

IMPLEMENTASI DATA MINING MENGGUNAKAN METODE DECISION TREE DALAM MEMPOLAKAN PENJUALAN PADA SHOWROOM MOTOR BEKAS

Shevti Arbekti Arman^{1*}, Rahayu Trisetyowati Untari², Erion³

Institut Teknologi dan Bisnis Ahmad Dahlan, Jakarta

e-mail: ¹shevtiarbekti@gmail.com

Abstract: *Basoka Motor in marketing activities is still done manually, has not been computerized and has not planned motorcycle sales using information technology. Basoka Motor needs to make sales data computerized and see sales patterns in the used motorcycle showroom. Sales pattern information is obtained using the classification method. The use of the decision tree method with the C.45 algorithm can categorize used motorcycle sales at Basoka Motor. The purpose of this research is for the manager to know the sales pattern of used motorcycles in one month as a reference for the next month. Processing motorcycle data using the C4.5 algorithm is feasible to use in predicting motorcycle sales because it produces 100% accuracy.*

Keywords: *C4.5 algorithm; decision tree; prediction; accuracy*

Abstrak: Basoka Motor dalam kegiatan pemasaran masih dilakukan secara manual, belum terkomputerisasi serta belum mempolakan penjualan motor menggunakan teknologi informasi. Basoka Motor perlu membuat data penjualan dapat terkomputerisasi dan melihat pola penjualan pada showroom motor bekas tersebut. Informasi pola penjualan didapatkan dengan menggunakan metode klasifikasi. Penggunaan metode decision tree dengan algoritma C.45 dapat mempolakan penjualan motor bekas pada Basoka Motor. Tujuan dari penelitian ini adalah agar manajer dapat mengetahui pola penjualan motor bekas dalam satu bulan sebagai acuan untuk bulan berikutnya. Penelitian ini menghasilkan 27 data motor sesuai dengan rules prediksi setelah dilakukan pencocokan rule prediksi dengan dataset Pengolahan data motor menggunakan algoritma C4.5 layak digunakan dalam mempolakan penjualan motor karena menghasilkan akurasi sebesar 100%.

Kata kunci: algoritma C4.5; decision tree; prediksi; akurasi.

PENDAHULUAN

Informasi merupakan bentuk data yang sudah melalui proses pengolahan sehingga memiliki arti bagi penerima serta dapat berupa fakta yang bermanfaat. Jadi data menjadi sebuah informasi sudah melalui suatu proses transformasi. Perkembangan teknologi komputer yang selalu memenuhi kebutuhan manusia di berbagai bidang salah satunya. Kebutuhan manusia di berbagai bidang salah satunya di bidang informasi dan pengolahan data dapat terpenuhi dikarenakan perkembangan teknologi yang sangat pesat (Eska, 2016). Dahulunya data

hanya tertumpuk dan tidak diolah lebih lanjut. Akan tetapi saat sekarang data telah diolah sehingga terdapat sebuah informasi baru. Dalam pengolahan data dibutuhkan sistem terkomputerisasi menggunakan database.

Sistem terkomputerisasi saat ini digunakan pada beberapa perusahaan untuk mendapatkan informasi penting mengenai kegiatan pemasaran. Salah satu pencarian informasi dari kumpulan data besar adalah menggunakan Knowledge Discovery in Database (KDD).

Data mining merupakan tahapan dalam KDD, kegiatannya yaitu mengumpulkan data dan memakai histori

data untuk agar ditemukan hubungan set data yang berukuran besar. (Maulana, Winanjaya, & Rizki, 2022). Basoka Motor belum menerapkan sistem tersebut. Saat ini Basoka Motor dalam kegiatan pemasarannya masih melakukannya secara manual dan belum mempolakan penjualan motor menggunakan teknologi informasi.

Tujuan dari penelitian ini adalah agar manajer dapat mengetahui pola penjualan motor bekas dalam satu bulan sebagai acuan untuk bulan berikutnya. Perusahaan Basoka Motor memerlukan agar data penjualan dapat terkomputerisasi dan melihat pola penjualan pada showroom motor bekas tersebut. Informasi pola penjualan didapatkan dengan menggunakan metode klasifikasi.

Klasifikasi diterapkan pertama kali di bidang tanaman yang dilakukan Carolus von Linne dalam melakukan klasifikasi spesies berdasarkan karakteristik fisiknya. Pohon keputusan merupakan metode yang bisa mengubah fakta yang besar menjadi bentuk pohon keputusan yang merepresentasikan aturan. Pohon keputusan merupakan salah satu metode-metode atau model-model yang telah dikembangkan dalam menyelesaikan kasus klasifikasi (Mardi, 2107).

C4.5 merupakan algoritma yang dapat digunakan dalam membentuk pohon keputusan untuk klasifikasi dan prediksi (Wahyu, Farazi, Mahendra, & Hapsari, 2023). Penggunaan algoritma C4.5 dengan aplikasi rapidminer akan menghasilkan pohon keputusan yang dapat digunakan anggota subbidang aset dalam memprediksi peralatan dan mesin kantor yang nantinya akan digunakan (Amalia & Resad, 2023).

Algoritma C4.5 menghasilkan decision tree dan terdapatnya rules prediksi, tingkat akurasi yang dihasilkan adalah sebesar 70% dengan menggunakan pengujian Confusion Matrix dalam memprediksi bimbingan (Nugraha, Defit, & Nurcahyo, 2023)

METODE

Algoritma C4.5

Algoritma C4.5 menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi (Prasetyo, 2014). Membangun pohon keputusan ada beberapa langkah yang harus dilakukan, yaitu: 1) Menentukan akar, 2) Membuat cabang berdasarkan kasus-kasus, 3) Mengulangi kembali proses pada setiap cabang sampai setiap cabang memiliki kelas yang sama (Kusrini & Emha, 2009).

Entropy

Entropy digunakan sebagai penentu node mana yang akan menjadi pemecah data latih berikutnya. Nilai yang lebih tinggi pada entropy dapat meningkatkan potensi klasifikasi. Jika entropy untuk node bernilai 0, maka semua data vektor yang berada pada label kelas sama dan node tersebut menjadi daun yang berisi keputusan (label kelas).

Jika salah satu dari elemen W_i jumlahnya 0 maka entropi dipastikan 0 juga. Jika proporsi semua elemen W_i sama jumlahnya maka dipastikan entropi bernilai 1 (Prasetyo, 2014).

Perhitungan nilai entropy menggunakan persamaan:

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

Ket:

S : Himpunan kasus

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

Dimana $\log_2 p_i$ dapat dihitung menggunakan persamaan:

$$\log(X) = \frac{\ln(X)}{\ln(2)}$$

Gain Ratio

Gain ratio merupakan kriteria yang paling banyak digunakan untuk memilih fitur sebagai pemecah dalam algoritma

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

Persamaan menghitung gain (Kusrini & Emha, 2009):

Ket:

S : himpunan kasus

A : atribut

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

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

$$Split_{info}(S) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \frac{|S_i|}{|S|}$$

Persamaan menghitung SplitEntropy:

Ket:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

Confussion Matrix

Nilai akurasi dihitung menggunakan *confussion matrix*. Terdapat empat istilah dari hasil klasifikasi untuk pengukuran kinerja menggunakan *confussion matrix*:

1. *False Positive* (FP): data negatif yang terprediksi sebagai data positif.
2. *False Negative* (FN): data positif yang terprediksi sebagai data negatif.
3. *True Positive* (TP): data positif yang terprediksi benar.
4. *True Negative* (TN): data negatif yang terprediksi dengan benar

Persamaan menghitung *confussion*

$$Akurasi = \frac{TP + FN}{TP + FN + FP + TN} \times 100\%$$

HASIL DAN PEMBAHASAN

Perhitungan Algoritma C4.5

Perhitungan manual menggunakan algoritma C4.5 dalam mempolakan penjualan pada showroom motor bekas.

Tabel 1. Data Motor

No	Pabrikan	Jenis	Tahun	Harga	Statu
1	Suzuki	Bebek	2020	12000000	Tidak Laris
2	Suzuki	Bebek	2019	11500000	Tidak Laris
3	Suzuki	Bebek	2020	9100000	Laris
4	Yamaha	Bebek	2020	10500000	Laris
5	Honda	Matic	2020	10600000	Laris
6	Honda	Matic	2019	10500000	Laris
7	Honda	Matic	2019	11000000	Laris
8	Yamaha	Bebek	2020	12000000	Tidak Laris
9	Honda	Matic	2020	11000000	Laris
10	Yamaha	Matic	2020	10500000	Laris
11	Yamaha	Matic	2019	11000000	Laris
12	Yamaha	Bebek	2019	9200000	Laris
13	Suzuki	Matic	2020	9900000	Laris
14	Yamaha	Bebek	2019	11000000	Tidak Laris
15	Suzuki	Bebek	2020	11500000	Tidak Laris
16	Suzuki	Bebek	2019	11100000	Tidak Laris
17	Suzuki	Bebek	2020	9500000	Laris
18	Yamaha	Bebek	2020	10000000	Laris
19	Honda	Matic	2020	10200000	Laris
20	Honda	Matic	2019	10300000	Laris
21	Honda	Matic	2019	11200000	Laris
22	Yamaha	Bebek	2020	11500000	Tidak Laris
23	Honda	Matic	2020	11500000	Laris
24	Yamaha	Matic	2020	10500000	Laris
25	Yamaha	Matic	2020	12000000	Laris
26	Yamaha	Bebek	2019	9500000	Laris
27	Suzuki	Matic	2020	9800000	Laris

Berdasarkan permasalahan pada tabel 1, maka dibentuklah pohon keputusan menggunakan algoritma C4.5 menggunakan beberapa langkah, yaitu:

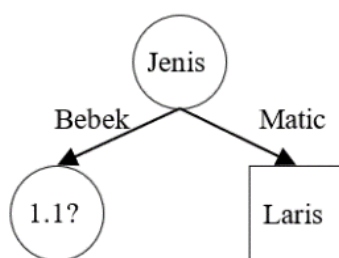
Menentukan akar

Menghitung jumlah kasus untuk keputusan laris dan tidak laris, entropy dari semua kasus dan kasus dibagi berdasarkan atribut jumlah barang, barang terjual, dan sisa barang. Kemudian lakukan penghitungan gain untuk setiap atribut.

Tabel 2. Perhitungan Node 1

NODE		Jml Kasus (S)	Tidak laris (S1)	Laris (S2)	Entropi	Gain
1	TOTAL	27	7	20	0.8256	
	Harga					0.3404
	A	7	0	7	0.0000	
	B	11	1	10	0.4394	
	C	9	6	3	0.9182	
	Pabrikan					0.1849
	Honda	8	0	8	0.0000	
	Suzuki	8	4	4	1.0000	
	Yamaha	11	3	8	0.8453	
	Jenis					0.3462
	Bebek	13	7	6	0.9957	
	Matic	14	0	14	0.0000	
	Tahun					0.0036
	2020	17	4	13	0.7871	
	2019	10	3	7	0.8812	

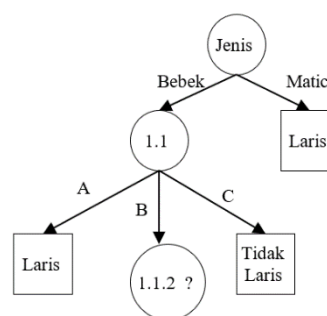
Dari hasil tabel 2. diketahui bahwa atribut dengan Gain tertinggi adalah jenis, yaitu 0.3462 sehingga harga menjadi node akar. Pohon keputusan dapat digambarkan

**Gambar 1. Pohon Keputusan Node 1****Tabel 3 . Perhitungan Node 1.1**

NODE		Jml Kasus (S)	Tidak laris (S1)	Laris (S2)	Entropy	Gain
1.1	Jenis-Bebek	13	7	6	0.9957	
	Harga					0.8418
	A	5	0	5	0.0000	
	B	2	1	1	1.0000	
	C	6	0	6	0.0000	
	Pabrikan					0.0413
	Honda	0	0	0	0.0000	
	Suzuki	6	4	2	0.9182	
	Yamaha	7	3	4	0.9852	
	Tahun					0.0069
	2020	8	4	4	1.0000	
	2019	5	3	2	0.9709	

Dari hasil tabel 3. diketahui bahwa atribut dengan Gain tertinggi adalah Harga, yaitu sebesar 0.8418. Maka harga

menjadi node cabang dari nilai atribut bebek. Ada tiga nilai atribut dari harga yaitu A, B, dan C. Dari ketiga nilai atribut tersebut, nilai atribut A sudah mengklasifikasikan kasus menjadi 1, yaitu keputusannya laris, dan nilai atribut C sudah mengklasifikasikan kasus menjadi tidak laris, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut B masih perlu dilakukan perhitungan lagi. Pohon keputusan dapat digambarkan sebagai berikut:

**Gambar 2. Pohon Keputusan Node 1.1**

Mengulangi kembali proses pada setiap cabang sampai setiap cabang memiliki kelas yang sama.

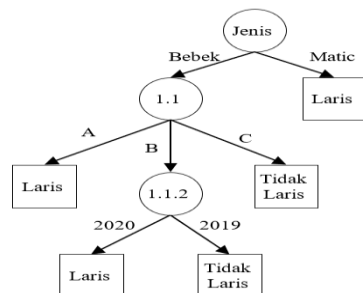
Menghitung jumlah kasus untuk keputusan laris dan tidak laris, entropy dari semua kasus dan kasus yang dibagi berdasarkan atribut pabrikan dan tahun yang dapat menjadi node cabang dari nilai atribut B. Kemudian, lakukan penghitungan gain tiap-tiap atribut.

Tabel 4. Perhitungan Node 1.1.2

NODE		Jml Kasus (S)	Tidak laris (S1)	Laris (S2)	Entropy	Gain
1.1.2	Jenis-Bebek dan Harga B	2	1	1	1.0000	
	Pabrikan					0
	Honda	0	0	0	0.0000	
	Suzuki	0	0	0	0.0000	
	Yamaha	2	1	1	1.0000	
	Tahun					1
	2020	1	0	1	0.0000	
	2019	1	1	0	0.0000	

Dari hasil tabel 4. diketahui bahwa atribut dengan gain tertinggi adalah tahun, yaitu 2019 dan 2020. Dari kedua nilai

tersebut, nilai atribut 2020 sudah mengklasifikasikan kasus menjadi satu yaitu keputusannya laris dan nilai atribut 2019 sudah mengklasifikasikan kasus menjadi satu dengan keputusan tidak laris. Sehingga tidak diperlukan.



Gambar 3. Pohon Keputusan Node 1.1.2

Perhitungan Confussion Matrix

Berdasarkan pohon keputusan maka didapatkanlah *rules* prediksi. Pencocokan rule dengan dataset akan menghasilkan nilai prediksi. Terdapat true positive jumlah prediksi positif yang benar terdapat 11 data, true negative jumlah prediksi negatif yang benar sebanyak 6 data, false positive jumlah prediksi positif yang salah sebanyak 1 data dan false negative yang salah sebanyak 2 data.

Tabel 5. Confussion Matrix

Outcome	Prediksi	
	Laris	Tidak laris
Laris	20	0
Tidak laris	0	7

Dari hasil pencocokan rule prediksi dengan dataset maka didapatkanlah hasil bahwa 27 data motor sesuai dengan *rules* prediksi. Pengolahan data motor menggunakan alogitima C4.5 layak digunakan dalam mempolakan penjualan motor karena menghasilkan akurasi sebesar 100%.

Perbandingan dengan menggunakan WEKA

Langkah-langkah dalam pengolahan data menggunakan WEKA:

Data yang diinputkan merupakan data yang formatnya .csv.

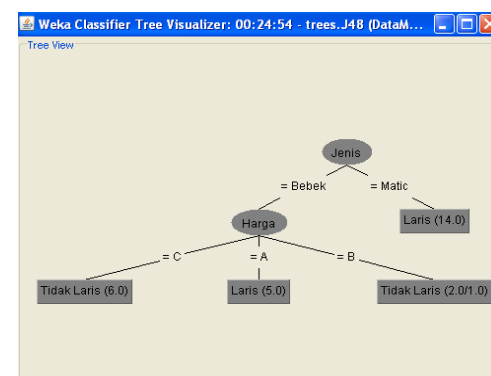
Gambar 4. Data Bulan Agustus 2021

Dari data yang didapat, peneliti melakukan penyeleksian data yang akan digunakan agar tidak menyimpang dari tujuan penulis.

Gambar 5. Data Yang Telah Diseleksi

Setelah data diseleksi, maka data disimpan dalam format “csv” (*comma delimited*), karena hanya data dalam format “csv” yang dapat dibaca pada program WEKA.

Setelah melakukan analisa, maka muncul pohon keputusan (*decision tree*).



Gambar 6. Hasil Output Dari Pengolahan Data

Berdasarkan pengujian secara manual maupun dengan menggunakan WEKA pada dasarnya sama, namun perbedaannya terlihat pada keputusan dari harga. Dimana keputusannya harga A,B dinyatakan laris sedangkan C dinyatakan tidak laris. Pada dasarnya diperlukan perhitungan lagi untuk harga B karena Ratio nya masih bisa dihitung.

SIMPULAN

Aplikasi data mining menggunakan metode decision tree dengan algoritma C.45 dapat mempolakan penjualan motor bekas pada Basoka Motor dengan. Aplikasi mampu menjadi media efektif dalam pengembangan Basoka Motor karena data telah terkomputerisasi sehingga manajer dapat mengetahui pola penjualan motor bekas dalam satu bulan sebagai acuan untuk bulan berikutnya. Dalam pengelolaan data, data dapat diinputkan melalui database dan aplikasi yang telah dirancang. Data menjadi lebih aman dengan perhitungan yang lebih akurat

DAFTAR PUSTAKA

- A. I., & R. S. (2023). Penerapan Algoritma C.45 Untuk Analisis Pengadaan Peralatan dan Mesin Kantor. *Journal of Information System Research* , 434-442.
- Aldy Gustiannur Rachmat, B. S. (2018). Perbandingan Metode SAW dan TOPSIS untuk penentuan Dosen Terbaik pada Jurusan Teknologi

- Informasi Politeknik Negeri Samarinda. Prosiding Seminar Ilmu Komputer dan Teknologi Informasi, (pp. 91-97). Samarinda.
- Eska, J. (2016). Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper menggunakan Algoritma C4.5. *Jurnal Teknologi dan Sistem Informasi*, 9-13.
- Kusrini & Emha, T. L. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Mardi, Y. (2107). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 213-219.
- Maulana, Y., Winanjaya, R., & Rizki, F. (2022). Penerapan Data Mining dengan Algoritma C4.5 Dalam Memprediksi Penjualan Tempe. *Bull.Comput.Sci.Res.*, 53-58.
- Nugraha, B. D., Defit, S., & Nurcahyo, G. W. (2023). Algoritma C4.5 Untuk Prediksi Bimbingan Siswa Berdasarkan Tipologi Hippocrates-Galenus. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 1-8.
- Prasetyo. (2014). *Data Mining : Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Wahyu, B. R., Farazi, A. F., Mahendra, C. P., & Hapsari, R. K. (2023). Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree Menggunakan Algoritma C4.5. *Journal of Information Technology*, 80-89.