



---

## PERANCANGAN APLIKASI DATA MINING UNTUK MEMPOLAKAN PENJUALAN PADA BASOKA MOTOR

Shevti Arbekti Arman<sup>1(\*)</sup>

<sup>1</sup>ITB Ahmad Dahlan, Jakarta

---

### *Abstract*

*Basoka Motor in its marketing activities is still done manually and has not yet idolized the sale of its motorbikes with the support of information technology. Designing an application can see the sales patterns that occur at Basoka Motor. To get information, one of the classification methods used in data mining is the decision tree method using the C.45 algorithm. Tests conducted between manual calculations and data mining applications on 27 motorcycle data obtained from Basoka Motor get the same results with an accuracy level of 100% using Confusion Matrix. Application development using the waterfall method with testing, namely Black Box Testing.*

**Kata Kunci:** C4.5, Confusion Matrix, Waterfall, Blackbox.

Januari – Juni 2023, Vol 4 (1) : hlm 39-50  
©2023 Institut Teknologi dan Bisnis Ahmad Dahlan.  
All rights reserved.

---

(\*) Korespondensi: [shevtiarbekti@gmail.com](mailto:shevtiarbekti@gmail.com) (Shevti Arbekti Arman)

## PENDAHULUAN

Informasi adalah data yang telah diproses menjadi bentuk yang memiliki arti untuk penerima dan dapat berupa fakta, suatu nilai yang bermanfaat. Jadi ada suatu proses transformasi data menjadi suatu informasi. Kemajuan teknologi yang pesat memungkinkan manusia untuk memenuhi kebutuhan mereka dalam berbagai bidang, termasuk informasi dan pengolahan data (Eska, 2016). Data dulunya hanyalah sesuatu data yang tertumpuk dan tidak diolah lebih lanjut. Namun saat ini data umumnya telah diolah sehingga mendapatkan suatu informasi baru. Untuk mengolah data tersebut dibutuhkan sebuah sistem terkomputerisasi dengan menggunakan database.

Saat ini sistem terkomputerisasi sudah digunakan pada berbagai perusahaan guna mendapatkan informasi-informasi penting mengenai kegiatan pemasaran. Salah satu pencarian informasi dari kumpulan data besar adalah menggunakan *Knowledge Discovery in Database* (KDD). Data mining merupakan salah satu langkah dalam proses Knowledge Discovery in Databases (KDD), dimana tujuannya adalah untuk mengumpulkan data dan menggunakan data historis untuk menemukan hubungan atau pola dalam kumpulan data yang besar (Maulana, Winanjaya, & Rizki, 2022).

Saat ini perusahaan Basoka Motor dalam kegiatan pemasarannya masih dilakukan secara manual dan belum mempolakan penjualan motornya dengan didukung teknologi informasi. Oleh karena itu penelitian ini akan membangun sebuah aplikasi dimana data penjualan dari Basoka Motor dapat terkomputerisasi dan bisa melihat pola penjualan yang terjadi pada Basoka Motor tersebut. Untuk mendapatkan informasi digunakan salah satu metode klasifikasi yang ada di data mining yaitu metode decision tree menggunakan algoritma C.45.

Dengan menggunakan algoritma C4.5 melalui aplikasi RapidMiner, anggota subbidang aset dapat memperoleh pohon keputusan yang berguna dalam melakukan prediksi terkait peralatan dan mesin kantor yang akan digunakan di masa depan (Amalia & Resad, 2023). C4.5 adalah sebuah algoritma yang digunakan untuk membentuk pohon keputusan dalam konteks klasifikasi dan prediksi (Wahyu, Farozzi, Mahendra, & Hapsari, 2023). Algoritma C4.5 menghasilkan decision tree dan terdapatnya *rules* prediksi, tingkat akurasi yang dihasilkan adalah sebesar 70% dengan menggunakan pengujian *Confussion Matrix* dalam memprediksi bimbingan (Nugraha, Defit, & Nurcahyo, 2023). Perusahaan Basoka Motor memerlukan sebuah aplikasi yang dapat membantu mereka agar data penjualan dapat terkomputerisasi dan dapat melihat pola penjualan pada *showroom* motor tersebut.

## METODE

Pengembangan perangkat lunak yang dilakukan dalam penelitian ini adalah menggunakan model waterfall. Model waterfall adalah model yang sering digunakan dalam pengembangan perangkat lunak (Wahid, 2020). Tahapan-tahapan pengembangan perangkat lunak menggunakan model waterfall (S. Rosa, 2013): 1) Analisis, pada tahapan ini peneliti melakukan analisa kebutuhan terhadap sistem yang baru berdasarkan hasil analisa yang didapatkan dari sistem yang lama dan hasil wawancara serta observasi yang dilakukan, 2) Desain, pada tahapan ini peneliti membuat desain interface dan model sistem menggunakan *Unified Modeling Language* (UML), 3) Pengkodean, aplikasi dibuat menggunakan bahasa pemrograman Borland Delphi dan desain UML menggunakan StarUML, 4) Pengujian, pengujian pada penelitian ini menggunakan black box testing.

## Algoritma C4.5

Dalam algoritma C4.5, digunakan kriteria gain untuk menentukan fitur yang akan digunakan sebagai pemecah node pada pohon keputusan yang dihasilkan (Prasetyo, 2014). Proses pembangunan pohon keputusan melibatkan beberapa langkah, yaitu sebagai berikut: 1) Menentukan akar pohon, 2) Membuat cabang-cabang pada pohon berdasarkan kasus-kasus yang ada, 3) Mengulangi proses ini pada setiap cabang sampai setiap cabang memiliki kelas yang sama (Kusrini & Emha, 2009).

### 1. Entropy

Entropy digunakan sebagai kriteria untuk menentukan pemecah data latih selanjutnya pada pohon keputusan. Nilai entropy yang lebih tinggi menunjukkan potensi klasifikasi yang lebih tinggi. Jika entropy pada suatu node memiliki nilai 0, berarti semua vektor data dalam node tersebut memiliki label kelas yang sama, dan node tersebut akan menjadi daun pada pohon keputusan dengan keputusan yang sesuai (label kelas).

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

Persamaan entropy:

Ket:

$S$  : Himpunan kasus

$n$  : Jumlah partisi  $S$

$p_i$  : Proporsi dari  $S_i$  terhadap  $S$

Persamaan  $\log_2 p_i$ :

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

### 2. Gain Ratio

Dalam algoritma C4.5, gain ratio merupakan kriteria yang paling umum digunakan untuk memilih fitur yang akan menjadi pemecah node.

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

Persamaan gain (Kusrini & Emha, 2009):

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

Ket:

$S$  : himpunan kasus

$A$  : atribut

$n$  : jumlah partisi atribut  $A$

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

$|S|$  : jumlah kasus dalam  $S$

Persamaan *SplitEntropy*:

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

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 *confusion matrix*:

Persamaan menghitung *confusion matrix*:

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

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

## HASIL DAN PEMBAHASAN

Tahapan-tahapan pengembangan perangkat lunak:

### 1) Analisis

Pengumpulan data dilakukan dengan menggunakan wawancara langsung kepada pemilik Basoka Motor. Selain menggunakan teknik wawancara, peneliti juga menggunakan teknik observasi langsung bagaimana selama penyimpanan data pada Basoka Motor.

### 2) Desain

Sebelum dilakukan perancangan sistem. Dilakukan perhitungan manual Algoritma C4.5 terlebih dahulu terhadap data yang telah diperoleh.

### Perhitungan Algoritma C4.5

Tabel 1. Data Basoka Motor

No	Pabrikan	Jenis	Tahun	Harga	Status
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

...	...	...	...	...	...
27	Suzuki	Matic	2020	9800000	Laris

Sumber: Basoka Motor, 2022

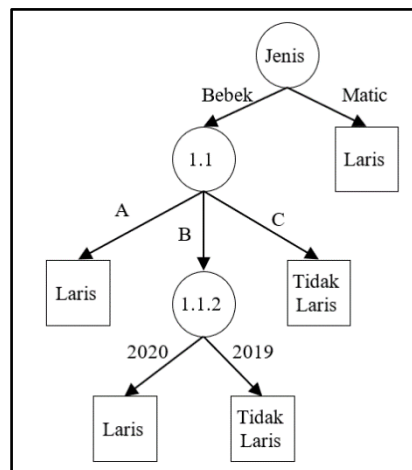
Dibentuklah pohon keputusan berdasarkan tabel 1 dimulai dari membuat akar, membuat cabang berdasarkan kasus-kasus hingga mengulangi kembali proses pada setiap cabang sampai setiap cabang memiliki kelas yang sama. Penerapan Algoritma C4.5 pada penelitian ini dimulai dari Node 1 sampai node terakhir yaitu Node 1.1.2.

Tabel 2. Node 1.1.2

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

Sumber: Analisis Data, 2022

Pohon keputusan yang dihasilkan dari Node 1.1.2 sebagai berikut:



Sumber: Analisis Data, 2022

Gambar 1. Pohon Keputusan Node 1.1.2

### Perhitungan Confussion Matrix

setelah didapatkan pohon keputusan, maka dicarilah *Confussion Matrix* untuk melihat akurasi dari hasil yang didapatkan.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP+TN}{TP+TN+FP+FN} * 100\% \\
 &= \frac{20+7}{20+7+0+0} * 100\% \\
 &= 100\%
 \end{aligned}$$

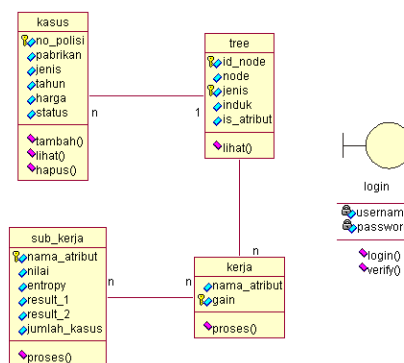
Outcome	Prediksi	
	Laris	Tidak laris
Laris	20	0
Tidak laris	0	7

Sumber: Analisis Data, 2022

Berdasarkan *Confussion Matrix* dapat disimpulkan bahwa dari 27 data motor yang dilakukan pencocokan rule prediksi dengan dataset menghasilkan akurasi sebesar 100%.

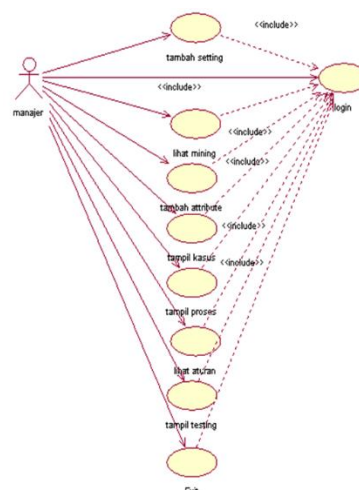
## UML

*Unified Modelling Language* digunakan sebagai visualisasi sistem piranti lunak. UML yang dirancang dalam pembuatan sistem ini sebagai berikut :



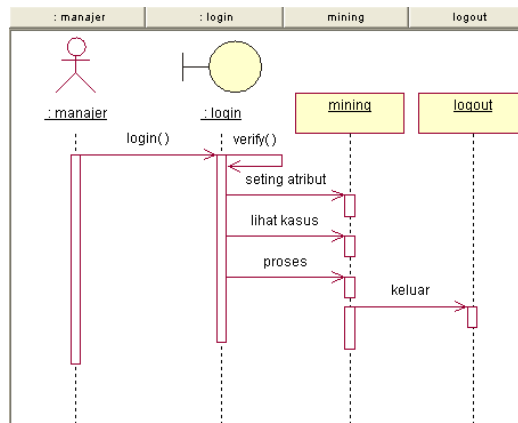
Sumber: Analisis Data, 2022

Gambar 2. Class Diagram



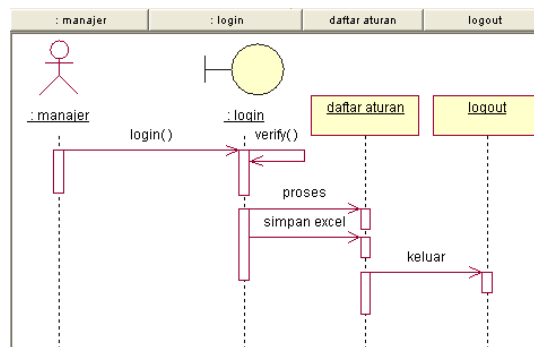
Sumber: Analisis Data, 2022

Gambar 3. Use Case Diagram



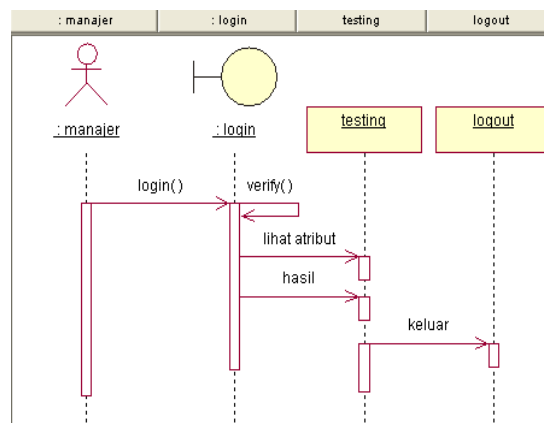
Sumber: Analisis Data, 2022

Gambar 4. Sequence Mining



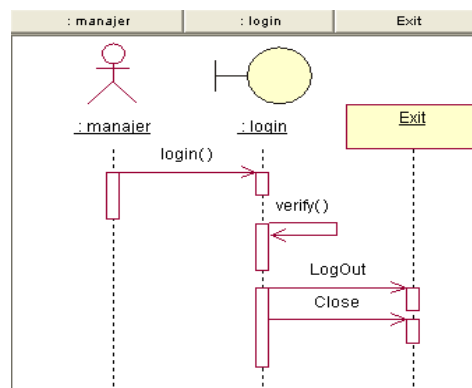
Sumber: Analisis Data, 2022

Gambar 5. Sequence Daftar Aturan



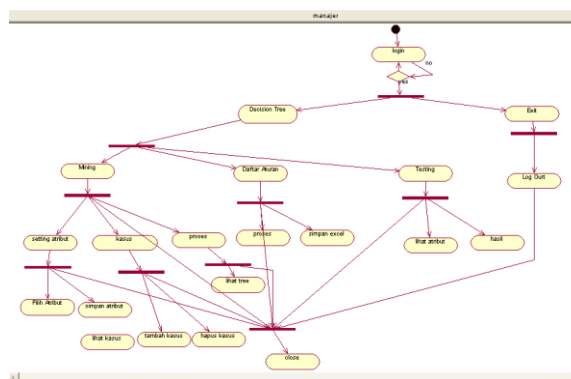
Sumber: Analisis Data, 2022

Gambar 6. Sequence Testing



Sumber: Analisis Data, 2022

Gambar 7. Sequence Exit



Sumber: Analisis Data, 2022

Gambar 8. Activity Diagram

## Interface

Desain user interface merupakan desain tampilan sistem yang akan dibuat, mulai dari desain halaman utama atau index sampai desain menu help.

The image shows a user interface design for a system titled 'WELCOME TO ALGORITHM DECISION TREE'. The interface includes two input fields: 'Username' and 'Password'. Below these fields are two buttons: 'Login' and 'Exit'. The text is centered, and the layout is simple and functional.

Sumber: Analisis Data, 2022



Gambar 9. Perancangan Antarmuka Login

A login interface design within a rectangular frame. It contains two columns of buttons. The left column has four buttons: 'Decision Tree', 'Mining', 'Aturan', and 'Testing'. The right column has three buttons: 'Close', 'Log Out', and another 'Close' button.

*Sumber: Analisis Data, 2022*

Gambar 10. Perancangan Antarmuka Halaman Utama

A main page interface design within a rectangular frame. At the top left is the title 'Mining'. Below it is a vertical stack of four buttons: 'Setting Atribut', 'Kasus', 'Proses', and 'Keluar'. To the right of these buttons is a large, empty rectangular area.

*Sumber: Analisis Data, 2022*

Gambar 11. Perancangan Antarmuka Mining

A mining interface design within a rectangular frame. At the top left is the title 'Setting Atribut'. Below it are two input fields. To the right of the second input field is a checkbox labeled 'Aktif'. To the right of the checkbox are two buttons: 'Simpan' and 'Tutup'.

*Sumber: Analisis Data, 2022*

Gambar 12. Perancangan Antarmuka Setting Atribut

A setting attribute interface design within a rectangular frame. At the top left is the title 'Testing'. Below it are two input fields and a dropdown menu. To the right of the dropdown menu are two buttons: 'selanjutnya' and 'Tutup'. To the right of the 'Tutup' button is a label 'Hasil'.

*Sumber: Analisis Data, 2022*

Gambar 13. Perancangan Antarmuka Testing

### 3) Implementasi

Aplikasi Data Mining dibuat menggunakan bahasa pemrograman Borland Delphi.



*Sumber: Analisis Data, 2022*

Gambar 14. Tampilan Login



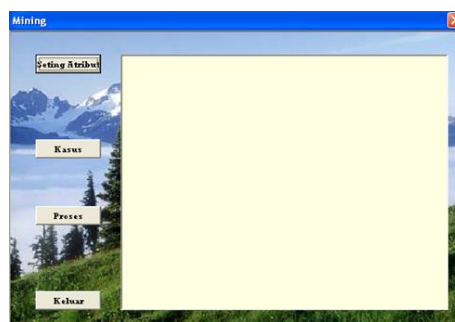
*Sumber: Analisis Data, 2022*

Gambar 15. Tampilan Menu Utama



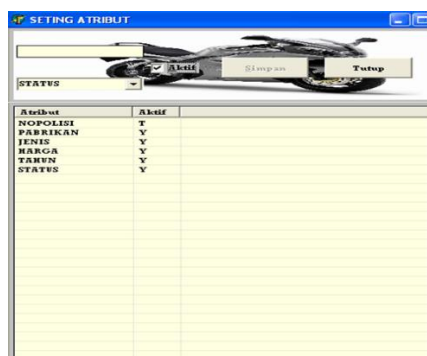
*Sumber: Analisis Data, 2022*

Gambar 16. Tampilan Decision Tree



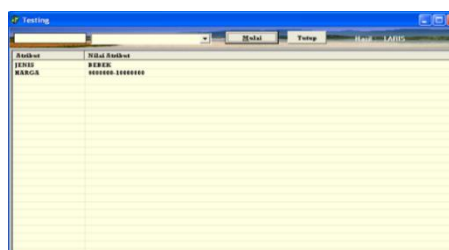
Sumber: Analisis Data, 2022

Gambar 17. Mining



Sumber: Analisis Data, 2022

Gambar 18. Setting Atribut



Sumber: Analisis Data, 2022

Gambar 19. Testing

#### 4) Pengujian

Pengujian yang dilakukan adalah dengan menggunakan Black Box Testing.

Tabel 4. Hasil Pengujian

No	Nama	Luaran yang diharapkan	Hasil
1.	Login	Username, password sesuai dengan validasi	Berhasil
2.	Decision Tree	Menu utama proses data mining	Berhasil
3.	Mining	Menampilkan kasus dari atribut	Berhasil
4.	Kasus	Menampilkan data motor	Berhasil
5.	Proses	Menampilkan decision tree atau pohon keputusan dari kasus	Berhasil
6.	Aturan	Menampilkan proses yang telah terjadi	Berhasil
7.	Testing	Menampilkan pengujian terhadap data yang telah di proses.	Berhasil
8.	Exit	Keluar dari aplikasi	Berhasil

Sumber: Analisis Data, 2022

## KESIMPULAN

Pengujian yang dilakukan antara perhitungan manual dengan aplikasi data mining terhadap 27 data motor yang di peroleh dari Basoka Motor mendapatkan hasil yang sama dengan tingkat akurasi sebesar 100% menggunakan *Confussion Matrix*. Dapat disimpulkan bahwa jenis, dalam hal ini jenis motor merupakan akar dari pohon yang memiliki nilai gain

tertinggi. Motor matic statusnya laris, sedangkan motor bebek perlu dilakukan perhitungan ulang. Dimana harga mempunyai gain tertinggi. Harga A yang bernilai 9.000.000 s/d 10.000.000 statusnya laris, harga C yang bernilai 11.100.000 s/d 12.000.000 statusnya tidak laris. Sedangkan harga B yang bernilai 10.100.000 s/d 11.000.000 perlu dilakukan perhitungan lagi. Dan di dapat yang mempunyai gain tertinggi yaitu tahun. Tahun 2019 dengan harga B dinyatakan tidak laris, namun tahun 2020 dengan harga B dinyatakan laris. Aplikasi dapat mempolakan penjualan pada Basoka Motor.

## 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.
- 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.
- S. Rosa, M. S. (2013). *Rekayasa Perangkat Lunak Edisi Revisi*. Bandung: Informatika.
- Wahid, A. A. (2020). Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi. *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK*, 1-5.
- Wahyu, B. R., Farozzi, 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.