

IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI JUMLAH PEREDARAN KENDARAAN RODA EMPAT DI KOTA TERNATE MENGGUNAKAN METODE C.45

Sahriar Hamza

Program Studi Teknik Informatika Fakultas Teknik

Universitas Muhammadiyah Maluku Utara

Jl. K.H Ahmad Dahlan No. 100, Kel.Sasa Kota Ternate Selatan

Abstrak

The purpose of the research is to determine the growth rate of Four-Wheel Vehicles in Ternate. The implementation of data mining using the C.45 Algorithm aims to predict the circulation of four-wheeled vehicles in Ternate through the year as an attribute in determining a value. The results are obtained from the gain and ratio through the classification process toward brand and number of vehicles based on the year. The data shows that there was an increase in the number of circulation of four-wheeled vehicles in Ternate in 2013 to 2016, which was 0.91%. This will predict to increase in the following year.

Key Words: *Data Mining, Algorithm C.45, four-wheel vehicles*

1. PENDAHULUAN

Pemanfaatan data yang ada di dalam sistem informasi untuk menunjang kegiatan pengambilan keputusan, tidak cukup hanya mengandalkan data operasional saja, diperlukan suatu analisis data untuk memnfaatkan gudang data yang sudah dimiliki untuk menggali informasi yang berguna dalam membantu pengambilan keputusan, hal ini mendorong munculnya cabang ilmu baru untuk mengatasi masalah pengalihan informasi atau pola yang penting atau menarik dari data dalam jumlah besar, yang disebut dengan data mining. Dengan menggunakan data mining, setiap kumpulan data dapat memberikan pengetahuan-pengetahuan yang sebelumnya tersembunyi dalam database sehingga menjadi informasi yang berharga bagi suatu organisasi untuk memperoleh informasi yang menunjang setiap kegiatan pada pengambilan suatu keputusan. Data mining dengan menggunakan Algoritma C.45 dapat digunakan untuk menyusun sistem yang mempunyai kemampuan melihat pola peredaran kendaraan

yang ada di kota ternate. Untuk dapat mengetahui pola tersebut, dibutuhkan data yang sesuai dengan kasus tersebut. Berdasarkan paper yang dipresentasikan pada International Conference on Data Mining (ICDM), mengidentifikasi ada 10 metode data mining terbaik yang terbaik yang bisa di gunakan diantaranya Algoritma C.45. Dengan menerapkan teknik ini akan dibangun sebuah pohon keputusan atau decision tree untuk mendapatkan informasi yang akurat. Salah satu teknik data mining dengan menggunakan decision tree dimana algoritma C.45 merupakan algoritma klasifikasi data dengan teknik pohon keputusan yang dapat mengolah data numerik atau kontinyu dan diskrit. Algoritma ini dapat menangani nilai atribut yang hilang, dan menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma- algoritma yang lain.

2. METODOLOGI

Data Mining sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang

bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik tertentu dalam pengenalan pola seperti Statistik dan Matematika (Kusrini, 2009).

Algoritma C.45

Algoritma C.45 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 digunakan sebagai parameter dalam melakukan klasifikasi data (Sunjana, 2010).

Secara umum Algoritma C4.5 membangun pohon keputusan dengan mengikuti proses sebagai berikut:

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.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada maka digunakan rumus sebagai berikut:

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

total *entropy* dapat dihitung dengan persamaan 1 sebagai berikut :

Entropy (Total)

$$\left(-\frac{4}{12} * \log_2\left(\frac{4}{12}\right)\right) + \left(-\frac{8}{12} * \log_2\left(\frac{8}{12}\right)\right) = 0.918296.$$

$$Entropy(Total) = 0.918296$$

Entropy (tahun).

$$Entropy(Total, Tahun, 2013) \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.918$$

$$Entropy(Total, Tahun, 2014) \left(-\frac{0}{3} * \log_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3} * \log_2\left(\frac{3}{3}\right)\right) = 0$$

$$Entropy(Total, Tahun, 2015) \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0.918$$

$$Entropy(Total, Tahun, 2016) \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.918$$

Entropy Merek

$$Entropy (Total, Merek Suzuki) \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) = 0.811$$

$$Entropy (Total, Merek Honda) \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) = 0.811$$

$$Entropy (Total, Merek Toyota) \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} * \log_2\left(\frac{2}{4}\right)\right) = 1.0000$$

Entropy Jumlah

$$Entropy (Total, Jumlah, A) \left(-\frac{3}{8} * \log_2\left(\frac{3}{8}\right)\right) + \left(-\frac{5}{8} * \log_2\left(\frac{5}{8}\right)\right) = 0.9988$$

$$Entropy (Total, Jumlah, B) \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) = 0$$

Proses Klasifikasi

Proses klasifikasi tiap-tiap *field* tabel peredaran Kendaraan roda empat berdasarkan Tahun, Merek dan Jumlah adalah sebagai berikut :

Tabel 1.1 tabel klasifikasi berdasarkan tahun

Klasifikasi Tahun	Naik	Turun
2013	2	1
2014	3	0
2015	1	2
2016	2	1

Tabel 1.2 tabel klasifikasi berdasarkan merek

Klasifikasi Merek	Naik	Turun
Toyota	3	1
Suzuki	3	1
Honda	2	2

Tabel 1.3 tabel klasifikasi berdasarkan Jumlah

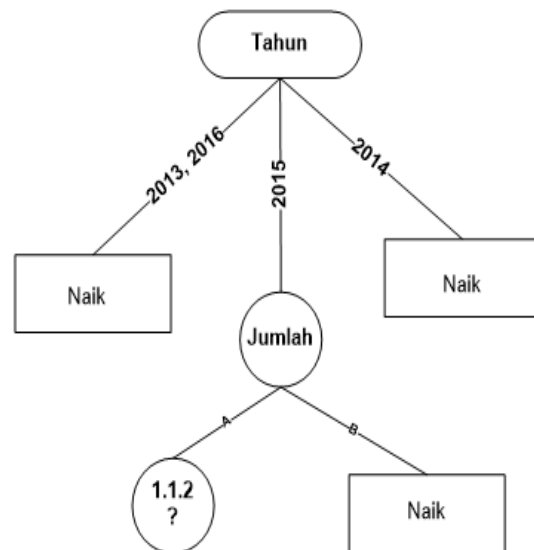
Klasifikasi Jumlah	Kode	Naik	Turun
20 - 70	A	5	3
70 - 140	B	4	0

Berikut ini adalah langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C.45 dalam menyelesaikan permasalahan pada tabel-tabel di atas maka perlu dilakukan perhitungan jumlah kasus untuk keputusan Naik, dan jumlah kasus untuk keputusan Turun. Untuk itu di hitung *Entropy* dari semua kasus dan dibagi berdasarkan atribut tahun, merek, jumlah. Setelah itu, lakukan penghitungan *gain* untuk setiap atribut. Hasil perhitungan perhitungan dapat dilihat pada tabel di bawah ini.

Tabel 1.4 hasil perhitungan node 1

Node		Jumlah Kasus (S)	Turun (S ₁)	Naik (S ₂)	Entropy	Gain
1	TOTAL	12	4	8	0.918296	
	Tahun					0.918296
	2013	3	1	2	0.918296	
	2014	3	0	3	0	
	2015	3	2	1	0.918296	
	2016	3	1	2	0.918296	
	Merek					0.044110
	Toyota	4	1	3	0.811278	
	Suzuki	4	1	3	0.811278	
	Honda	4	2	2	1	
	Jumlah					0,282006
	A	8	3	5	0.54434	
	B	4	0	4	0	

Dari hasil tabel diatas dapat diketahui bahwa atribut dengan Gain tertinggi adalah adalah 0.918296. Ada 4 nilai atribut dari tahun 2013-2016. nilai atribut tersebut sudah mengklasifikasikan kasus yaitu keputusannya adalah Naik sehingga proses Pohon keputusan yang terbentuk sampai tahap ini ditunjukkan sebagai berikut :



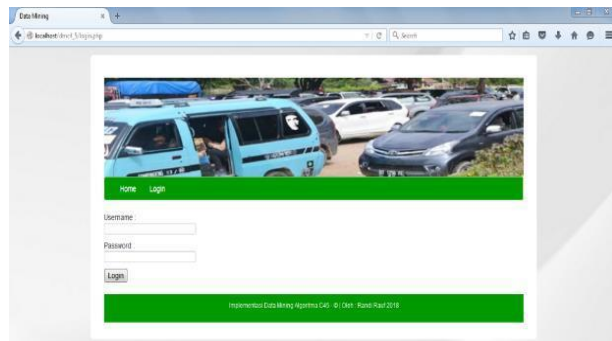
Gambar 1.1 pohon Keputusan

3. Hasil dan Pembahasan

Dalam pembahasan ini akan dijelaskan mengenai pengujian Aplikasi dengan menggunakan *form-form* yang telah di desain sesuai dengan hasil perhitungan yang telah di lakukan.

Form Login Aplikasi

Form ini merupakan halaman login sistem dari implementasi data mining untuk peredaran kendaraan roda empat. Halaman ini berisi hak akses user atau pengguna dan administrator.



Form Menu Utama Aplikasi

Jika proses login sukses, maka admin dapat mengakses halaman utama. Halaman ini berisi menu pengelolaan semua data yang ada dalam database aplikasi. Fasilitas ini menggunakan dialog antarmuka yang userfriendly agar komunikasi user dengan aplikasi menjadi lebih mudah. Menu ini adalah menu untuk melakukan identifikasi user yang akan root ke dalam Aplikasi kemudian user atau admin dapat melihat menu apa saja yang dimiliki oleh aplikasi ini.



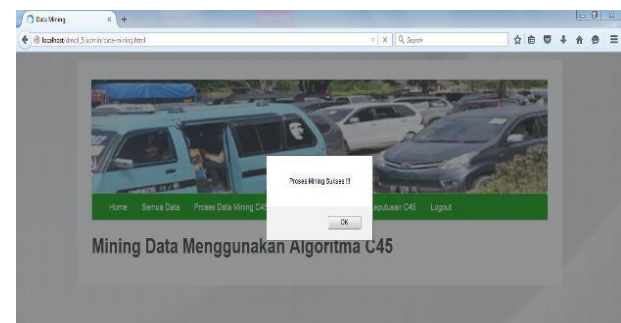
Form Data Survey

Form Semua data survey merupakan form yang digunakan untuk mengelola data kendaraan roda empat.



Form Proses Data Mining

Form proses klasifikasi merupakan form yang digunakan untuk mengklasifikasikan data peredaran kendaraan roda empat.



Form Perhitungan Algoritma C.45

Form perhitungan C.45 merupakan form yang digunakan untuk perhitungan kendaraan roda empat yang telah diklasifikasikan sebelumnya dengan menggunakan field-field yang ada pada tabel. Sehingga di peroleh hasil sebagai berikut



Form Pohon Keputusan

Form pohon keputusan merupakan form yang digunakan untuk menampilkan hasil keputusan dari data perhitungan peredaran kendaraan roda empat dengan menggunakan Algoritma C.45



4. KESIMPULAN

Berdasarkan analisa, perancangan dan implementasi yang telah dilakukan pada aplikasi untuk peredaran kendaraan roda empat Kota Ternate dengan implementasikan data mining metode C4.5, dapat diambil kesimpulan sebagai berikut :

1. data untuk peredaran kendaraan empat dapat diimplementasikan dengan baik menggunakan metode algoritma C45.
2. Pengujian yang dilakukan terhadap data tahun 2013 sampai 2016 peningkatan hanya 0.91% hal ini akan terus miningkat di tahun-tahun yang akan datang.

DAFTAR PUSTAKA

David Hartanto, dkk, 2014. *Implementasi Data Mining dengan Algoritma C.45 untuk Memprediksi Tingkat Kelulusan Mahasiswa*. Teknik Informatika. Universitas Multimedia Nusantara. ISSN 2085-4552.

Hakim, Lukmanul dan Musalini, Uus. 2004. *Cara Cerdas menguasai Layout, Desain dan Aplikasi Web*. Jakarta. Penerbit PT. Elex Media Komputindo.

Jogiyanto HM. (2005). *Analisis dan Desain Sistem Informasi*. Andi, Yogyakarta.

Kusrini & Emha Taufiq Luthfi, 2009, *Algoritma Data Mining*, CV Andi, Yogyakarta.

Supriansyah, dkk, 2006. *30 menit menjadi webmaster*. Abndang : Oase Media Sunjana, 2010, *Seminar Nasional Aplikasi Teknologi Informasi 2010*.

Snati 2010. *Aplikasi Mining Data Mahasiswa Dengan Metode Klasifikasi Decision Tree*, 24-29.