

## **Data Mining Menentukan Minat Konsumen Memilih Sepeda Motor Idaman Dengan Algoritma C4.5**

**Tio Widodo<sup>1</sup>, Dedi Setiawan<sup>2</sup>, Astri Syahputri<sup>3</sup>**

<sup>1,3</sup> Sistem Informasi, STMIK Triguna Dharma

<sup>2</sup> Teknik Komputer, STMIK Triguna Dharma

Email: <sup>1</sup>tiowidodvr46@gmail.com, <sup>2</sup>setiawandedi07@gmail.com, <sup>3</sup>astri.syahputri29@gmail.com

Email Penulis Korespondensi: tiowidodvr46@gmail.com

### **Abstrak**

PT. Alfa Scorpii merupakan perusahaan pendistribusian kendaraan roda 2 atau sepeda motor bermerk Yamaha. Banyak jenis dan tipe sepeda motor yang di jual oleh PT. Alfa Scorpii. PT. Alfa Scorpii harus mencoba mempengaruhi konsumen dengan merek yang ditawarkannya, bahkan yang semula tidak ingin menjadi ingin membeli produk tersebut. Karena pada prinsipnya konsumen yang menolak hari ini belum tentu menolak hari berikutnya. PT. Alfa Scorpii juga harus dapat memprediksi konsumen dalam memilih tipe sepeda motor idaman yang sesuai dengan kebutuhan mereka.

Karena alasan itu maka dibutuhkan suatu sistem yang mampu memberikan informasi atau data kepada PT. Alfa Scorpii untuk melihat minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan mereka. Dari berbagai riset yang sudah dilakukan, ternyata data mining mampu untuk mengatasi masalah tersebut. Dengan alasan itulah maka diangkat penelitian dengan judul “Data Mining Untuk Menentukan Jumlah Minat Konsumen Dalam Memilih Tipe Sepeda Motor Idaman Sesuai Kebutuhan Para Konsumen Pada PT. Alfa Scorpii Gatot Subroto Medan Dengan Menggunakan Algoritma C4.5”.

Penelitian tersebut bertujuan untuk menciptakan suatu sistem berbasis komputerisasi, kemudian dengan diterapkannya sistem tersebut maka hasil yang didapatkan akan benar-benar akurat dan cepat kemudian dapat membantu PT. Alfa Scorpii. Dapat dikatakan bahwa dengan pengujian sistem berdasarkan nilai atribut yang diberikan pada setiap alternatif akan memberikan jawaban tipe sepeda motor idaman sesuai kebutuhan para konsumen berdasarkan nilai akhir yang didapat. Hal ini karena penerapan algoritma yang di masukkan ke dalam coding program sehingga sistem ini dapat membantu PT. Alfa Scorpii.

Kata Kunci : Konsumen, PT. Alfa Scorpii, Data Mining, Algoritma C4.5

## **1. PENDAHULUAN**

Pentingnya peran transportasi dalam memberikan akses terhadap barang dan jasa yang sangat esensial bagi kehidupan seseorang, terutama kelompok marginal, semakin mendapat perhatian dalam kajian transportasi. Solusi umum untuk meningkatkan aksesibilitas masyarakat adalah dengan memberikan pelayanan transportasi publik. Namun dalam kondisi rendahnya pelayanan transportasi publik, kepemilikan kendaraan pribadi menjadi salah satu solusi yang tidak dapat dihindarkan. Sepeda motor, dengan harganya yang relatif murah dan fleksibilitas yang ditawarkan kendaraan pribadi, menjadi pilihan banyak masyarakat dengan pendapatan menengah rendah.

Secara nasional jumlah kendaraan bermotor meningkat setiap tahun terutama sepeda motor. Berdasarkan data dari gabungan industri sepeda motor Indonesia (AISI) populasi kendaraan bermotor di Indonesia pada tahun 2010 mencapai 50.824.128 unit dan pada tahun 2014 meningkat menjadi 63.530.160 unit atau meningkat sebesar 25 persen [1]. Kepemilikan sepeda motor merupakan suatu daya dorong atau kekuatan yang timbul dari dalam diri masyarakat untuk memberikan kemudahan mobilitas dalam mencapai tujuan yang diinginkan. Kepemilikan sepeda motor merupakan respons masyarakat berpendapatan rendah terhadap keinginan yang terdapat dalam masyarakat untuk memiliki sepeda motor, karena peran sepeda motor yang begitu penting bagi kelangsungan perekonomian mereka dan mereka berupaya untuk mencapainya [2].

Kondisi di atas tersebut menjadi peluang bagi penyedia kendaraan bermotor untuk terus berinovasi menarik minat konsumen agar tertarik membeli kendaraan bermotor mereka, salah satunya yaitu PT. Alfa Scorpii. PT. Alfa Scorpii merupakan perusahaan pendistribusian kendaraan roda 2 atau sepeda motor bermerk Yamaha. Banyak jenis dan tipe sepeda motor yang di jual oleh PT. Alfa Scorpii. PT. Alfa Scorpii harus mencoba mempengaruhi konsumen dengan merek yang ditawarkannya, bahkan yang semula tidak ingin menjadi ingin membeli produk tersebut. Karena pada prinsipnya konsumen yang menolak hari ini belum tentu menolak hari berikutnya. PT. Alfa Scorpii juga harus dapat memprediksi konsumen dalam memilih tipe sepeda motor idaman yang sesuai dengan kebutuhan mereka, hal ini agar perusahaan dapat mengetahui tipe sepeda motor yang bagaimana yang menjadi idaman konsumen.

Karena alasan itu maka dibutuhkan suatu sistem yang mampu memberikan informasi atau data kepada PT. Alfa Scorpii untuk melihat minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan mereka. Dari berbagai riset yang sudah dilakukan, ternyata data mining mampu untuk mengatasi masalah tersebut.

Penelitian ini akan menjelaskan bagaimana data mining dalam menentukan minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan mereka.

Data Mining merupakan proses semi *otomatic* yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang berguna dan bermanfaat yang tersimpan di dalam database besar. Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [3]. Konsep data mining merupakan upaya dalam mengenali informasi yang tersembunyi didalam kumpulan data yang berjumlah besar. Data mining bukanlah suatu bidang ilmu yang baru, data mining muncul sekitar tahun 1990, disebabkan terjadinya penumpukan data maka diperlukan data mining untuk menganalisis, mengestimasi dan mengelompokkan data tersebut [2].

Dalam data mining terdapat banyak algoritma yang dapat dijadikan alternatif penyelesaian masalah, yaitu salah satunya adalah algoritma C4.5. Algoritma C4.5 merupakan kelompok algoritma *decision tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data [3]. Bahwa algoritma C4.5 dalam memprediksi jumlah pinjaman hasil yang didapatkan bersifat klasifikasi dengan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Artinya algoritma C4.5 dapat digunakan dalam menentukan jumlah minat konsumen untuk memilih tipe sepeda motor idaman sesuai kebutuhan[4].

## 2. METODOLOGI PENELITIAN

### 2.1 Minat

Minat adalah suatu keadaan dimana seseorang mempunyai perhatian terhadap sesuatu dan disertai keinginan untuk mengetahui dan mempelajari maupun membuktikan lebih lanjut.

Minat adalah suatu perhatian yang kuat dan mendalam disertai dengan perasaan senang terhadap suatu kegiatan sehingga mengarahkan seseorang untuk melakukan kegiatan tersebut dengan kemauan sendiri .

### 2.2 Konsumen

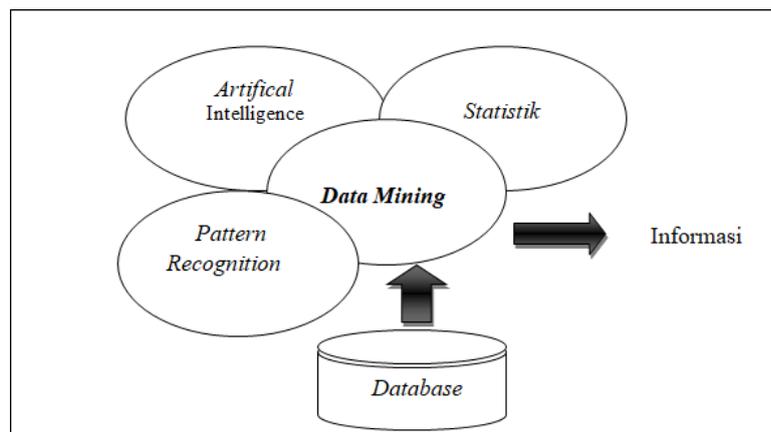
Pengertian konsumen secara umum adalah setiap orang yang memakai barang atau jasa yang tersedia dalam masyarakat dengan maksud untuk memenuhi kebutuhannya, orang lain maupun makhluk hidup lain, untuk berbagai kepentingan tanpa memperdagangkannya kembali .

Perilaku konsumen adalah studi tentang bagaimana individu, kelompok dan organisasi memilih, membeli, menggunakan dan menempatkan barang, jasa, ide atau pengalaman untuk memuaskan keinginan dan kebutuhan mereka .

Menurut Schiffman dan Kanuk (2010) dalam perilaku yang diperlihatkan konsumen dalam mencari, membeli, menggunakan, mengevaluasi, dan menghabiskan produk dan jasa yang mereka harapkan akan memuaskan kebutuhan mereka. Perilaku konsumen adalah tindakan bagaimana perorangan dalam memperoleh, menggunakan serta membuang barang dan jasa ekonomi, termasuk proses pengambilan Keputusan sebelum menetapkan tindakan.

### 2.3 Data Mining

Istilah *Data Mining* sudah muncul sejak tahun 1990-an dan mulai populer sejak awal tahun 2000-an. *Data Mining* adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode *Data Mining* ini dapat digunakan untuk mengambil keputusan di masa depan [5]. *Data Mining* ini juga dikenal dengan istilah *pattern recognition* [6]. Data mining dapat digunakan untuk pengelompokan Kasus dan Kematian COVID-19 di Asia Tenggara[7]. *Data Mining* merupakan proses semi *otomatic* yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang berguna dan bermanfaat yang tersimpan di dalam database besar [8]. Di samping itu, ada istilah lain yang mempunyai makna yang sama dengan *Data Mining* yaitu *Knowledge Discovery in Database* (KDD) [9]. Data mining adalah metode untuk menemukan informasi baru yang berguna dari kumpulan data yang besar dan dapat membantu dalam pengambilan keputusan [10]. Di dalam *Data Mining* maupun KDD bertujuan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna [11].



Gambar 2.1 Akar Ilmu Data Mining

**2.4 Algoritma C4.5**

Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut [4]:

1. Pilih atribut sebagai akar
2. Buat cabang untuk masing-masing nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus dan jumlah responden sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*.

$$Entropy(S) = \sum_{i=0}^n -pi * \log^2 pi$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya:

*S* : Himpunan kasus

*n* : Jumlah partisi *S*

*pi* : Jumlah kasus pada partisi ke-*i*

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

Rumus (1) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya:

*S* : Himpunan kasus

*A* : Atribut

*n* : Jumlah partisi atribut *A*

*|S<sub>i</sub>|* : Jumlah kasus pada partisi ke-*i*

*|S|* : Jumlah kasus dalam *S*

Algoritma C4.5 yaitu sebuah algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3[12]. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continuu data*, dan *pruning* [4].

**3. HASIL DAN PEMBAHASAN**

**3.1 Deskripsi Data Penilaian**

Pengumpulan data pada penelitian ini menggunakan teknik observasi. Kegiatan tersebut dilakukan dengan melakukan tinjauan langsung ke tempat studi kasus dimana akan di lakukan penelitian yaitu di PT. Alfa Scorpii Gatot Subroto Medan. Dari pengumpulan data yang dilakukan diperoleh data penjualan sepeda motor dari ahan sebagai berikut :

Tabel 3.1 Data Penjualan Sepeda Motor

No	Sepeda Motor	Jenis	Tahun	Harga	Status
1	Yamaha Mio M3 125	Matic	2014	Rp. 17.725.000	Laris
2	Yamaha Mio Z	Matic	2016	Rp. 15.800.000	Laris
3	Yamaha Fino 125	Matic	2012	Rp. 20.300.000	Laris

4	Yamaha Fazzio	Scooter Matic	2022	Rp. 22.000.000	Laris
5	Yamaha All New Nmax 155	Matic Sport	2015	Rp. 34.299.000	Laris
6	Yamaha Xmax	Matic Sport	2017	Rp. 61.475.000	Laris
7	Yamaha Tmax DX	Matic Sport	2022	Rp. 323.000.000	Tidak Laris
8	Yamaha Lexi S	Matic	2018	Rp. 28.350.000	Tidak Laris
9	Yamaha Aerox 155 Connected	Matic Sport	2016	Rp. 30.535.000	Laris
10	Yamaha Lexi	Matic	2018	Rp. 24.999.000	Laris
11	Yamaha YZ 125 X	Trail Sport	2020	Rp. 78.300.000	Tidak Laris
12	Yamaha MT-15	Naked Bike Sport	2019	Rp. 37.930.000	Laris
13	Yamaha MT-25	Naked Bike Sport	2015	Rp. 58.460.000	Tidak Laris
14	Yamaha XSR 155	Naked Bike Sport	2019	Rp. 40.435.000	Laris
15	Yamaha YZF R15	Sport		Rp. 38.700.000	Tidak Laris
16	Yamaha WR 155 R	Trail Sport	2018	Rp. 40.419.000	Laris
17	Yamaha Vixion	Sport	2007	Rp. 33.034.000	Laris
18	Yamaha MX King 150	Bebek Sport	2021	Rp. 28.000.000	Laris
19	Yamaha MT09	Sport	2022	Rp. 250.000.000	Laris
20	Yamaha R15M Connected	Sport	2014	Rp. 44.541.000	Laris
21	Yamaha Gear 125	Matic	2020	Rp. 17.530.000	Laris
22	Yamaha Freego	Matic	2018	Rp. 26.100.000	Laris
23	Yamaha Mio S Smart & Sophisticated	Matic	2012	Rp. 18.467.000	Laris
24	Yamaha XRide 125	Matic	2013	Rp. 18.582.000	Laris
25	Yamaha All New Soul GT AKS	Matic	2015	Rp. 19.000.000	Tidak Laris

Sebelum proses algoritma dilakukan, dilakukan dahulu proses klasifikasi untuk tiap-tiap *field* tabel Penjualan yaitu sebagai berikut :

1. Klasifikasi Jenis Sepeda Motor

Tabel 3.2 Klasifikasi Jenis Sepeda Motor

No	Klasifikasi Jenis	Laris	Tidak Laris
1	Matic	12	0
2	Sport	7	6

2. Klasifikasi Tahun

Tabel 3.3 Klasifikasi Tahun

No	Klasifikasi Tahun	Kode	Laris	Tidak Laris
1	2016 – 2022	A	5	3
2	2007 – 2016	B	14	3

3. Klasifikasi Harga

Tabel 3.4 Klasifikasi Harga

No	Klasifikasi Harga	Kode	Laris	Tidak Laris
1	Rp. 15.000.000 s/d Rp. 30.000.000	A	5	2
2	Rp 30.000.000 s/d Rp. 100.000.000	B	14	2
3	Rp. 100.000.000 s/d Rp. 350.000.000	C	0	2

### 3.2 Melakukan Perhitungan

Dalam data training *set atribut sample* dan minimal harus memiliki satu atribut target yang nilainya merupakan kesimpulan sementara permasalahan dari setiap *instance (record)*, dalam penelitian ini nilai dari *atribut target* adalah: laris atau tidak laris. *Atribut input* yang memiliki *gain ratio* yang terbesar adalah *atribut* yang menjadi akar, dalam penelitian ini nilai dari *atribut input* adalah: jenis kendaraan, tahun dan harga.

Perancangan data mining ini penulis menggunakan algoritma C4.5. Proses pada pohon keputusan adalah mengubah bentuk data (*table*) menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule*.

Dalam kasus yang tertera pada gambar 3.1 akan dibuat pohon keputusan untuk menentukan sepeda motor yang terjual atau tidak dengan melihat jenis sepeda motor, tahun, dan harga. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- Pilih atribut akar.
- Buat cabang untuk tiap-tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain*, seharusnya kita harus mencari nilai *entropy* terlebih dahulu.

Rumusnya seperti persamaan berikut.

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

Keterangan :

S : himpunan kasus

A : atribut

n : jumlah partisi S

$p_i$  : proporsi dari  $S_i$  terhadap S

Sementara itu, penghitungan nilai *gain* dapat dilihat pada persamaan 2 berikut.

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

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

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

$|S|$  : jumlah kasus dalam S

Berikut ini adalah penjelasan lebih terperinci mengenai tiap-tiap langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan pada tabel 3.1.

- Menghitung jumlah kasus untuk keputusan laris, jumlah kasus untuk keputusan tidak laris, *Entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut jumlah sepeda motor, sepeda motor terjual, dan sisa sepeda motor. Setelah itu, lakukan penghitungan *gain* untuk setiap atribut. Hasil perhitungan ditunjukkan oleh tabel 3.5.

Tabel 3.5 Perhitungan Node 1

NODE			Jumlah Kasus (S)	Tidak laris (S <sub>1</sub> )	Laris (S <sub>2</sub> )	Entropy	Gain
1	TOTAL		25	6	19	0.795040	
	Harga						0.203917
		A	7	2	5	0,868721	
		B	16	2	14	0.543564	
		C	2	2	0	0	
	Jenis						0.277443
		Matic	12	0	12	0	
		Sport	13	6	7	0.995378	
	Tahun						0.143717
		A	8	3	5	0.950672	
		B	17	3	14	0.671247	

Baris total *entropy* pada tabel 3.5 dihitung dengan persamaan 1 sebagai berikut :

$$Entropy(Total) = (-\frac{6}{25} * \log_2(\frac{6}{25})) + (-\frac{19}{25} * \log_2(\frac{19}{25})) = 0.795040$$

$$Entropy(Total, Harga A) = (-\frac{7}{25} * \log_2(\frac{7}{25})) + (-\frac{18}{25} * \log_2(\frac{18}{25})) = 0.868721$$

$$Entropy(Total, Harga B) = (-\frac{16}{25} * \log_2(\frac{16}{25})) + (-\frac{9}{25} * \log_2(\frac{9}{25})) = 0.543564$$

$$Entropy(Total, Harga C) = (-\frac{2}{25} * \log_2(\frac{2}{25})) + (-\frac{23}{25} * \log_2(\frac{23}{25})) = 0.868721$$

**Entropy Jenis**

$$Entropy(Total, Matic) = (-\frac{0}{12} * \log_2(\frac{0}{12})) + (-\frac{12}{12} * \log_2(\frac{12}{12})) = 0$$

$$Entropy(Total, Sport) = (-\frac{13}{13} * \log_2(\frac{13}{13})) + (-\frac{0}{13} * \log_2(\frac{0}{13})) = 0.995378$$

**Entropy Tahun**

$$Entropy(Total, Tahun A) = (-\frac{3}{8} * \log_2(\frac{3}{8})) + (-\frac{5}{8} * \log_2(\frac{5}{8})) = 0.950672$$

$$Entropy(Total, Tahun B) = (-\frac{8}{17} * \log_2(\frac{8}{17})) + (-\frac{9}{17} * \log_2(\frac{9}{17})) = 0.680077$$

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

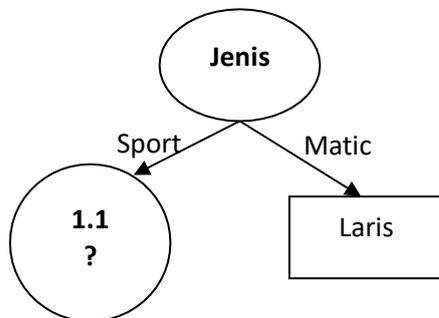
$$Gain(Total, Harga) = 0.795040 - (\frac{7}{25} * 0.868721) + (\frac{16}{25} * 0.543564) + (\frac{2}{25} * 0) = 0.203917$$

$$Gain(Total, Jenis) = 0.795040 - (\frac{12}{25} * 0) + (\frac{13}{25} * 0.995378) = 0.277443$$

$$Gain(Total, Tahun) = 0.795040 - (\frac{8}{25} * 0.950672) + (\frac{17}{25} * 0.680077) = 0.143717$$

Dari hasil tabel 3.5 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah jenis, yaitu 0.277443. dan nilai atribut Matic sudah mengklasifikasikan kasus menjadi laris, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut Sport masih perlu dilakukan perhitungan lagi. dengan demikian Sport dapat menjadi node cabang dari nilai atribut Jenis.

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada gambar 3.1.



Gambar 3.1 Pohon Keputusan Hasil Perhitungan Node 1

- b. Menghitung jumlah kasus, jumlah kasus untuk keputusan laris, jumlah kasus untuk keputusan tidak laris, dan *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut harga, merek, dan tahun yang dapat menjadi node akar dari nilai atribut adalah Sport. Setelah itu, lakukan penghitungan *Gain* untuk tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh tabel 4.7.

Tabel 3.6 Perhitungan Node 1.1

NODE			Jumlah Kasus (S)	Tidak laris (S <sub>1</sub> )	Laris (S <sub>2</sub> )	Entropy	Gain
1.1	Jenis-Sport		13	6	7	0.995727	
	Harga						0.459566
		A	7	4	3	0.985815	
		B	4	0	4	0.000000	
		C	2	2	0	0.000000	
	Tahun						0.032720
		A	8	3	5	0.958042	
		B	5	3	2	0.970950	

Baris total *entropy* pada tabel 3.6 dihitung dengan persamaan sebagai berikut.

$$Entropy (Total, Sport) = (-\frac{7}{13} * \log_2(\frac{7}{13})) + (-\frac{6}{13} * \log_2(\frac{6}{13})) = 0.995727$$

$$Entropy (Sport, Harga A) = (-\frac{4}{7} * \log_2(\frac{4}{7})) + (-\frac{3}{7} * \log_2(\frac{3}{7})) = 0.985815$$

$$Entropy (Sport, Harga B) = (-\frac{0}{4} * \log_2(\frac{0}{4})) + (-\frac{4}{4} * \log_2(\frac{4}{4})) = 0$$

$$Entropy (Sport, Harga C) = (-\frac{2}{2} * \log_2(\frac{2}{2})) + (-\frac{0}{2} * \log_2(\frac{0}{2})) = 0$$

$$Entropy (Sport, Tahun A) = (-\frac{3}{8} * \log_2(\frac{3}{8})) + (-\frac{5}{8} * \log_2(\frac{5}{8})) = 0.958042$$

$$Entropy (Sport, Tahun B) = (-\frac{2}{5} * \log_2(\frac{2}{5})) + (-\frac{3}{5} * \log_2(\frac{3}{5})) = 0.970950$$

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

$$= 0.995727 - ((\frac{7}{13} * 0.985815) + (\frac{4}{13} * 0) + (\frac{2}{13} * 0))$$

$$= 0.459566$$

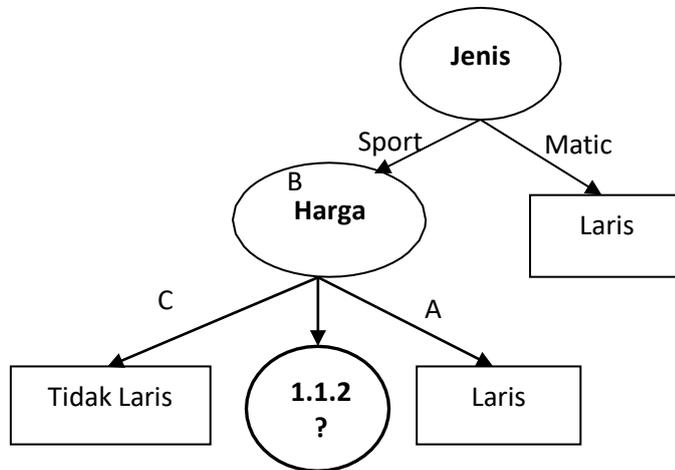
$$Gain (Sport, Tahun) = Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_n)$$

$$= 0.995727 - ((\frac{8}{13} * 0.958042) + (\frac{5}{13} * 0.970950))$$

$$= 0.032720$$

Dari hasil pada tabel 3.6 dapat diketahui bahwa atribut dengan *Gain* tertinggi adalah Harga, yaitu sebesar 0.459566. Ada tiga nilai atribut dari Harga yaitu A, B dan C. Dari ketiga nilai atribut tersebut, nilai atribut A sudah mengklasifikasikan kasus yaitu keputusannya laris dan nilai atribut C sudah mengklasifikasikan kasus yaitu tidak laris, sehingga tidak perlu dilakukan perhitungan lebih lanjut, tetapi untuk nilai atribut B masih perlu dilakukan perhitungan lagi. dengan demikian harga dapat menjadi node cabang dari nilai atribut sport.

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada gambar 3.2 berikut.



Gambar 3.2 Pohon Keputusan Hasil Perhitungan Node 1.1

- c. Menghitung jumlah kasus, jumlah kasus untuk keputusan laris, jumlah kasus untuk keputusan tidak laris, dan *entropy* dari semua kasus dan kasus yang dibagi berdasarkan atribut harga, dan tahun yang dapat menjadi node akar dari nilai atribut adalah Jenis. Setelah itu, lakukan penghitungan *Gain* untuk tiap-tiap atribut. Hasil perhitungan ditunjukkan oleh tabel 4.8.

Tabel 3.7 Perhitungan Node 1.1.2

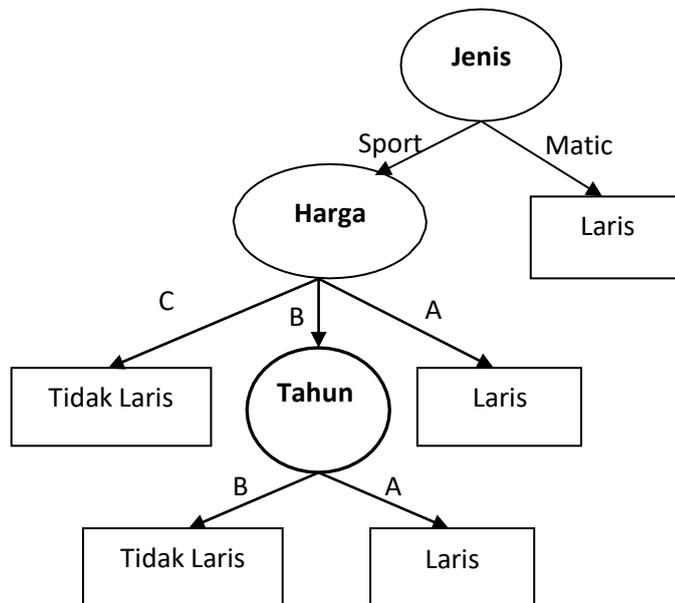
NODE		Jml Kasus (S)	Tidak laris (S1)	Laris (S2)	Entropy	Gain
1.1.2	Jenis Sport dan Harga B	4	0	4	0.000000	
	Tahun					1
	A	3	0	3	0.000000	
	B	3	3	0	0.000000	

Baris total *entropy* pada tabel 3.7 dihitung dengan persamaan sebagai berikut :

$$\begin{aligned}
 \text{Entropy (Total harga B)} &= \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0 \\
 \text{Entropy (B, Tahun A)} &= \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1}\right)\right) + \left(-\frac{1}{0} * \log_2 \left(\frac{1}{0}\right)\right) = 0 \\
 \text{Entropy (B, Tahun B)} &= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0}\right)\right) = 0 \\
 \text{Gain (B, Tahun)} &= \text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \\
 &= 1 - \left(\left(\frac{1}{2} * 0\right) + \left(\frac{1}{2} * 0\right)\right) \\
 &= 1
 \end{aligned}$$

Dari hasil pada tabel 3.7 dapat diketahui bahwa atribut dengan *Gain* tertinggi sama, yaitu sebesar 1. dengan demikian tahun dapat menjadi node cabang dari nilai atribut Harga B. Karena atribut yang lebih berpengaruh. Ada 2 nilai atribut dari tahun yaitu A dan B. Dari kedua nilai atribut tersebut, nilai atribut A sudah mengklasifikasikan kasus menjadi laris dan nilai atribut B sudah mengklasifikasikan kasus menjadi tidak laris sehingga tidak perlu dilakukan perhitungan lebih lanjut,

Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan pada gambar 3.3 berikut.



Gambar 3.3 Pohon Keputusan Hasil Perhitungan Node 1.1.2

**3.3 Mengambil Kesimpulan**

Dari hasil perhitungan Node 1.1.2 di dapat kesimpulan yaitu sebagai berikut:

- Jenis sepeda motor matic memiliki target laris
- Jenis sepeda motor sport melihat dari atribut harga, jika harga A maka target laris, jika harga C maka target tidak laris.
- Jenis sepeda motor sport dengan harga B melihat dari atribut tahun, jika tahun A maka target laris, jika tahun B maka target tidak laris

**3.4 Hasil**

Hasil penelitian dari data mining dalam menentukan jumlah minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan para konsumen yaitu berupa suatu aplikasi berbasis *web programming* yang menerapkan algoritma C4.5 yang mampu memberikan jawaban pasti minat konsumen terhadap sepeda motor berdasarkan perhitungan algoritma C4.5.

### 3.4.1 Hasil Tampilan Antar Muka

Hasil tampilan antarmuka merupakan kegiatan akhir dari proses perancangan sistem, di mana sistem ini akan dioperasikan secara menyeluruh. Berikut ini adalah hasil dari tampilan antarmuka dari data mining dalam menentukan jumlah minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan para konsumen menggunakan algoritma C4.5:

#### 1. Menu Login

*Menu login* berguna untuk mengamankan sistem dari *user – user* yang tidak bertanggung jawab. Berikut tampilan dari *menu login* adalah sebagai berikut :



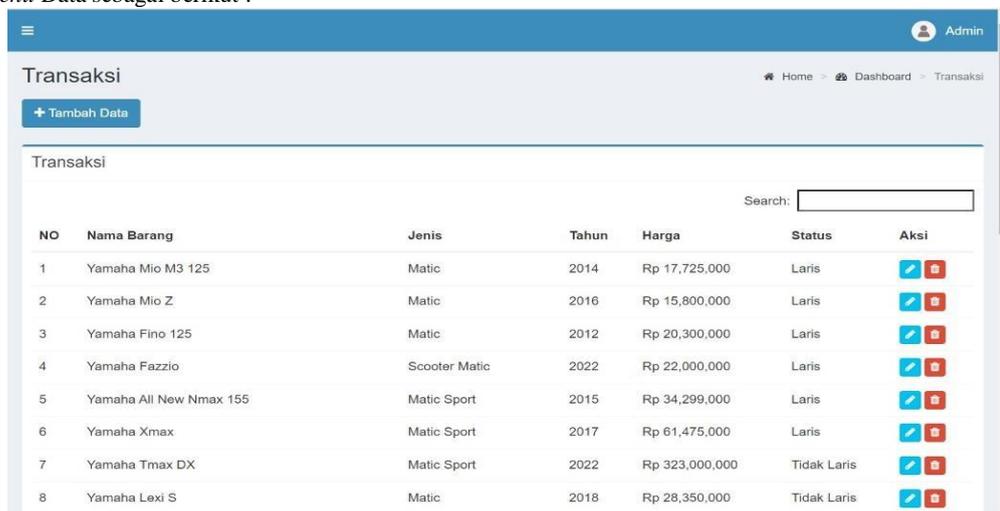
Gambar 3.1 Tampilan *Form Login*



Gambar 3.2 Tampilan *Menu Utama*

#### 2. Menu Data

*Menu Data* digunakan untuk pengolahan data transaksi, simpan data, ubah data, dan penghapusan data. Berikut tampilan pada *menu Data* sebagai berikut :

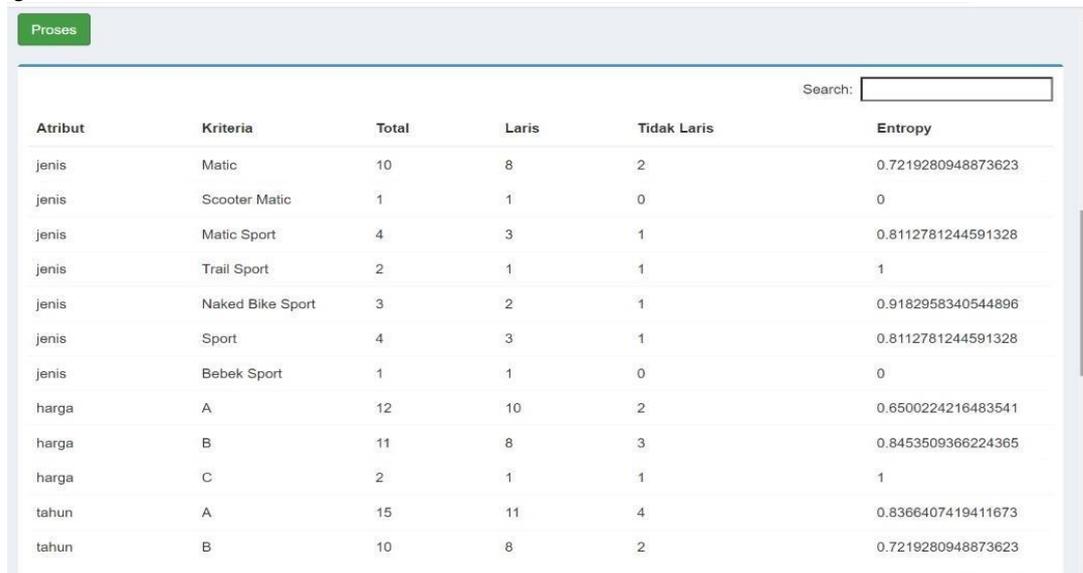


NO	Nama Barang	Jenis	Tahun	Harga	Status	Aksi
1	Yamaha Mio M3 125	Matic	2014	Rp 17,725,000	Laris	 
2	Yamaha Mio Z	Matic	2016	Rp 15,800,000	Laris	 
3	Yamaha Fino 125	Matic	2012	Rp 20,300,000	Laris	 
4	Yamaha Fazzio	Scooter Matic	2022	Rp 22,000,000	Laris	 
5	Yamaha All New Nmax 155	Matic Sport	2015	Rp 34,299,000	Laris	 
6	Yamaha Xmax	Matic Sport	2017	Rp 61,475,000	Laris	 
7	Yamaha Tmax DX	Matic Sport	2022	Rp 323,000,000	Tidak Laris	 
8	Yamaha Lexi S	Matic	2018	Rp 28,350,000	Tidak Laris	 

Gambar 3.3 Tampilan *Form data Transaksi*

3. *Menu Proses*

*Menu Proses* digunakan untuk proses perhitungan berupa penginputan data, proses data. Berikut tampilan pada *menu proses* sebagai berikut :



Atribut	Kriteria	Total	Laris	Tidak Laris	Entropy
jenis	Matic	10	8	2	0.7219280948873623
jenis	Scooter Matic	1	1	0	0
jenis	Matic Sport	4	3	1	0.8112781244591328
jenis	Trail Sport	2	1	1	1
jenis	Naked Bike Sport	3	2	1	0.9182958340544896
jenis	Sport	4	3	1	0.8112781244591328
jenis	Bebek Sport	1	1	0	0
harga	A	12	10	2	0.6500224216483541
harga	B	11	8	3	0.8453509366224365
harga	C	2	1	1	1
tahun	A	15	11	4	0.8366407419411673
tahun	B	10	8	2	0.7219280948873623

Gambar 3.4 Tampilan *Form Proses Perhitungan*

4. *Laporan hasil Perhitungan*

Kemudian adapun tampilan hasil laporan dari proses program sebagai berikut :



**PT. Alfa Scorpii Gatot Subroto**

---

**Laporan Hasil Pola Penjualan Sepeda Motor**

No	Kesimpulan
1	Jenis sepeda motor matic memiliki target laris
2	Jenis sepeda motor sport melihat dari atribut harga, jika harga A maka target laris, jika harga C maka target tidak laris.
3	Jenis sepeda motor sport dengan harga B melihat dari atribut tahun, jika tahun A maka target laris, jika tahun B maka target tidak laris

Medan, 13 September 2022

Gambar 3.5 Hasil Laporan Program

## 4. KESIMPULAN

Berdasarkan perumusan dan pembahasan bab-bab sebelumnya dapat diambil kesimpulan sebagai berikut :

1. Dalam menganalisa masalah yang terjadi dalam menentukan jumlah minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan para konsumen yaitu pertama dengan menentukan atribut terlebih dahulu, kemudian data alternative diberikan nilai atribut untuk dilakukan proses perhitungan, selanjutnya dari

- hasil perhitungan akan didapatkan minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan.
2. Dalam merancang aplikasi data mining dalam menentukan jumlah minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan para konsumen yaitu dengan menggunakan pemodelan UML (*Unified Modeling Language*). Adapun UML yang digunakan diantaranya adalah *use case diagram*, *activity diagram*, dan *class diagram*. Penerapan UML ini sangat membantu dalam merancang dan membangun aplikasi data mining mulai dari awal sampai akhir apa yang akan di kerjakan pada sistem sudah tergambar jelas pada UML. Jadi dalam merancang dan membangun aplikasi sistem pakar tidak lari dari konsep dan melebar kemana-mana.
  3. Dalam mengimplementasikan algoritma C4.5 dalam menentukan jumlah minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan para konsumen menunjukkan hasil yang sangat baik, terbukti dengan melakukan perhitungan berdasarkan nilai atribut di dapat hasil yang akurat.
  4. Dari hasil penelitian, sebelum digunakannya aplikasi, maka pengujian algoritma C4.5 dengan cara menghitung manual. Hal ini tentunya memerlukan banyak waktu, karena dalam menghitung untuk menentukan minat konsumen dalam memilih tipe sepeda motor idaman sesuai kebutuhan harus di hitung persatu atribut. Kemudian setelah algoritma C4.5 diterapkan di aplikasi data mining maka hasil yang didapatkan semakin cepat dan akurat. Hasil nilai yang didapatkan juga sesuai dengan perhitungan manual.

### UCAPAN TERIMAKASIH

Terima Kasih diucapkan kepada kedua orang tua serta keluarga yang selalu memberi motivasi, Doa dan dukungan moral maupun materi, serta pihak-pihak yang telah mendukung dalam proses pembuatan jurnal ini yang tidak dapat disebutkan satu persatu. Kiranya jurnal ini bisa memberi manfaat bagi pembaca dan dapat meningkatkan kualitas jurnal selanjutnya.

### DAFTAR PUSTAKA

- [1] R. Natalia, "Pengaruh Jumlah Dan Jenis Kendaraan Terhadap Pajak Kendaraan Bermotor (PKB) dan Pendapatan Asli Daerah (PAD) di Provinsi Kalimantan Barat," *J. Ekon. Drh.*, 2017.
- [2] U. Brawijaya *et al.*, "PERAN SEPEDA MOTOR BAGI MASYARAKAT kekuatan yang timbul dari dalam diri masyarakat untuk memberikan kemudahan masyarakat berpendapatan rendah terhadap sejumlah pernyataan mengenai mengacu kepada karakter dan gaya hidup seseorang serta nilai-nilai yang aktifitas yang dilakukan bukan ditempat tinggalnya . Artinya keterkaitan antar," hal. 19–20, 2018.
- [3] J. T. Samudra, B. H. Hayadi, dan P. S. Ramadhan, "Komparasi 3 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Kenaikan Jabatan," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 2, hal. 127, 2022, doi: 10.53513/jsk.v5i2.5642.
- [4] R. Girsang, E. F. Ginting, dan M. Hutasuhut, "Penerapan Algoritma C4 . 5 Pada Penentuan Penerima Program Bantuan Pemerintah Daerah," vol. 1, hal. 449–459, 2022.
- [5] N. Ndruru, Y. Syahra, dan E. Elfiriani, "Penerapan Metode Fp-Growth Untuk Penjualan Produk Seni Ukir Pada Buulolo Galery," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 5, no. 1, hal. 45, 2022, doi: 10.53513/jsk.v5i1.4770.
- [6] L. K. Sihombing, T. Tugiono, dan U. F. S. Sitorus Pane, "Implementasi Data Mining Dalam Menganalisa Pola Penjualan Roti," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 3, hal. 228–238, 2022.
- [7] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [8] S. N. Br Sembiring, H. Winata, dan S. Kusnasari, "Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 1, hal. 31, 2022, doi: 10.53513/jursi.v1i1.4784.
- [9] M. G. Suryanata, D. H. Pane, dan M. Hutasuhut, "Implementasi Algoritma K-Means Dalam Mengukur Tingkat Kepuasan Siswa Terhadap Pelayanan Sekolah," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 2, no. 2, hal. 118–125, 2019.
- [10] J. Hutagalung and F. Sonata, "Penerapan Metode K-Means Untuk Menganalisis Minat Nasabah Asuransi," *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 3, pp. 1187–1194, 2021, doi: 10.30865/mib.v5i3.3113.
- [11] E. F. Ginting, F. Sonata, dan M. Hutasuhut, "Penerapan Data Mining Untuk Penyusunan Tata Letak Obat Pada Rumah Sakit Bhayangkara Tk-II Medan Dengan Algoritma Fp-Growth," *SNASTIKOM*, 2019.
- [12] M. Hutasuhut, D. Octaviana, dan J. Halim, "Penerapan Data Mining dalam Menganalisa Pola Kelayakan Siswa Pada Kelas Unggulan Menggunakan Algoritma Iterative Dichotomiser 3 ( ID3 ) pada," vol. 18, no. 2, hal. 154–160, 2019.