



Analisis Faktor Penentu Harga Laptop Menggunakan Algoritma Decision Tree

Zainul Arifin Alwi Al Qayyis^{1*}, Zaehol Fatah²

¹ Teknologi Informasi, Universitas Ibrahimy

² Sistem Informasi, Universitas Ibrahimy

^{1*} alwie2743@gmail.com, ² zaeholfatah@gmail.com

Article History:

Received Nov 24th, 2024

Revised Nov 25th, 2024

Accepted Nov 26th, 2024

Abstrak

Beberapa dekade terakhir, laptop menjadi kebutuhan dasar bagi masyarakat untuk berbagai aktivitas sehari-hari, mulai dari belajar, mengajar, bekerja, hingga berbelanja. Hal ini mendorong permintaan laptop meningkat. Untuk menganalisis faktor-faktor yang memengaruhi harga laptop dengan menggunakan algoritma Decision Tree, memberikan informasi kepada calon pembeli dalam memilih laptop sesuai anggaran dan kriteria yang diinginkan, serta memberi wawasan kepada produsen terkait strategi penetapan harga. Metode yang digunakan adalah algoritma C4.5, salah satu metode Decision Tree yang memanfaatkan struktur pohon untuk klasifikasi, serta untuk data berukuran besar dengan banyak variabel, serta mampu menganalisis berbagai jenis data. Hasil penelitian menunjukkan bahwa model pohon keputusan yang dihasilkan memiliki beberapa tingkat kedalaman, dengan variabel yang memengaruhi harga laptop seperti prosesor, tipe penyimpanan, dan RAM, di mana prosesor menjadi faktor yang paling berpengaruh. Dengan gain sebesar 0,4983. Model ini mencapai akurasi sebesar 87.63%. Dengan mempertimbangkan segmen pasar, pengguna dan kualitas produk.

Kata Kunci : Decision Tree, harga laptop, algoritma C4.5, klasifikasi, data mining, kualitas produk

Abstract

In recent decades, laptops have become a basic necessity for people in various daily activities, including studying, teaching, working, and shopping. This has driven an increase in laptop demand. To analyze the factors affecting laptop prices, a Decision Tree algorithm is employed, providing potential buyers with information to choose laptops according to their budget and desired criteria, and giving producers insights into pricing strategy. The method used is the C4.5 algorithm, one of the Decision Tree methods that utilizes a tree structure for classification, is capable of handling large datasets with many variables, and can analyze various types of data. The research results show that the decision tree model generated has multiple levels of depth, with variables affecting laptop prices such as the processor, storage type, and RAM, with the processor being the most influential factor with a gain of 0.4983. The model achieves an accuracy of 87.63%, considering market segments, users, and product quality.

Keyword : Decision Tree, laptop prices, C4.5 algorithm, classification, data mining, product quality

PENDAHULUAN

Dalam beberapa dekade terakhir, perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan di berbagai sektor kehidupan. Salah satu perangkat penting yang menjadi kebutuhan utama di era digital adalah laptop. Laptop tidak hanya berfungsi sebagai alat komputasi sehari-hari oleh individu, tetapi juga oleh perusahaan, institusi pendidikan, dan organisasi lainnya untuk berbagai keperluan, mulai dari pekerjaan, studi, hingga hiburan (Erliani et al., 2023). Meningkatnya permintaan terhadap laptop membuat harga menjadi salah satu faktor penting yang dipertimbangkan oleh konsumen dalam proses pembelian.

Harga laptop dipengaruhi oleh berbagai faktor, seperti spesifikasi perangkat keras (hardware), jenis dan ukuran layar, kapasitas penyimpanan (storage), merek, serta fitur tambahan lainnya. Di tengah banyaknya pilihan laptop dengan spesifikasi yang beragam, proses pengambilan keputusan untuk memilih laptop yang sesuai dengan kebutuhan dan anggaran menjadi semakin kompleks. Memahami faktor-faktor yang memengaruhi harga laptop menjadi penting bagi



konsumen yang ingin mendapatkan produk dengan harga dan kualitas terbaik serta bagi produsen yang menetapkan strategi harga yang tepat (Jayen et al., 2022).

Dalam konteks ilmu data, teknik data mining semakin banyak digunakan untuk menggali pola dan pengetahuan dari kumpulan data yang besar dan kompleks. Data mining adalah proses untuk memperoleh informasi berguna dari basis data besar dan mengekstraksi informasi baru guna membantu pengambilan keputusan (SLN, 2023). Melalui proses ini, data mining dapat menemukan korelasi, pola, dan tren baru dengan menganalisis sejumlah besar data dalam suatu repository, yang menghasilkan wawasan berharga untuk pengambilan keputusan yang lebih baik (Uddin & Fatah, 2024). Selain itu, data mining memungkinkan identifikasi karakteristik dan informasi tak terduga yang mungkin muncul di masa depan (Sekar Setyaningtyas et al., 2022)

Salah satu teknik yang umum dalam data mining adalah decision tree, yaitu algoritma yang efektif untuk tugas-tugas klasifikasi dan prediksi (Assauqi, 2024). Algoritma Decision Tree sering digunakan untuk mengklasifikasikan masalah dan mencari solusi (Permana et al., 2021). Dengan struktur pohon biner, decision tree memudahkan visualisasi proses pengambilan keputusan berdasarkan fitur-fitur dalam data. Dalam hal ini, algoritma ini digunakan untuk mengidentifikasi faktor-faktor yang memengaruhi harga laptop, di mana setiap simpul merepresentasikan fitur penting seperti prosesor, RAM, dan kapasitas penyimpanan.

Selain itu, algoritma C4.5 merupakan salah satu varian algoritma decision tree yang digunakan untuk klasifikasi berbasis pohon keputusan (Pada et al., 2023). Proses pembentukan pohon keputusan pada Algoritma C4.5 melibatkan beberapa langkah, seperti memilih atribut sebagai akar, membuat cabang untuk setiap nilai, dan membagi instance pada cabang tersebut hingga seluruh kasus dalam cabang tersebut termasuk dalam satu kelas yang sama (Syahputra, 2021). Untuk memilih atribut sebagai akar, digunakan nilai gain tertinggi dari atribut-atribut yang ada, yang dihitung menggunakan rumus gain dan entropy:

Rumus ini digunakan untuk menghitung **Gain** dari suatu atribut **A** dalam himpunan kasus **S**

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Sedangkan rumus ini digunakan untuk menghitung rumus Entropy :

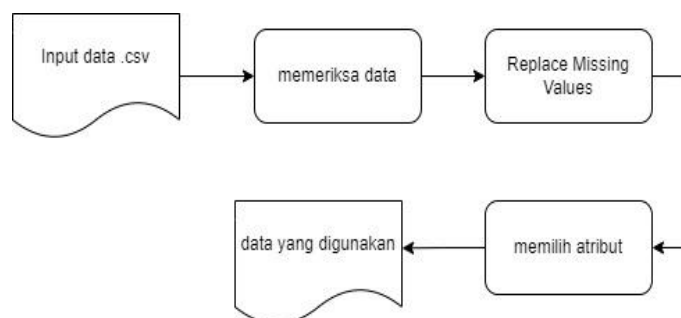
$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

METODE

Pengumpulan Data

Data penelitian ini diambil dari dataset public yang tersedia di platform Kaggle (Airi et al., 2023), yang berisi data penjualan laptop yang mengandung informasi spesifikasi laptop seperti merek, ukuran layar, prosesor, RAM, penyimpanan, serta harga dalam mata. Meskipun data asli dikumpulkan dalam konteks global, penelitian ini berfokus pada pasar Indonesia.

Preprocessing Data



Gambar 1. Proses Prosesing Data

Langkah pertama dalam preprocessing data adalah mengimpor dataset dari file CSV yang berisi data mentah seperti harga, spesifikasi, dan merek laptop. Setelah itu, dilakukan pemeriksaan data untuk mendeteksi nilai yang hilang (missing values), data yang tidak sesuai atau outlier, serta memeriksa data di setiap kolom (Khotimah et al., 2021). Jika ada nilai yang hilang, mereka ditangani melalui Teknik imputasi dengan mengganti nilai yang hilang menggunakan rata-rata, median, atau modus, atau dihapus jika tidak signifikan. Selanjutnya, atribut yang paling relevan dengan analisis, seperti RAM, prosesor, ukuran layar, dan jenis penyimpanan, dipilih, sementara atribut yang tidak relevan diabaikan.



Decision Tree

Decision tree mengevaluasi fitur – fitur seperti merek, RAM, prosesor, tipe penyimpanan, ukuran layar untuk menentukan mana yang paling berpengaruh dalam prediksi harga laptop. Dataset akan dibagi secara berulang berdasarkan fitur yang dipilih, serta node disetiap daun mewakili keputusan akhir, dimana rata – rata laptop dalam kelompok tersebut diprediksi. Proses ini didasarkan pada perhitungan entropy dan information gain(Jannah et al., 2023), yang dilakukan secara berulang kali sampai setiap atribut pohon mencapai kelas yang tepat dan tidak memungkinkan lagi dilakukan perhitungan lebih lanjut. Kolaborasi antara pemilihan fitur, pembagian dataset, serta perhitungan entropy dan information gain memastikan bahwa model decision tree dapat menghasilkan prediksi yang akurat dalam menentukan faktor yang paling berpengaruh terhadap harga laptop.

Analisis Hasil Pengujian

Analisis hasil pengujian dilakukan menggunakan confusion matrix untuk memvisualisasikan kinerja model decision tree dalam memprediksi harga laptop. Confusion matrix menampilkan informasi actual dan prediksi yang kemudian digunakan untuk menghitung akurasi, presisi, dan recall(SLN, 2023). Akurasi pengukuran jumlah prediksi yang benar, persisi emlihat ketepatan prediksi kelas positif, dan recall menghitung kemampuan model dalam mendeteksi kelas positif. Perhitungan manual dilakukan untuk memastikan hasil pengujian sesuai dengan nilai yang dihasilkan oleh modilm sehingga hasil dapat diandalkan

HASIL DAN PEMBAHASAN

Analisis Data

Data yang digunakan merupakan dataset yang ada di Kaggle yang datanya sudah di cleaning, mencakup data merek, RAM, prosesor, tipe penyimpanan, ukuran layar, dan klasifikasi harga dengan dua kelas Low = RP 0 - RP 10,000,000, High = RP 10,000,001 – Keatas.

Tabel 1. Data Penjualan Laptop

No.	Merek	RAM	Prosesor	Tipe Penyimpanan	Ukuran Layar(Inci)	Harga klasifikasi
1	Apple	8 GB	Core i5	SSD	13,3	High
2	Apple	8 GB	Core i5	HDD	13,3	High
3	HP	8 GB	Core i5 7200U	SSD	15,6	Low
4	Apple	8 GB	Core i5	SSD	13,3	High
5	Acer	4 GB	A9-Series 9420	HDD	15,6	Low
6	Apple	8 GB	Core i5	HDD	13,3	High
7	Asus	16 GB	Core i7 8550U	SSD	14	High
8	Acer	8 GB	Core i5 8250U	SSD	14	High
9	HP	4 GB	Core i5 7200U	HDD	15,6	Low
10	HP	4 GB	Core i3 6006U	HDD	15,6	Low
11	Dell	4 GB	Core i3 6006U	SSD	15,6	Low
12	Apple	8 GB	Core i5	SSD	13,3	High
13	Dell	8 GB	Core i7 7500U	SSD	15,6	High
14	Lenovo	8 GB	Core i3 7100U	HDD	15,6	Low
15	Dell	8 GB	Core i5 8250U	SSD	13,3	High
16	Lenovo	8 GB	Core i5 7300HQ	SSD	15,6	High
17	HP	4 GB	E-Series E2-9000e	HDD	15,6	Low
18	Dell	8 GB	Core i5 8250U	SSD	13,3	High
19	HP	8 GB	Core i7 8550U	SSD	15,6	High
20	Apple	8 GB	Core i5	SSD	13,3	High
...
969	Lenovo	4 GB	Core i7 6500U	SSD	14	High
970	Lenovo	16 GB	Core i7 6500U	SSD	13,3	High
971	Lenovo	2 GB	Celeron Dual Core N3050	HDD	14	Low
972	HP	6 GB	Core i7 6500U	HDD	15,6	High

Data penjualan kemudian dikelompokkan menjadi enam kategori, yaitu klasifikasi berdasarkan merek, RAM, prosesor, tipe penyimpanan, ukuran, dan klasifikasi harga.



Tabel 2. Klasifikasi Merek

Merek	Jml Penjualan
Dell	251
Lenovo	244
HP	211
Asus	100
Acer	83
MSI	26
Toshiba	47
Apple	10

Tabel 3. Klasifikasi RAM

RAM	Jml Penjualan
2 GB	29
4 GB	302
6 GB	143
8 GB	498
12 GB	18
16 GB	116
24 GB	2
32GB	3

Tabel 4. Klasifikasi Prosesor

Prosesor	Jml Penjualan
Core I3	120
Core I5	363
Core I7	363
A-Series	39
Celeron	48
Core M	3
E-Series	9
FX	2
Pentium	23
Xeon	4

Tabel 5. Klasifikasi Tipe Penyimpanan

Tipe Penyimpanan	Jml Penjualan
HDD	360
SSD	835

Tabel 6. Klasifikasi Ukuran Layar

Ukuran Layar (Inci)	Jml Penjualan
13	152
14	182
15,6	638

Tabel 7. Klasifikasi Harga

Harga	Klasifikasi Harga	Jml Penjualan
Rp 0 - Rp 10,000,000	Low	256
Rp 10,000,001 - Keatas	High	716

Perhitungan Algoritma C4.5

Perhitungan pada data menggunakan algoritma C4.5. Proses pohon keputusan meliputi mengubah data menjadi model pohon, mengonversi model pohon menjadi aturan (rule), dan menyederhanakan aturan tersebut. Data penjualan pada akan digunakan untuk membentuk pohon keputusan yang menentukan status pada harga laptop low dan high. Berikut adalah hasil penghitungan dalam pembentukan pohon keputusan dengan algoritma C4.5:

Tabel 8. Pehitungan Node

Node	JML	Klasifikasi Harga		Entropy	Nilai Gain
		Low	High		
Total	972	256	290	1,02754526	
Merek					0,225636236
Dell	251	42	209	0,65156192	
Lenovo	244	75	169	0,89012247	
HP	211	54	157	0,82053339	
Asus	100	36	64	0,94268319	
Acer	83	44	39	0,99738066	
MSI	26	0	26	0	
Toshiba	47	23	19	1,03276987	
Apple	10	0	10	0	
RAM					0,436676389
2 GB	8	8	0	0	
4 GB	300	190	110	0,94807824	
6 GB	27	17	10	0,95095605	
8 GB	498	41	457	0,41033397	
12 GB	18	0	18	0	
16 GB	116	0	116	0	
24 GB	2	0	2	0	
32GB	3	0	3	0	
Prosesor					0,498280233
Core I3	120	90	30	0,81127812	
Core I5	483	142	341	0,87382717	



Core I7	363	3	360	0,06905438
A-Series	38	32	6	0,62924922
Celeron	48	48	0	0
Core M	3	0	3	0
E-Series	9	9	0	0
FX	2	0	2	0
Pentium	23	22	1	0,25801867
Xeon	3	0	3	0
Tipe Penyimpanan				0,356483556
HDD	337	186	151	0,99220521
SSD	635	70	565	0,50062806
Ukuran Layar				0,108332089
13	152	2	84	0,55504673
14	182	28	74	0,94334927
15,6	638	255	132	0,99908861

Hasil Confusion Matriks

accuracy: 87.63%

	true High	true Low	class precision
pred. High	128	9	93.43%
pred. Low	15	42	73.68%
class recall	89.51%	82.35%	

Gambar 2. Hasil Confusion Matriks

Hasil confusion matrix menunjukkan akurasi model sebesar 87.63%, dengan performa yang lebih baik dalam memprediksi kelas "High" dibandingkan "Low". Untuk kelas "High", precision mencapai 93.43% dengan recall 89.51%, menunjukkan bahwa model sangat baik dalam mengenali data "High" dengan sedikit kesalahan. Namun, untuk kelas "Low", precision lebih rendah, yaitu 73.68%, dengan recall 82.35%, menandakan model kurang akurat dalam memprediksi kelas ini dibandingkan kelas "High". Hal ini menunjukkan model cenderung lebih optimal dalam menangani data kelas "High".

Nilai Atribut Tertinggi Menggunakan Rapidminer

attribute	weight
Ukuran Layar	0.269
Prosesor	0.291
Tipe Penyimpanan	0.108
Merek	0.184
RAM	0.147

Gambar 3. Nilai Atribut Tertinggi Menggunakan Rapidminer

Bobot atribut dalam menentukan harga laptop. Atribut yang paling berpengaruh adalah Prosesor (0.291), diikuti oleh Ukuran Layar (0.269), Merek (0.184), dan RAM (0.147). Tipe Penyimpanan (0.108) memiliki pengaruh terendah. Hal ini menandakan bahwa spesifikasi inti seperti prosesor dan ukuran layar lebih dominan dalam menentukan harga dibandingkan atribut lainnya

KESIMPULAN

Berdasarkan hasil metode decision tree, bahwa salah satu faktor utama yang mempengaruhi harga laptop adalah Prosesor dengan nilai gain sebesar 0,4983 Model ini juga memiliki akurasi yang baik, yaitu 87.63%, dengan kinerja terbaik pada kelas harga Low (presisi 73.68%), High(presisi 93.43%) dan kemampuan tinggi dalam mengidentifikasi



kelas "High" (recall 89.51%). Hasil ini menunjukkan bahwa laptop memiliki beberapa segmen pasar, pengguna dan kualitas produk.

DAFTAR PUSTAKA

- Airi, F. A. H., Suprapti, T., & Bahtiar, A. (2023). Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke. *E-Link: Jurnal Teknik Elektro Dan Informatika*, 18(1), 73. <https://doi.org/10.30587/e-link.v18i1.5271>
- Assauqi, M. F. R., & Fatah, Z. (2024). PREDIKSI RISIKO DEMAM BERDARAH MENGGUNAKAN DECISION TREE BERDASARKAN GEJALA KLINIS DAN DATA LABORATORIUM. *JURNAL ILMIAH SAINS TEKNOLOGI DAN INFORMASI*, 2(4), 42-53.
- Erliani, N., Suryowati, K., & Jatipaningrum, M. T. (2023). Klasifikasi Tingkat Penjualan Laptop Di E-Commerce Menggunakan Algoritma Classification and Regression Tree (Cart). *Jurnal Statistika Industri Dan Komputasi*, 8(2), 40-47. <https://doi.org/10.34151/statistika.v8i2.4455>
- Jannah, E., Sihombing, V., & Masrizal, M. (2023). Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5. *MEANS (Media Informasi Analisa Dan Sistem)*, 8(1), 1-7. <https://doi.org/10.54367/means.v8i1.2569>
- Jayen, F., Kadir, A., Amrulloh, R., & Hadi, A. (2022). Pengaruh Kualitas Produk dan Harga Terhadap Keputusan Pembelian Laptop Lenovo Pada Mahasiswa Sekolah Tinggi Ilmu Ekonomi Pancasetia Banjarmasin. *Jurnal Aplikasi Pelayaran Dan Kepelabuhanan*, 13(1), 110-121. <https://doi.org/10.30649/japk.v13i1.87>
- Khotimah, B. K., Syarief, M., Miswanto, M., & Suprajitno, H. (2021). Optimasi Bobot K-Means Clustering untuk Mengatasi Missing Value dengan Menggunakan Algoritma Genetica. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(4), 745. <https://doi.org/10.25126/jtiik.2021844912>
- Hasibuan, T. H., & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4. 5 Pada Uin Syarif Hidayatullah Jakarta. *SKANIKA: Sistem Komputer dan Teknik Informatika*, 6(1), 61-74.
- Permana, A. P., Ainiyah, K., & Holle, K. F. H. (2021). Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 6(3), 178-188. <https://doi.org/10.14421/jiska.2021.6.3.178-188>
- Sekar Setyaningtyas, Indarmawan Nugroho, B., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52-61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- SLN, F. (2023). *Buku Dasar Data Mining from A to Z*. January, 15-15.
- Syahputra, A. K. (2021). Search Engine & Web Browser. *JUSIM (Jurnal Sistem Informasi Musirawas)*, 2, 1-23. https://www.academia.edu/38517150/Internet_and_Website_Search_Engine_dan_Web_Browser
- Uddin, M. B., & Fatah, Z. (2024). *Gudang Jurnal Multidisiplin Ilmu Penerapan Data Mining Clustering K-Means Dalam Mengelompokkan Data Penduduk Penyandang Disabilitas*. 2(November), 86-94.