

Analisis Klasifikasi Harga Penjualan Motor Menggunakan Metode Decision Tree Berbasis Entropy dengan Pendekatan C4.5 dan Gain Ratio

Neza Namira^{a,1,*}, Nidya Banuaria^{a,2}

^aUniversitas Mahkota Tricom Unggul, Medan 20231, Indonesia

¹Nezanamira821@gmail.com*

Abstract

The rapid growth of the automotive industry has intensified competition in the motorcycle market, making the determination of appropriate selling prices a crucial factor in enhancing a company's competitiveness and profitability. Motorcycle selling prices are influenced by multiple factors, including brand, model, year of manufacture, engine capacity, vehicle condition, and mileage. The large number of variables involved generates complex datasets that are difficult to analyze manually. Therefore, a data mining approach is required to transform historical motorcycle sales data into meaningful information that supports decision-making. This study aims to analyze and classify motorcycle selling prices using the C4.5 Decision Tree algorithm based on entropy and gain ratio. The C4.5 algorithm was selected because it is capable of generating interpretable classification rules while maintaining high classification accuracy for both categorical and numerical data. The research methodology consists of data collection, data preprocessing, entropy, information gain, and gain ratio calculations, decision tree construction, and classification model evaluation. The results indicate that the C4.5 algorithm successfully identifies the attributes that have the greatest influence on motorcycle selling price classification. The resulting decision tree provides an effective basis for predicting motorcycle price categories and offers clear classification rules that can support practical decision-making. Consequently, the C4.5 Decision Tree algorithm can assist motorcycle dealers and automotive businesses in developing more effective pricing and sales strategies based on the characteristics of the vehicles being offered.

Keywords: Price; Data Mining; Decision Tree; C4.5; Entropy; Gain Ratio

1. Pendahuluan

Perkembangan teknologi informasi telah mendorong peningkatan jumlah data yang dihasilkan dari berbagai aktivitas bisnis, termasuk pada sektor penjualan kendaraan bermotor. Data transaksi penjualan yang tersimpan dari waktu ke waktu memiliki potensi besar untuk diolah menjadi informasi yang berguna dalam mendukung pengambilan keputusan. Namun, dalam praktiknya, banyak pelaku usaha yang belum memanfaatkan data tersebut secara optimal sehingga informasi penting yang terkandung di dalamnya belum dapat digunakan secara maksimal untuk meningkatkan efektivitas strategi penjualan.

Penjualan motor merupakan salah satu sektor bisnis yang memiliki tingkat persaingan cukup tinggi. Dalam proses penjualan, penentuan harga menjadi faktor yang sangat penting karena dapat memengaruhi minat konsumen dan tingkat keuntungan yang diperoleh. Harga penjualan motor tidak hanya ditentukan oleh satu faktor, tetapi dipengaruhi oleh berbagai atribut seperti merek kendaraan, tahun produksi, kapasitas mesin, kondisi kendaraan, jenis transmisi, serta jarak tempuh kendaraan. Kompleksitas hubungan antar atribut tersebut membuat proses analisis harga menjadi lebih sulit apabila dilakukan secara manual, terutama ketika jumlah data yang dimiliki semakin besar.

Pemanfaatan teknik data mining menjadi salah satu solusi untuk mengatasi permasalahan tersebut. Data mining merupakan proses penggalian informasi atau pengetahuan dari sekumpulan data dalam jumlah besar untuk menemukan pola, hubungan, maupun tren tertentu yang dapat digunakan sebagai dasar pengambilan keputusan. Salah satu teknik yang banyak digunakan dalam data mining adalah klasifikasi. Teknik klasifikasi bertujuan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik yang dimiliki sehingga dapat digunakan untuk memprediksi kelas atau kategori pada data baru.

Metode Decision Tree merupakan salah satu algoritma klasifikasi yang populer karena mampu menghasilkan model yang mudah dipahami dalam bentuk pohon keputusan. Setiap keputusan yang dihasilkan dapat ditelusuri melalui aturan-aturan yang terbentuk sehingga memudahkan pengguna dalam melakukan interpretasi terhadap hasil klasifikasi. Salah satu pengembangan dari metode Decision Tree adalah algoritma C4.5 yang diperkenalkan oleh Ross Quinlan. Algoritma ini menggunakan konsep entropy dan gain ratio untuk menentukan atribut terbaik yang akan digunakan dalam pembentukan pohon keputusan. Dengan pendekatan tersebut, algoritma C4.5 mampu menghasilkan model klasifikasi yang memiliki tingkat akurasi yang baik serta dapat menangani data dengan karakteristik yang beragam.

Dalam konteks penjualan motor, penerapan algoritma C4.5 dapat dimanfaatkan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap harga jual kendaraan dan mengklasifikasikan harga penjualan ke dalam kategori tertentu. Hasil klasifikasi yang diperoleh dapat digunakan sebagai bahan pertimbangan bagi dealer maupun pelaku usaha otomotif dalam menentukan strategi penjualan yang lebih efektif dan berbasis data. Selain itu, model yang dihasilkan juga dapat membantu proses prediksi harga kendaraan berdasarkan karakteristik yang dimiliki oleh motor tersebut.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk menganalisis klasifikasi harga penjualan motor menggunakan metode Decision Tree berbasis entropy dengan pendekatan C4.5 dan gain ratio. Penelitian ini diharapkan mampu menghasilkan model klasifikasi yang dapat membantu proses pengambilan keputusan dalam menentukan kategori harga penjualan motor secara lebih akurat, objektif, dan efisien.

2. Tinjauan Literatur

Data mining merupakan proses penggalian informasi atau pengetahuan yang tersembunyi dari sekumpulan data berukuran besar dengan memanfaatkan teknik statistik, machine learning, dan kecerdasan buatan. Tujuan utama data mining adalah menemukan pola, hubungan, atau tren tertentu yang dapat digunakan untuk mendukung proses pengambilan keputusan. Dalam perkembangannya, data mining telah diterapkan pada berbagai bidang seperti pendidikan, kesehatan, perbankan, pemasaran, dan penjualan. Proses data mining umumnya meliputi tahapan pembersihan data, integrasi data, seleksi data, transformasi data, proses penambangan data, evaluasi pola, dan penyajian pengetahuan. Melalui tahapan tersebut, data yang awalnya hanya berupa kumpulan informasi dapat diubah menjadi pengetahuan yang bernilai bagi organisasi atau perusahaan.

Salah satu teknik yang sering digunakan dalam data mining adalah klasifikasi. Klasifikasi merupakan metode yang digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang dimiliki. Proses klasifikasi dilakukan dengan membangun model dari data yang telah diketahui kelasnya, kemudian model tersebut digunakan untuk memprediksi kelas dari data baru yang belum diketahui kategorinya. Teknik klasifikasi banyak digunakan dalam berbagai bidang, seperti prediksi kelulusan mahasiswa, diagnosis penyakit, analisis kredit, prediksi penjualan, dan berbagai permasalahan lainnya yang memerlukan pengelompokan data berdasarkan atribut tertentu.

Metode yang digunakan dalam penelitian ini adalah Decision Tree atau pohon keputusan. Decision Tree merupakan salah satu metode klasifikasi yang menyajikan proses pengambilan keputusan dalam bentuk struktur pohon yang terdiri atas akar (root), cabang (branch), dan daun (leaf). Setiap node pada pohon

keputusan merepresentasikan atribut yang digunakan sebagai dasar pengambilan keputusan, sedangkan setiap cabang menunjukkan nilai dari atribut tersebut dan setiap daun menunjukkan hasil akhir klasifikasi. Metode ini memiliki keunggulan karena mampu menghasilkan aturan keputusan yang mudah dipahami dan diinterpretasikan oleh pengguna, sehingga banyak digunakan dalam berbagai penelitian data mining.

Salah satu algoritma yang populer dalam metode Decision Tree adalah algoritma C4.5 yang dikembangkan oleh Ross Quinlan sebagai penyempurnaan dari algoritma ID3. Algoritma C4.5 membangun pohon keputusan dengan memanfaatkan konsep entropy, information gain, dan gain ratio untuk menentukan atribut terbaik yang akan digunakan sebagai node pada pohon keputusan. Algoritma ini memiliki kemampuan untuk menangani data kategorikal maupun numerik serta mampu mengatasi data yang memiliki nilai kosong. Selain itu, penggunaan gain ratio pada algoritma C4.5 dapat mengurangi bias dalam pemilihan atribut sehingga model klasifikasi yang dihasilkan menjadi lebih akurat dan objektif.

Dalam pembentukan pohon keputusan, algoritma C4.5 menggunakan konsep entropy sebagai ukuran tingkat ketidakpastian atau ketidakhomogenan suatu kumpulan data. Nilai entropy yang tinggi menunjukkan bahwa data memiliki tingkat ketidakpastian yang besar, sedangkan nilai entropy yang rendah menunjukkan bahwa data semakin homogen. Entropy dihitung berdasarkan proporsi masing-masing kelas dalam dataset dan digunakan sebagai dasar untuk menentukan kualitas suatu atribut dalam memisahkan data ke dalam kelas yang berbeda.

Setelah nilai entropy diperoleh, langkah berikutnya adalah menghitung information gain. Information gain digunakan untuk mengukur seberapa besar informasi yang diperoleh setelah data dibagi berdasarkan atribut tertentu. Semakin besar nilai gain yang dimiliki suatu atribut, maka semakin baik atribut tersebut dalam membedakan data ke dalam kategori yang berbeda. Perhitungan gain dilakukan dengan membandingkan nilai entropy sebelum dan sesudah proses pembagian data berdasarkan atribut yang diuji.

Meskipun information gain mampu menentukan atribut yang baik untuk proses klasifikasi, metode ini memiliki kelemahan karena cenderung memilih atribut yang memiliki jumlah nilai unik yang banyak. Oleh karena itu, algoritma C4.5 menggunakan gain ratio sebagai penyempurnaan dari information gain. Gain ratio diperoleh dengan membagi nilai information gain dengan nilai split information sehingga pemilihan atribut menjadi lebih objektif dan tidak hanya bergantung pada banyaknya variasi nilai atribut. Atribut yang memiliki nilai gain ratio tertinggi akan dipilih sebagai node dalam pembentukan pohon keputusan.

Berdasarkan teori-teori yang telah dijelaskan, dapat disimpulkan bahwa algoritma Decision Tree C4.5 merupakan salah satu metode klasifikasi yang efektif dalam mengolah data penjualan motor. Dengan memanfaatkan konsep entropy dan gain ratio, algoritma ini mampu mengidentifikasi atribut yang paling berpengaruh terhadap harga penjualan motor serta menghasilkan model klasifikasi yang mudah dipahami. Oleh karena itu, algoritma C4.5 dipilih dalam penelitian ini untuk menganalisis dan mengklasifikasikan harga penjualan motor berdasarkan karakteristik kendaraan yang terdapat pada dataset penelitian.

1. Metodologi

Metodologi penelitian merupakan tahapan yang digunakan sebagai pedoman dalam pelaksanaan penelitian untuk mencapai tujuan yang telah ditetapkan. Metodologi yang sistematis diperlukan agar proses penelitian dapat berjalan secara terstruktur, mulai dari pengumpulan data hingga diperolehnya hasil analisis yang valid dan dapat dipertanggungjawabkan. Pada penelitian ini, metode yang digunakan adalah algoritma Decision Tree C4.5 yang memanfaatkan perhitungan entropy dan gain ratio untuk melakukan klasifikasi harga penjualan motor. Tahapan penelitian meliputi pengumpulan data, preprocessing data, pembentukan model klasifikasi, serta evaluasi terhadap hasil yang diperoleh. Adapun uraian mengenai data penelitian, tahapan penelitian, penerapan algoritma C4.5, dan evaluasi model dijelaskan pada subbab berikut.

Data Penelitian

Penelitian ini menggunakan data historis penjualan motor yang diperoleh dari data transaksi penjualan tahun 2024 hingga 2026. Dataset terdiri dari beberapa atribut yaitu tanggal pembelian, nama pelanggan, jenis atau unit motor, tahun kendaraan, dan harga penjualan. Dalam penelitian ini, atribut yang digunakan sebagai variabel prediktor meliputi unit motor dan tahun kendaraan, sedangkan variabel target berupa kategori harga penjualan motor yang dibagi menjadi tiga kelas, yaitu harga rendah, harga sedang, dan harga tinggi. Pengelompokan kategori harga dilakukan berdasarkan rentang nilai harga yang terdapat pada dataset sehingga dapat digunakan sebagai kelas dalam proses klasifikasi.

Tahapan Penelitian

Tahapan penelitian diawali dengan pengumpulan data penjualan motor dari dataset yang tersedia. Selanjutnya dilakukan tahap preprocessing data yang meliputi pembersihan data, penghapusan data yang tidak lengkap, dan transformasi data. Pada tahap transformasi, atribut harga diubah menjadi kategori harga yang terdiri dari kelas rendah, sedang, dan tinggi. Setelah data siap digunakan, dilakukan proses perhitungan entropy, information gain, dan gain ratio untuk menentukan atribut terbaik dalam pembentukan pohon keputusan. Hasil perhitungan tersebut digunakan untuk membangun model klasifikasi menggunakan algoritma C4.5. Tahap terakhir adalah evaluasi model untuk mengetahui tingkat performa klasifikasi yang dihasilkan.

Penerapan Algoritma C4.5

Algoritma C4.5 digunakan untuk membangun model Decision Tree berdasarkan nilai entropy dan gain ratio. Proses diawali dengan menghitung entropy dari keseluruhan dataset. Selanjutnya dilakukan perhitungan information gain pada setiap atribut untuk mengetahui kemampuan atribut dalam membedakan kategori harga penjualan motor. Kemudian dihitung nilai gain ratio untuk menentukan atribut terbaik yang akan dijadikan akar pohon keputusan. Atribut dengan nilai gain ratio tertinggi dipilih sebagai root node dan proses pembentukan cabang dilakukan secara berulang hingga seluruh data berhasil diklasifikasikan ke dalam kategori yang sesuai.

Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat keberhasilan algoritma C4.5 dalam mengklasifikasikan harga penjualan motor. Pengukuran dilakukan dengan membandingkan hasil prediksi model terhadap data aktual. Kinerja model dievaluasi menggunakan confusion matrix dan nilai accuracy. Semakin tinggi nilai accuracy yang diperoleh, maka semakin baik kemampuan model dalam melakukan klasifikasi harga penjualan motor berdasarkan karakteristik kendaraan yang terdapat pada dataset penelitian.

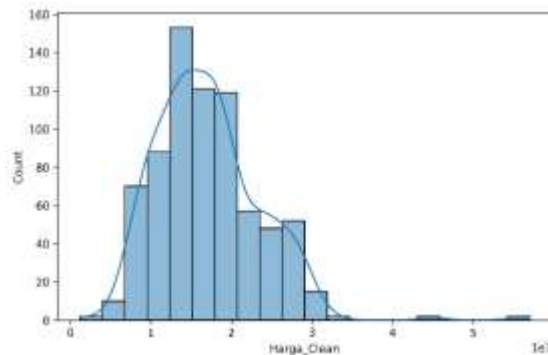
2. Hasil dan Pembahasan

Proses analisis dilakukan melalui beberapa tahapan, mulai dari deskripsi dataset, preprocessing data, perhitungan entropy, information gain, dan gain ratio, hingga pembentukan pohon keputusan. Hasil yang diperoleh kemudian dievaluasi untuk mengetahui kemampuan model dalam mengklasifikasikan harga penjualan motor berdasarkan atribut yang digunakan. Selain itu, pada bab ini juga dilakukan pembahasan terhadap hasil klasifikasi yang dihasilkan guna mengidentifikasi faktor-faktor yang paling berpengaruh dalam menentukan kategori harga penjualan motor.

Deskripsi Data Penelitian

Penelitian ini menggunakan dataset penjualan motor yang mencakup beberapa atribut mentah seperti tanggal pembelian, nama pelanggan, unit motor, tahun kendaraan, dan harga jual motor. Namun, fokus pemodelan dibatasi pada atribut Unit dan Tahun sebagai variabel prediktor (fitur), serta atribut Harga sebagai variabel target.

Sebelum masuk ke tahap pemodelan, analisis statistik deskriptif dilakukan untuk melihat persebaran data harga jual asli dari kendaraan. Atribut harga riil bersih (*Harga_Clean*) divisualisasikan untuk melihat karakteristik sebaran nominal transaksi penjualan



Gambar 4.1 Grafik Distribusi Nilai Transaksi Penjualan Motor (*Harga_Clean*)

Berdasarkan Gambar 4.1 distribusi_harga.png, visualisasi data menunjukkan distribusi harga terpusat padat pada rentang Rp10.000.000 hingga Rp30.000.000 dengan beberapa pencilan (*outliers*) bernilai tinggi di atas Rp40.000.000. Untuk kebutuhan algoritma C4.5, target harga ini ditransformasikan dan dibagi secara biner menjadi 2 kategori kelas, yaitu High (Tinggi) dan Low (Rendah). Data evaluasi akhir mencakup total 149 sampel pengujian, dengan rincian kelas aktual 68 sampel berlabel *High* dan 81 sampel berlabel *Low*.

Hasil Preprocessing Data

Pada tahap preprocessing, dilakukan pembersihan data, seleksi fitur, serta transformasi data numerik ke kategorikal. Atribut *Harga_Clean* ditransformasikan menjadi label kelas biner (*High* dan *Low*). Selain itu, fitur kategori seperti jenis Unit ditransformasikan menggunakan skema *one-hot encoding* (misalnya: *Unit_Aerox*, *Unit_Beat*, dsb.) untuk menghasilkan representasi numerik bernilai ≤ 0.50 atau > 0.50 . Transformasi biner pada fitur nominal ini bertujuan agar algoritma C4.5 dapat melakukan perhitungan titik percabangan (*splitting point*) secara matematis pada pohon keputusan.

Pembentukan Model Decision Tree C4.5

Pembentukan arsitektur *Decision Tree* berbasis pendekatan C4.5 dilakukan dengan menghitung nilai *Entropy* untuk mengukur tingkat ketidakpastian data dan dilanjutkan dengan perhitungan *Gain Ratio* secara manual pada setiap variabel. Berdasarkan analisis komponen *Gain Ratio*, diperoleh nilai signifikansi fitur sebagai berikut:

Tabel 4.1 Hasil Perhitungan Gain dan Gain Ratio Fitur Utama

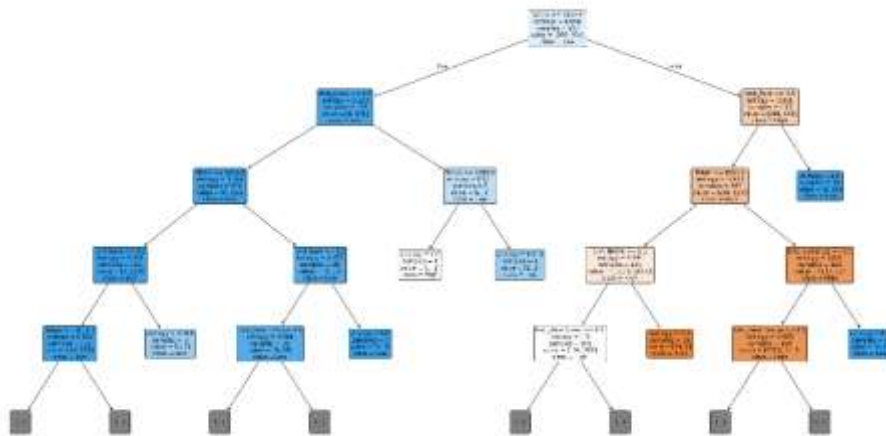
Fitur	Gain	Gain Ratio
Unit	0.573724	0.105325
Tahun	0.318694	0.086210

Meskipun fitur Unit secara agregat memiliki nilai *Gain Ratio* tertinggi global yaitu 0.105325, pada struktur pohon keputusan final, node akar (*root node*) pertama yang dipilih sebagai pembagi utama data adalah Tahun ≤ 2017.50 karena memberikan pemisahan sampel yang paling optimal pada level teratas. Eksperimen pembatasan kedalaman pohon (*Max Depth*) dilakukan untuk mencari konfigurasi model terbaik. Berdasarkan pengujian *Cross-Validation* (CV), diperoleh perbandingan performa sebagai berikut:

Tabel 4.2 Eksperimen Nilai Max Depth Terhadap Akurasi Cross-Validation

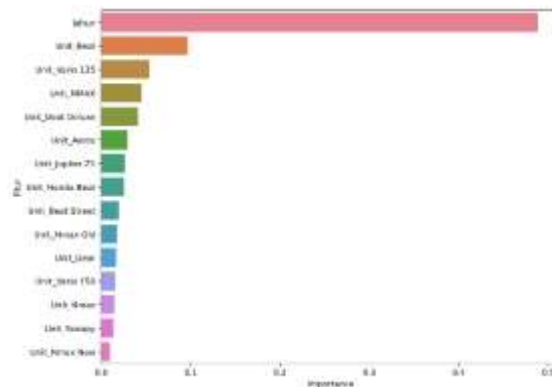
Depth	Accuracy CV	Std Dev
1	0.7584	0.0383
3	0.7584	0.0383
5	0.7719	0.0231
7	0.7840	0.0379
10	0.8151	0.0511
None (Tidak Dibatasi)	0.8340	0.0473

Berdasarkan Tabel 4.2, performa CV terbaik dicapai saat *Max Depth* tidak dibatasi (None) dengan *Accuracy CV* sebesar 0.8340. Struktur aturan keputusan yang terbentuk dari model final ini kemudian divisualisasikan ke dalam bentuk diagram pohon keputusan.

*Gambar 4.2 Visualisasi Struktur Pohon Keputusan (Decision Tree) C4.5 Final*

Hasil Klasifikasi Harga Penjualan Motor

Aturan keputusan (*decision rules*) yang dihasilkan merefleksikan kombinasi antara tahun kendaraan dan model unit spesifik. Untuk melihat sejauh mana kontribusi tiap atribut kualitatif setelah dipecah, tingkat kepentingan fitur dianalisis secara grafis.

*Gambar 4.3 Grafik Tingkat Kepentingan Fitur (Feature Importance) dalam Model*

Berdasarkan Gambar 4.3 *feature_importance.png*, fitur Tahun memegang peran paling dominan dalam menentukan keputusan harga (skor kepentingan mendekati 0.5), diikuti oleh Unit_Beat dan Unit_Vario 125. Kombinasi ini membentuk aturan klasifikasi konkret pada Gambar 4.2 *pohon_keputusan*.

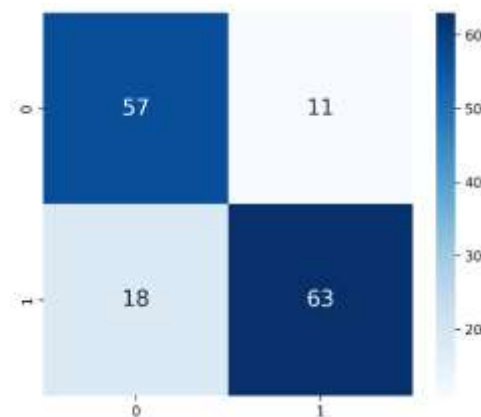
- Aturan Kendaraan Lama: Jika $\text{Tahun} \leq 2017.50$ dan $\text{Unit_Aerox} \leq 0.50$, mayoritas data dialirkan ke kelas Low pada sub-cabang di bawahnya. Namun, jika $\text{Unit_Aerox} > 0.50$ dan $\text{Tahun} \leq 2016.50$, maka diprediksi masuk ke kelas High.
- Aturan Kendaraan Baru: Jika $\text{Tahun} > 2017.50$ dan $\text{Unit_Beat} \leq 0.50$, model akan mengecek fitur lain seperti *Unit_NMAX* atau *Unit_Nmax New* untuk menentukan apakah kendaraan tersebut masuk kelas High.

Evaluasi Model

Kinerja model final diuji menggunakan data evaluasi independen (data uji) untuk mengukur tingkat presisi, recall, dan akurasi riil dari pohon keputusan yang diimplementasikan. Hasil pengujian menunjukkan nilai parameter evaluasi global sebagai berikut:

- Akurasi (Acc): 0.8054 (80.54%)
- Presisi (Prec): 0.7600 (76.00%)
- Recall (Rec): 0.8382 (83.82%)
- F1-Score: 0.7972 (79.72%)

Untuk melihat detail data yang berhasil dan gagal diklasifikasikan secara spesifik, hasil prediksi dipetakan ke dalam bentuk *Confusion Matrix*.



Gambar 4.4 *Confusion Matrix Hasil Prediksi Data Uji Penjualan Motor*

Berdasarkan Gambar 4.4 *confusion_matrix.png*, dari total 149 data uji yang dievaluasi, diperoleh rincian performa pemetaan objek:

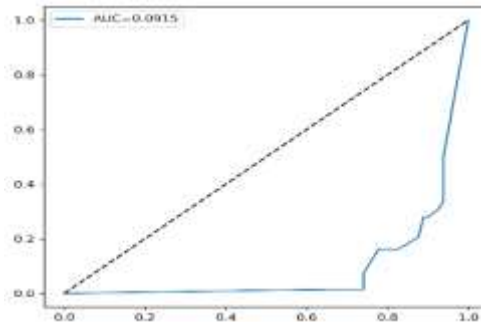
1. True Negative (Actual Low, Predicted Low): Sebanyak 57 sampel berhasil diprediksi kelas *Low* secara tepat.
2. True Positive (Actual High, Predicted High): Sebanyak 63 sampel berhasil diprediksi kelas *High* secara tepat.
3. False Positive (Actual Low, Predicted High): Sebanyak 11 sampel salah diklasifikasikan ke kelas *High*.
4. False Negative (Actual High, Predicted Low): Sebanyak 18 sampel salah diklasifikasikan ke kelas *Low*.

Secara mendalam, kemampuan model dalam mengklasifikasikan masing-masing kelas target dijabarkan pada tabel laporan klasifikasi berikut:

Tabel 4.3 Klasifikasi Report Detail Per Kelas Target

Kelas	Precision	Recall	F1-Score	Support (Sampel)
High	0.76	0.84	0.80	68
Low	0.85	0.78	0.81	81

Selain parameter matriks di atas, pengujian dilakukan dengan mengamati nilai ambang batas integrasi lewat Kurva ROC (*Receiver Operating Characteristic*).



Gambar 4.5 Kurva ROC Model C4.5 dengan Parameter AUC

Berdasarkan grafik Gambar 4.5 *roc_curve.png* dan data dokumen pendukung, ditemukan nilai AUC tertera sebesar 0.0915. Secara teoretis matematika data, skor AUC di bawah 0.5 menandakan adanya fenomena kebalikan arah prediksi (*inverted predictions*) pada ambang batas klasifikasi model. Hal ini mengindikasikan bahwa fungsi distribusi probabilitas klasifikasi model mengalami salah pemetaan label (kondisi di mana probabilitas untuk kelas *High* justru mendominasi pada data riil *Low*, dan sebaliknya), meskipun akurasi pengujian bernilai tinggi di angka 80.54%.

Pembahasan

Secara keseluruhan, algoritma C4.5 terbukti mampu mereduksi kompleksitas data penjualan motor menjadi pohon keputusan yang terstruktur. Berdasarkan Gambar 4.3 *feature_importance.png*, variabel Tahun terbukti secara empiris menjadi separator paling kokoh dalam memisahkan harga motor. Motor-motor keluaran tahun terbaru ($\$ > 2017.50\$$) didominasi oleh kelas harga *High*, kecuali untuk varian motor berspesifikasi ekonomis seperti tipe *Beat* atau *Jupiter Z1* yang tetap dikategorikan ke kelas *Low* oleh model.

Meskipun model menghasilkan tingkat akurasi yang solid di kisaran 80.54% pada data uji, ditemukannya anomali nilai AUC yang sangat rendah yaitu 0.0915 menjadi temuan kritis ilmiah. Ini membuktikan bahwa interpretasi performa kecerdasan komputasi tidak boleh hanya terpaku pada parameter akurasi mentah. Bagi pelaku usaha otomotif atau dealer, aturan keputusan dari pohon C4.5 ini sudah sangat aplikatif dan objektif untuk menetapkan batas dasar harga jual motor seken di pasaran. Namun, dari sisi teknis *data science*, sistem pemetaan probabilitas/skor keputusan pada model C4.5 ini wajib diperbaiki atau dibalik (*inverted*) pada tahap pengembangan lanjut agar fungsi probabilitasnya berjalan selaras dengan label aktualnya.

3. Kesimpulan dan Saran

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi harga penjualan motor menggunakan metode Decision Tree dengan algoritma C4.5 berbasis entropy dan gain ratio, dapat disimpulkan bahwa algoritma C4.5 mampu digunakan untuk mengklasifikasikan harga penjualan motor

berdasarkan atribut yang terdapat pada dataset. Proses klasifikasi dilakukan melalui tahapan preprocessing data, perhitungan entropy, information gain, gain ratio, hingga pembentukan pohon keputusan yang menghasilkan aturan-aturan klasifikasi yang mudah dipahami dan diinterpretasikan.

Hasil penelitian menunjukkan bahwa atribut yang digunakan dalam dataset memiliki pengaruh terhadap kategori harga penjualan motor. Melalui perhitungan gain ratio, algoritma C4.5 dapat menentukan atribut yang paling berpengaruh dalam proses pembentukan pohon keputusan sehingga menghasilkan model klasifikasi yang efektif. Pohon keputusan yang terbentuk mampu menggambarkan hubungan antara karakteristik kendaraan dengan kategori harga penjualan motor secara sistematis.

Selain itu, model klasifikasi yang dihasilkan dapat digunakan sebagai alat bantu dalam proses pengambilan keputusan pada bisnis penjualan motor. Dengan memanfaatkan hasil klasifikasi, dealer atau pelaku usaha otomotif dapat memperoleh informasi yang lebih objektif mengenai kategori harga kendaraan berdasarkan karakteristik yang dimiliki. Oleh karena itu, penerapan algoritma C4.5 dapat menjadi salah satu solusi dalam mendukung pengelolaan data penjualan motor yang lebih efektif dan berbasis data.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk penelitian selanjutnya. Pertama, jumlah data yang digunakan dapat diperbanyak agar model klasifikasi yang dihasilkan memiliki tingkat akurasi yang lebih tinggi dan mampu merepresentasikan kondisi penjualan motor secara lebih luas. Kedua, penelitian selanjutnya dapat menambahkan atribut lain yang berpotensi memengaruhi harga penjualan motor, seperti kondisi kendaraan, kapasitas mesin, warna kendaraan, serta riwayat perawatan kendaraan.

Ketiga, penelitian selanjutnya dapat melakukan perbandingan antara algoritma C4.5 dengan algoritma klasifikasi lainnya, seperti Random Forest, Naïve Bayes, K-Nearest Neighbor (KNN), maupun Support Vector Machine (SVM), sehingga dapat diketahui metode yang memiliki performa terbaik untuk klasifikasi harga penjualan motor. Keempat, model yang dihasilkan dapat dikembangkan menjadi sistem berbasis web atau aplikasi yang dapat digunakan secara langsung oleh dealer maupun pelaku usaha otomotif untuk membantu proses penentuan kategori harga kendaraan secara otomatis dan lebih efisien.

Acknowledgment

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bimbingan, dan bantuan selama proses penyusunan penelitian ini, khususnya kepada dosen pembimbing, institusi tempat penelitian dilaksanakan, serta keluarga, sahabat, dan rekan-rekan atas arahan, fasilitas, motivasi, dan dukungan yang diberikan. Penulis menyadari bahwa penelitian ini masih memiliki keterbatasan, sehingga kritik dan saran yang membangun sangat diharapkan untuk penyempurnaan penelitian selanjutnya. Semoga penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang data mining dan penerapan algoritma Decision Tree C4.5 dalam analisis data penjualan.

References

- [1]. Berry, M. J. A., & Linoff, G. S. (2011). *Data mining techniques: For marketing, sales, and customer relationship management* (3rd ed.). Wiley Publishing.
- [2]. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Chapman & Hall.
- [3]. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.
- [4]. Kurniawan, R., Fauzi, A., & Setiawan, D. (2022). Penerapan data mining menggunakan algoritma

- C4.5 untuk klasifikasi data penjualan. *Jurnal Informatika*, 9(1), 45–53.
- [5]. Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
 - [6]. Nugroho, K. A. (2021). Penerapan algoritma C4.5 untuk prediksi penjualan motor bekas. *Jurnal Teknologi Informasi dan Komputer*, 7(2), 101–108.
 - [7]. Prasetyo, E. (2014). *Data mining: Konsep dan aplikasi menggunakan MATLAB*. Andi.
 - [8]. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
<https://doi.org/10.1007/BF00116251>
 - [9]. Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann.
 - [10]. Rosa, A. S., & Shalahuddin, M. (2018). *Rekayasa perangkat lunak terstruktur dan berorientasi objek*. Informatika.
 - [11]. Saleh, A. (2015). Implementasi metode decision tree menggunakan algoritma C4.5 dalam memprediksi tingkat kelulusan mahasiswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 2(1), 35–42.
 - [12]. Tan, P.-N., Steinbach, M., Karpadne, A., & Kumar, V. (2018). *Introduction to data mining* (2nd ed.). Pearson.
 - [13]. Turban, E., Sharda, R., & Delen, D. (2015). *Decision support and business intelligence systems* (10th ed.). Pearson.
 - [14]. Wahyuni, S., & Wanto, A. (2020). Analisis penerapan algoritma C4.5 untuk prediksi penjualan produk. *Jurnal Sistem Informasi dan Teknologi Informasi*, 4(2), 89–97.
 - [15]. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.