

Klasifikasi Jenis Kendaraan Menggunakan Decision Tree Dan Evaluasi Akurasi Melalui Confusion Matrix

Samsul^{1*}, Zaehol Fatah²

¹Teknologi Informasi, Universitas Ibrahimy

²Sistem Informasi, Universitas Ibrahimy

^{1*}samsulsodung@email.com, ²zaeholfatah@gmail.com

ABSTRACT

Data mining merupakan salah satu metode yang paling efektif dalam menghasilkan klasifikasi yang akurat, efisien, dan relevan. Pengelompokan jenis kendaraan berdasarkan sistem transmisi dilakukan dengan menggunakan algoritma Decision Tree dan dievaluasi melalui confusion matrix. Dataset yang digunakan mencakup empat jenis kendaraan: Bebek, Skuter, Sport, dan Trail, dengan tiga jenis transmisi: Manual, Automatic, dan Kopling. Algoritma Decision Tree dipilih karena kemampuannya dalam membagi dataset secara rekursif untuk menghasilkan aturan klasifikasi yang jelas dan mudah dipahami. Model dilatih dan diuji untuk memprediksi jenis transmisi berdasarkan fitur kendaraan, dengan hasil akurasi mencapai 95%. Evaluasi menggunakan confusion matrix mengungkap distribusi prediksi benar dan salah pada setiap kategori. Hasilnya menunjukkan bahwa transmisi Automatic dan Kopling diklasifikasikan dengan akurasi tinggi, meskipun terdapat beberapa kesalahan pada prediksi transmisi Manual. Nilai Cohen's Kappa sebesar 0,913 mengindikasikan kesesuaian yang sangat baik antara prediksi dan data aktual. Algoritma Decision Tree terbukti efektif dalam klasifikasi jenis kendaraan, meskipun diperlukan perbaikan untuk meningkatkan akurasi pada kategori tertentu.

Keywords: Data mining, Decision Tree, Confusion matrix



This Is Open Access Article Under The CC Attribution-ShareAlike 4.0 License.



PENDAHULUAN

Klasifikasi jenis kendaraan merupakan komponen penting dalam berbagai aplikasi teknologi modern, mulai dari system manajemen transportasi cerdas (Intelligent Transportation Systems - ITS) hingga pengembangan teknologi kendaraan otonom [1]. Proses klasifikasi ini bertujuan untuk mengelompokkan kendaraan berdasarkan atribut tertentu, seperti jenis, ukuran, penggunaan bahan bakar, atau jenis transmisi. Klasifikasi yang akurat memungkinkan sistem transportasi bekerja dengan lebih efisien, baik dalam hal manajemen lalu lintas, pengelompokan di jalan raya, hingga penentuan pola perilaku kendaraan dalam lingkungan perkotaan [2].

Salah satu pendekatan umum dalam klasifikasi adalah berdasarkan transmisi kendaraan, terutama pada kendaraan roda dua yang populer di beberapa Negara, termasuk Indonesia. Jenis transmisi yang paling digunakan dalam klasifikasi ini adalah transmisi Manual, Automatic, dan Kopling [3]. Transmisi menjadi faktor penentu yang signifikan dalam pengelompokan kendaraan karena mempengaruhi pola operasional kendaraan, seperti akselerasi dan manuver. Penelitian yang berfokus pada klasifikasi kendaraan berdasarkan transmisi memiliki relevansi dalam berbagai konteks, seperti pemantauan kendaraan, manajemen armada, dan layanan perawatan kendaraan.

Untuk melakukan klasifikasi, berbagai metode dari domain pembelajaran mesin (Machine Learning) dapat diterapkan. Salah satu metode yang paling banyak digunakan adalah algoritma Decision Tree. Algoritma ini dikenal karena kemampuannya yang baik dalam mengenai data kategori dan numeric serta

kemampuannya untuk membuat keputusan yang mudah diinterpretasikan. Decision Tree bekerja dengan membagi dataset ke dalam subset yang lebih kecil secara rekursif berdasarkan kriteria tertentu, misalnya Information Gain atau Gini Index [4]. Pemisahan ini bertujuan untuk memaksimalkan homogenitas data dalam setiap cabang, sehingga setiap cabang mewakili satu keputusan yang didasarkan pada satu atau lebih pada atribut. Dalam kasus klasifikasi jenis kendaraan, dan tentunya jenis transmisi, memainkan peran penting dalam proses pemisahan data.

Keuntungan dari penggunaan algoritma Decision Tree antara lain adalah kemampuannya untuk menangani data tanpa perlu proses yang ekstensif, seperti normalisasi atau standarisasi. Selain itu, Decision Tree dapat digunakan untuk menghasilkan model prediksi yang dapat diinterpretasikan secara visual, yang memberikan pemahaman yang lebih jelas tentang bagaimana keputusan diambil oleh modal [5]. Namun kelemahan utama algoritma ini adalah kecenderungannya untuk overfitting pada data pelatihan jika pohon terlalu dalam atau terlalu kompleks. Oleh karena itu, diperlukan teknik peruning untuk memangkas pohon dan menghindari kompleksitas yang tidak perlu, yang pada gilirannya dapat meningkatkan generalisasi model terhadap data baru [6].

Meskipun algoritma Decision Tree dapat menghasilkan model yang akurat, evaluasi performa model tidak cukup hanya dengan mengandalkan metrik akurasi. Akurasi hanya menunjukkan proposal prediksi yang benar dari keseluruhan prediksi, terutama jika terdapat ketidakseimbangan kelas. Oleh karena itu, alat evaluasi lain yang lebih komprehensif seperti confusion matrix diperlukan. Confusion matrix memberikan pandangan yang lebih terperinci tentang bagaimana model mengklasifikasikan data, dengan menunjukkan jumlah prediksi yang benar dan salah untuk setiap kelas [7]. Confusion matrix terdiri dari empat komponen utama: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari komponen ini, berbagai metrik tambahan seperti presisi, recall, dan F1-score dapat dihitung untuk memberikan gambaran lebih jelas tentang performa model pada setiap kelas.

Selain itu, untuk memperkuat analisis, metrik seperti Cohen's Kappa sering digunakan untuk mengukur tingkat kesesuaian antara prediksi model dengan label actual dengan memperhitungkan kemungkinan kesesuaian yang terjadi secara kebetulan [8]. Nilai Kappa yang tinggi menunjukkan bahwa model memiliki kemampuan prediktif yang baik dan tidak hanya mengandalkan prediksi acak. Metrik ini sangat penting dalam aplikasi dengan ketidakseimbangan kelas, di mana akurasi saja mungkin tidak cukup menggambarkan performa model secara adil [9].

Dalam konteks klasifikasi jenis kendaraan, pengguna confusion matrix dan Cohen's kappa memberikan wawasan yang lebih dalam tentang seberapa baik model dapat membedakan antara jenis transmisi, khususnya dalam situasi di mana beberapa jenis kendaraan mungkin memiliki karakteristik yang mirip. Misalnya, kendaraan dengan transmisi kopling dan manual sering kali memiliki pola operasional yang serupa, sehingga penting untuk mengevaluasi seberapa sering kesalahan klasifikasi terjadi diantara kedua kelas tersebut [10]. Penggunaan confusion matrix memungkinkan identifikasi area di mana model perlu diperbaiki, misalnya dengan menambah fitur baru atau mengubah parameter model.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi jenis kendaraan menggunakan algoritma Decision Tree, dengan fokus pada prediksi jenis transmisi. Selain itu penelitian ini mengevaluasi kinerja model dengan menggunakan confusion matrix untuk mendapatkan pemahaman yang lebih baik tentang distribusi kesalahan prediksi dan menilai seberapa baik model dapat membedakan antara jenis transmisi yang ada. Diharapkan bahwa hasil penelitian ini dapat memberikan kontribusi dalam pengembangan sistem klasifikasi kendaraan yang lebih efisien dan akurat, baik untuk keperluan teknologi transportasi masa depan maupun aplikasi praktis di lapangan.

METODE PENELITIAN

a. Dataset

Dataset yang digunakan dalam penelitian ini mencakup informasi dari berbagai jenis kendaraan, khususnya empat jenis kendaraan bermotor, yaitu Bebek, Skuter, Sport, dan Trail. Masing-masing jenis kendaraan ini memiliki karakteristik yang berbeda, terutama dari segi transmisi, yang dibagi menjadi tiga jenis: Manual, Automatic, dan Kopling.

Fitur-fitur utama dalam dataset mencakup kapasitas mesin (CC), bobot kendaraan (Kg), tahun produksi, serta jenis transmisi. Fitur-fitur ini digunakan sebagai variabel independen dalam model klasifikasi, sedangkan jenis transmisi adalah variabel target atau variabel dependen.

b. Algoritma Decision Tree

Decision Tree adalah algoritma yang sering digunakan untuk masalah klasifikasi karena kesederhanaannya dalam representasi visual dan interpretasi [11]. Pada penelitian ini, Decision Tree digunakan untuk mengklasifikasi jenis transmisi kendaraan berdasarkan fitur yang ada. Algoritma ini bekerja dengan cara mempartisi dataset berdasarkan atribut yang memberikan informasi terbaik dalam memisahkan kelas.

Pada setiap titik keputusan dalam pohon, algoritma mengevaluasi semua fitur yang tersedia dan memilih fitur yang paling relevan untuk memisahkan dataset. Kriteria pemilihan fitur pada penelitian ini adalah Gini Index, yang mengukur ketidakmurnian (impurity) pada node.

c. Rumu Gini Index

$$Gini(p) = 1 - \sum_{i=1}^k p_i^2$$

Di mana :

- P_i adalah proporsi data pada kelas i di node tersebut.
- K adalah jumlah kelas, yaitu 3 kelas (Manual, Automatic, Kopling).

Misalnya, jika sebuah node berisi 60% data dengan kelas Manual, 30% Automatic, dan 10% Kopling, maka Gini Index untuk node tersebut adalah:

$$Gini = 1 - (0.6^2 + 0.3^2 + 0.1^2) = 1 - (0.36 + 0.09 + 0.01) = 0.54$$

Algoritma kemudian memilih atribut yang memberikan penurunan Gini Index terbesar (pemisahan terbaik).

d. Peruning pada Decision Tree

Untuk menghindari overfitting, pohon keputusan yang terbentuk dilakukan pruning. Pruning bertujuan untuk menghilangkan cabang yang tidak memberikan informasi tambah atau hanya memberikan peningkatan kecil pada performa. Ada dua jenis pruning :

- Pre-pruning, yang menghentikan pertumbuhan pohon sebelum menjadi terlalu dalam.
- Post-pruning, di mana pohon dipangkas setelah pohon tumbuh sepenuhnya.

Pada penelitian ini digunakan post-pruning untuk memastikan model tidak terlalu kompleks.

e. Evaluasi kerja dengan Confusion Matrix

Setelah model Decision Tree dibangun, kinerjanya dievaluasi dengan menggunakan confusion matrix. Confusion matrix adalah matrix kontingensi yang membandingkan hasil prediksi model dengan label sebenarnya. Matrix ini memberikan informasi penting mengenai distribusikesalahan yang dibuat oleh model.

Berikut adalah definisi dari elemen-elemen dalam confusion matrix:

- True Positive (TP): Jumlah instance yang benar-benar positif oleh model.
- True Negative (TN): Jumlah instansi yang benar-benar negatif dan diprediksi sebagai negative oleh model.
- False Positive (FP): Jumlah instance yang sebenarnya negative namun di prediksi sebagai positif.
- False Negatif (FN) : Jumlah Instance yang sebenarnya positif namun di prediksi sebagai negative.

Dari confusion matrix, dapat dihitung berbagai metrix evaluasi seperti akurasi, presisi, recall, F1-score, dan Cohen's Kappa.

f. Akurasi

Akurasi memberikan gambaran tentang seberapa baik model dapat mengklasifikasikan seluruh kelas dengan benar [12]. dihitung dengan rumus:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Metrik ini memberikan gambaran umum mengenai seberapa baik performa model secara keseluruhan.

g. Recall

Recall mengukur proporsi instance positif yang berhasil diidentifikasi dengan benar oleh model, dihitung oleh rumus:

$$Recall = \frac{TP}{TP + FN}$$

Recall berguna untuk mengetahui kemampuan model dalam menemukan instance dari kelas tertentu.

h. F1-Score

F1 score adalah rata-rata harmonis dari precision dan recall, memberikan gambaran seimbang antara keduanya:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

i. Cohen's Kappa

Cohen's Kappa mengukur kesesuaian antara hasil prediksi model dan label sebenarnya dengan mempertimbangkan kemungkinan kesepakatan yang terjadi secara acak. Rumus Cohen's Kappa adalah:

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

Di mana P_o adalah proporsi kesepakatan yang diamati, dan P_e adalah proporsi kesepakatan yang diharapkan terjadi secara kebetulan. Nilai Kappa berkisar antara -1 (kesepakatan buruk) hingga 1 (kesepakatan sempurna).

HASIL DAN PEMBAHASAN

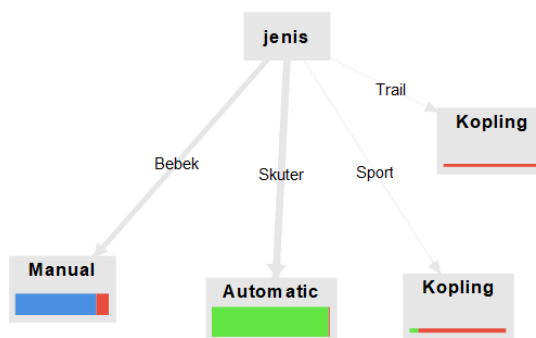
Pada penelitian ini, Decision Tree digunakan untuk mengklasifikasi jenis kendaraan berdasarkan transmisi menggunakan dataset yang terdiri dari empat jenis kendaraan: Bebek, Skuter, Sport, dan Trail. Algoritma dijalankan melalui platform RapidMiner, yang memfasilitasi penerapan model klasifikasi dan evaluasi performannya dengan confusion matrix serta berbagai metrik evaluasi lainnya.

a. Hasil dari RapidMiner

Setelah menerapkan algoritma Decision Tree menggunakan RapidMiner. Berikut adalah hasil visualisasi pohon keputusan yang dihasilkan:

- Node Utama (Root Node): variable yang dipilih sebagai node utama adalah jenis kendaraan.
- Cabang pertama: Cabang pertama pada pohon keputusan memisahkan kendaraan berdasarkan jenis Bebek, Skuter, Sport, dan Trail.
- Node daun (Leaf Node): pada setiap node daun, model memprediksi jenis transmisi (Manual, Automatic, atau Kopling) dengan menghitung distribusi kelas pada setiap subset data.

Berikut adalah contoh representasi pohon keputusan sederhana yang dihasilkan:



Gambar 1. Pohon Keputusan

Dari hasil diatas, terlihat bahwa setiap jenis kendaraan memiliki kecenderungan untuk menggunakan jenis transmisi tertentu :

- Bebek => Manual
- Skuter => Automatic
- Sport => Kopling
- Trail => Kopling

b. Confusion Matrix Hasil dari RapidMiner

Setelah model dilatih menggunakan data latih, hasil prediksi pada data uji dievaluasi menggunakan confusion matrix berikut yang dihasilkan oleh RapidMiner.

accuracy: 95.00%

	true Manual	true Automatic	true Kopling	class precision
pred. Manual	6	0	1	85.71%
pred. Automatic	0	11	0	100.00%
pred. Kopling	0	0	2	100.00%
class recall	100.00%	100.00%	66.67%	

Gambar 2. Cofusion Matrix

Dari table di atas, dapat disimpulkan bahwa model melakukan klasifikasi dengan cukup baik, namun terdapat satu kesalahan klasifikasi di mana kendaraan dengan transmisi Manual salah diprediksi sebagai kopling.

c. Evaluasi Kinerja

Matrix performayang dihitung dari hasil confusion matrix di atas menunjukkan performa klasifikasi yang cukup baik, dengan hasil sebagai berikut:

- Akurasi

$$Accuracy = \frac{6 + 11 + 2}{6 + 11 + 2 + 1 + 0 + 0} = \frac{19}{20} = 95\%$$

- Precision untuk Manual :

$$Precision = \frac{6}{6 + 0} = 1.0$$

- Recall untuk Manual :

$$Recall = \frac{6}{6 + 1} = 0.857$$

- F1-Score untuk Manual :

$$F1 = 2 \times \frac{1.0 \times 0.857}{1.0 + 0.857} = 0.923$$

Hasil ini menunjukkan bahwa model memiliki tingkat presisi dan recall yang tinggi untuk transmisi Manual, dengan F1-score mencapai 0.923. Model juga memiliki performa yang baik pada transmisi Automatic dan Kopling.

d. Nilai Cohen's Kappa

Selain akurasi dan matrix evaluasi lainnya, digunakan juga Cohen's Kappa untuk mengukur tingkat kesepakatan antara prediksi model dan data aktual. Nilai Kappa yang dihitung adalah sebagai berikut:

$$Kappa = \frac{P_o - P_e}{1 - P_e}$$

Di mana:

- P_o = persentase kesepakatan yang diamati.
- P_e = persentase kesepakatan yang di harapkan secara acak.

Dengan menggunakan data dari confusion matrix di atas, nilai Kappa yang dihasilkan adalah 0.913, yang menunjukkan kesesuaian yang sangat baik antara prediksi model dan label sebenarnya. Nilai ini berada di atas 0.8, yang biasanya diinterpretasikan sebagai kesesuaian yang sangat baik.

e. Pembahasan

Dari hasil evaluasi yang diperoleh melalui RapidMiner, kita dapat menarik beberapa kesimpulan penting:

- Akurasi Tinggi
Model Decision Tree berhasil mencapai akurasi sebesar 95%, yang menunjukkan bahwa algoritma ini mampu mengklasifikasikan kendaraan dengan sangat baik berdasarkan jenis transmisi.
- Kesalahan Minimal
Terdapat sedikit kesalahan klasifikasi, terutama pada transmisi Manual yang salah diprediksi sebagai Kopling. Kesalahan ini kemungkinan besar disebabkan oleh kemiripan antara fitur-fitur kendaraan dengan transmisi Manual dan Kopling.
- Kesepakatan yang Sangat baik
Nilai Cohen's Kappa sebesar 0.913 menunjukkan bahwa kesepakatan antara hasil prediksi dan data actual sangat baik, memperkuat keandalan model dalam membuat prediksi yang akurat.
- Visualisasi yang Mudah Dipahami:
Salah satu kelebihan utama dari Decision Tree adalah kemudahan interpretasi. Pohon keputusan yang dihasilkan oleh RapidMiner sangat intuitif, memungkinkan pengambil keputusan untuk dengan mudah memahami bagaimana setiap fitur mempengaruhi hasil klasifikasi.

KESIMPULAN

Algoritma Decision Tree telah berhasil mengklasifikasikan jenis kendaraan berdasarkan jenis transmisi dengan tingkat akurasi sebesar 95%, yang menunjukkan bahwa model ini efektif dalam melakukan prediksi berdasarkan dataset yang digunakan. Keberhasilan model ini menunjukkan bahwa klasifikasi menggunakan Decision Tree tidak hanya memberikan hasil yang akurat, tetapi juga memberikan interpretasi yang jelas dan dapat dimanfaatkan untuk pengambilan keputusan yang tepat.

Penggunaan algoritma Decision Tree dengan akurasi yang tinggi seperti ini dapat meningkatkan keyakinan dalam pemilihan jenis transmisi kendaraan oleh konsumen. Model ini memungkinkan identifikasi yang lebih tepat terhadap preferensi transmisi kendaraan, sehingga konsumen dapat membuat keputusan yang lebih terinformasi saat membeli sepeda motor. Dengan demikian, model ini memberikan keuntungan

baik bagi konsumen, dalam memilih kendaraan yang sesuai dengan kebutuhannya, maupun bagi produsen dan dealer kendaraan, yang dapat menyasar segmen pasar yang tepat. Hal ini juga berpotensi mengurangi risiko ketidakpuasan konsumen dan meningkatkan efisiensi dalam strategi penjualan kendaraan.

DAFTAR PUSTAKA

- [1] Z. Hariyanto and J. Matematika, “Klasifikasi jenis kendaraan bergerak berbasis,” 2015.
- [2] I. Najiyah and S. Topiq, “KLASIFIKASI JENIS KENDARAAN RODA EMPAT,” vol. 3, no. 2, pp. 199–206, 2021.
- [3] B. Pribadi, M. Naseer, J. Raya, and P. No, “Sistem Klasifikasi Jenis Kendaraan Melalui Teknik Olah Citra Digital,” vol. 3, no. 2, pp. 1–5, 2014.
- [4] M. I. U. Rosyidi and N. Rochmawati, “Teknik Bagging Pada Algoritma Klasifikasi Decision Tree dan SVM Untuk Klasifikasi SMS Berbahasa Indonesia,” vol. 05, pp. 265–271, 2023.
- [5] A. K. Dewantoro, I. Iwut, E. Susatio, F. Teknik, and U. Telkom, “SIMULASI DAN ANALISIS SISTEM PENGHITUNG KEPADATAN LALU LINTAS DAN KLASIFIKASI KENDARAAN BERBASIS WEBCAM DENGAN METODE BACKGROUND SUBTRACTION,” vol. 2, no. 2, pp. 2833–2840, 2015.
- [6] C. N. Syahputri and M. S. Hasibuan, “OPTIMASI KLASIFIKASI DECISION TREE DENGAN TEKNIK PRUNING UNTUK MENGURANGI OVERFITTING,” vol. 11, no. 2, pp. 87–96, 2024.
- [7] M. R. Darwis *et al.*, “Analisis Sentiment Gambar pada Media Sosial dengan Pendekatan Deep Learning,” vol. 7, no. 2, pp. 91–95, 2024.
- [8] “KLASIFIKASI JENIS KENDARAAN PADA JALAN RAYA MENGGUNAKAN YOLOV7,” vol. 5, no. 4, pp. 661–666, 2023.
- [9] F. Putri, S. Ndruru, and S. Amdayani, “Development of e-modules based on socio-scientific issues on Chemistry content,” vol. 10, no. 2, pp. 129–140, 2023.
- [10] A. Wasik *et al.*, “Implementasi data mining untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1,” vol. 1, no. 2, pp. 469–479, 2024.
- [11] D. Jollyta, A. Hajjah, E. Haerani, and M. Siddik, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish, 2023.
- [12] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.