

PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PENJUALAN LAPTOP

Muhammad Aqil Sadik¹, Ratih Friska Dwi Andini²

^{1,2,3}Pendidikan Teknik Informatika dan Komputer, Fakultas Keguruan dan Ilmu Pendidikan,
Universitas Sebelas Maret

Email: ¹aqilsadik12@student.uns.ac.id, ²ratihfriska@student.uns.ac.id

Abstract.

Laptops are electronic devices that are very often used today, both children and even adults can use laptops for aspects of work in offices, homes, schools and industry. With the development of current technology, the types and brands of laptops are becoming more and more. There are various types of laptops with different prices and specifications to choose from. Therefore the seller must be able to distinguish which laptop has a lot of demand, so it is necessary to recommend the type of laptop using the data mining classification method C 4.5 algorithm. The data needed is price, specifications, color, and quality. C4.5 is a standard classification method based on decision trees. The C4.5 algorithm is built in several stages which include selecting the attribute as the root, creating a branch for each value and dividing the cases into branches.

Keywords: Laptop, data mining, Algorithm C4.5

Abstrak.

Laptop merupakan alat elektronik yang sangat sering digunakan saat ini, baik dari kalangan anak-anak bahkan dewasa penggunaan laptop bisa untuk aspek pekerjaan di perkantoran, rumah, sekolah maupun industri dengan berkembangnya teknologi saat ini jenis dan merk laptop menjadi semakin banyak . bermacam-macam jenis laptop dengan harga dan spesifikasi yang berbeda yang dapat dipilih. Oleh karena itu penjual harus bisa membedakan laptop mana yang memiliki banyak peminatnya, sehingga perlu perekomendasi jenis laptop dengan menggunakan data mining metode klasifikasi algoritma C 4.5 data yang diperlukan adalah harga, spesifikasi, warna, dan kualitas. C4.5 adalah metode klasifikasi standar berdasarkan pohon keputusan pohon keputusan. Algoritma C4.5 dibangun dengan beberapa tahap yang meliputi pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai dan membagi kasus dalam cabang. Tahapan-tahapan ini akan diulangi untuk setiap cabang sampai semua kasus pada cabang memiliki

kelas yang sama. Dari penyelesaian pohon keputusan maka nantinya akan didapatkan beberapa rules

Kata Kunci: *laptop, data mining, prediksi, Algoritma C 4.5*

1. Pendahuluan

Perkembangan globalisasi saat ini sangat mempengaruhi berbagai bidang, salah satunya yaitu teknologi. Laptop merupakan barang elektronik yang banyak digunakan, karena merupakan sebuah komputer yang dapat dibawa kemana saja serta memiliki ukuran yang fleksibel dan juga enteng. Fungsi laptop mirip dengan komputer desktop pada umumnya yaitu memudahkan pekerjaan dan tugas yang sedang dikerjakan. Toko Sehati Solo beralamat di Gonilan no. 1 adalah toko yang menyediakan bermacam-macam merk laptop dengan berbagai kriteria.

Penjualan laptop juga sudah meningkat sebanyak rata-rata 1,4% dari tahun 2018 dilansir oleh lembaga riset Gartner setelah sebelumnya menurun selama 6 tahun terakhir dikarenakan banyaknya jenis merk dan spesifikasi laptop sehingga konsumen bingung dalam memilih laptop sesuai dengan kebutuhan, dari permasalahan yang ada ini dibutuhkan riset yang dapat membantu para konsumen untuk memberikan rekomendasi terhadap laptop yang ingin digunakan.

Toko Sehati Solo termasuk salah satu toko yang ramai dikunjungi oleh banyak orang yang ingin membeli atau sekedar bertanya karena toko ini sangat lengkap. Perkembangan teknologi komputer yang selalu berkembang dalam usaha memenuhi kebutuhan manusia di bidang komunikasi, informasi dan pengolahan data. Sebagai contoh penjualan laptop yang berkembang dan tumbuh sangat cepat. Dengan semakin meningkatnya konsumen yang membutuhkan laptop, maka semakin banyak pula permasalahan yang timbul di dalam memberikan pelayanannya kepada pembeli. Sistem persediaan laptop harus disiapkan dengan maksimal agar penjual dapat memenuhi keinginan pembeli.

Untuk mempermudah penjual memilih laptop mana yang banyak diminati pembeli agar disediakan stok untuk laptop tersebut, maka perlu diprediksi untuk penjualan laptop terbanyak dengan metode klasifikasi menggunakan metode data mining algoritma C4.5.

2. Tinjauan Pustaka

2.1 Data Mining

Menurut Hermawati (2013:3) data mining merupakan suatu proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstrak pengetahuan (*knowledge*) secara otomatis. Data mining merupakan salah satu langkah dari proses KDD (*Knowledge Discovery in Databases*) (Jules, G., 2001)

(1) teknik dan sifat data mining : (a) Classification, (b)Clustering, (c) Association Rule Discovery, (d) Regression, (e) Deviation Detection.

(2) tahapan data mining (a) pembersihan data, (b)integrasi data, (c)seleksi data, (d) transformasi data, (e) proses mining, (f) evaluasi pola, (g) representasi pengetahuan.

Data mining adalah sebuah proses, jadi prosesnya harus mengikuti prosedur yaitu proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*) secara keseluruhan, pra-pemrosesan data, pembangunan model, evaluasi model dan akhirnya model. penerapan.

2.2 Algoritma C4.5

Algoritma C4.5 adalah metode algoritma yang digunakan untuk membangun suatu pohon keputusan. Pohon keputusan adalah metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Bisa juga diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu.(Kusrini, 2018)

Cara algoritma C4.5 untuk membangun pohon keputusan yaitu: (a) Pilih atribut yang akan digunakan sebagai akar (b) Buatlah sebuah cabang untuk setiap nilai (c) Bagilah kasus dalam sebuah cabang (d) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Ada beberapa tahap dalam membangun pohon keputusan dengan algoritma C4.5(Rahmadya, 2013)

1. Siapkan data training. Data ini diambil dari data yang sudah pernah ada sebelumnya dan sudah dibagi dalam kelas-kelas tertentu. 2. Kemudian tentukan akar pohon tersebut. Pilih akar dari atribut, metode ini menghitung nilai gain dari semua atribut,yang menjadi akar pertama adalah nilai gain yang paling. Sebelum menentukan nilai gain, terlebih dahulu hitung nilai entropy.

Pendekatan C4. 5 merupakan bentuk perluasan dari algoritma ID3 yang diusulkan oleh Quinlan, J. R. (1993) dimana kriteria *gain ratio* telah diadopsi untuk klasifikasi pohon keputusan atau *decision tree classification* alih-alih kriteria pembagian gain informasi. *Decision tree* dibangun dengan menerapkan strategi *Divide* dan *Conquer*, termasuk data yang kontinyu dan dapat menangani data yang hilang. Fungsi gain informasi dan entropi diformulasikan sebagai:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{j=1}^v \frac{|D_j|}{|D|} * \text{Info}(D_j)$$

$$\text{Entropy}(S) = \sum_{i=1}^c -P_i \log_2 P_i$$

Ket:

D : partisi dataset

Entropy (S) : subset sebelum pemisahan

v : nilai dataset
 P_i : proporsi dari nilai target ke-i dalam subset

2.3 Decision Tree

Pohon (tree) adalah sebuah struktur data yang terdiri dari *node* dan *edge*. Node dalam pohon dibagi menjadi tiga bagian: node akar, node cabang/internal, dan node daun.

Mirip dengan pohon, *decision tree* dihubungkan oleh cabang-cabang yang bergerak dari atas ke bawah mulai dari simpul akar dan berakhir di simpul daun. Node root ditempatkan di bagian atas pohon keputusan dan semua atribut node dievaluasi dan dapat bercabang. Setiap cabang dapat memasuki simpul lain atau dipindahkan ke simpul daun. Setiap node mewakili atribut, cabangnya mewakili nilai atribut, dan node daun mewakili kelas.

Pohon keputusan adalah representasi sederhana dari teknik klasifikasi untuk jumlah kelas yang terbatas. Keduanya diberi label dengan atribut. Tepi diberi label dengan kemungkinan nilai atribut dan simpul daun diberi label dengan kelas yang berbeda. (Hermawati. F. Astuti, 2013).

3. Metode Penelitian

3.1 Menentukan Objek Observasi

kegiatan observasi ini dilakukan dengan tujuan untuk memprediksi penjualan laptop. metode ini dilakukan sebagai tolak ukur penjual dalam mempersiapkan stok laptop. Sehingga penjual kedepannya dapat menyediakan laptop yang banyak diminati agar selalu ada stok.

3.2 Pengumpulan Data

Data yang digunakan sebagai penelitian adalah data sampel yang diambil dari toko Sehati Solo

No.	Harga	Spesifikasi	Ukuran	Berat	Warna	Minat
1	Murah	Sedang	Sedang	Ringan	Hitam	Banyak
2	Murah	Sedang	Besar	Berat	Hitam	Banyak
3	Murah	Rendah	Besar	Berat	Silver	Banyak
4	Mahal	Sedang	Besar	Berat	Hitam	Sedikit
5	Murah	Rendah	Besar	Berat	Hitam	Banyak
6	Mahal	Sedang	Sedang	Ringan	Hitam	Banyak
7	Mahal	Tinggi	Besar	Berat	Hitam	Banyak
8	Mahal	Sedang	Besar	Ringan	Hitam	Banyak
9	Mahal	Tinggi	Besar	Ringan	Silver	Banyak
10	Mahal	Tinggi	Sedang	Berat	Cream	Sedikit
11	Murah	Sedang	Sedang	Ringan	Hitam	Banyak

No.	Harga	Spesifikasi	Ukuran	Berat	Warna	Minat
12	Mahal	Tinggi	Sedang	Ringan	Silver	Banyak
13	Mahal	Tinggi	Sedang	Ringan	Cream	Banyak
14	Mahal	Sedang	Besar	Ringan	Silver	Banyak
15	Mahal	Tinggi	Sedang	Berat	Hitam	Sedikit
16	Murah	Rendah	Sedang	Ringan	Hitam	Banyak
17	Murah	Sedang	Besar	Berat	Silver	Banyak
18	Mahal	Tinggi	Besar	Berat	Silver	Banyak
19	Murah	Sedang	Sedang	Berat	Hitam	Sedikit
20	Murah	Rendah	Sedang	Berat	Silver	Banyak
21	Murah	Rendah	Besar	Berat	Silver	Sedikit
22	Mahal	Tinggi	Besar	Ringan	Hitam	Banyak
23	Murah	Sedang	Sedang	Ringan	Silver	Banyak
24	Mahal	Sedang	Sedang	Ringan	Silver	Banyak
25	Mahal	Sedang	Besar	Ringan	Cream	Sedikit
26	Murah	Rendah	Sedang	Berat	Cream	Sedikit
27	Mahal	Tinggi	Sedang	Ringan	Hitam	Banyak
28	Murah	Rendah	Besar	Berat	Cream	Sedikit
29	Mahal	Sedang	Sedang	Ringan	Cream	Banyak
30	Mahal	Tinggi	Besar	Berat	Cream	Banyak

Tabel.1 Data Laptop

Pada tabel 1 terdapat 30 penilaian dari 6 penilaian yang menuntut minat konsumen terhadap laptop yang ingin dibeli.

4. Hasil dan Pembahasan

Node			Jumlah	Sedikit	Banyak	Entropy	Gain
1	Total		30	8	22	0,8366	
	Harga						0
		Mahal	17	4	13	0,7871	
		Murah	13	4	9	0,8905	
	Spesifikasi						0,03
		Tinggi	10	2	8	0,7219	
		Sedan g	13	3	10	0,7793	

		Rendah	7	3	4	0,9852	
	Ukuran						0
		Besar	15	4	11	0,8366	
		Sedang	15	4	11	0,8366	
	Berat						0,66
		Berat	15	7	8	0,9968	
		Ringan	15	1	14	0,3534	
	Warna						0,11
		Hitam	13	3	10	0,7793	
		Silver	10	1	9	0,469	
		Cream	7	4	3	0,9852	

Tabel 2. Perhitungan entropy dan gain

pada tabel 2 dapat diketahui bahwa gain tertinggi adalah berat sebesar 0,66. Tetapi, berat dapat menjadi node akar karena memiliki 2 atribut yaitu berat dan ringan sehingga diperlukan perhitungan kembali.

Berat	Harga	Spesifikasi	Ukuran	Warna	Minat
Berat	Murah	Rendah	Besar	Silver	Banyak
Berat	Murah	Rendah	Besar	Silver	Sedikit
Berat	Murah	Sedang	Besar	Silver	Banyak
Berat	Mahal	Tinggi	Besar	Silver	Banyak
Berat	Murah	Rendah	Sedang	Silver	Banyak
Berat	Murah	Rendah	Besar	Hitam	Banyak
Berat	Mahal	Sedang	Besar	Hitam	Sedikit
Berat	Murah	Sedang	Besar	Hitam	Banyak
Berat	Mahal	Tinggi	Besar	Hitam	Banyak
Berat	Murah	Sedang	Sedang	Hitam	Sedikit
Berat	Mahal	Tinggi	Sedang	Hitam	Sedikit
Berat	Murah	Rendah	Besar	Cream	Sedikit
Berat	Mahal	Tinggi	Besar	Cream	Banyak
Berat	Murah	Rendah	Sedang	Cream	Sedikit
Berat	Mahal	Tinggi	Sedang	Cream	Sedikit

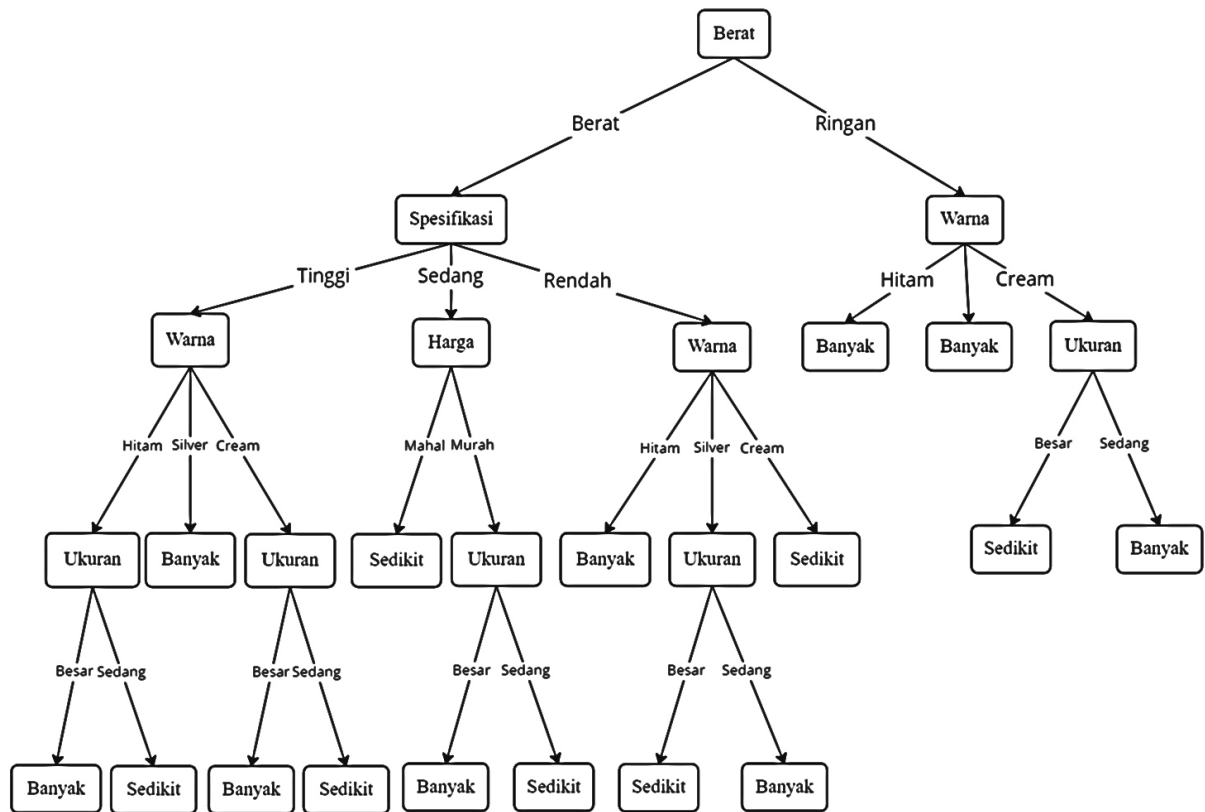
Tabel 3. Data dari atribut "berat"

Nod e			Jumla h	Sediki t	Banya k	Entrop y	Gai n
1.1	Total		15	7	8	0,9968	
	Harga						0,19
		Mahal	6	1	5	0,65	
		Murah	9	6	3	0,9183	
	Spesifikas i						0,23
		Tinggi	5	4	1	0,7219	
		Sedang	4	2	2	1	
		Renda h	6	1	5	0,65	
	Ukuran						0,17
		Besar	10	3	7	0,8813	
		Sedang	5	4	1	0,7219	
	Warna						0,14
		Hitam	6	3	3	1	
		Silver	5	4	1	0,7219	
		Cream	4	3	1	0,8113	

Tabel 4. Hasil perhitungan atribut "berat"

pada tabel 4 dapat diketahui bahwa gain tertinggi adalah spesifikasi sebesar 0,23. Namun di dalam atribut spesifikasi memiliki 3 atribut lagi yang diperlukan perhitungan kembali. Begitu Pula dengan tabel-tabel selanjutnya yang diperlukan penghitungan untuk atribut dengan nilai gain tertinggi dan atribut yang masih memiliki atribut lagi di dalamnya.

Dari perhitungan diatas dapat dibuat sebuah *decision tree* sebagai berikut



4.1 Mengubah tree menjadi rules

R1 : IF berat = berat ^ spesifikasi = tinggi ^ warna = hitam ^ ukuran = besar

THEN minat = banyak

R2: IF berat = berat ^ spesifikasi = tinggi ^ warna = hitam ^ ukuran = sedang

THEN minat = sedikit

R3: IF berat = berat ^ spesifikasi = tinggi ^ warna = silver

THEN minat = banyak

R4: IF berat = berat ^ spesifikasi = tinggi ^ warna = cream ^ ukuran = besar

THEN minat = banyak

R5: IF berat = berat ^ spesifikasi = tinggi ^ warna = cream ^ ukuran = sedang

THEN minat = sedikit

R6: IF berat = berat ^ spesifikasi = sedang ^ harga = mahal

THEN minat = sedikit

R7: IF berat = berat ^ spesifikasi = sedang ^ harga = murah ^ ukuran = besar

THEN minat = banyak

R8: IF berat = berat ^ spesifikasi = sedang ^ harga = murah ^ ukuran = sedang

THEN minat = sedikit

R9 : IF berat = berat ^ spesifikasi = rendah ^ warna = hitam

THEN minat = banyak

R10: IF berat = berat ^ spesifikasi = rendah ^ warna = silver ^ ukuran = besar

THEN minat = sedikit

R11: IF berat = berat ^ spesifikasi = rendah ^ warna = silver ^ ukuran = sedang

THEN minat = banyak

R12: IF berat = berat ^ spesifikasi = rendah ^ warna = cream

THEN minat = sedikit

R13: IF berat = ringan ^ warna = hitam

THEN minat = banyak

R14: IF berat = ringan ^ warna = silver

THEN minat = banyak

R15: IF berat = ringan ^ warna = cream ^ ukuran = besar

THEN minat = sedikit

R16: IF berat = ringan ^ warna = cream ^ ukuran = sedang

THEN minat = banyak

5. Kesimpulan

Berdasarkan penelitian yang sudah dilakukan, maka dapat disimpulkan pembelian laptop menggunakan metode data mining algoritma C4.5 bermanfaat dalam proses untuk mengambil keputusan untuk menentukan sebuah laptop dengan kriteria tertentu memiliki peminat yang banyak atau sedikit. Hasil yang diperoleh dapat diimplementasikan untuk menyusun strategi kriteria-kriteria laptop yang akan diperjualbelikan.