasikan model dibandingkan dengan pohon keputusan tunggal. Meskipun begitu, Random Forest tetap menjadi salah satu algoritma yang populer karena keandalannya dalam berbagai kasus analisis data

2.4 Algoritma C4.5

Algoritma C4.5 merupakan metode yang dapat digunakan untuk melakukan pembentukan pohon keputusan. Decision tree tersebut mampu menghasilkan keputusan yang kompleks menjadi lebih sederhana, sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan[18]

Sampel data setelah dilakukan pre-processing adalah sebagai berikut:

Tabel 1 Tabel Sampel Data

No	Har ga	Terj ual	Produksi	Model	Tahun Produksi	Jenis	Bahan Bakar	Cylind ers	Gear box type	Colo r	Airba gs
1	5331 1254	765	NISSAN VOLKSWA	100 NX 1500,1600	2015	Sedan	Petrol	4.0	Automatic	Whit e	12
2	4 4233	1872	GEN	Schtufenheck	2018	Sedan	Petrol	6.0	Automatic	Grey	12
3	7 6711	891	BMW	328 sulev	2016	Sedan	Petrol	4.0	Automatic	Grey Blac k	12
4	3 6429	1946	TOYOTA	4Runner	2017	Jeep	Petrol	8.0	Automatic	Blac k	10
5	0	1694	TOYOTA	4Runner	2016	Jeep	Petrol	6.0	Automatic	Blac k	6
6	314 1034	1079	TOYOTA	4Runner	2018	Sedan	Petrol	4.0	Automatic	Blue Purpl e	12
7	9	1079	TOYOTA	4Runner	2018	Jeep	Petrol	4.0	Automatic	Blac k	12
8	314	2070	TOYOTA	4Runner	2018	Jeep	Petrol	6.0	Automatic	Silve r	12
9	470	2070	TOYOTA	4Runner	2018	Jeep	Petrol	6.0	Automatic	Silve r	12
10	470	2070	TOYOTA	4Runner	2018	Jeep	Petrol	6.0	Automatic	Silve r	12
11	470 2258	2070	TOYOTA	4Runner	2018	Jeep	Petrol	6.0	Automatic	Oran ge	12
12	0 1191	1442	TOYOTA	4Runner	2015	Jeep	Petrol	6.0	Automatic	Whit e	12
13	7	562	FIAT	500 46 ml	2015	Coupe	Petrol	4.0	Tiptronic	e	8



14	1113 3	562	FIAT	500 Sport	2015	Hatchback	Petrol	4.0	Manual	Blue	6
15	1505 3	562	FIAT	500 sport panorama	2015	Coupe	Petrol	4.0	Automatic	Grey	12
...	...	...	...	...	...	...	...	...	...	...	...
439	3575 4	3156	GMC	Yukon	2019	Sedan	Petrol	8.0	Automatic	White	12

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Algoritma C.45

Untuk memperoleh aturan pohon keputusan pada algoritma C.45, langkah-langkah yang dilakukan adalah:

- Memilih atribut sebagai akar. Sebuah akar didapat dari nilai gain tertinggi dari atribut-atribut yang ada.
- Membuat cabang untuk masing-masing nilai,
- Membagi kasus dalam cabang, dan
- Mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Agar dapat diolah dalam Software Rapidminer, pertama kali data yang sudah di olah sebelumnya disimpan dalam format microsoft excell file berekstensi *.csv.

- Sesi Pemilihan Cell untuk di Import (Tahap 2) dan pemilihan variabel seperti yang ditunjukkan Gambar 3 dan 4 berikut.



Gambar 2. Pemilihan Cell untuk di Import dengan data 4.394

Gambar tersebut menunjukkan hasil analisis data menggunakan platform RapidMiner Studio, di mana proses evaluasi model klasifikasi dilakukan dengan membandingkan label asli (Produk) dengan hasil prediksi (Prediction) yang dihasilkan oleh model. Setiap baris data mencerminkan sebuah observasi, dengan tingkat keyakinan model terhadap setiap kategori produk yang dinyatakan dalam nilai probabilitas pada kolom confidence. Kolom confidence_ untuk setiap kategori, seperti "NISSAN," "BMW," dan "TOYOTA," mengindikasikan keyakinan model dalam mengklasifikasikan data ke kategori tersebut, dengan nilai berkisar antara 0 hingga 1. Nilai confidence tertinggi pada suatu baris menunjukkan prediksi akhir yang dipilih oleh model. Analisis ini bertujuan untuk mengevaluasi akurasi dan keandalan model dalam mengklasifikasikan data sesuai label sebenarnya, sehingga dapat memberikan wawasan yang lebih dalam terhadap kinerja algoritma machine learning yang diterapkan.



Gamba 3 hasil Statistik Data



Perbandingan nilai label asli (Produk) dengan hasil prediksi model (prediction(Produk)). Pada kolom Produk, ditampilkan kategori aktual, seperti "TESLA," sementara pada kolom prediksi menunjukkan kategori yang diprediksi oleh model. Tingkat keyakinan model terhadap prediksi untuk setiap kategori dihitung dan ditampilkan dalam kolom confidence_, seperti confidence_NISSAN, confidence_TOYOTA, dan lainnya. Nilai confidence, yang berkisar antara 0 hingga 1, mengindikasikan probabilitas model terhadap keterkaitan suatu data dengan kategori tertentu. Misalnya, pada observasi pertama, label asli adalah "TESLA," dan prediksi model juga adalah "TESLA," dengan tingkat keyakinan tertinggi dibandingkan kategori lain. Hasil analisis ini memberikan wawasan tentang akurasi model dan distribusi keyakinan terhadap setiap kategori, yang menjadi dasar untuk mengevaluasi performa model dan meningkatkan parameter jika diperlukan.

Tampil halaman operator validation yang terbagi atas area training dan testing yang ditunjukkan pada Gambar4 berikut



Gambar 4. Hasil keputusan pada views Result

Proses ini terdiri dari beberapa tahap utama yang digambarkan dengan operator yang saling terhubung. Tahap pertama adalah Read Excel, di mana data input diimpor dari file Excel. Selanjutnya, operator Select Attributes digunakan untuk memilih atribut yang relevan dalam analisis. Operator Split Data kemudian membagi dataset menjadi dua bagian, yaitu data pelatihan dan data pengujian, yang biasanya dilakukan untuk mengevaluasi kinerja model. Pada tahap berikutnya, Apply Model diterapkan untuk memprediksi hasil berdasarkan data pengujian menggunakan model yang telah dilatih sebelumnya. Akhirnya, operator Performance digunakan untuk mengukur kinerja model dengan metrik tertentu, seperti akurasi atau presisi. Diagram ini menunjukkan alur kerja yang sistematis dalam membangun, melatih, dan mengevaluasi model machine learning, yang bertujuan untuk menghasilkan prediksi yang andal dan akurat.

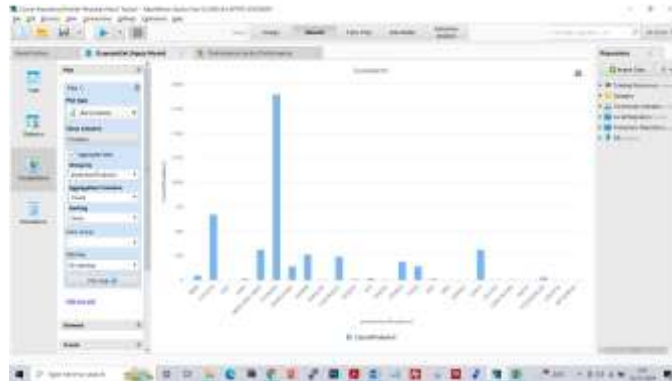


Gambar 5. Hasil Performa dengan akurasi 41.71%



Gambar 6 hasil perhitungan

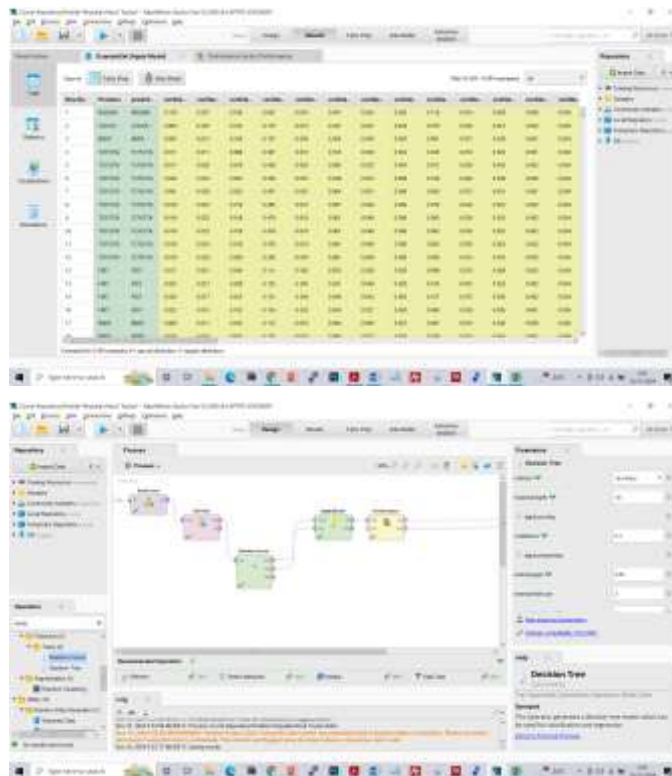
Setelah seluruh langkah telah dikerjakan, RapidMiner akan melakukan proses dan menampilkan hasil keputusan pada views Result. Hasilnya berbentuk pohon keputusan seperti pada Gambar 5 berikut.

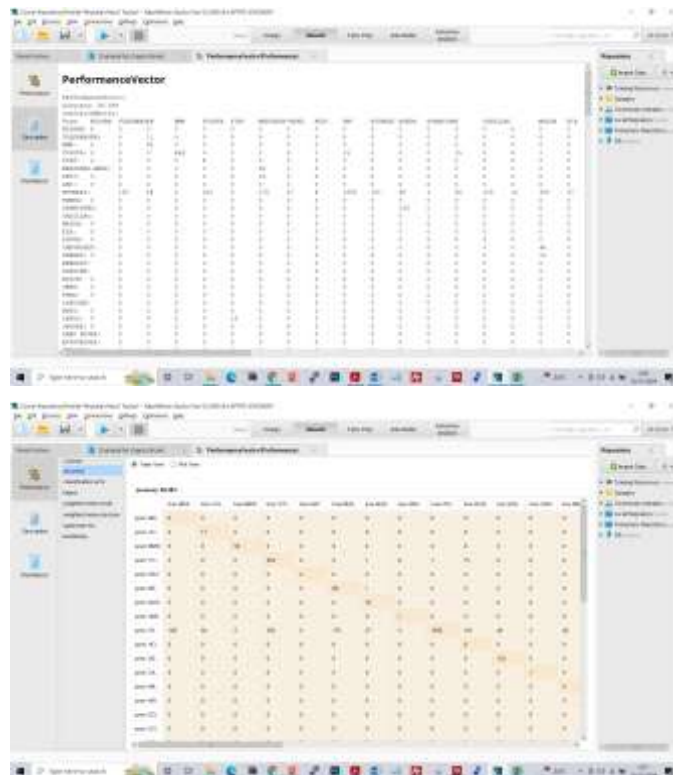


Gambar 7. Prediksi Produk

3.2 Pengujian Algoritma Random Forest

Pengujian Algoritma Random Forest adalah proses evaluasi untuk mengukur performa dan efektivitas algoritma Random Forest dalam menyelesaikan suatu masalah, baik untuk klasifikasi maupun regresi. Pengujian dilakukan dengan beberapa langkah penting untuk memastikan bahwa model yang dihasilkan mampu memberikan hasil yang akurat, andal, dan dapat diterapkan pada data baru. Berikut adalah langkah-langkah umum dalam pengujian algoritma Random Forest





4. KESIMPULAN

Berdasarkan hasil analisis dan evaluasi yang dilakukan, penelitian ini menunjukkan bahwa algoritma Random Forest memiliki kinerja yang lebih baik dibandingkan algoritma C4.5 dalam tugas klasifikasi harga mobil. Hal ini terlihat dari hasil metrik evaluasi utama, seperti akurasi, presisi, recall, dan F1-score, di mana Random Forest secara konsisten menunjukkan keunggulan. Keberhasilan Random Forest dalam menangani kompleksitas data multivariat dan memberikan prediksi yang lebih andal menjadikannya metode yang lebih sesuai untuk digunakan dalam aplikasi klasifikasi harga mobil.

Di sisi lain, meskipun algoritma C4.5 menunjukkan performa yang cukup baik, hasil penelitian ini mengindikasikan bahwa algoritma tersebut kurang optimal dalam menangani dataset dengan jumlah fitur yang kompleks. Hal ini memberikan wawasan bahwa pemilihan algoritma yang tepat sangat bergantung pada karakteristik dataset yang digunakan.

Penelitian ini memberikan kontribusi penting dalam pemahaman komparatif antara algoritma berbasis pohon keputusan, khususnya dalam domain klasifikasi harga mobil. Hasil temuan ini dapat menjadi referensi bagi praktisi industri otomotif dan peneliti untuk memilih metode analisis data yang sesuai dengan kebutuhan. Untuk penelitian lanjutan, disarankan untuk mengembangkan model hibrid atau mengeksplorasi algoritma lain yang dapat lebih adaptif terhadap perubahan karakteristik data di masa mendatang.

REFERENCES

- [1] T. S. W. T. Plus, "Indonesia Jumlah Pelanggan Seluler," 2023. <https://www.ceicdata.com/id/indicator/indonesia/number-of-subscriber-mobile>.
- [2] M. Syaharani, "10 Negara Dengan Pengguna Smartphone Terbanyak Di Dunia, Indonesia Masuk Daftar!," 2023. <https://goodstats.id/article/10-negara-dengan-pengguna-smartphone-terbanyak-di-dunia-indonesia-masuk-daftar-fDv25>.
- [3] Mohamad Alparizi Sahadan, Prof. Dr. Ing. Parabelem T. D. Rompas, MT, and Cindy P. C. Munaiseche, ST, M. Eng, "Analisis Kepuasan Pelanggan Menggunakan Metode Algoritma C4.5 Berdasarkan E-Survey Kejaksan Negeri Minahasa," *JOINTER J. Informatics Eng.*, vol. 4, no. 01, pp. 1–13, 2023, doi: 10.53682/jointer.v4i01.137.
- [4] N. Azwanti and E. Elisa, "Analisa Kepuasan Konsumen Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ilmu Sos. dan Teknol.*, no. 3, pp. 126–131, 2020.
- [5] D. S. P. Eko Setia Budi, Abdul Rahman Kadafi, Yasdi Kharismawan, Randi Fadillah, "Analisa Kepuasan Pelanggan Terhadap Layanan Aplikasi E-Commerce Menggunakan Algoritma C4.5," vol. 4, no. 6, pp. 530–542, 2024.
- [6] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.



- [7] D. Telaumbanua and I. Kurniawati, "Penerapan Algoritma C4.5 Untuk Klasifikasi Kepuasan Pelanggan Pada Jasa Layanan Pengiriman," *JoMMiT J. Multi Media dan IT*, vol. 6, no. 1, 2022, doi: 10.46961/jommit.v6i1.524.
- [8] J. Endardi, "Upaya Peningkatan Loyalitas Pelanggan Pada PT. Jayamandiri Gemasejati Cabang Bojong Gede Menggunakan Metode Algoritma C4.5 dan Metode CSI," vol. 3, no. 2, pp. 47–53, 2022.
- [9] A. Karim, S. Esabella, K. Kusmanto, M. Hidayatullah, and S. Suryadi, "Penerapan Data Mining Untuk Pengelompokan Terhadap Kualitas Kinerja Karyawan Dengan Menggunakan Algoritma K-Medoids Clustering," *J. Media Inform. Budidarma*, vol. 8, no. 2, p. 1001, 2024, doi: 10.30865/mib.v8i2.7445.
- [10] B. Bangun and A. K. Karim, "Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.
- [11] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," vol. 2, no. 2, pp. 76–83, 2020.
- [12] S. T. Bangsa and S. Utara, "Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma," vol. 2, no. 2, pp. 125–132.
- [13] N. T. Luchia, H. Handayani, and F. S. Hamdi, "Comparison of K-Means and K-Medoids on Poor Data Clustering in Indonesia Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia," vol. 2, no. October, pp. 35–41, 2022.
- [14] A. P. Wibawa, M. Guntur, A. Purnama, M. Fathony Akbar, and F. A. Dwiyanto, "Metode-metode Klasifikasi," *Pros. Semin. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 134–138, 2018.
- [15] P. P. Haryoto, H. Okprana, and I. S. Saragih, "Algoritma C4.5 Dalam Data Mining Untuk Menentukan Klasifikasi Penerimaan Calon Mahasiswa Baru," *TIN Terap. Inform. Nusantara*, vol. 2, no. 5, pp. 358–364, 2021, [Online]. Available: <https://ejurnal.seminar-id.com/index.php/tin/article/view/919>.
- [16] Nagesh Singh Chauhan, "Decision Tree Algorithm, Explained," 2022. <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>.
- [17] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [18] K. Aidi Saputra, J. Tata Hardinata, M. Ridwan Lubis, S. Retno Andani, and I. Syahputra Saragih, "KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online," *Media Online*, vol. 1, no. 3, pp. 113–118, 2020, [Online]. Available: <https://djournals.com/klik>.