



KLASIFIKASI HARGA SMARTPHONE BERDASARKAN SPESIFIKASI TEKNIS MENGGUNAKAN ALGORITMA C4.5

Zio Fahreliandro¹, Muhammad Nugi Sehpudin Saputra², Saiful Islam³, Muhammad Sony Maulana⁴

^{1,2,3}Program Studi Sistem Informasi, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika

Jln. Abdul Rahman Saleh No.18, Kec Pontianak Tenggara, Kota Pontianak, Kalimantan Barat

¹ ziofahreliandro@gmail.com, ² egisaputra1n2@gmail.com, ³ ipungfast26@gmail.com, ⁴ sony@bsi.ac.id

Abstract

Smartphones have become essential devices in modern society, offering diverse specifications and price ranges. Price differences are influenced by factors such as screen type, RAM capacity, network support, and additional features. This study applies the C4.5 algorithm to classify smartphone prices and identify the most influential attributes. Using a Kaggle dataset with 168 entries, the model achieved an accuracy of 81.25%. Screen type, 5G support, and screen resolution were identified as the primary determinants of price. These findings provide clearer insights into the relationship between technical specifications and smartphone prices, serving as a useful reference for market analysis and consumer decision-making.

Keywords: *C4.5, Price classification, Smartphone, Decision tree*

Abstrak

Smartphone merupakan bagian penting dari kehidupan modern dengan spesifikasi dan harga yang beragam. Perbedaan harga dipengaruhi oleh faktor seperti jenis layar, kapasitas RAM, dukungan jaringan, dan fitur tambahan. Penelitian ini menerapkan algoritma C4.5 untuk mengklasifikasikan harga smartphone serta mengidentifikasi atribut yang paling berpengaruh. Dataset Kaggle dengan 168 entri menghasilkan akurasi sebesar 81,25%. Jenis layar, dukungan 5G, dan resolusi layar terbukti sebagai faktor utama penentu harga. Temuan ini memberikan pemahaman mengenai hubungan spesifikasi teknis dengan harga smartphone, serta dapat dijadikan acuan dalam analisis pasar dan keputusan pembelian.

Kata kunci: *C4.5, Klasifikasi harga, Smartphone, Pohon keputusan*

1. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi telah menjadikan *smartphone* sebagai perangkat multifungsi yang tidak lagi sekedar alat komunikasi, melainkan juga berperan sebagai sarana hiburan, penunjang produktivitas, dan platform untuk menciptakan konten digital. Ragam fitur yang tersedia membuat spesifikasi teknis menjadi pertimbangan utama konsumen dalam memilih

dan membeli *smartphone* [1]. Tingkat kepemilikan telepon seluler mencapai 67,88% dari total populasi, menggambarkan bahwa *smartphone* telah menjadi kebutuhan yang nyaris tak terpisahkan dalam kehidupan masyarakat [2]. Situasi ini memicu persaingan yang semakin ketat di antara produsen *smartphone* untuk menghadirkan produk dengan performa dan fitur yang terus



ditingkatkan, demi mempertahankan eksistensi dan daya saing di pasar perangkat mobile [3].

Sebagai pengguna *smartphone*, konsumen dihadapkan pada banyak pilihan model perangkat yang menawarkan berbagai macam harga dan spesifikasi [4]. Harga sebuah *smartphone* ditentukan oleh berbagai aspek, seperti merek dan model perangkat, toko tempat produk dijual, kapasitas penyimpanan, besaran RAM, tahun peluncuran, ukuran dan dimensi fisik, bobot, jenis layar, ukuran layar, resolusi tampilan, serta faktor-faktor teknis lainnya [5]. Selain itu, ulasan dari pengguna juga berfungsi sebagai sumber informasi berharga yang membantu calon pembeli dalam menilai kualitas produk serta memahami pengalaman bertransaksi dengan penjual [6]. Dalam praktiknya, calon pembeli biasanya menilai kewajaran harga *smartphone* dengan membandingkan spesifikasi teknis yang tersedia, sehingga penting bagi mereka untuk memahami hubungan antara spesifikasi perangkat dan kategori harganya [7].

Seiring berkembangnya bidang data mining dan machine learning, berbagai penelitian telah dilakukan untuk mengklasifikasikan harga *smartphone* berdasarkan spesifikasi teknis yang dimilikinya. Ningrum et al. (2023) menerapkan algoritma *K-Nearest Neighbors* (KNN) dengan memanfaatkan atribut RAM, memori internal, dan kamera sebagai variabel utama. Meskipun metode ini cukup sederhana dan mudah diimplementasikan, KNN memiliki kelemahan dalam hal sensitivitas terhadap pemilihan nilai k dan ketergantungan pada jarak antar data, sehingga kinerjanya dapat menurun pada dataset dengan distribusi yang kompleks [8]. Arisusanto et al. (2023) menggunakan algoritma *Random Forest* yang dioptimalkan dengan *Grid Search* dan berhasil meningkatkan akurasi klasifikasi harga. Namun, model *Random Forest* cenderung bersifat kompleks dan kurang interpretatif, sehingga sulit digunakan untuk memahami atribut spesifikasi mana yang paling berpengaruh terhadap harga *smartphone* [9].

Penelitian lain oleh Najwa Ardelia et al. (2024) menerapkan algoritma *Logistic Regression* dengan atribut RAM, kamera, dan memori internal sebagai variabel input. Meskipun metode ini mampu memberikan hasil klasifikasi yang cukup baik, *Logistic Regression* memiliki keterbatasan dalam menangkap hubungan nonlinier antar atribut, yang sering

kali muncul pada data spesifikasi *smartphone*. Di sisi lain, Gustiana (2024) menegaskan bahwa algoritma C4.5 memiliki keunggulan dalam menghasilkan pohon keputusan yang akurat, mudah diinterpretasikan, serta mampu menangani atribut numerik, kategorikal, dan data yang tidak lengkap (*missing values*) [10].

Berdasarkan analisis terhadap penelitian terdahulu tersebut, dapat disimpulkan bahwa sebagian besar penelitian masih berfokus pada peningkatan akurasi model, namun kurang menekankan aspek interpretabilitas dan kemudahan analisis terhadap faktor-faktor penentu harga. Selain itu, belum banyak penelitian yang secara eksplisit membandingkan peran atribut spesifikasi dalam pembentukan kategori harga melalui aturan keputusan yang mudah dipahami. Oleh karena itu, diperlukan pendekatan yang tidak hanya mampu menghasilkan klasifikasi harga *smartphone* yang akurat, tetapi juga memberikan pemahaman yang jelas mengenai kontribusi masing-masing atribut spesifikasi

Algoritma C4.5 dipilih dalam penelitian ini karena kemampuannya dalam menyederhanakan permasalahan kompleks ke dalam bentuk pohon keputusan yang bersifat transparan dan interpretatif. Dengan memanfaatkan *information gain ratio*, algoritma ini mampu menentukan atribut yang paling berpengaruh dalam proses klasifikasi, sehingga hasil analisis dapat digunakan sebagai dasar pengambilan keputusan oleh konsumen maupun pelaku industri [11]. Kebaruan dari penelitian ini terletak pada penerapan algoritma C4.5 untuk mengklasifikasikan harga *smartphone* dengan menitikberatkan pada interpretasi aturan keputusan serta identifikasi atribut spesifikasi yang paling signifikan, bukan hanya pada nilai akurasi semata.

Permasalahan utama dalam penelitian ini adalah bagaimana membangun model klasifikasi harga *smartphone* yang akurat, mampu menangani berbagai jenis atribut spesifikasi, serta memberikan interpretasi yang jelas terhadap faktor-faktor teknis yang berpengaruh. Berdasarkan permasalahan tersebut, hipotesis yang diajukan adalah bahwa algoritma C4.5 mampu menghasilkan model klasifikasi harga *smartphone* yang akurat dan interpretatif, serta dapat mengidentifikasi atribut spesifikasi yang paling dominan dalam menentukan kategori harga.



Tujuan penelitian ini adalah menerapkan algoritma C4.5 dalam proses klasifikasi harga smartphone berdasarkan spesifikasi teknis, serta mengidentifikasi faktor-faktor teknis yang memiliki pengaruh terbesar terhadap pembentukan kategori harga. Hasil penelitian diharapkan dapat memberikan kontribusi ilmiah berupa penguatan kajian klasifikasi harga smartphone berbasis pohon keputusan, serta kontribusi praktis bagi konsumen dan pelaku industri dalam memahami hubungan antara spesifikasi smartphone dan tingkat harga yang ditawarkan di pasar.

2. TINJAUAN PUSTAKA

2.1. Algoritma C4.5

Algoritma C4.5 merupakan salah satu metode pembelajaran mesin yang digunakan untuk membangun pohon keputusan dalam proses klasifikasi. Algoritma ini bekerja dengan memilih atribut yang memiliki nilai *gain ratio* tertinggi sebagai simpul akar, kemudian membagi data ke dalam cabang-cabang hingga diperoleh simpul daun yang merepresentasikan kelas keputusan [12]. Proses ini memungkinkan permasalahan klasifikasi yang kompleks disederhanakan ke dalam struktur pohon yang mudah dipahami dan dianalisis [13].

Dibandingkan dengan algoritma pohon keputusan sebelumnya, seperti ID3, C4.5 memiliki beberapa keunggulan penting. Algoritma ini mampu menangani atribut numerik dan kategorikal, mengelola data yang memiliki nilai kosong (*missing values*), serta melakukan proses pemangkasan (*pruning*) untuk mengurangi risiko *overfitting*. Keunggulan tersebut menjadikan C4.5 lebih fleksibel dan stabil ketika diterapkan pada data dunia nyata yang cenderung heterogen, seperti data spesifikasi smartphone.

Dalam konteks penelitian terdahulu, algoritma C4.5 sering dibandingkan dengan metode klasifikasi lain seperti K-Nearest Neighbors, Logistic Regression, dan Random Forest. Metode KNN relatif mudah diimplementasikan, namun memiliki kelemahan dalam hal sensitivitas terhadap skala data dan pemilihan parameter jarak. Logistic Regression menawarkan model yang sederhana, tetapi kurang mampu menangkap hubungan nonlinier antar atribut. Sementara itu, Random Forest umumnya menghasilkan akurasi yang tinggi,

tetapi memiliki tingkat kompleksitas yang lebih besar dan sulit diinterpretasikan. Oleh karena itu, C4.5 dipandang sebagai solusi yang seimbang antara akurasi dan interpretabilitas, khususnya ketika tujuan penelitian tidak hanya memprediksi kelas, tetapi juga memahami faktor-faktor yang memengaruhi keputusan klasifikasi.

2.2. Klasifikasi Harga Smartphone

Klasifikasi merupakan proses pembentukan model atau fungsi yang digunakan untuk memetakan objek data ke dalam kelas atau kategori tertentu berdasarkan atribut yang dimilikinya [14]. Dalam penelitian ini, klasifikasi digunakan untuk mengelompokkan smartphone ke dalam kategori harga tertentu berdasarkan spesifikasi teknis yang dimiliki.

Berbagai penelitian sebelumnya telah menerapkan teknik klasifikasi untuk memprediksi atau mengelompokkan harga smartphone. Sebagian penelitian berfokus pada peningkatan akurasi prediksi dengan memanfaatkan algoritma kompleks, seperti Random Forest atau metode ensemble. Namun, pendekatan tersebut sering kali menghasilkan model yang sulit diinterpretasikan, sehingga kurang memberikan informasi mengenai alasan di balik pengelompokan harga yang dihasilkan. Di sisi lain, pendekatan klasifikasi yang bersifat interpretatif masih relatif terbatas, terutama yang secara eksplisit menampilkan aturan keputusan yang mudah dipahami oleh pengguna nonteknis.

Oleh karena itu, penggunaan algoritma C4.5 dalam klasifikasi harga smartphone memiliki nilai tambah, karena mampu menghasilkan aturan keputusan yang transparan dan logis. Aturan-aturan tersebut dapat digunakan untuk menjelaskan mengapa suatu smartphone diklasifikasikan ke dalam kategori harga tertentu, sehingga hasil penelitian tidak hanya bersifat prediktif, tetapi juga analitis.

2.3. Spesifikasi Smartphone

Spesifikasi smartphone dapat dipahami sebagai kumpulan informasi teknis yang menggambarkan kemampuan dan karakteristik suatu perangkat. Spesifikasi ini mencakup berbagai aspek, mulai dari identitas produk hingga komponen teknis seperti kapasitas RAM, memori internal, jenis dan ukuran layar, resolusi

kamera, prosesor, sistem operasi, kapasitas baterai, serta fitur pendukung lainnya [15].

Dalam berbagai penelitian terdahulu, spesifikasi smartphone sering digunakan sebagai variabel input utama dalam proses klasifikasi harga. Namun, sebagian besar penelitian hanya memilih atribut tertentu tanpa analisis mendalam mengenai kontribusi relatif masing-masing spesifikasi terhadap pembentukan kategori harga. Selain itu, interaksi antar spesifikasi sering kali diabaikan, padahal kombinasi atribut tertentu dapat memberikan pengaruh yang signifikan terhadap harga perangkat.

Penelitian ini menempatkan spesifikasi smartphone tidak hanya sebagai variabel input, tetapi juga sebagai objek analisis untuk mengetahui atribut mana yang paling dominan dalam menentukan kategori harga. Dengan memanfaatkan struktur pohon keputusan yang dihasilkan oleh algoritma C4.5, hubungan antara spesifikasi smartphone dan harga dapat dianalisis secara lebih sistematis dan mudah dipahami.

2.4. Sintesis Penelitian Terdahulu dan Kebaruan Penelitian

Berdasarkan kajian terhadap penelitian-penelitian sebelumnya, dapat disimpulkan bahwa sebagian besar studi terkait klasifikasi harga smartphone lebih menitikberatkan pada peningkatan akurasi model dengan menggunakan algoritma yang kompleks. Namun, aspek interpretabilitas dan pemahaman terhadap faktor-faktor teknis penentu harga masih belum banyak dibahas secara mendalam.

Kebaruan penelitian ini terletak pada penerapan algoritma C4.5 untuk klasifikasi harga smartphone dengan fokus pada analisis aturan keputusan dan identifikasi atribut spesifikasi yang paling berpengaruh. Dengan pendekatan ini, penelitian tidak hanya menghasilkan model klasifikasi yang akurat, tetapi juga memberikan wawasan yang jelas mengenai hubungan antara spesifikasi smartphone dan kategori harga. Pendekatan tersebut diharapkan dapat memberikan kontribusi yang lebih komprehensif dibandingkan penelitian-penelitian sebelumnya, baik dari sisi akademis maupun praktis.

3. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan klasifikasi. Tujuan utama dari penelitian ini adalah mengidentifikasi faktor-faktor yang mempengaruhi kisaran harga ponsel dan mengelompokkan harga ke dalam kategori tertentu menggunakan algoritma C4.5. Penelitian ini dilakukan melalui tahapan pengumpulan data spesifikasi smartphone, dilanjutkan dengan pra-pemrosesan untuk membersihkan dan menyiapkan data. Data kemudian dibagi menjadi data latih dan data uji sebelum algoritma C4.5 diterapkan untuk membentuk model klasifikasi harga. Model yang dihasilkan dievaluasi menggunakan metrik akurasi dan dianalisis bersama visualisasi hasil untuk mengetahui faktor spesifikasi yang paling berpengaruh, hingga akhirnya ditarik kesimpulan mengenai kinerja algoritma C4.5. secara umum, tahapan penelitian digambarkan pada gambar 1



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari situs Kaggle dan terdiri dari lebih dari 100 entri data spesifikasi smartphone dari berbagai merek yang beredar di pasar Indonesia. Dataset mencakup sejumlah atribut teknis, antara lain kapasitas RAM, penyimpanan internal, kapasitas baterai, ukuran dan resolusi layar, jenis layar, fitur tambahan seperti dukungan 5G, NFC, eSIM, dual SIM, dan slot memori, dimensi fisik perangkat (panjang, lebar, ketebalan, dan berat), serta harga smartphone yang dinyatakan dalam satuan ribuan rupiah (*os_price*). Dataset ini dipilih karena merepresentasikan variasi spesifikasi dan kisaran harga smartphone yang cukup beragam.



3.2. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum digunakan dalam pembentukan model klasifikasi. Proses ini meliputi pembersihan data dengan menghapus entri yang memiliki nilai kosong, transformasi atribut kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma klasifikasi, serta pemilihan atribut yang relevan dengan tujuan penelitian. Selain itu, dilakukan proses pembentukan kategori harga sebagai variabel target klasifikasi [16]. Kategori harga dibagi menjadi tiga kelas, yaitu Low, Medium, dan High, berdasarkan distribusi nilai harga dalam dataset. Penentuan batas masing-masing kategori dilakukan dengan pendekatan statistik menggunakan pembagian rentang harga ke dalam tiga interval yang relatif seimbang. Pendekatan ini bertujuan untuk menghasilkan distribusi kelas yang proporsional sehingga model klasifikasi dapat belajar secara optimal dan mengurangi potensi bias akibat ketidakseimbangan data.

3.3. Pembagian Data

Setelah melalui tahap pra-pemrosesan, dataset dibagi menjadi dua bagian, yaitu data pelatihan (*training data*) dan data pengujian (*testing data*) [17]. Sebanyak 80% data digunakan sebagai data pelatihan untuk membangun model klasifikasi, sedangkan 20% sisanya digunakan sebagai data pengujian untuk mengevaluasi kinerja model. Pembagian ini dilakukan untuk memastikan bahwa model yang dihasilkan mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya.

3.4. Penerapan Algoritma C4.5

Model klasifikasi dibangun menggunakan algoritma C4.5 yang diimplementasikan melalui *Decision Tree Classifier* dengan kriteria entropy. Algoritma ini memanfaatkan konsep *information gain ratio* dalam menentukan atribut terbaik sebagai simpul pada pohon keputusan. Proses pembentukan pohon keputusan bertujuan untuk mempelajari pola hubungan antara atribut spesifikasi smartphone dan kategori harga yang dihasilkan, sekaligus

menghasilkan aturan keputusan yang mudah diinterpretasikan.

3.5. Evaluasi Model

Evaluasi model dilakukan untuk menilai kinerja algoritma C4.5 dalam mengklasifikasikan harga smartphone berdasarkan spesifikasi teknis. Kinerja model diukur menggunakan metrik akurasi untuk mengetahui tingkat ketepatan prediksi model terhadap data uji. Selain itu, analisis terhadap struktur pohon keputusan dilakukan untuk mengidentifikasi atribut spesifikasi yang memiliki pengaruh paling signifikan dalam menentukan kategori harga. Hasil evaluasi ini digunakan sebagai dasar dalam menarik kesimpulan mengenai efektivitas algoritma C4.5 dalam klasifikasi harga smartphone.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari seluruh proses penelitian yang telah dilakukan, mulai dari penerapan tahapan pra-pemrosesan data, pembagian data, pembangunan model menggunakan algoritma C4.5, hingga evaluasi performa model dalam mengklasifikasikan faktor-faktor yang mempengaruhi harga smartphone.

4.1. Hasil Pengumpulan Data

Penelitian ini menggunakan dataset publik *Kaggle* yang berisi spesifikasi smartphone berbagai merek yang beredar di pasar Indonesia. Dataset terdiri dari 168 baris data dengan 19 atribut yang merepresentasikan karakteristik teknis perangkat, seperti kapasitas memori, penyimpanan internal, fitur jaringan, dimensi fisik, kapasitas baterai, jenis dan ukuran layar, serta harga smartphone.

Kelengkapan atribut dalam dataset ini dinilai relevan untuk digunakan dalam klasifikasi harga smartphone, karena mencerminkan faktor-faktor teknis yang secara teoritis dan empiris memengaruhi nilai jual perangkat. Variasi spesifikasi dan kisaran harga yang cukup luas juga memungkinkan algoritma C4.5 membentuk aturan keputusan yang representatif.

4.2. Hasil Pra-pemrosesan Data

Pada langkah awal, dilakukan pemeriksaan terhadap nilai kosong (*missing values*) yang terdapat dalam dataset. Nilai-nilai kosong tersebut dapat mengganggu proses pelatihan model apabila tidak ditangani dengan baik. Pada penelitian ini, baris yang memiliki nilai kosong dihapus menggunakan fungsi `dropna()`, sehingga hanya data yang lengkap dan valid yang digunakan untuk analisis lebih lanjut.

Selanjutnya pembentukan kategori harga menggunakan kolom `os_price` yang harga smartphone dalam ribuan rupiah, awalnya berupa nilai numerik kontinu. Agar algoritma C4.5 dapat mengelompokkan harga ke dalam kelas tertentu, data harga diubah menjadi tiga kategori, yaitu low, medium, dan high. Pengelompokan ini dilakukan berdasarkan rentang harga tertentu sehingga model dapat melakukan klasifikasi harga secara lebih terstruktur, berikut hasilnya:

Tabel 1. Price Category

High	53
Low	52
Medium	52

Selanjutnya yaitu menkonversi atribut kategorikal ke bentuk numerik. Dataset berisi beberapa atribut bertipe teks, seperti `display_type` dan `device`. Karena algoritma C4.5 yang diimplementasikan melalui `DecisionTreeClassifier` membutuhkan data numerik, maka seluruh atribut kategorikal diubah menjadi bentuk numerik menggunakan `LabelEncoder()`. Proses ini memastikan bahwa model dapat mengenali pola dari setiap atribut tersebut tanpa kehilangan informasi esensial.

Atribut bertipe kategorikal, seperti `display_type` dan `device`, kemudian dikonversi ke dalam bentuk numerik agar dapat diproses oleh algoritma C4.5 yang diimplementasikan melalui `Decision Tree Classifier`. Setelah proses ini, dataset dipisahkan menjadi variabel fitur (X) dan variabel target (Y) untuk memudahkan proses pelatihan model.

4.3. Hasil Pembagian Data

Dataset yang telah melalui tahap pra-pemrosesan dibagi menjadi data pelatihan dan data pengujian dengan perbandingan 80% untuk data latih dan 20% untuk data uji. Pembagian ini

bertujuan untuk memastikan bahwa model mampu melakukan generalisasi dengan baik terhadap data yang belum pernah dilihat sebelumnya. Proporsi pembagian tersebut umum digunakan dalam penelitian klasifikasi dan dianggap cukup untuk menghasilkan model yang stabil.

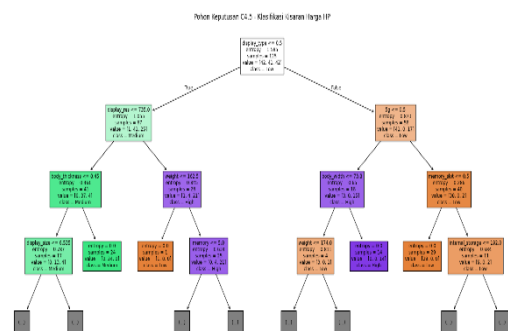
```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42, stratify=y
)
```

Gambar 2. Pembagian Data

4.4. Hasil Penerapan Algoritma C4.5

Model klasifikasi harga smartphone dibangun menggunakan algoritma C4.5 dengan kriteria *entropy*. Model dilatih menggunakan data pelatihan untuk mempelajari pola hubungan antara spesifikasi smartphone dan kategori harga. Untuk menghindari *overfitting* dan menjaga interpretabilitas model, kedalaman maksimum pohon keputusan dibatasi hingga lima tingkat.

Visualisasi pohon keputusan menunjukkan bahwa atribut tertentu, khususnya yang berkaitan dengan tampilan layar dan dukungan jaringan, sering muncul sebagai simpul pemisah utama. Hal ini mengindikasikan bahwa atribut-atribut tersebut memiliki kontribusi yang signifikan dalam membedakan kategori harga smartphone. Temuan ini sejalan dengan teori bahwa kualitas tampilan dan teknologi jaringan merupakan salah satu indikator utama dalam penentuan harga perangkat mobile modern.



Gambar 3. Hasil Visualisasi Pohon Keputusan

4.5. Hasil Evaluasi

Evaluasi model dilakukan menggunakan data pengujian, dan hasil pengujian menunjukkan bahwa algoritma C4.5 mampu mencapai tingkat

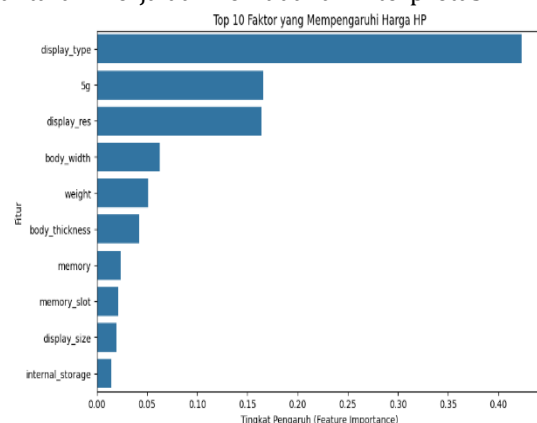
akurasi sebesar **81,25%**. Nilai ini menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan ke dalam kategori harga yang sesuai, sehingga model dapat dikatakan memiliki kinerja yang cukup baik.

```
accuracy = model.score(X_test, y_test)
print(f"Akurasi Model: {accuracy*100:.2f}%")

Akurasi Model: 81.25%
```

Gambar 3. Hasil Akurasi

Jika dibandingkan dengan penelitian terdahulu, tingkat akurasi ini berada pada kisaran yang kompetitif. Beberapa penelitian yang menggunakan algoritma KNN dan Logistic Regression melaporkan akurasi yang bervariasi, namun sering kali menghadapi keterbatasan dalam hal interpretabilitas model. Sementara itu, meskipun Random Forest mampu menghasilkan akurasi yang lebih tinggi, kompleksitas modelnya menyulitkan analisis faktor-faktor penentu harga. Dalam konteks ini, algoritma C4.5 menawarkan keseimbangan antara kinerja dan kemudahan interpretasi.



Gambar 4. Hasil Faktor yang Mempengaruhi Harga HP

Analisis *feature importance* menunjukkan bahwa atribut seperti *display_type*, dukungan jaringan 5G, dan resolusi layar (*display_res*) merupakan faktor yang paling berpengaruh dalam menentukan kategori harga smartphone. Temuan ini konsisten dengan penelitian sebelumnya yang menyatakan bahwa teknologi layar dan jaringan menjadi salah satu faktor utama yang memengaruhi persepsi nilai dan

harga perangkat smartphone. Hal ini juga mencerminkan tren pasar smartphone saat ini, di mana kualitas visual dan konektivitas menjadi daya tarik utama bagi konsumen.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa algoritma C4.5 tidak hanya mampu mengklasifikasikan harga smartphone dengan tingkat akurasi yang baik, tetapi juga memberikan pemahaman yang jelas mengenai atribut spesifikasi yang paling berpengaruh. Dengan demikian, model yang dihasilkan dapat digunakan sebagai alat analisis yang informatif bagi konsumen maupun pelaku industri dalam memahami hubungan antara spesifikasi smartphone dan kategori harga.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma C4.5 yang diimplementasikan melalui *DecisionTreeClassifier* berbasis *entropy* mampu mengklasifikasikan kisaran harga smartphone ke dalam kategori Rendah, Sedang, dan Tinggi dengan performa yang cukup baik. Model yang dibangun menghasilkan tingkat akurasi sebesar 81,25%, yang menunjukkan bahwa spesifikasi teknis smartphone memiliki keterkaitan yang kuat terhadap pembentukan kategori harga. Hasil analisis *feature importance* menunjukkan bahwa tipe layar (*display_type*) merupakan faktor yang paling berpengaruh dalam menentukan harga smartphone, diikuti oleh dukungan jaringan 5G dan resolusi layar (*display_res*). Temuan ini mengindikasikan bahwa aspek tampilan visual dan teknologi konektivitas menjadi komponen utama yang memengaruhi nilai jual smartphone di pasar Indonesia.

Sebagai saran untuk pengembangan penelitian selanjutnya, disarankan untuk melakukan optimasi parameter pada *DecisionTreeClassifier* guna meningkatkan performa model serta meminimalkan kemungkinan *overfitting*. Selain itu, penelitian lanjutan dapat membandingkan algoritma C4.5 dengan metode klasifikasi lain seperti Random Forest, Support Vector Machine (SVM), atau algoritma berbasis ensemble untuk memperoleh model dengan akurasi dan stabilitas yang lebih baik. Dari sisi pra-pemrosesan data, penggunaan metode imputasi data sebagai alternatif penghapusan nilai kosong juga dapat dipertimbangkan agar pemanfaatan dataset



menjadi lebih optimal dan informasi yang berharga tidak hilang. Dengan pengembangan tersebut, diharapkan model klasifikasi harga smartphone yang dihasilkan dapat semakin akurat dan aplikatif dalam mendukung analisis pasar maupun pengambilan keputusan pembelian.

6. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada pihak-pihak yang telah menyediakan sumber data sehingga proses pengolahan dan analisis dapat dilakukan dengan baik. Apresiasi khusus diberikan kepada platform Kaggle sebagai penyedia data sekunder yang digunakan dalam penelitian ini. Penulis juga berterima kasih kepada rekan-rekan yang memberikan dukungan dan masukan selama penyusunan penelitian sehingga tugas ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA:

- [1] A. E. Syaputra and Y. S. Eirlangga, "Implentasi Metode Simple Additive Weighting dalam Memberikan Rekomendasi Smartphone Terbaik Kepada Pelanggan," *J. Sistim Inf. dan Teknol.*, vol. 5, no. 2, pp. 103–109, 2023, doi: 10.37034/jsisfotek.v5i1.215.
- [2] M. R. R. Saelan and A. Subekti, "K-Best Selection Untuk Meningkatkan Kinerja Artificial Neural Network Dalam Memprediksi Range Harga Ponsel," *INTI Nusa Mandiri*, vol. 19, no. 1, pp. 10–16, 2024, doi: 10.33480/inti.v19i1.5554.
- [3] A. Arisusanto, N. Suarna, and G. Dwilestari, "Jurnal Teknologi Ilmu Komputer Analisa Klasifikasi Data Harga Handphone Menggunakan Jurnal Teknologi Ilmu Komputer," *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 43–47, 2023, doi: 10.56854/jtik.v1i2.51.
- [4] R. Amperawansyah and D. P. Putri, "Pemilihan Jenis Smartphone Sesuai dengan Kebutuhan Menggunakan Metode Forward Chaining dan Decicion Tree," *INOVTEK Polbeng - Seri Inform.*, vol. 7, no. 1, p. 156, 2022, doi: 10.35314/isi.v7i1.2402.
- [5] E. T. A. A'fena, A. Ristyawan, and A. Nugroho, "Pengaruh Normalisasi Fitur Pada Perbandingan Algoritma Random Forest Regressor Dan KNN Regressor Dalam Menentukan Harga Smartphone," *KRESNA J. Ris. dan Pengabdi. Masy.*, vol. 5, no. 1, pp. 74–83, 2025, doi: 10.36080/kresna.v5i1.226.
- [6] A. Junaedi and D. Kurniawan, "Analisis sentimen ulasan produk menggunakan algoritma Naive Bayes," *J. Inform. dan Komput.*, vol. 12, no. 2, pp. 56–63, 2021.
- [7] A. Karim, F. Nurhadi, I. K. O. Setiawan, I. A. Rizky, and R. Br. Manurung, "Pengaruh Normalisasi Data pada Klasifikasi Harga Ponsel Berdasarkan Spesifikasi Menggunakan Klasifikasi Naive Bayes dan Multinomial Logistic Regression," *J. Rekayasa Elektro Sriwij.*, vol. 4, no. 1, pp. 8–16, 2023, doi: 10.36706/jres.v4i1.59.
- [8] D. F. Ningrum, S. P. Ramadhani, I. Paryudi, I. Veritawati, S. R. Candra, and N. Nursari, "Prediksi Harga Smartphone berdasarkan Spesifikasi menggunakan K-Nearest Neighbors," *J. Informatics Adv. Comput.*, vol. 4, no. 2, pp. 101–106, 2023, doi: 10.35814/jiac.v4i2.6293.
- [9] D. Najwa Ardelia, H. Desfianty Arifin, S. Daniswara, and A. Puspita Sari, "Klasifikasi Harga Ponsel Menggunakan Algoritma Logistic Regression," *J. Informatics Electron. Eng.*, vol. 04, no. 01, pp. 37–43, 2024.
- [10] Z. Gustiana, "Performance Evaluation Algoritma C 4.5 Pada Klasifikasi Data," *Djtechno J. Teknol. Inf.*, vol. 5, no. 2, pp. 289–296, 2024, doi: 10.46576/djtechno.v5i2.4654.
- [11] K. A. Saputra, J. T. Hardinata, M. R. Lubis, S. R. Andani, and I. S. Saragih, "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>
- [12] Ade Izyuddin and Setyawan Wibisono, "Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree Dengan Algoritma C4.5," *J. Manaj. Inform. Sist. Inf.*, vol. 3, no. 2, pp. 146–156, 2020.
- [13] Sugiyanti and Muhammad Fauzi Firdaus, "Analisis Data Mining Dalam Prediksi Pemilihan Jurusan Pada Sma Manggala Menggunakan Algoritma C 4.5," *J. Manaj. Inform. dan Sist. Inf.*, vol. 8, no. 2, pp. 304–311, 2025, doi: 10.36080/kresna.v5i1.226.



- 10.36595/misi.v8i2.1648.
- [14] A. Husnah Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk Laris," *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021.
- [15] A. S. Silalahi and Q. M. H. S. Subekti, "Pengaruh Harga Dan Spesifikasi Brand Infinix Terhadap Keputusan Pembelian Smartphone," *J. Econ. Manag. Business, Entrep.*, vol. 1, no. 2, pp. 98–105, 2023, doi: 10.31537/jembe.v2i2.1553.
- [16] S. Febriani, "Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5 Siska Febriani Sistem Informasi *) Rohmansyah@gmail.com," *J. Ilmu Data*, vol. 2, no. 9, pp. 1–9, 2022, [Online]. Available: <http://www.ilmudata.org/index.php/ilmudata/article/view/213%0Ahttp://www.ilmudata.org/index.php/ilmudata/article/download/213/212>
- [17] M. Irsyad, A. Silvy, and Y. M. Ardi, "Prediksi Harga Ponsel Berdasarkan Spesifikasinya Menggunakan Algoritma Linear Regression," *INTI Nusa Mandiri*, vol. 19, no. 2, pp. 251–258, 2025, doi: 10.33480/inti.v19i2.6292.