
Penerapan data Mining dalam Pemilihan Mobil menggunakan Algoritma C4.5

Fitriana Harahap
UNIVERSITAS POTENSI UTAMA
Jl. Kl. Yos Sudarso KM 6,5 No 3 A Tj. Mulia Medan
Email : fitriana@potensi-utama.ac.id

Abstrak

Untuk memilih kendaraan yang tepat sesuai kebutuhan dan dana yang dimiliki oleh konsumen, memerlukan suatu analisa cermat yang mempertimbangkan banyak kriteria dan factor. Kriteria-kriteria yang dijadikan patokan dalam memilih suatu kendaraan antara lain, harga, suku cadang, volume silinder, daya jual kendaraan tersebut. Untuk mengolah semua kriteria tersebut diperlukan suatu sistem yang dapat memilih dan mengelompokkan kriteria-kriteria yang dipilih oleh konsumen, sehingga dapat membantu konsumen dalam memilih kendaraan yang paling tepat, oleh karena itu dibutuhkan suatu sistem untuk pengambilan keputusan dalam melakukan pembelian mobil. Algoritma C4.5 adalah algoritma klasifikasi data bertipe 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 akan didapatkan beberapa rule. Dalam hal ini penulis mengklasifikasikan pembelian mobil berdasarkan penjualan mobil pada PT. XYZ. Penerapan Algoritma C4.5 ini dapat membantu pembeli untuk memberi bahan pertimbangan pembeli untuk menentukan mobil yang akan dibelinya.

Kata kunci: Pemilihan mobil, Data Mining, Decision Tree, Algoritma C4.5

Abstract

To choose the right vehicle according to the needs and funds owned by consumers, requires a careful analysis that takes into account many criteria and factors. The criteria used as a benchmark in choosing a vehicle, among others, price, spare parts, cylinder volume, the power of the vehicle. To process all these criteria required a system that can select and classify criteria chosen by consumer, so that can assist consumer in choosing the most appropriate vehicle, therefore needed a system for decision making in making car purchase. Algorithm C4.5 is a data classification algorithm of type of decision tree. The decision tree The C4.5 algorithm is constructed with several stages including the selection of attributes as roots, creating branches for each value and dividing instances in branches. The stages of this stage will be repeated for each branch until all the cases on the branch have the same class. From the completion of the decision tree there will be some rules. In this case the authors classify car purchases based on car sales at PT. XYZ. Application of Algorithm C4.5 can help buyers to give buyer consideration to determine the car to be bought.

Keywords: Car selection, Data Mining, Decision Tree, Algorithm C4.5

1. PENDAHULUAN

PT. XYZ adalah dealer yang dipercaya untuk memasarkan kendaraan dengan merk Toyota, PT. XYZ sering kali mengalami kesulitan dalam membantu konsumen menentukan jenis dan tipe kendaraan yang akan mereka beli. Hal ini disebabkan begitu banyaknya merk dan harga bersaing yang ditawarkan oleh pihak produsen otomotif. Untuk memilih kendaraan yang tepat

sesuai kebutuhan dan dana yang dimiliki oleh konsumen, memerlukan suatu analisa cermat yang mempertimbangkan banyak kriteria dan factor. Kriteria-kriteria yang dijadikan patokan dalam memilih suatu kendaraan antara lain, harga, suku cadang, volume silinder, daya jual kendaraan tersebut. Untuk mengolah semua kriteria tersebut diperlukan suatu sistem yang dapat memilih dan mengelompokkan kriteria-kriteria yang dipilih oleh konsumen, sehingga dapat membantu konsumen dalam memilih kendaraan yang paling tepat.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Salah satu teknik yang ada pada data mining adalah klasifikasi [1][2].

Dalam penelitian sebelumnya dengan judul “ Penerapan data mining dalam Memprediksi Pembelian Cat”. Penelitian tersebut dilakukan untuk memprediksi pembelian cat menggunakan algoritma pohon keputusan C4.5 serta untuk mengukur kinerja algoritma C4.5 dari sisi keakuratan hasil prediksi [3]. Dalam studi yang lain, penelitian terkait sudah dilakukan dalam menentukan kemungkinan pengunduran diri mahasiswa di STMIK AMIKOM Yogyakarta, hasil penelitian yang diperoleh adalah kinerja algoritma C4.5 lebih cepat dan akurat dibandingkan dengan algoritma *K-Nearest* (Kusrini dkk,2009) [2].

2. METODE PENELITIAN

Metodologi penelitian akan sangat membantu penulis dalam proses kerja penyelesaian masalah. Penelitian ini dibagi atas tiga tahapan. Tahap pertama adalah studi literatur yang menghasilkan sebuah proposal. Tahap kedua adalah pemodelan dan perancangan sistem, berisi tentang proses pengumpulan data dan bagaimana analisa data terhadap algoritma yang digunakan. Tahap ketiga adalah tahap implementasi, berisi tentang *tools* yang digunakan dan uji coba hasil penelitian.

2.1 Kerangka Kerja Penelitian

1. Mendefinisikan Masalah
Pada tahap ini dilakukan peninjauan ke sistem yang akan diteliti untuk mengamati serta melakukan eksplorasi lebih dalam dan menggali permasalahan yang ada pada sistem yang berjalan saat ini. Tahap ini adalah langkah awal untuk menentukan rumusan masalah dari penelitian.
 2. Menganalisa Masalah
Permasalahan yang ditemukan kemudian akan di analisa. Langkah dalam proses analisa masalah adalah langkah untuk memahami masalah yang telah ditentukan. Dengan menganalisa permasalahan yang telah ditentukan tersebut, maka diharapkan masalah tersebut dapat dipahami dengan baik.
 3. Menentukan Tujuan
Berdasarkan pemahaman dari permasalahan yang telah di analisa, langkah berikutnya adalah menentukan tujuan yang akan dicapai dalam penelitian ini. Tujuan yang akan dicapai merupakan pengetahuan yang dapat digunakan untuk pengambilan keputusan dalam pembelian mobil.
 4. Mempelajari Literatur
Penelitian ini dilakukan untuk melengkapi perbendaharaan kaidah, konsep, teori-teori yang mendukung dalam penyelesaian masalah dalam penelitian ini. Penelitian juga dilakukan melalui buku-buku, jurnal-jurnal, yang ada hubungannya dengan proposal tesis maupun referensi yang lain. Penelitian ini bertujuan untuk mengumpulkan data, baik data pokok maupun data pendukung, dimana semua data tersebut sangat dibutuhkan dalam penelitian.
-

5. Mengumpulkan Data
Mengumpulkan data untuk pelatihan dan pengujian *C4.5*. Semakin banyak data diperoleh, semakin baik dalam menyelesaikan masalahnya. Mengumpulkan data yang akurat dan membagi data tersebut ke dalam kriteria yang sudah ditentukan. Pembagian kriteria digunakan untuk mempermudah dalam pengelompokan data.
6. Menganalisa Algoritma *C4.5*
Setelah data dikumpulkan dilakukan analisa data untuk menyesuaikan proses data yang akan diolah pada Algoritma *C4.5*, dimana tahapan yang dilakukan pada Algoritma *C4.5* dimulai dengan menghitung *entropy*, menghitung *gain*, membuat pohon keputusan dan membuat *rule*.
7. Mengolahan Data
Setelah proses analisa metode pada tahap ini dilakukan pengolahan data yang terdiri dari beberapa tahapan untuk menghasilkan pohon keputusan yaitu Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada.
8. Mengimplementasi
Sesuai dengan pengolahan data maka pada tahap implementasi adalah tentang bagaimana pengolahan datanya diterapkan dalam sebuah tools.
9. Menguji
Pada tahap ini, dilakukan penilaian apakah perangkat lunak yang dikembangkan telah sesuai dengan tujuan yang diharapkan.

2.2 Landasan Teori

2.2.1 KDD (*Knowledge discovery in database*)

Menurut Fayyad Istilah *data mining* dan *knowledge discovery in database* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah *data mining*. Proses KDD secara garis besar dapat dijelaskan sebagai berikut :

1. *Data Selection*
Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing / Cleaning*
Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi focus KDD. Proses pembersihan mencakup antara lain membuang *duplikasi data*, memeriksa data yang *inkosisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).
3. *Transformation*
Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.
4. *Data mining*
Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation / Evaluation*
Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah

pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya .

2.2.2 Data Mining

Data mining (DM) adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah suatu proses yang menggunakan teknik statistik, matematika, kecerdasan tiruan dan *machine-learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [1].

Menurut Yusuf Elmande dan Prabowo ,2012, *data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam *data mining* sangat bervariasi [5].

2.2.3 Klasifikasi

Klasifikasi adalah proses penemuan model yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bias digunakan untuk memprediksi kelas dari obyek yang label kelasnya tidak diketahui. Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (*fase training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi [5]. Proses klasifikasi didasarkan pada empat komponen :

1. *Kelas*
Variabel dependen yang berupa kategorikal yang merepresentasikan “*label*” yang terdapat pada obyek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa.
2. *Predictor*
Variabel independen yang direpresentasikan oleh karakteristik (*atribut*) data. Contohnya : merokok, minum alkohol, tekanan darah, tabungan, asset, gaji.
3. *Training dataset*
Satu *set* data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
4. *Testing dataset*
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

2.2.4 Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain *ID3*, *CART*, dan *C4.5* . Algoritma *C4.5* merupakan pengembangan dari algoritma *ID3* [4].

Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang memprediksikan aturan. Aturan dapat dengan mudah dipahami dengan alami. Dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language* untuk mencari record pada kategori tertentu .

Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Dengan masing-masing rangkaian pembagian, anggota himpunan hasil menjadi mirip satu dengan yang lain [2].

2.2.5 Decision Tree

Pohon keputusan atau dikenal dengan Decision Tree adalah salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node mempresentasikan atribut, cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Node yang paling atas dari decision tree disebut sebagai root [6].

3. HASIL DAN PEMBAHASAN

Proses pada pohon keputusan adalah mengubah bentuk data (tabel) menjadi model pohon, mengubah model pohon menjadi *rule* dan menyederhanakan *rule*. Secara umum algoritma *C4.5* untuk membangun pohon keputusan adalah sebagai berikut [4] :

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap nilai
3. Bagi kasus dalam cabang
4. 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* digunakan rumus di bawah ini :

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

Di mana :

S = Himpunan kasus

A = Atribut

n = Jumlah partisi atribut *A*

$|S_i|$ = Jumlah kasus pada partisi ke-*i*

$|S|$ = Jumlah kasus dalam *S*

Sementara itu, perhitungan nilai *entropy* adalah seperti persamaan 2 di bawah ini :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Di mana :

S = Himpunan kasus *n* = Jumlah partisi *S* *A* = Fitur

P_i = Proporsi dari $|S_i|$ terhadap *S*

Tabel 1. Data Hasil Pra-Proses

No	Harga	Suku Cadang	Volume Silinder	Harga Jual Kembali	Beli
1	Mahal	Rendah	Sedang	Mahal	Ya
2	Mahal	Sedang	Sedang	Mahal	Ya
3	Mahal	Tinggi	Rendah	Mahal	Ya
4	Mahal	Sedang	Sedang	Mahal	Ya
5	Terjangkau	Rendah	Sedang	Terjangkau	Ya
6	Terjangkau	Rendah	Tinggi	Terjangkau	Ya
7	Mahal	Rendah	Sedang	Mahal	Tidak
8	Mahal	Rendah	Sedang	Mahal	Ya
9	Normal	Tinggi	Sedang	Normal	Ya
10	Terjangkau	Rendah	Sedang	Terjangkau	Ya
11	Terjangkau	Rendah	Sedang	Terjangkau	Tidak
12	Terjangkau	Sedang	Sedang	Terjangkau	Tidak
13	Mahal	Rendah	Rendah	Mahal	Tidak
14	Mahal	Sedang	Rendah	Mahal	Tidak
15	Terjangkau	Rendah	Rendah	Terjangkau	Tidak
16	Terjangkau	Sedang	Rendah	Terjangkau	Tidak

17	Mahal	Rendah	Sedang	Mahal	Ya
18	Mahal	Sedang	Sedang	Mahal	Ya
19	Mahal	Tinggi	Tinggi	Mahal	Tidak
20	Normal	Sedang	Tinggi	Normal	Tidak

Perhitungan Entropy dan Gain.

Nilai pada Entropy baris harga dihitung dengan menggunakan persamaan 2 :

$$\text{Entropy (Total)} = \left(-\frac{9}{20} * \log_2\left(\frac{9}{20}\right)\right) + \left(-\frac{11}{20} * \log_2\left(\frac{11}{20}\right)\right) = 0,99277$$

$$\text{Entropy (Harga, Mahal)} = \left(-\frac{4}{11} * \log_2\left(\frac{4}{11}\right)\right) + \left(-\frac{7}{11} * \log_2\left(\frac{7}{11}\right)\right) = 0,94566$$

$$\text{Entropy (Harga, Normal)} = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy (Harga, Terjangkau)} = \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) = 0,98523$$

$$\text{Entropy (Suku Cadang, Tinggi)} = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0,91830$$

$$\text{Entropy (Suku Cadang, Sedang)} = \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) = 0,98523$$

$$\text{Entropy (Suku Cadang, Rendah)} = \left(-\frac{4}{10} * \log_2\left(\frac{4}{10}\right)\right) + \left(-\frac{6}{10} * \log_2\left(\frac{6}{10}\right)\right) = 0,97095$$

$$\text{Entropy (Volume Silinder, Tinggi)} = \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0,91830$$

$$\text{Entropy (Volume Silinder, Sedang)} = \left(-\frac{3}{12} * \log_2\left(\frac{3}{12}\right)\right) + \left(-\frac{9}{12} * \log_2\left(\frac{9}{12}\right)\right) =$$

0,81128

$$\text{Entropy (Volume Silinder, Rendah)} = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) = 0,71193$$

$$\text{Entropy (Daya Jual, Mahal)} = \left(-\frac{4}{11} * \log_2\left(\frac{4}{11}\right)\right) + \left(-\frac{7}{11} * \log_2\left(\frac{7}{11}\right)\right) = 0,94566$$

$$\text{Entropy (Daya Jual, Normal)} = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\text{Entropy (Daya Jual, Terjangkau)} = \left(-\frac{4}{7} * \log_2\left(\frac{4}{7}\right)\right) + \left(-\frac{3}{7} * \log_2\left(\frac{3}{7}\right)\right) = 0,98523$$

Sementara itu, nilai Gain pada baris harga dihitung dengan menggunakan persamaan 1 :

$$\begin{aligned} \text{Gain (Total, Harga)} &= 0,99277 - \left(\left(\frac{11}{20} * 0,94566\right) + \left(\frac{2}{20} * 1\right) + \left(\frac{7}{20} * 0,98523\right)\right) \\ &= 0,02783 \end{aligned}$$

$$\begin{aligned} \text{Gain (Total, Suku Cadang)} &= 0,99277 - \left(\left(\frac{3}{20} * 0,91830\right) + \left(\frac{7}{20} * 0,98523\right) + \right. \\ &\left. \left(\frac{10}{20} * 0,97095\right)\right) = 0,02472 \end{aligned}$$

$$\begin{aligned} \text{Gain (Total, Volume Silinder)} &= 0,99277 - \left(\left(\frac{3}{20} * 0,91830\right) + \left(\frac{12}{20} * 0,81128\right) + \right. \\ &\left. \left(\frac{5}{20} * 0,72193\right)\right) = 0,18778 \end{aligned}$$

$$\text{Gain (Total, Daya Jual)} = 0,99277 - \left(\left(\frac{11}{20} * 0,94566 \right) + \left(\frac{2}{20} * 1 \right) + \left(\frac{7}{20} * 0,98523 \right) \right) = 0,01581$$

Tabel 2. Perhitungan Node 1

Node			Jlm Kasus	Tidak	Ya	Entropy	Gain
1	Total		20	9	11	0.99277	
	Harga						0.02783
		Mahal	11	4	7	0.94566	
		Normal	2	1	1	1.00000	
		Terjangkau	7	4	3	0.98523	
	Suku cadang						0.02472
		Tinggi	3	1	2	0.91830	
		Sedang	7	4	3	0.98523	
		Rendah	10	4	6	0.97095	
	Volume Silinder						0.18778
		Tinggi	3	2	1	0.91830	
		Sedang	12	3	9	0.81128	
		Rendah	5	4	1	0.72193	
	Daya Jual						0.01581
		Mahal	11	4	7	0.94566	
		Normal	2	1	1	1.00000	
		Terjangkau	7	4	3	0.98523	

Dari hasil pada tabel 2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Volume Silinder, yaitu sebesar 0.18778. Pengujian terhadap analisa, sangat penting dilakukan untuk menentukan dan memastikan apakah hasil analisa tersebut telah sesuai dengan keputusan yang diharapkan. Untuk menguji kebenaran dari hasil pengolahan data yang dilakukan secara manual, maka dapat menggunakan salah satu software aplikasi WEKA 3.5.5 *knowledge Explorer*.

3.1 Langkah - langkah Implementasi

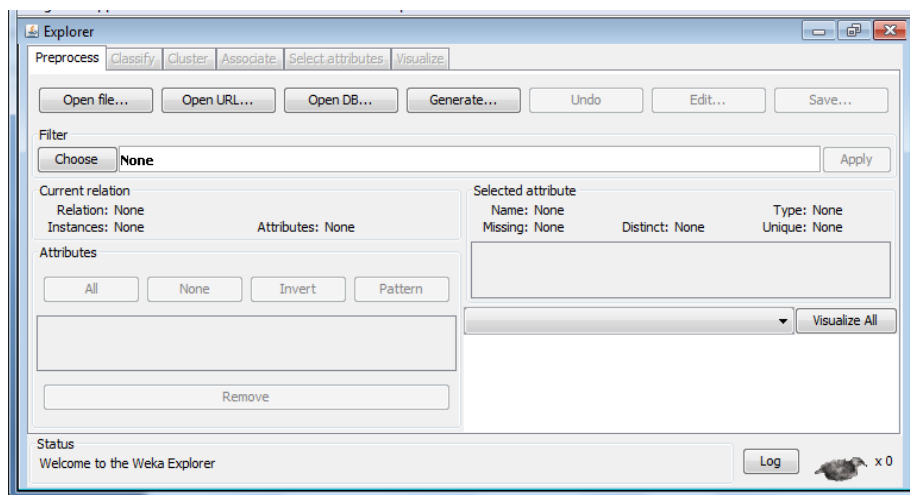
Seluruh variabel yang terdiri dari atribut kondisi dan atribut keputusan yang digunakan untuk menentukan pembelian cat disimpan pada Microsoft excel dengan nama file datamobil.xls (yang berisi kasus atau kriteria dalam menghasilkan rule). Selanjutnya proses transformation data dimana File datamobil.xls kemudian disimpan dengan extension csv, selanjutnya file dibuka dengan notepad atau editor teks lainnya dan data sudah berubah dalam format command separated.

Langkah selanjutnya adalah membuka aplikasi *Weka5.5*, Buka Software WEKA 3.5.5, kemudian muncul tampilan awal seperti gambar 1 di bawah ini :



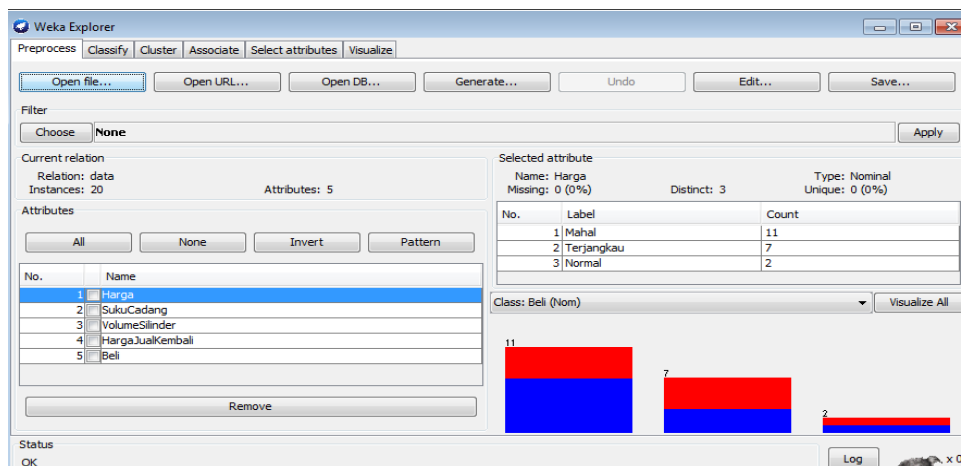
Gambar 1. Tampilan Depan

Langkah selanjutnya adalah membuka aplikasi *Weka 3.5.5*, *Weka5.5*, dimulai dari proses *Explorer* data yang akan dianalisa menggunakan *Weka 3.5.5*. *Explorer* data dilakukan dengan cara memilih menu *applications* dan pilih menu *explorer*. Tampilan halaman *Explorer* data dapat dilihat pada gambar 2 berikut:



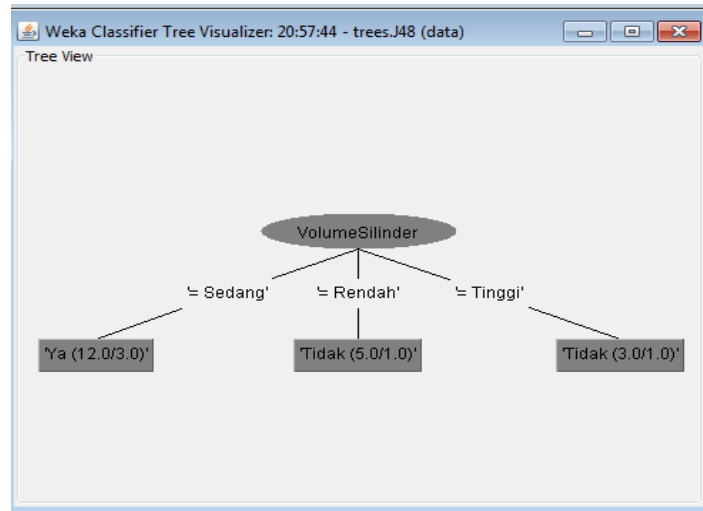
Gambar 2. Tampilan Weka Explorer

Klik *Open File*, lalu pilih *directori* mana data *.csv* tersebut disimpan kemudian kemudian tampil hasil setelah dokumen *csv* dibuka, data dapat dilihat pada gambar 3 berikut:



Gambar 3. Hasil Open File

Kemudian data disesuaikan dengan menambahkan informasi awal dan data tersebut sudah dapat digunakan sebagai inputan dalam WEKA 3.5.5. Selanjutnya Klik buttom Choose, pilih J48 dalam pembentukan pohon keputusan dan klik menu Start . Pada tahap ini proses data mining dilakukan dengan memilih algoritma yang akan dipakai dalam menghasilkan sebuah pohon keputusan, algoritma C4.5.



Gambar 4. Tree View

4. KESIMPULAN

Berdasarkan hasil penelitian yang penulis lakukan pada PT. XYZ, maka penulis dapat menarik kesimpulan bahwa pembelian mobil dengan menggunakan metode Data Mining khususnya Algoritma C4.5 akan bermanfaat sekali dalam proses pengambilan keputusan dalam pembelian mobil.

1. Yang menjadi faktor tertinggi yang mempengaruhi pembelian mobil pada PT. XYZ adalah faktor volume silinder.
2. Volume Silinder mobil Tinggi maka pembeli tidak minat membeli mobil pada PT. XYZ.
3. Volume Silinder mobil Sedang maka pembeli berminat membeli mobil pada PT. XYZ.
4. Ternyata Harga tidak mempengaruhi pembelian mobil pada PT. XYZ
Selanjutnya penulis menyarankan agar dapat membandingkan metode pengambilan keputusan dengan metode Data Mining dengan metode lainnya.

5. SARAN

Berdasarkan hasil penelitian ini, ada beberapa saran yang dalam menentukan pembelian mobil yaitu:

1. Diharapkan Metode *Data Mining* khususnya *Algoritma C4.5* dapat diterapkan dalam proses pembelian mobil dan proses lainnya yang berhubungan dengan pengambilan keputusan khususnya pada PT. XYZ Medan.
2. Hasil uji coba Decision Tree dengan software WEKA membantu menguatkan hasil penelitian yang dilakukan.

3. Penulis menyarankan agar dapat membandingkan metode pengambilan keputusan dengan metode *Data Mining* dengan metode lainnya.

DAFTAR PUSTAKA

- [1] Efraim Turban, Jay E. Aronson, Ting Peng Liang, 2005. Decision Support System and Intelligent Systems Edisi 7 Jilid 1, Andi Yogyakarta.
 - [2] Kusri, (2009). Algoritma Data Mining ,Andi Yogyakarta
 - [3] Fitriana Harahap, 2015. “ Penerapan Data Mining dalam Memprediksi Pembelian Cat (Konferensi Nasional Sistem & Informatika 2015 9 – 10 Oktober 2015)”.
 - [4] Muhammad Syahril, 2011 .”Konversi Data Training Tentang Penyakit Hipertensi Menjadi Bentuk Pohon Keputusan dengan Teknik Klasifikasi Menggunakan Tools Rapid Miner 4.1 ,Vol 10, No.2,Mei.
 - [5] Yusuf Elmande and Prabowo Pudjo Widodo, "Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras : Studi Kasus Perum Bulog Divre Lampung," TELEMATIKA MKOM, Maret 2012.
 - [6] Eka Budi Rahayu, "Pemilihan ALGORITMA C4.5 UNTUK PENJURUSAN SISWA SMA NEGERI 3 PATI," Universitas Dian Nuswantoro Semarang.
-