

Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur

Yahya¹, Reni Zuliana²

Fakultas Teknik Universitas Hamzanwadi
ayhay7078@gmail.com, liareny72@gmail.com

Abstrak

Dalam dunia kerja ataupun dunia pendidikan, bahkan instansi pemerintah ataupun umum telah menggunakan alat elektronik yang canggih terutama komputer. komputer banyak dimanfaatkan untuk membantu pekerjaan manusia dalam menyelesaikan pekerjaannya agar lebih mudah dan efisien. Karena dengan komputer kita dapat dipermudah dalam membuat informasi yang lebih dimengerti oleh banyak orang. Selain komputer, software (perangkat lunak) juga sangat diperlukan dalam mengolah sebuah data guna mempermudah seseorang dalam memproses datanya menjadi data yang lebih bermanfaat. Kantor Dinas Lingkungan Hidup dan Kebersihan Kabupaten Lombok Timur Kecamatan Selong merupakan salah satu kantor yang menerima subsidi BBM dari pemerintah pusat. Oleh sebab itu prediksi menggunakan algoritma Decition Tree (C4.5) dalam menghitung jumlah penggunaan BBM perlu dilakukan guna mengoptimalkan penggunaan BBM dengan sebaik-baiknya. Karena banyak petugas ataupun oknum yang tidak bertanggung jawab terkena kasus karena penyalahgunaan penggunaan BBM bersubsidi. Dengan prediksi ini maka penggunaan dapat diatur secara optimal serta penggunaan yang berlebihan dapat diatasi.

Kata kunci : *Algoritma, BBM, Decition Tree (C4.5).*

Abstract

In the world of work or the world of education, even government or public agencies have used sophisticated electronic devices, especially computers. computers are widely used to help human work in completing their work to make it easier and more efficient. Because with our computers can be made easier in making information that is more understood by many people. Besides computers, software (software) is also very necessary in processing a data to make it easier for someone to process their data into more useful data. Environmental and Hygiene Office East Lombok District Selong District is one of the offices that receives fuel subsidies from the central government. Therefore, the prediction using the Decition Tree algorithm (C4.5) in calculating the amount of BBM usage needs to be done in order to optimize the use of BBM as well as possible. Because many officers or individuals are not responsible for cases due to misuse of subsidized fuel. With this prediction, usage can be adjusted optimally and excessive use can be overcome.

Keywords: *Algorithm, BBM, Decition Tree (C4.5)*

1. Pendahuluan

Teknologi informasi telah berkembang pesat di kalangan masyarakat. Banyak orang yang memanfaatkan kecanggihan teknologi untuk mempermudah pekerjaan mereka dan contoh salah satunya adalah teknologi komputer. Kecanggihan komputer dimanfaatkan untuk menyampaikan informasi yang berasal dari data yang telah diolah sehingga orang lain dipermudah dalam menerima informasi. Dinas Lingkungan Hidup dan Kebersihan merupakan salah satu instansi pemerintah yang memiliki banyak data. Salah satunya adalah data penggunaan Bahan Bakar Minyak (BBM). Setiap bulannya Kantor Dinas Lingkungan Hidup dan Kebersihan membuat laporan untuk mengetahui jumlah penggunaan BBM yang telah dipergunakan untuk melaksanakan kegiatan kantor. Prediksi jumlah penggunaan BBM setiap bulannya harus diketahui, karena BBM tersebut merupakan subsidi dari pemerintah yang harus dipergunakan dengan baik. Oleh sebab itu penggunaan BBM harus sesuai dengan ketersediaan yang telah ditetapkan oleh pemerintah pusat. Informasi prediksi penggunaan BBM diperlukan oleh kepala dinas dan pengelola petugas bagian BBM. Salah satu metode yang dapat digunakan untuk mengatasi permasalahan di atas adalah dengan menggunakan algoritma *decition tree* (C4.5) atau sering kali disebut dengan pohon keputusan. Algoritma *decition tree* (C4.5) merupakan salah satu jenis algoritma

pengelompokan data berdasarkan kriteria tertentu. Dengan menggunakan algoritma *decition tree* (C4.5) ini, penggunaan BBM akan dikelompokkan menjadi 4 kelompok, (penggunaan banyak, penggunaan standar, penggunaan cukup, penggunaan kurang). Berdasarkan latar belakang yang telah dikemukakan diatas maka penulis mengambil judul “Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur”.

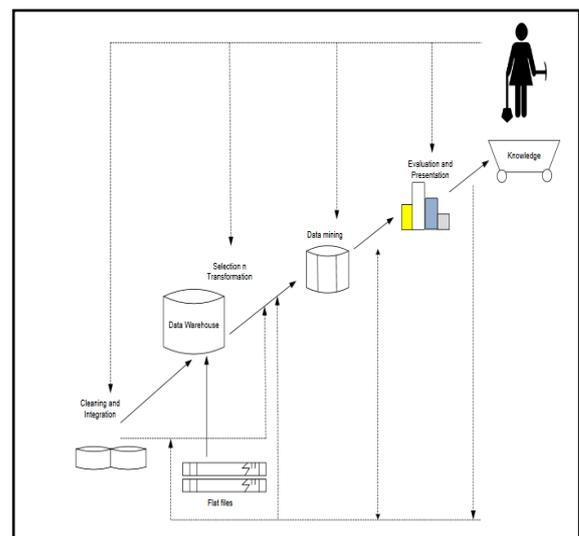
2. Tinjauan Pustaka

Dalam penelitian ini, diperlukan beberapa definisi dan pengertian. Beberapa hal diantaranya : Data Mining dan Algoritma Decition Tree (C4.5).

2.1. Data Mining

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi Data Mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011). Data Mining meramalkan tren dan sifat perilaku bisnis yang sangat berguna untuk mendukung pengambilan keputusan penting. Analisis yang diotomatisasi yang dilakukan oleh Data Mining melebihi yang dilakukan oleh sistem pendukung keputusan tradisional yang sudah banyak

digunakan (Moertini, 2002). Secara khusus, koleksi metode yang dikenal sebagai Data Mining menawarkan metodologi dan solusi teknis untuk mengatasi analisis data medis dan konstruksi prediksi model (Bellazzi & Zupanb, 2008). Secara umum, tugas Data Mining dapat diklasifikasikan menjadi dua kategori yaitu, deskriptif dan prediktif. Tugas pertambangan deskriptif mengkarakterisasi sifat umum data dalam database pertambangan prediktif, tugas data pada saat ini untuk membuat prediksi (Han & Kamber, 2007). Penerapan metode machine learning untuk database besar disebut Data Mining, dalam Data Mining volume besar data diproses untuk membangun sebuah model sederhana, misalnya memiliki akurasi prediksi yang tinggi. Area aplikasi Data Mining melimpah, selain retail, bank, menganalisis data masa lalu untuk membangun model aplikasi kredit, deteksi penipuan dan pasar saham. Dalam manufaktur, model pembelajaran digunakan untuk optimasi, kontrol, dan pemecahan masalah. Dalam pengobatan, program pembelajaran adalah digunakan untuk diagnosis medis. Dalam telekomunikasi, pola panggilan dianalisis untuk optimasi jaringan dan memaksimalkan kualitas layanan. Dalam ilmu pengetahuan, sejumlah besar data dalam fisika, astronomi, dan biologi dapat hanya dianalisis cukup cepat oleh komputer (Alpaydin, 2010). Model Data Mining memberikan contoh penerapannya pada berbagai algoritma dan dataset yang besar (Larose, 2007).



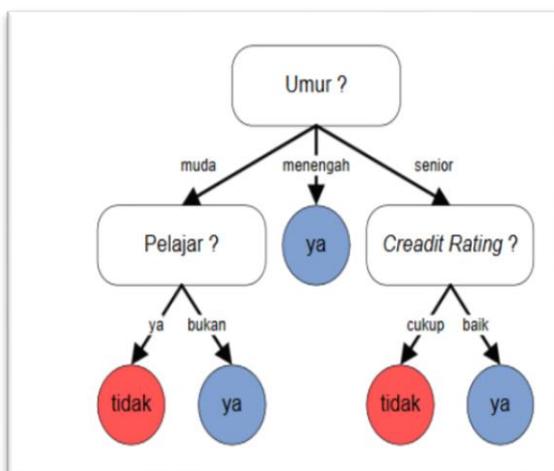
Gambar 2.1 Tahapan Data Mining

2.2. Algoritma Decision Tree (C4.5)

Pohon keputusan mirip sebuah struktur pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan attribut, setiap cabang menggambarkan hasil dari attribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat dikonversi ke aturan klasifikasi. Secara umum, keputusan pengklasifikasi pohon memiliki akurasi yang baik, namun keberhasilan penggunaan tergantung pada data yang akan diolah. (C4.5) didesain oleh J. Ross Quinlan, dinamakan (C4.5) karena merupakan keturunan dari pendekatan ID3 untuk membangun pohon keputusan. (C4.5) merupakan algoritma yang cocok digunakan untuk masalah klasifikasi pada machine learning dan Data Mining (Wu & Kumar, 2009). (C4.5) memetakan atribut dari kelas sehingga dapat digunakan untuk menemukan prediksi terhadap data yang belum muncul.

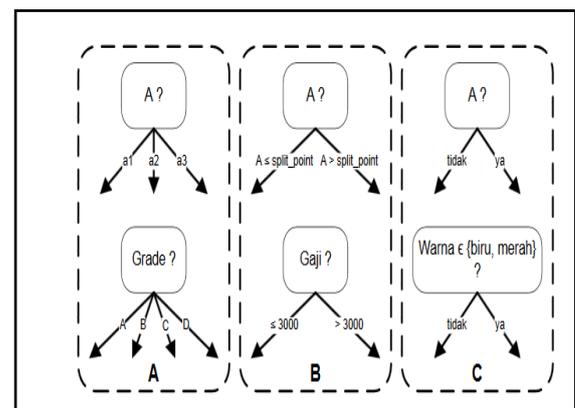
Pohon keputusan sendiri merupakan pendekatan “divide and conquer” dalam mempelajari masalah dari sekumpulan data independen yang digambarkan dalam bagan pohon (Witten, Frank, & Hall, 2011). Pohon keputusan juga merupakan sekumpulan pertanyaan yang tersusun secara sistematis, dimana setiap pertanyaan yang ada menentukan percabangan berdasarkan nilai atribut dan berhenti pada daun dari pohon yang merupakan prediksi dari kelas variabel. Semua metode yang menghasilkan pohon keputusan bekerja mulai dari akar yang paling atas, dengan menguji data yang ada dan dipecah kedalam setiap cabang dari pohon yang lebih kecil. Keputusan dipilih berdasarkan kesesuaian antara data yang diuji dengan aturan dari cabang pohon keputusan yang ada hingga mencapai daun dan nilai kelas yang terdapat pada daun diberikan pada tupel data yang diperiksa tersebut.

terdiri dari umur, pekerjaan dan credit rating. Node internal digambarkan dalam bentuk persegi dan node daun digambarkan dalam bentuk oval. Didalam pohon keputusan, node pusat merupakan atribut dari data yang diuji (tuple), cabang merupakan hasil dari pengujian atribut, dan daun merupakan kelas yang terbentuk (Han & Kamber, 2006). Jika atribut yang diuji merupakan nilai nominal, biasanya jumlah cabang yang terbentuk akan sejumlah dengan jumlah kemungkinan dari nilai atribut yang dipilih sesuai dengan metode pemilihan atribut. Metode pemilihan atribut akan membuat daftar atribut yang paling berpengaruh dalam penyusunan cabang pohon keputusan agar cabang yang terbentuk terdiri dari data yang sejenis (Han & Kamber, 2006). Oleh karena itu, semakin mirip data dalam cabang yang terbentuk akan membuat pohon keputusan yang semakin akurat.



Gambar 2.3 Contoh pohon keputusan

Pada gambar 2.3, digambarkan konsep dari buys komputer, yang digunakan untuk prediksi pelanggan yang ingin membeli komputer. Data



Gambar 2.4 Pembuatan cabang pohon keputusan

Dalam pembuatan cabang, ada tiga kemungkinan yang akan terjadi sesuai dengan jenis data. Jika A merupakan salah satu atribut

dari data yang diuji, maka kemungkinan yang akan terjadi akan bergantung pada nilai A :

- a. Jika A bernilai diskrit, maka cabang akan terbentuk untuk setiap nilai A. Atribut A akan dikeluarkan dari daftar atribut yang perlu diperiksa karena setelah cabang A terbentuk nilai atribut A pada setiap cabang akan selalu sama.
- b. Jika A bernilai kontinyu, maka akan terbentuk 2 buah cabang, dimana $A \leq$ dari nilai perpecahan, dan $A >$ dari nilai perpecahan. Nilai perpecahan ditentukan oleh metode pemilihan atribut saat membuat daftar atribut.
- c. Jika A bernilai Diskrit dan biner, maka akan terbentuk 2 cabang yaitu cabang untuk nilai benar, dan cabang untuk nilai salah.

Dalam pembuatan pohon keputusan, setiap algoritma menerapkan ukuran pemilihan atribut yang berbeda-beda. Ukuran pemilihan atribut merupakan ukuran yang digunakan dalam menentukan kriteria yang terbaik untuk mengelompokkan tuple. Ukuran pemilihan atribut ini juga disebut sebagai *splitting rules* karena menentukan bagaimana data akan dipisahkan kesetiap cabang. (C4.5) yang merupakan pengembangan dari ID3 menggunakan *Information gain* untuk ukuran pemilihan atribut. *Information gain* diciptakan oleh Claude Shannon dengan mempelajari nilai informasi dari data, dan menggunakan nilai tersebut sebagai acuan dalam menentukan atribut yang akan digunakan dalam menyusun

pohon keputusan (Han & Kamber, 2006). Atribut yang dipilih akan menghasilkan partisi dengan data yang lebih seragam, dan dapat menghasilkan pohon keputusan yang sesederhana mungkin dengan perulangan yang sedikit. Pohon keputusan menjadi sangat populer karena dalam pembuatan klasifikasi pohon keputusan tidak memerlukan pengetahuan yang banyak ataupun pengaturan yang rumit. Pohon keputusan mampu mengolah data dengan dimensi yang besar, serta menghasilkan pengetahuan dalam diagram pohon yang mudah dimengerti oleh manusia. Metode pembelajaran dalam membuat pohon keputusan tergolong mudah dan cepat, serta memiliki akurasi tinggi. Pohon keputusan banyak diaplikasikan dalam berbagai bidang seperti obat-obatan, produksi, analisa keuangan, astronomi, dan biologi.

Beberapa kelebihan dari pohon keputusan :

- a. Hasil analisa berupa diagram pohon yang sangat mudah dimengerti.
- b. Mudah untuk dibangun, serta membutuhkan data percobaan yang lebih sedikit dibandingkan algoritma klasifikasi lainnya.
- c. Mampu mengolah data nominal dan kontinyu.
- d. Model yang dihasilkan dapat dengan mudah dimengerti, beberapa teknik klasifikasi yang berbeda seperti neural network menyajikan model dengan informasi logis yang tersirat, sehingga perlu dipelajari.
- e. Menggunakan teknik statistik sehingga dapat divalidasikan.

- f. Cepat dan handal dalam mengolah dataset besar.
- g. Akurasi yang dihasilkan mampu menandingi teknik klasifikasi yang lainnya.

Berikut tahapan pembuatan algoritma Decision Tree (C4.5):

- Tahap 1
Perhatikan label pada data, jika sama semua, maka akan dibentuk daun dengan nilai label data keseluruhan.
- Tahap 2
Menghitung nilai informasi dengan menggunakan semua data yang ada dengan formula.

$$\text{Info}(D) = - \sum_{i=1}^m p_i \log_2 (p_i)$$

Dimana p_i merupakan probabilitas tuple dalam D yang menjadi kelas C_i dengan asumsi $|C_{i,D}|/|D|$. $\text{info}(D)$ atau disebut juga *entropy* dari D merupakan rata-rata informasi yang diperlukan untuk identifikasi tuple dalam D (Alpaydın, 2010). Jika Nilai A adalah nilai diskrit, maka data D akan dipisahkan sejumlah nilai data A sehingga nilai setiap cabang akan murni dan sejenis. Setelah percabangan pertama, jumlah percabangan yang mungkin terjadi diukur dengan persamaan.

- Tahap 3
Menghitung nilai informasi dengan formula untuk setiap atribut dengan memperhatikan isi data dari atribut.

$$\text{Info}_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{info}(D_j)$$

Dimana $\frac{|D_j|}{|D|}$ merupakan bobot dari partisi j . $\text{info}_A(D)$ merupakan informasi yang diperlukan untuk mengklasifikasikan tuple dari D dari partisi A . Semakin kecil hasil persamaan ini, semakin baik pula partisi yang dihasilkan. Nilai dari sebuah atribut menentukan penting tidaknya atribut tersebut dalam penyusunan pohon keputusan. Jika atribut bernilai kontinyu, maka akan dicari *splitpoint* dengan cara mengurutkan seluruh data menurut atribut tersebut dari kecil ke besar, lalu dirata-rata antar satu data dengan data setelahnya. Nilai informasi akan dihitung satu persatu calon *splitpoint* yang akan dipilih yang terkecil.

- Tahap 4
Nilai *gain* untuk setiap atribut akan diperhitungkan, nilai dengan *gain* tertinggi akan dijadikan cabang dalam pohon keputusan.

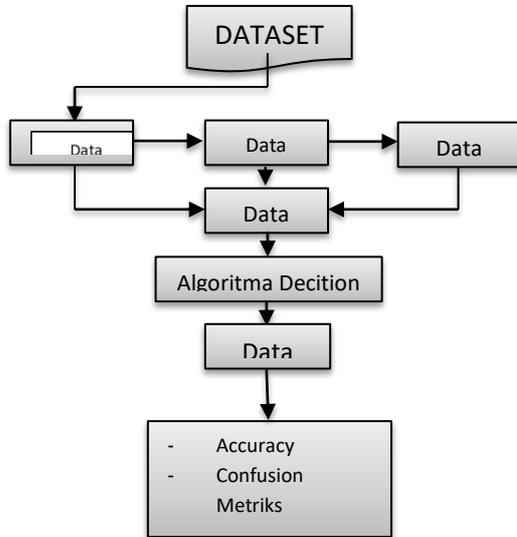
$$\text{gain}(A) = \text{info}(D) - \text{info}_A(D)$$

- Tahap 5
Setelah cabang pohon keputusan terbentuk, perhitungan dilakukan kembali seperti tahap 1 sampai 4. Namun jika cabang telah mencapai maksimal yang diperbolehkan, daun akan terbentuk dengan nilai mayoritas dari nilai data.

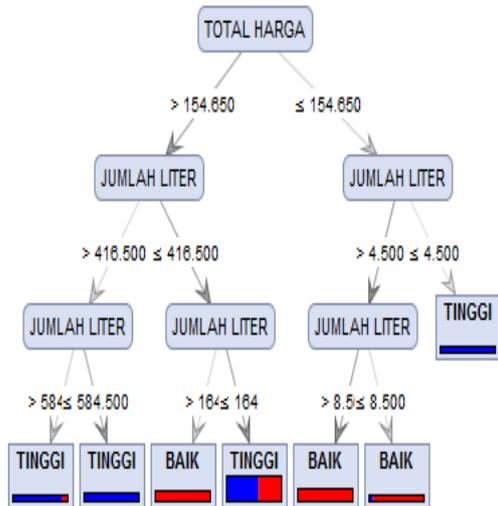
3. Metodologi Penelitian

Dalam melakukan penelitian ini, tahapan-tahapan proses yang dilalui ditunjukkan seperti pada alur proses di bawah ini :

ALUR PROSES PENELITIAN



MODEL ALGORITMA DECISION TREE (C4.5)



4. Hasil Penelitian dan Pembahasan

Tujuan dari penelitian ini adalah untuk menguji keakuratan analisa prediksi penggunaan BBM dengan menggunakan Algoritma Decition Tree (C4.5). Data yang dianalisa adalah data penggunaan BBM, yaitu semua data tersebut telah disetujui oleh pihak Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur. Dari uji coba 1, 2, dan 3, ditemukan nilai accuracy, sensitivy, specitivity, ppv, dan npv pada persamaan dibawah ini.

Tabel 4.1 : Hasil uji coba

	Uji Coba 1	Uji Coba 2	Uji Coba 3
Accuracy	65.67%	66.67%	67.33%
Sensitiviy	56.96%	57.64%	58.01%
Specitivity	94.29%	95.77%	98.55%
Ppv	97.04%	97.78%	99.26%
Npv	40.00%	41.21%	41.21%

Berdasarkan tabel diatas, dapat diketahui nilai akurasi dari uji coba 1 dengan K-Fold Validation 3 adalah 65.67%. Dan nilai akurasi dari uji coba 2 dengan K-Fold Validation 5 adalah 66.67%. Sedangkan hasil dari nilai akurasi uji coba 3 dengan K-Fold Validation 10 adalah 67.33%. Dari hasil tersebut dapat disimpulkan bahwa semakin besar nilai K-Fold Validation yang digunakan maka semakin tinggi nilai akurasi yang diperoleh.

5. Kesimpulan

Dari hasil penelitian yang telah dilakukan, maka dapat disimpulkan bahwa penggunaan BBM dapat diprediksi dan dievaluasi dengan memanfaatkan teknik Data Mining menggunakan algoritma Decision Tree (C4.5), dari data penggunaan BBM dengan data training yang telah diperoleh. Metode algoritma Decision Tree (C4.5) yang digunakan pada prediksi ini cukup baik karena menghasilkan nilai akurasi 67.33%. Dari tiga kali percobaan yang telah dilakukan dapat diketahui hasil nilai K-Fold Validation yaitu, semakin besar nilai K-Fold Validation yang digunakan maka semakin tinggi nilai akurasi yang akan diperoleh. Metode algoritma Decision Tree (C4.5) yang digunakan dapat diketahui pula tingkat penggunaan yang tinggi dan yang baik, sehingga dalam menganalisa data metode ini cukup akurat untuk tetap mengoptimalkan pemakaian BBM.

dengan Seleksi Atribut berbasis Algoritma Genetika dalam Diagnosa Penyakit Jantung.

- [4] Siti Maspirah, (2011). Algoritma Klasifikasi C4.5 berbasis Particle Swarm Optimization untuk Evaluasi Penentuan kelayakan Pemberian Kredit Koperasi Syariah.
- [5] Mahindra Suryaning Praja, S.kom. Erna Zuni Astuti, M.kom. Penerapan Data Mining untuk Rekomendasi Beasiswa pada SMAN 1 Mlonggo Menggunakan Algoritma C4.5.
- [6] Desiyana Lasut, (2012). Prediksi Loyalitas Pelanggan pada Perusahaan Penyedia Layanan Multimedia dengan Algoritma C4.5 berbasis Particle Swarm Optimization.
- [7] Moehamad Rizki Ilham. Implementasi Data Mining menggunakan Algoritma C4.5 untuk Prediksi Kepuasan Pelanggan Taksi *Kosti*.

Daftar Pustaka

- [1] Susanto Hariyanto, (2012). Segmentasi dan Klasifikasi Perilaku Pembayaran Pelanggan Pada Perusahaan Penyedia Layanan Multimedia Dengan Algoritma K-Means dan C4.5.
- [2] Evicienna, (2012). Penerapan Algoritma C4.5 berbasis Particle Swarm Optimization untuk Prediksi Hasil Pemilihan Legislatif DPRD DKI Jakarta.
- [3] Edy, (2012). Penerapan Algoritma C4.5