

Clustering dan Pemodelan Arsitektur E-Government Menggunakan Metode C4.5

Sefri Yansyah^{1✉}

¹Independent Researcher

sefriyansyah90@gmail.com

Abstract

E-Government is a form of government service effort to present complete, easily accessible, updated and high-accuracy data to the public. Good Governance is the aspiration of the Pelalawan Regency Government so that the service becomes a government administration that prioritizes aspects of accountability, transparency and accountability. participatory, Assist in identification efforts and then become a solution to solving problems where so far the Pelalawan District Government programs are often not on target, data gaps and the dynamic nature of data from time to time. This is caused by the large number of information systems facilitated by Ministries, Agencies and other Institutions for use by Regional Apparatus Organizations that are not relevant to the variables and data criteria that are suitable to be applied to Pelalawan Regency, considering the data that has been processed so far is very large and large. Solution Method C4.5 can assist in identification efforts and then become a problem solving solution where so far the Pelalawan District Government program is often not right on target, data gaps and the dynamic nature of data from time to time. The data used in this study are 3 information systems facilitated by the Ministry of Social Affairs, Visualization and Analysis, a collection of data with different and growing formats is then clustered using the C.45 method by carrying out the Big Data principles, namely volume, velocity, variety, varicity and value. This research will later become input to the Regional Government in determining decisions on the implementation of assistance to the poor which refers to the indicators from this data mining.

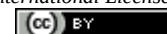
Keywords: Big Data, Data Mining, Decicion Three, Clustering, Algorithm C4.5.

Abstrak

E-Government adalah bentuk upaya layanan pemerintah menyajikan data dengan lengkap, mudah diakses, update dan memiliki tingkat akurasi yang tinggi kepada masyarakat, *Good Governance* merupakan cita-cita Pemerintah Daerah Kabupaten Pelalawan agar layanan tersebut menjadi sebuah penyelenggaraan Pemerintahan yang mengedepankan aspek akuntabilitas, transparan dan parsitipatif, Membantu dalam upaya pengidentifikasian kemudian menjadi solusi pemecahan masalah dimana selama ini program Pemerintah Daerah Kabupaten Pelalawan sering tidak tepat sasaran, kesenjangan data serta sifat data yang dinamis antar waktu ke waktu. hal tersebut diakibatkan oleh banyaknya sistem informasi yang difasilitasi oleh Kementerian, Badan dan Lembaga lainnya untuk digunakan oleh Organisasi Perangkat Daerah tidak relevan dengan variabel dan kriteria data yang cocok untuk diterapkan pada kabupaten pelalawan, mengingat data yang diolah selama ini sangat banyak dan besar. Solusi Metode C4.5 dapat membantu dalam upaya pengidentifikasian kemudian menjadi solusi pemecahan masalah dimana selama ini program Pemerintah Daerah Kabupaten Pelalawan sering tidak tepat sasaran, kesenjangan data serta sifat data yang dinamis antar waktu ke waktu, Data yang digunakan dalam penelitian ini adalah 3 sistem informasi yang difasilitasi oleh Kementerian Sosial, Visualisasi dan Analisa, sekumpulan data dengan format berbeda dan berkembang kemudian diklasterisasi menggunakan metode C.45 dengan mengusung prinsip *Big Data* yaitu *volume*, *velocity*, *variety*, *varicity* dan *value*. Penilitan ini nantinya menjadi masukan kepada Pemerintah Daerah dalam penentuan keputusan pada pelaksanaan bantuan masyarakat miskin yang mengacu pada indikator dari penggalan data ini.

Kata kunci: Big Data, Data Mining, Decicion Three, Clustering, Algoritma C4.5.

JIdT is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Data mining, sering juga disebut Knowledge Discovery In Database (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [1]. KDD adalah suatu metode ekstraksi non-trivial yang tersirat suatu informasi tidak diketahui sebelumnya tetapi hasil yang diterima dari data tersebut terkandung potensi informasi [2]. Proses yang

dilakukan dalam KDD adalah Data Mining, merupakan cara mendapatkan pengetahuan (*Knowledge Discovery*) yang digali dari himpunan data dengan volume sangat besar, dan dapat mendukung dalam mengambil keputusan [3]. Data mining adalah sebuah algoritma untuk menggali informasi yang berharga pada database sehingga akan ditemukan suatu pola yang menarik dimana sebelumnya pola tersebut tidak diketahui [4]. Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Sedangkan samples merupakan field-field data yang nantinya akan kita

gunakan sebagai parameter dalam melakukan klasifikasi data[5]. Salah satu algoritma yang dipakai dalam data mining adalah algoritma C4.5 [6]. Algoritma C4.5 adalah algoritma yang digunakan untuk membuat *rule* dan suatu pohon keputusan. Algoritma C4.5 menggambarkan nilai atribut menjadi kelas yang dapat diterapkan untuk klasifikasi baru [7]. Penerapan Algoritma C4.5 dapat digunakan untuk mengklasterisasi kelayakan bantuan masyarakat miskin. Sebagai salah satu upaya dalam meningkatkan pelayanan dan efisiensi penyelenggaraan Pemerintahan, konsep otomatisasi administrasi dengan mendayagunakan Teknologi Informasi yang dikenal dengan *e-Government* merupakan cita-cita yang ingin dicapai sebagai upaya menuju terwujudnya *good governance* di Kabupaten Pelalawan. Dimana selama ini program Pemerintah Daerah Kabupaten Pelalawan sering tidak tepat sasaran, kesenjangan data serta sifat data yang dinamis antar waktu ke waktu. hal tersebut diakibatkan oleh banyaknya sistem informasi yang difasilitasi oleh Kementerian, Badan dan Lembaga lainnya untuk digunakan oleh Organisasi Perangkat Daerah. Dalam penelitian sebelumnya mengenai data mining untuk menentukan parameter kunci syarat tumbuh optimal beberapa tanaman industri dentifikasi calon pegawai suatu perusahaan. Algoritma C4.5 dapat digunakan untuk melakukan prediksi dan klasifikasi terhadap calon pegawai yang berpotensi untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data yang sudah ada dan melakukan prediksi terhadap calon pegawai baru yang ingin masuk ke perusahaan. Berdasarkan metode pengukuran akurasi ten-fold cross validation telah didapatkan hasil pengukuran tingkat keberhasilan prediksi calon pegawai baru sebesar 71% [8]. Data Mining sering juga disebut KDD meliputi kegiatan pengumpulan, pemakaian data historis menemukan pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [9]. Dalam penelitian sebelumnya mengenai klasifikasi penderita penyakit diabetes menggunakan algoritma C4.5 *Decision Tree* adalah algoritma yang bisa menghasilkan keputusan dengan cara membentuk pohon keputusan. Dari hasil Pengujian menghasilkan akurasi yang cukup besar yaitu 97,12% Precision sebesar 93,02% %, dan *Recall* sebesar 100,00%[10]. Dalam penelitian sebelumnya mengenai prediksi potensi kebakaran hutan dengan algoritma klasifikasi C4.5 Penelitian menghasilkan kesimpulan bahwa C4.5 dengan percentage split 80% data training dan 20% data testing menghasilkan akurasi tertinggi yaitu 89,7859% [11]. Dalam penelitian sebelumnya mengenai *Rapid Miner* digunakan untuk eksperimen data teks dengan algoritma KNN dengan tujuan mencari nilai akurasi, presisi dan recall. Dari hasil penelitian diperoleh nilai sebesar 87.00% untuk accuracy dan 0.916 untuk nilai AUC. Nilai-nilai yang cukup tinggi

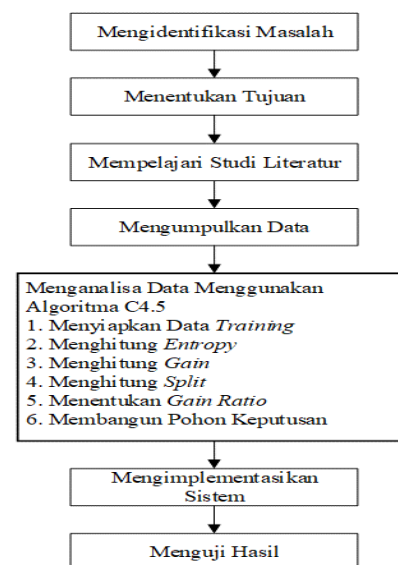
untuk klasifikasi opini mahasiswa terhadap pandemi ini sehingga penelitian ini digolongkan sebagai *Excellent Classification* [12].

2. Metodologi Penelitian

Susunan dari aktivitas penelitian untuk mengetahui hasil pada permasalahan spesifik yang bertujuan untuk menjawab pertanyaan dari rumusan masalah, sehingga bisa dikerjakan tahapan-tahapan yang akan digunakan serta waktu yang dipakai sebagai acuan dalam mendapatkan analisa data. Metode penelitian berfungsi sebagai acuan untuk memudahkan penulisan agar bisa terarah dalam mengatasi masalah sesuai dengan permasalahan yang diteliti.

2.1 Kerangka Penilitan

Kerangka penelitian merupakan gambaran langkah-langkah yang dikerjakan penulis dalam menyelesaikan masalah penelitian yang bertujuan agar tidak melenceng dari pokok pembahasan, maka urutan langkah-langkah di buat secara sistematis sehingga dapat digunakan sebagai acuan yang jelas dan mudah untuk menyelesaikan permasalahan yang ada. Urutan langkah-langkah yang akan dibuat pada penelitian ini dapat dilihat pada gambar 2.1.



Gambar 1. Flowchart Penelitian

Berdasarkan kerangka kerja penelitian pada Gambar 2.1 di atas, maka akan dijelaskan alur penelitian tersebut sebagai berikut:

a. Mengidentifikasi Masalah

mendefinisikan masalah diperlukan data-data yang berkaitan dengan masalah dalam melakukan penelitian.

b. Menentukan Tujuan

Langkah berikutnya setelah mengidentifikasi masalah, ditentukan tujuan yang akan diinginkan

dari penelitian ini sehingga dapat memberikan solusi dari masalah yang ada. Manfaat pada tujuan ini dapat berguna bagi Kabupaten Pelalawan dalam mengidentifikasi bantuan masyarat miskin sehingga bantuan tersebut tepat sasaran. Diharapkan dengan tujuan yang telah ditentukan dapat menganalisa permasalahan dengan baik.

c. Mempelajari Studi Literatur

Mempelajari literatur sangat penting dilakukan sebagai pedoman untuk melakukan penelitian baik secara teoritis dalam memperoleh informasi tertentu yang telah dijelaskan oleh para peneliti sebelumnya.

d. Mengumpulkan data

Dalam tahapan mengumpulkan data ini menjelaskan tentang kegiatan pengumpulan data yang diperlukan dalam penelitian ini.

e. Menganalisa Data Menggunakan Algoritma C4.5

Setelah semua data yang dibutuhkan terkumpul dan dianalisa, selanjutnya dilakukan pengelompokan data sehingga dapat memudahkan dalam tahap Analisa masalah. Data yang dipakai dalam penelitian ini adalah bantuan masyarat miskin di Kabupaten Pelalawan.

f. Mengimplementasi Sistem

Pada tahap ini untuk mengidentifikasi bantuan masyarat miskin sehingga dapat tersalurkan tepat sasaran di lakukan perhitungan secara manual setiap elemen menggunakan Algoritma C4.5 dan elemen tersebut dimasukkan kedalam form-form yang terdapat pada aplikasi Rapidminer. Implementasi ini dilakukan untuk membandingkan hasil analisa yang didapatkan secara manual dengan hasil menggunakan sistem.

g. Menguji Hasil

Pengujian bertujuan untuk mengetahui apakah knowledge yang diperoleh berhasil dan dapat diterapkan dengan baik untuk membantu menyelesaikan masalah bantuan masyarat miskin di Kabupaten Pelalawan sehingga bantuan tersebut dapat terealisasi dengan tepat.

2.2. Analisa Data

Analisa data dilakujan dengan mengambil data berupa hasil *export* sistem informasi Data Terpadu Kesejahteraan Sosial (DTKS) adalah basis data yang digunakan pemerintah untuk menyalurkan bansos Program Keluarga Harapan (PKH) dan Bantuan Pangan Non Tunai (BPNT) di Dinas Sosial Kabupaten Pelalawan.

2.3. Clustering

Clustering adalah pengelompokan atau penggabungan data dengan kelas data yang memiliki karakteritik serupa dijadikan satu kelompok dan yang berbeda

berada pada kelompok lain [13]. Clustering merupakan suatu metode pengelompokan data, observasi, atau mengelompokan kelas yang mempunyai kemiripan objek. *Clustering* tidak memiliki target variable dalam melakukannya, berbeda dengan metode klasifikasi. Clustering sering dijadikan langkah awal dalam proses Data Mining [14]. Clustering berarti cluster atau pengelompokan adalah salah satu proses dalam sebuah pengelompokan Data Mining yang dipakai dalam mengelompokan suatu data yang banyak untukn memprediksi setiap data mempunyai kesamaan dari tiap-tiap nilai yang bersifat homogen dan diperoleh menggunakan cara pengamatan terhadap objek-objek, data-data yang ingin diketahui cluster dari setiap data, dalam metode clustering umumnya sebuah data dibantu dengan memakai pendekatan sebuah algoritma bermaksud untuk lebih memperinci dan target dalam pengelempokan data jauh lebih terstruktur dalam sistem matematis [15].

2.4. Data Mining

Data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan Teknik pengenalan pola seperti teknik statistik dan matematika [16]. Aktivitas yang melingkupi pengumpulan, penggunaan data historis untuk mendapatkan keruntutan atau ikatan dalam set data berukuran besar. Kaidah yang mengungkapkan blok antara beberapa atribut kerap disebut *affinity analysis* atau *market basket analysis* [17]. Data Mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut [18]. Proses *Data Mining* mendapatkan hasil dimulai dari data mentah dan berakhir dengan *knowledge* atau informasi yang telah diolah diperoleh dari tahapan-tahapan berikut [19].

a. Data *Cleansing*, berperan dalam penggalian

pengetahuan (*discovery knowledge*) dimana data-data tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih relevan dapat digunakan untuk diproses ulang.

b. Data Integration, pada bagian ini berlangsung peleburan data, dimana sumber-sumber data yang berulang (multiple data), file-file yang iteratif (multiple file), bisa diafiliasikan dan dikombinasikan kedalam suatu sumber.

c. Data Selection, pada tahap ini data yang relevan terhadap kajian bisa dipilih dan diperoleh dari koleksi data yang ada.

d. Data Transformation, juga diketahui sebagai data consolidation. Pada proses ini dilakukan normalisasi dan agregasi data. Data-data yang telah terseleksi, ditransformasikan kedalam format yang

cocok untuk prosedur penggalian (meaning procedure).

- e. Data Mining, pada bagian ini menggunakan teknik-teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna, tahapan ini merupakan tahap yang paling penting.
- f. Pattern Evaluation, pada proses ini pola-pola menarik dengan jelas memaparkan knowledge yang sudah diidentifikasi berlandaskan measure yang telah diberikan.
- g. *Knowledge Representation*, tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari Data Mining. Tahap ini merupakan tahap penting dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user.

2.5. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal dipahami dengan bahasa alami [20]. Tahapan dari proses algoritma C4.5 diuraikan sebagai berikut:

- a. Menyiapkan data training
Data berasal dari data yang sudah dikumpulkan sebelumnya dan di kelompokkan kedalam kelas tertentu
- b. Menghitung nilai Entropy total dari semua atribut
- c. Menghitung nilai Gain untuk setiap nilai atribut
- d. Membuat node cabang untuk semua atribut sampai membuat pohon keputusan
- e. Menentukan rule-rule dari pohon keputusan
- f. Mengulangi proses hingga semua cabang mempunyai kelas yang sama. Proses percabangan akan berhenti jika:
 - i. Semua kasus dalam simpul n mendapat kelas yang sama
 - ii. Tidak ada variabel independen di dalam kasus yang dipartisi lagi
 - iii. Tidak ada kasus di dalam cabang yang kosong

Untuk perhitungan mencari nilai entropy menggunakan rumus disajikan pada Persamaan (1).

$$Entropy(N) = \sum_{i=1}^t -qr * \log_2 qr \quad (1)$$

Dimana N adalah set kasus atau jumlah kasus. t merupakan jumlah partisi N dan qr merupakan proporsi Nr ke N.

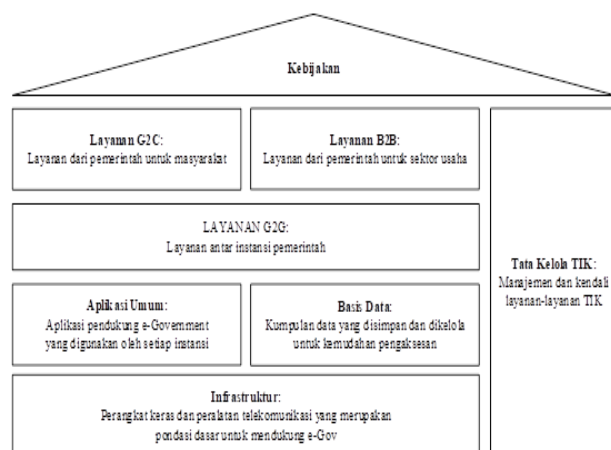
Rumus perhitungan untuk mencari nilai gain yang disajikan pada Persamaan (2).

$$Gain(N,M) = Entropy(N) - \sum_{o=1}^t \frac{|No|}{|N|} * Entropy(No) \quad (2)$$

Dimana N adalah set kasus atau jumlah kasus. M merupakan fitur atau atribut yang digunakan. T merupakan jumlah partisi pada atribut M. |No| merupakan jumlah pada partisi ke N dan |N| adalah jumlah kasus di N.

2.6 E-Government

E-Government merupakan suatu metode pemerintahan dalam memajukan kemampuan, efektivitas, keterbukaan, dan responsibilitas pelaksanaan pemerintahan dengan penggunaan teknologi informasi yang bertujuan untuk mempermudah dalam transaksi pelayanan publik dengan biaya yang dapat dijangkau bagi masyarakat, tahapan ini merupakan tahapan proses pengambilan data yang sebahagian data tersebut sudah terintegrasi. Adapun Arsitektur *e-Government* sesuai dengan Inpres No 3/2003: Kebijakan Payung hukum yang mengatur pelaksanaan *e-Government* dapat dilihat pada Gambar 2.



Gambar 2. Arsitektur E-Government

3. Hasil dan Pembahasan

Penilitan ini nantinya menjadi masukan kepada Pemerintah Daerah dalam penentuan keputusan pada pelaksanaan bantuan masyarakat miskin yang mengacu pada indikator dari penggalian data ini.

3.1 Analisa Data

Representasi pengetahuan yang didapatkan adalah angka garis kemiskinan tahun 2021 oleh Badan Pusat Statistik Kabupaten Pelalawan, kemudian klasterisasi peraturan bupati pelalawan berbentuk kriteria dan indikator fakir miskin yang dirangkum dalam tiga kriteria dasar, untuk mempermudah dalam membaca data yang akan disajikan, akan dijabarkan pada Tabel 1 dan Tabel 2.

Tabel 1. Garis Kemiskinan Kabupaten Pelalawan

Wilayah	Tahun (Rupiah)		
	2019	2020	2021
Pelalawan	534.078	585.728	618.739

Tabel 2. Pedoman Penetapan Kriteria dan Indikator Fakir Miskin di Kabupaten Pelalawan Tahun 2022

NO	KRITERIA	INDIKATOR		
		A	B	C
1	Tidak mempunyai sumber mata pencaharian dan/atau mempunyai sumber mata pencaharian tetapi tidak mempunyai kemampuan memenuhi kebutuhan dasar	Penghasilan tetap tetapi tidak mempunyai kemampuan memenuhi kebutuhan dasar	Penghasilan tidak tetap dan tidak mempunyai kemampuan memenuhi kebutuhan dasar	Tidak memiliki penghasilan dan tidak mempunyai kemampuan memenuhi kebutuhan dasar
2	Mempunyai pengeluaran sebagian besar digunakan untuk memenuhi konsumsi makanan pokok dengan sangat sederhana	Pengeluaran 50 % besar digunakan untuk memenuhi konsumsi makanan pokok dengan sangat sederhana	Pengeluaran 75 % besar digunakan untuk memenuhi konsumsi makanan pokok dengan sangat sederhana	Pengeluaran 100% besar digunakan untuk memenuhi konsumsi makanan pokok dengan sangat sederhana
3	Luas lantai rumah kecil kurang dari 8 m2/orang	Luas lantai rumah kecil kurang dari 8 m2/orang	Luas lantai rumah kecil kurang dari 6 m2/orang	Luas lantai rumah kecil kurang dari 4 m2/orang

Data yang diperoleh berupa data mentah berupa file Microsoft Excel dari masing-masing file yang berbeda. Sehingga dilakukan pra-proses data dilakukan dengan berbagai metode pemilihan diantaranya menggunakan *Data Reduction* untuk menentukan status atau kelas kelayakan agar tidak memakan waktu yang lama untuk memasuki proses *Data Transformation*, dengan m pendekengkasifikasikan kriteria berdasarkan Tabel 2 adapun parameteranya adalah:

a. *Feature Selection*

Prose pemilihan fitur yang ditemukan pada sebuah data set awal

b. *Feature Composition*

Faktor penentu dalam menghasilkan kelas agar pengalihan data lebih berkualitas

3.2 Implementasi Algoritma C4.5

Proses impelementasi algoritma C4.5 dalam proses data mining melalui beberapa tahap dibawah ini:

- Menyiapkan Data *Training*
- Mancari Nilai *Entropy*
- Menghitung *Gain*
- Menghitung *Split*
- Menentukan *Gain Ratio*
- Membangun Pohon Keputusan

Data yang akan diolah menggunakan algortima C4.5 adalah Data Terpadu Kesejahteraan Sosial milik Kementerian Sosial Republik Indonesia yang di ekstrak oleh Dinas Sosial Kabupaten Pelalawan.

Keterangan:

- Jika Jumlah Penghasilan > 2.986.000 Maka Tidak Layak
- Jika Jumlah Penghasilan <= 2986000 dan Jumlah Pengeluaran > 1.714.833 maka Layak

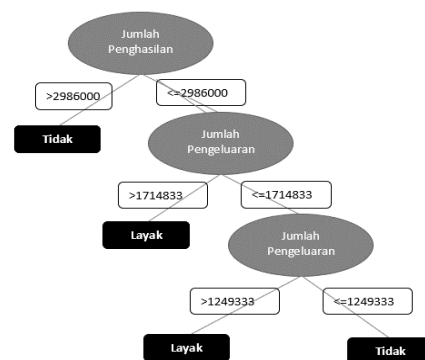
3. Jika Jumlah Penghasilan <= 2986000 dan Jumlah Pengeluaran <= 1714833 dan >1249333 maka Layak

4. Jika Jumlah Penghasilan <= 2986000 dan Jumlah Pengeluaran <= 1714833 dan <=1249333 maka Tidak Layak.

3.2 Pembentukan Pohon Keputusan

Pada tahap ini untuk mengidentifikasi bantuan masyarat miskin sehingga dapat tersalurkan tepat sasaran di lakukan perhitungan secara manual setiap elemen menggunakan Algoritma C4.5 dan elemen tersebut dimasukkan kedalam form-form yang terdapat pada aplikasi *Rapidminer* maupun sistem informasi yang dibangun. Tahapan pohon keputusan (*decision tree*) dibangun adalah sebagai berikut:

- Menentukan atribut sebagai akar.
- Membuat cabang pada setiap nilai
- Membagi kasus dalam cabang
- Pada Setiap Cabang, prosesnya diulangi hingga semua kasus pada setiap cabang memiliki kelas yang sama untuk menentukan atribut sebagai akar, disesuaikan pada nilai *gain* paling tinggi dari atribut-atribut yang ada yang disajikan pada Gambar 3.



Gambar 3. Pohon Keputusan

Dilihat dari pohon keputusan yang dihasilkan, terlihat tidak ada lagi terdapat percabangan yang harus dilakukan perhitungan, karena itu dapat disimpulkan dari berbagai atribut sudah berhasil diciptakan.

3.3 Evaluasi Peforma Model Algoritma C4.5

Pada teknik klasifikasi data mining tahap evaluasi sering digunakan untuk mengukur tingkat akurasi perhitungan dari algoritma tersebut. Pada tahap ini model evaluasi yang diterapkan adalah area under curve (AUC). AUC adalah area dibawah kurva untuk meningkatkan lintas mempelajari komparabilitas dari perhitungan algoritma setelah melewati tahap cross validation yang diklasifikasikan antara nilai 0.00 hingga 1.00 disajikan pada Tabel 3.

Tabel 3. Klasifikasi Nilai AUC

AUC	Hasil
0.90 – 1.00	Sangat Baik
0.80 – 0.90	Baik
0.70 – 0.80	Sedang
0.60 – 0.70	Buruk
<0.60	Gagal

Selain menggunakan AUC pada tahap metode evaluasi juga sering menggunakan Confusion Matrik yang merupakan tabulasi yang unik dari contoh prediksi yang benar dan tidak benar untuk setiap kelas. Berikut adalah rumus yang digunakan dalam mengevaluasi *accuracy*, *precision* dan *recall* yang disajikan pada Rumus (3), Rumus (4), Rumus (5), Rumus (6), dan Rumus (7).

$$\text{Accuracy} = \frac{Tp+Tn}{\text{Total Semua Data}} \quad (3)$$

$$\text{Precision} = \frac{Tp}{Tp+Fp} \quad (4)$$

$$\text{Recall} = \frac{Tp}{Tp+Fn} \quad (5)$$

$$\text{Spesifisity} = \frac{Tn}{Tn+Fp} \quad (6)$$

$$\text{AUC} = \frac{\text{Recall}+\text{Spesifisitas}}{2} \quad (7)$$

Pengujian menggunakan tiga metode yang berbeda yaitu:

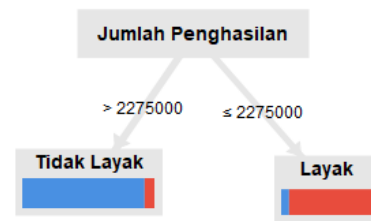
a. 10-Fold Cross Validation.

b. Pengujian selanjutnya menentukan tipe sampling data meliputi *Linear Sampling*, *Shuffled Sampling*, dan *Stratified Sampling*.

c. Perbandingan *Accuracy*, *Precision*, *Recall* dan AUC.

Dengan Skenario Uji Coba menggunakan 10-Fold Cross Validation dan menggunakan tipe pengambilan dengan *Linear Sampling Data*, *Shuffled Sampling Data*, *Stratified Sampling Data* Penentuan pohon keputusan Algoritma C4.5 ditentukan dengan kriteria Gain Ratio, Minimal Gain adalah 0.01, teknik *Pruning*

dan *Prepuning*. Hasil pohon keputusan bisa dilihat pada Gambar 4.



Gambar 4. Pohon Keputusan

Hasil evaluasi berdasarkan performa dari ketiga model skenario yang dilakukan disajikan pada Tabel 4.

Tabel 4. Hasil evaluasi

Pengujian	Akurasi	Presisi	Recall	AUC
Skenario 1	92.00 %	91.30 %	85.00%	0.825
Skenario 2	92.00 %	92.50 %	94.17 %	0.908
Skenario 3	90.00 %	93.33 %	91.67 %	0.892

4. Kesimpulan

Berdasarkan hasil pengujian dan implementasi Data Mining menggunakan algoritma c4.5 pada usulan bantuan masyarakat miskin di Dinas Sosial Kabupaten Pelalawan, maka didapatkan kesimpulan bahwa proses pengujian pada data test, sebanyak 3 kali pengujian, dimana terhadap 50 data, dengan tingkat akurasi antara 90% - 92%, dengan atribut akar (*root*) adalah kategori dan atribut menjadi cabang adalah satuan, sedangkan atribut yang akan menjadi ranting layak dan tidak layak.

Daftar Rujukan

- [1] Panggabean, D. S. O., Buulolo, E., & Silalahi, N. (2020). Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda. *JURIKOM (Jurnal Riset Komputer)*, 7(1), 56-62. <http://dx.doi.org/10.30865/jurikom.v7i1.1947>
- [2] Subroto, G., Sulistiyowati, N., & Ridha, A. A. (2022). Klasifikasi Jenis Kekerasan Pada Perempuan Dan Anak Dengan Algoritma Multinomial Naïve Bayes. *INTECOMS: Journal of Information Technology and Computer Science*, 5(1), 104-113. <https://doi.org/10.31539/intecom.v5i1.3598>
- [3] Dewi, S., Defit, S., & Yuhandri, Y. (2021). Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means. *Jurnal Sistim Informasi dan Teknologi*, 28-33. <https://doi.org/10.37034/jsisfotek.v3i1.40>
- [4] Lestari, Y. I. (2021). Prediksi Tingkat Kepuasan Pelayanan Online Menggunakan Metode Algoritma C. 45. *Jurnal Informatika Ekonomi Bisnis*, 148-154. <https://doi.org/10.37034/infec.v3i4.104>
- [5] Umam, K., Puspitasari, D., & Nurhadi, A. (2020). Penerapan Algoritma C4. 5 Untuk Prediksi Loyalitas Nasabah PT Erdika Elit Jakarta. *Jurnal Media Informatika Budidarma*, 4(1), 65-71. <http://dx.doi.org/10.30865/mib.v4i1.1652>
- [6] Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4. 5. *Jurnal Informasi dan Teknologi*, 23-29. <https://doi.org/10.37034/jidt.v2i1.22>
- [7] Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *Jurnal*

- SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan), 4(1), 32-39. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- [8] Rintiasti, A., & Fauzi, R. Data Mining Parameter Syarat Tumbuh Optimal Lahan Perkebunan untuk Rule Sistem Pakar Menggunakan Metode C45 dan Forward Chaining. *Indonesian Journal of Industrial Research*, 13(2), 95-105. [10.33104/jihp.v13i2.4139](https://doi.org/10.33104/jihp.v13i2.4139)
- [9] Matondang, M. R., Lubis, M. R., & Tambunan, H. S. (2021). Analisis Data mining dengan Metode C. 45 pada Klasifikasi Kenaikan Rata-Rata Volume Perikanan Tangkap. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 2(2), 74-81. <https://doi.org/10.30645/brahmana.v2i2.68>
- [10] Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, 4(1), 32-39. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- [11] Primajaya, A., Sari, B. N., & Khusaeri, A. (2020). Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4. 5 Studi Kasus Provinsi Kalimantan Barat. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 6(2), 188-192. <http://dx.doi.org/10.26418/jp.v6i2.37834>
- [12] Mara, A. T., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (DARING) Di Universitas Kristen Wira Wacana Sumba. *Jointer-Journal of Informatics Engineering*, 2(01), 24-31. <https://doi.org/10.53682/jointer.v2i01.30>
- [13] Sudarsono, B. G., & Lestari, S. P. (2021). Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means. *Jurnal Media Informatika Budidarma*, 5(1), 258-263. <http://dx.doi.org/10.30865/mib.v5i1.2670>
- [14] Harahap, F. (2021). Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita. *TIN: Terapan Informatika Nusantara*, 2(4), 191-197. <https://ejurnal.seminar-id.com/index.php/tin>
- [15] Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020, February). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritma C4. 5. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 3, pp. 64-71).
- [16] Windarto, A. P., Irawan, E., Saputra, W., & Okprana, H. (2020, July). Penerapan Data Mining Menggunakan Algoritma C4. 5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS. In *Prosiding Seminar Nasional Riset Information Science (SENARIS)* (Vol. 2, pp. 376-385). <http://dx.doi.org/10.30645/senaris.v2i0.185>
- [17] Maulana, Y., Winanjaya, R., & Rizki, F. (2022). Penerapan Data Mining dengan Algoritma C. 45 Dalam Memprediksi Penjualan Tempe. *Bulletin of Computer Science Research*, 2(2), 53-58. <https://doi.org/10.47970/siskom-kb.v4i1.173>
- [18] Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, 2(3). <https://doi.org/10.22437/jkam.v2i1.5430>
- [19] Ordila, R., Wahyuni, R., Irawan, Y., & Sari, M. Y. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda). *Jurnal Ilmu Komputer*, 9(2), 148-153. <https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181>
- [20] Desyanti, D. (2021). Implementasi Metode C. 45 dalam Menganalisa Tingkat Kecemasan Mahasiswa Menyusun Tugas Akhir. *JURNAL UNITEK*, 14(1), 17-29. <https://doi.org/10.52072/unitek.v14i1.175>