

Sistem Informasi Penjualan Sepeda Motor Bekas Menggunakan Algoritma C4.5

Wimmie Handiwidjojo¹, Ferry Kristian Sukendro²

Abstrak— Sepeda motor sebagai salah satu alat transportasi memiliki jumlah yang cukup tinggi dibandingkan dengan kendaraan bermotor lainnya. Di dunia ini banyak sekali sepeda motor yang diproduksi dengan variasi merk dan jenis yang berbeda-beda. Tidak heran jika pemilik toko sepeda motor cenderung bingung untuk menentukan atribut sepeda motor yang banyak diminati oleh orang-orang.

Oleh karena itu, dibutuhkan suatu sistem yang dapat melakukan pengelolaan untuk menghitung nilai dari atribut-atribut yang telah ditentukan untuk mengetahui atribut yang paling berpengaruh dalam penjualan sepeda motor berdasarkan data yang telah tersedia, dalam penelitian ini penulis menggunakan algoritma C4.5 untuk diimplementasikan ke dalam sistem. Sistem ini akan menghitung jumlah nilai gain dari tiap atribut pada data penjualan sepeda motor.

Penelitian ini menghasilkan sebuah sistem yang dapat mencatat transaksi penjualan maupun pembelian untuk digunakan dalam perhitungan nilai gain pada algoritma C4.5 dari tiap atribut sepeda motor yang telah ditentukan yaitu merk, jenis, warna, dan tahun. Pada keluaran sistem ini akan ditampilkan daftar nilai gain dari tiap atribut. Posisi atribut yang memiliki nilai gain paling tinggi yang ditampilkan menunjukkan kebutuhan pasar.

Kata Kunci: Algoritma C4.5, Gain..

Abstract— The motorcycle as transportation vehicles have a fairly high number compared to other transportation vehicles. A lot of motorcycles are manufactured with a variety of brands, types, color, etc. No wonder, the motorcycle shop owners tend to be confused determining the attributes of motorcycles such as brand, type, color and production year which meet the market demand.

To solve the problem, we build a system that can calculate the value of the attributes of the most influential in the sale of used motorcycles based on the available data. In this research, the authors used C4.5 algorithms to be implemented into the system. This system will calculate the total value of the gain of each attribute on motorcycle sales data.

This research resulted a system that can record sales and purchases transactions to be used in the calculation of the gain of each attribute of motorcycle using C4.5 algorithms. The output of this system will show the value of the gain of each attribute (brand, type, color and production year). The highest position of the attribute means the highest value of

the gain that also means it meets the market demand.

Keywords: Algorithm C4.5, Gain.

I. PENDAHULUAN

Membeli motor bekas dikarenakan kemampuan daya beli masyarakat yang lemah telah memunculkan bisnis baru yaitu jual beli motor bekas. Didukung oleh buruknya transportasi publik menyebabkan masyarakat memutuskan untuk menggunakan transportasi yang merakyat ini. Tidak jarang satu keluarga memiliki lebih dari 1 motor karena kebutuhan transportasi yang berbeda-beda. Toko yang menyediakan bisnis jual beli motor bekas tumbuh seperti jamur di musim hujan. Tetapi toko juga perlu berhati-hati dalam pengadaan barang yang dalam hal ini berupa motor bekas, jika tidak melakukan perhitungan yang cermat, maka bukan hal tidak mungkin tidak ada pembeli yang melirik dagangannya. Toko perlu melakukan analisis pasar, tentang jenis motor apa, merek apa serta tahun berapa paling diminati masyarakat. Bahkan tidak jarang warnapun juga jadi bahan pertimbangan. Salah dalam melakukan analisis dan prediksi kebutuhan konsumen berdampak besar bagi untung dan rugi toko. Dalam penelitian ini data diambil dari data penjualan sepeda motor “Surya Mandiri Motor” selama 1 tahun terakhir.

Sistem informasi yang akan dibangun akan membantu toko sepeda motor bekas untuk menganalisa atribut yang paling berpengaruh dalam penjualan sepeda motor dan dikeluarkan dalam bentuk hasil perhitungan information gain dari algoritma C4.5. Dengan algoritma C4.5 akan diurutkan atribut yang paling tinggi nilainya. Data yang akan diolah adalah jenis motor, merek motor, tahun produksi dan warna motor. Dengan adanya sistem tersebut, diharapkan proses analisa atribut dalam penjualan sepeda motor bekas ini dapat dijadikan menolong pengelola toko untuk meningkatkan penjualan sepeda motornya.

II. LANDASAN TEORI

Penelitian ini berjudul “Sistem Informasi Penjualan Sepeda Motor Bekas Menggunakan Algoritma C4.5”. Sebagai sebuah sistem informasi, karena merupakan kerangka kerja untuk mengkoordinasikan sumber daya manusia dan komputer demi untuk mencapai sasaran perusahaan. Sebagai sebuah sistem informasi maka aktivitas yang terjadi menunjukkan adanya kegiatan mengumpulkan, memproses, menyimpan, menganalisis, dan menyebarkan informasi untuk tujuan yang khusus sebagai mana yang didefinisikan oleh Turban referensi [5].

¹ Program Studi Sistem Informasi, Universitas Kristen Duta Wacana, Yogyakarta, 55224 – Indonesia, Tlp. 0274 563929 ext. 323, email: whanz@staff.ukdw.ac.id.

² Program Studi Sistem Informasi, Universitas Kristen Duta Wacana, Yogyakarta, 55224 – Indonesia, email: ferryks_93@yahoo.com

Pemanfaatan sistem informasi menurut referensi [3] akan menolong mengkoordinasikan semua sumber daya yang dimiliki kemudian memroses dan menganalisis data yang dimiliki sehingga dapat dihasilkan sebuah luaran yang berupa saran untuk menolong toko mengambil keputusan dalam membeli motor bekas yang akan dijual lagi sehingga diharapkan tujuan toko yaitu mencapai keuntungan yang maksimal dapat dicapai.

A. Algoritma C4.5

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, dan kecerdasan buatan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar menurut referensi [2].

Penelitian ini menggunakan salah satu metoda *Data Mining* yaitu *Algoritma C4.5*. *Algoritma C4.5* merupakan algoritma yang dapat digunakan untuk membentuk pohon keputusan. Sedang pohon keputusan adalah salah satu metode klasifikasi yang cukup mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. Manfaat utamanya adalah *mem-break down* proses pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambil keputusan akan lebih mudah menemukan solusi dari permasalahan yang ada menurut referensi [4].

Dalam *algoritma C4.5* untuk membangun pohon keputusan hal pertama yang dilakukan yaitu memilih atribut sebagai akar. Kemudian dibuat cabang untuk tiap-tiap nilai didalam akar tersebut. Langkah berikutnya yaitu membagi kasus dalam cabang. Kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [4].

Untuk memilih atribut dengan akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan rumus sebagai berikut [1].

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

Dimana:

S : himpunan kasus

A : Atribut

N : jumlah partisi atribut *A*

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

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

Gain : Penguatan

Sementara Rumus untuk menghitung Entropy dapat dilihat sebagai berikut [1]:

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

Dimana:

S : himpunan Kasus

n : jumlah partisi *S*

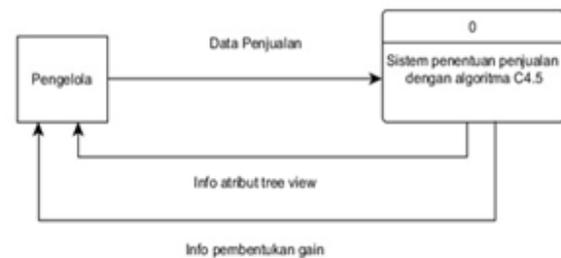
P_i : Jumlah kasus pada partisi ke-*i*

III. PERANCANGAN SISTEM

A. Data Flow Diagram dan Flow Chart

Data Flow Diagram menggambarkan proses aliran data yang terjadi dalam sistem yang dibangun sehingga jelas batasan kerja, interaksi, dan transformasi data dalam sistem tersebut. Simbol kotak menunjukkan entitas luar, arah panah menunjukkan aliran data, lingkaran menunjukkan proses, dan kotak tanpa garis sisi menunjukkan berkas atau penyimpanan data.

Penggambaran Data flow diagram dimulai dengan penggambaran konteks diagram atau bisa juga disebut DFD level 0 yang merupakan penggambaran umum dari alur data yang terdapat pada sistem.



Gambar 1. Data Flow Diagram Level 0

Sementara itu rangkaian proses penentuan penjualan dengan implementasi algoritma c4.5 secara keseluruhan digambarkan dengan flowchart pada Gambar 2 berikut ini.

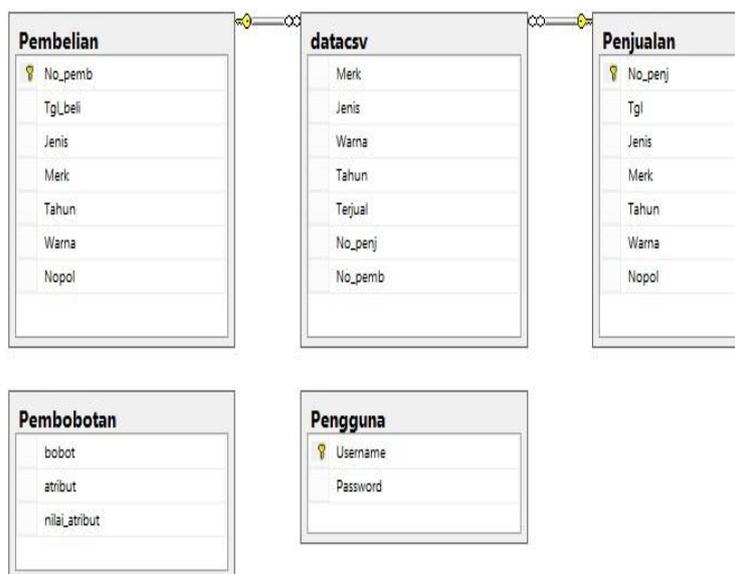


Gambar 2. Flowchart Sistem

B. Rancangan Basis Data

Berdasarkan data flow yang ada maka dibangunlah rancangan basis data untuk sistem. Tabel yang dibuat

dalam rancangan basis data penelitian ini yaitu tabel Pembelian, Penjualan, datacsv, Pembobotan, dan Pengguna. Rancangan basis data dapat dilihat pada gambar 3 dibawah ini.



Gambar 3. Rancangan Basis Data

C. Perhitungan Algoritma C4.5

Penelitian ini menggunakan data tahun 2013 dari data pembelian sepeda motor toko “Surya Mandiri Motor”. Sebagai contoh sederhana untuk perhitungan Algoritma C4.5: adalah sebagai berikut:

Tabel 1 Contoh Kasus Penjualan Motor

No	Merk	Jenis	Tahun	Warna	Terjual
1	Yamaha	Bebek	<2010	Hitam	Ya
2	Yamaha	Matic	<2010	Hitam	Ya
3	Yamaha	Sport	<2010	Hitam	Ya
4	Yamaha	Sport	2010-2013	Merah	Ya
5	Yamaha	Matic	<2010	Merah	Ya
6	Honda	Bebek	<2010	Merah	Ya
7	Honda	Bebek	2010-2013	Merah	Ya
8	Honda	Sport	<2010	Hitam	Ya
9	Honda	Sport	2010-2013	Hitam	Ya
10	Honda	Matic	<2010	Merah	Ya
11	Yamaha	Matic	2010-2013	Hitam	Tidak
12	Yamaha	Sport	<2010	Merah	Tidak
13	Honda	Sport	2010-2013	Merah	Tidak
14	Yamaha	Bebek	2010-2013	Merah	Tidak

Dari hasil perhitungan pada tabel 2 dapat diketahui atribut dengan *information gain* tertinggi yang akan menjadi node akar adalah **Tahun** dengan nilai *information gain* sebesar 0.123940887. Jika perhitungan dilanjutkan akan didapat data pada node 1 seperti pada tabel 2.

Tabel 2 Hasil Perhitungan Data

Merek	JK	T	Y	Entropy	Gain
	(S)	(S1)	(S2)		
Total	14	4	10	0.863120569	
Merk					0.039148672
	Yamaha	8	3	5	0.954434003
	Honda	6	1	5	0.650022422
Jenis					0.005977711
	Bebek	4	1	3	0.811278124
	Sport	6	2	4	0.918295834
	Matic	4	1	3	0.811278124
Tahun					0.123940887
	<2010	8	1	7	0.543564443
	2010-2013	6	3	3	1
Warna					0.039148672

	Hitam	6	1	5	0.650022422
	Merah	8	3	5	0.954434003

Baris TOTAL kolom Entropy pada tabel 2 dihitung dengan cara sebagai berikut:

$$Gain(Total, Tahun) = Entropy(total) - \sum_{i=1}^n \frac{|Tahun_i|}{|Total|} * Entropy(Tahun_i)$$

$$Gain(Total, Tahun) = 0.863120569 - \left(\left(\frac{8}{14} * 0.543564443 \right) + \left(\frac{5}{14} * 1 \right) \right)$$

$$Gain(Total, Tahun) = 0.123940887$$

IV. HASIL IMPLEMENTASI

Tahap implementasi merupakan tahap penerapan sistem, sehingga sistem siap digunakan. Tahap ini menjelaskan mengenai penerapan sistem berikut proses dan fungsi-fungsi yang ada di dalam sistem. Sistem ini dibangun menggunakan bahasa pemrograman VB.NET dan menggunakan basis data SQL Server 2012.

Seperti yang sudah disampaikan dibagian depan bahwa data yang dimasukkan adalah data pembelian tahun 2013 dari toko sepeda motor bekas “Surya Mandiri Motor”. Sistem menyediakan interface untuk memasukkan data pembelian sepeda motor di bagian kiri gambar 4, sementara data yang sudah masuk kedalam basis data dapat dilihat pada tampilan daftar pembelian diposisi sebelah kanan gambar 5.

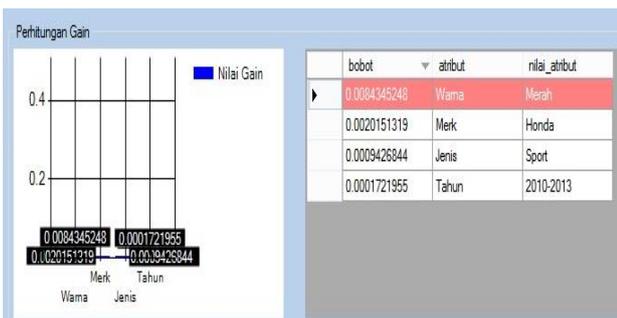


Gambar 4. Antarmuka Input Pembelian

No_pemb	Tgl_beli	Jenis	Merk	Tahun	Warna
B000000001	2/7/2013	Sport	Yamaha	<2010	Merah
B000000002	3/1/2013	Sport	Yamaha	<2010	Putih
B000000003	4/18/2013	Sport	Yamaha	<2010	Putih
B000000004	5/3/2013	Sport	Yamaha	<2010	Hitam
B000000005	6/1/2013	Sport	Yamaha	<2010	Hitam
B201550001	5/21/2015	Bebek	Honda	<2010	Putih
B201550002	5/24/2015	Bebek	Yamaha	<2010	Putih
B201550003	5/24/2015	Sport	Yamaha	2010-2013	Biru
B201550004	5/24/2015	Sport	Kawasaki	2010-2013	Kuning
B201550005	5/24/2015	Bebek	Suzuki	<2010	Merah
B201550006	5/24/2015	Bebek	Yamaha	<2010	Putih
B201550007	5/24/2015	Matic	Yamaha	<2010	Kuning
B201550008	5/24/2015	Matic	Yamaha	<2010	Putih

Gambar 5. Tampilan Basisdata Pembelian

Sistem ditujukan untuk mengetahui atribut yang paling tinggi atau faktor apa yang mempengaruhi penjualan motor, karena user atau pemilik toko seringkali bingung motor apa yang akan dibeli supaya cepat laku. Sistem akan menghitung information gain (penguatan informasi) dengan menerapkan Algoritma C4.5 pada keempat atribut yaitu merk, tahun produksi, warna dan jenis.



Gambar 6. Hasil Perhitungan Gain

Gunakan satuan SI (MKS) atau CGS sebagai satuan utama (satuan SI sangat dianjurkan). Hindar penggunaan gabungan satuan SI dan CGS, seperti arus dalam ampere dan medan magnet dalam oersteds. Hal ini dapat membingungkan karena persamaan tidak imbang secara dimensi. Jika anda ingin menggunakan unit secara campuran, sebutkan dengan jelas untuk setiap kuantitas di dalam persamaan.

Nilai *information gain* tertinggi pada *field* atribut akan menunjukkan sepeda motor yang diminati pasar pada periode waktu tersebut sehingga direkomendasikan untuk layak dibeli oleh toko. Hasil luarannya dapat dilihat pada gambar 6 sebagai hasil perhitungan gain (penguatan).

Dari hasil perhitungan didapatkan bahwa atribut “Warna” yang memiliki nilai *gain* tertinggi dimana nilai atributnya adalah “Merah” sehingga dapat diambil kesimpulan bahwa atribut yang paling tinggi nilainya adalah “Warna” dengan nilai “Merah” dan memiliki nilai

gain (penguatan): 0,0069318179. Sementara nilai dari atribut “Merk” dengan nilai “Honda” hasilnya 0,0032423952. Atribut “Jenis” dengan nilai “Sport” memiliki nilai *gain* 0,0009417054, dan terakhir atribut “Tahun” dengan nilai “2010-2013” memiliki hasil 0,0002745057. Hasil perhitungan ini berdasarkan jumlah data yang sudah terdapat di dalam basisdata yang berjumlah 263 data.

Setelah proses ini dikerjakan, maka langkah selanjutnya diadakan uji coba untuk memasukkan data penjualan apakah nilai *gain* bisa berubah atau tidak. Untuk melakukan uji coba ini pengguna memasukkan data dari halaman Penjualan agar data klasifikasi untuk kasus bernilai “Ya” bertambah 1 untuk membuktikan nilai bertambah atau tidak.

Nopol : AD2898BA
 Jenis : Sport
 Merk : Kawasaki
 Tahun : 2010-2013
 Warna : Merah

Gambar 7. Memasukkan Data Penjualan

Setelah satu data penjualan ditambahkan oleh *user* ke dalam basis data, sehingga jumlah kasus “Ya” untuk data klasifikasi bertambah 1, maka pengguna kembali melihat hasil perhitungan pada halaman Perhitungan apakah ada perubahan atau tidak. Ternyata hasil perhitungannya seperti tabel 3 berikut:

Tabel 3.: Hasil Perhitungan Setelah Ditambahkan Data Penjualan

bobot	atribut	nilai_atribut
0.0066325268	Warna	Merah
0.0044925840	Jenis	Sport
0.0032938396	Merk	Honda
0.0001814219	Tahun	2010-2013

Tabel diatas menunjukkan perubahan *gain* setelah dimasukkan 1 data Penjualan dengan kasus “Ya”. Dari hasil tersebut dapat disimpulkan bahwa setelah data ditambahkan maka nilai *gain* akan berubah sesuai jumlah perbandingan data yang dimasukkan karena dilakukan perhitungan ulang. Sehingga sebelumnya atribut “Warna” bernilai “Merah” dengan nilai *gain* 0,0069318179 sekarang menjadi 0,0066325268 dan tetap menempati urutan pertama. Selanjutnya terjadi perubahan posisi, dimana semula posisi kedua adalah atribut “Merk” bernilai “Honda” dengan nilai *gain* 0,0032423952, sekarang berubah menjadi 0,0032938396 menempati urutan ketiga dan posisi kedua diambil alih oleh atribut “Jenis” bernilai “Sport” karena nilai *gain*-nya berubah

dari 0,0009417054 menjadi 0,0044925840. Demikian pula dengan atribut “Tahun” mengalami perubahan *gain* menjadi 0,0001814219.

Dengan demikian setiap data penjualan baru dimasukkan, akan mempengaruhi besarnya *gain* dan itu artinya nilai *gain* juga ikut berubah dan akibatnya menyebabkan perubahan urutan atribut tergantung dari data baru yang dimasukkan.

Nilai *Gain* atau dapat diistilahkan dengan nilai “penguatan” dapat dijadikan masukan bagi pemilik toko sepeda motor bekas “Surya Mandiri Motor” dalam menentukan atribut mana yang harus dipertimbangkan apakah merek, jenis, warna atau tahun produksi untuk membeli sepeda motor bekas. *Gain* (Penguatan) dengan nilai terbesar menunjukkan bahwa atribut itulah yang sedang diminati pasar (*Market Demand*). Dengan melihat hasil perhitungan *gain*, maka pemilik toko ditolong untuk membuat keputusan sepeda motor apa yang sebaiknya dibeli supaya cepat laku dijual kepada konsumen.

V. KESIMPULAN

Berdasarkan hasil penelitian dapat diambil kesimpulan bahwa:

1. *Algortima C4.5* dapat digunakan untuk menolong pengambilan keputusan pembelian sepeda motor bekas oleh toko dengan memperhatikan nilai *gain* (penguatan) tertinggi dari empat atribut seperti Merek, jenis, warna dan tahun produksi.
2. *Gain* atribut dengan nilai tertinggi menunjukkan besarnya *market demand* (diminati pasar) dalam arti atribut tersebut memiliki pengaruh yang paling besar bagi konsumen pembeli sepeda motor bekas.
3. Nilai *gain* akan berubah manakala ada perubahan pencatatan data penjualan ataupun pembelian.
4. Hasil prediksi akan lebih akurat jika data yang di olah dalam jumlah lebih besar dalam arti tidak hanya setahun.

DAFTAR PUSTAKA

- [1] Bramer, Max. (2007). *Principles of Data Mining*. London: Springer.
- [2] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Andi Yogyakarta.
- [3] Kadir, A. (2014). *Pengenalan Sistem Informasi Edisi Revisi*. Yogyakarta : Andi Yogyakarta.
- [4] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta : Andi Yogyakarta.
- [5] Turban E., McLean E., Wetherbe J. (2001). *Information Technology for Management : Making Connections for Strategic Advantage*: Wiley