

Penerapan Algoritma Data Mining C4.5 Untuk Klasifikasi Kualitas Layanan Publik Kantor Dinas Kependudukan dan Pencatatan Sipil Kabupaten Manokwari Selatan

Arianti Adahati¹, Syahril Amin², Edizon Inyomusi³

^{*)}Dosen Program Studi Ilmu Komputer, STIMIK Kreatindo Manokwari
e-mail: syahril.amin.ari@gmail.com

Abstraksi

Penelitian ini bertujuan menerapkan algoritma data mining C4.5 untuk mencari penilaian kualitas layanan publik pada kantor Dinas Kependudukan dan Pencatatan Sipil Kabupaten Manokwari Selatan. Penelitian ini menggunakan 6 (enam) Atribut Kualitas Layanan yaitu: (1) Bukti Langsung; (2) Keandalan; (3) Daya Tanggap; (4) Jaminan dan (5) Empati. Metode yang digunakan pada penelitian ini meliputi beberapa tahapan yaitu (1) pengumpulan data; (2) Preprocessing (3) Penentuan data Latih dan Data Uji; (4) Pembuatan Model Klasifikasi; dan (5) Pengujian Model Klasifikasi. Hasil penelitian yang diperoleh menunjukkan bahwa atribut yang paling memiliki pengaruh paling besar adalah daya tanggap dan jaminan. Dari pengujian model yang dilakukan didapatkan tingkat akurasi confusion matrix sebesar 94,12 %

Kata Kunci: Kualitas Layanan, Pohon Keputusan, Algoritma C4.5

Abstract

This study aims to apply the C4.5 data mining algorithm to look for assessments of the quality of public services at the Office of Population and Civil Registration Office of South Manokwari Regency. This study uses 6 (six) Service Quality Attributes, namely: (1) Direct Evidence; (2) Reliability; (3) Responsiveness; (4) Guarantees and (5) Empathy. The method used in this study includes several stages, namely (1) data collection; (2) Preprocessing; (3) Determination of Training data and Test Data; (4) Making a Classification Model; and (5) Testing the Classification Model. The results of the research obtained indicate that the attributes that have the greatest influence are responsiveness and assurance. From the model test, the accuracy of the confusion matrix is 94.12%.

Keywords: Service Quality, Decision Tree, C4.5 Algorithm

1. PENDAHULUAN

Melalui perkembangan teknologi saat ini, hadir sebuah metode data mining yang dapat digunakan untuk proses prediksi, klasifikasi. Data mining dapat digunakan untuk melihat kemungkinan-kemungkinan kejadian yang akan datang. Salah satu metode yang paling sering digunakan dalam model klasifikasi adalah algoritma C4.5. Dalam banyak penelitian yang telah dilakukan, algoritma C4.5 dapat mengklasifikasikan dengan baik. Penelitian ini akan menerapkan algoritma C4.5 untuk mendapatkan model klasifikasi kualitas layanan publik di Kantor Dinas Kependudukan dan Pencatatan Sipil Kabupaten Manokwari Selatan.

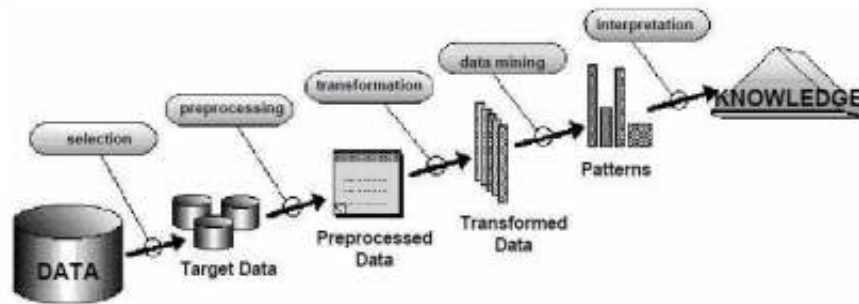
Konsep kualitas bersifat relatif, karena penilaian kualitas sangat ditentukan dari perspektif yang digunakan. Menurut Trilestari[1] pada dasarnya terdapat tiga orientasi kualitas yang seharusnya konsisten antara yang satu

dengan yang lain, yaitu persepsi pelanggan, produk, dan proses. Kata "kualitas" sendiri mengandung banyak pengertian, menurut Kamus Bahasa Indonesia, kualitas berarti: (1) tingkat baik buruknya sesuatu; (2) derajat atau taraf (kepandaian, kecakapan, dsb); atau mutu. Konsep kualitas pelayanan dapat dipahami melalui perilaku konsumen (*consumer behavior*), yaitu suatu perilaku yang dimainkan oleh konsumen dalam mencari, membeli, menggunakan, dan mengevaluasi suatu produk maupun pelayanan yang diharapkan dapat memuaskan kebutuhan mereka.

Data mining merupakan teknologi yang menggabungkan metode analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar [2]. Data mining atau *Knowledge Discovery in Databases* (KDD) adalah pengambilan informasi yang tersembunyi, dimana informasi tersebut

sebelumnya tidak dikenal dan berpotensi bermanfaat. Proses ini meliputi sejumlah pendekatan teknis yang berbeda, seperti

clustering, data summarization, learning classification rules.



Gambar 1. Proses KDD

Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan pola dan *trend* baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Menurut Larose (2005), data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

- a. Deskripsi
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari untuk menggambar pola dan kecenderungan yang terdapat dalam data.
- b. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target yang dibuat berdasarkan nilai variabel prediksi.
- c. Klasifikasi
Dalam klasifikasi terdapat target variabel kategori, sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam 3 (tiga) kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.
- d. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Contoh prediksi dalam bisnis dan penelitian adalah:
 - 1) Prediksi harga beras dalam 3 (tiga) bulan mendatang;

- 2) Prediksi presentase kenaikan kecelakaan lalu lintas tahun depan jika batas bawah kecepatan dinaikan.

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. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi kontinu data, dan pruning [3].

Pohon keputusan menggunakan representasi struktur pohon (*tree*) di mana setiap node merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut dan daun merepresentasikan kelas. Node yang paling atas dari pohon keputusan disebut sebagai *root*. Salah satu keuntungan yang paling signifikan dari pohon keputusan adalah kenyataan bahwa pengetahuan dapat diekstrak dan direpresentasikan dalam bentuk klasifikasi aturan if-then [4]. Pohon keputusan merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami [5]. Pohon keputusan bekerja dengan membentuk pohon keputusan yang dapat disimpulkan aturan-aturan klasifikasi tertentu, salah satunya adalah C4.5 [6].

Data dalam metode *decision tree* (pohon keputusan) biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat

sebagai kriteria dalam pembentukan *tree*. Contoh data dalam aplikasi keputusan bermain tennis

menggunakan metode *decision tree* dapat dilihat pada gambar 2.

| Nama | Cuaca | Angin | Temperatur | Main |
|-------|---------|--------|------------|-------|
| Ali | cerah | keras | panas | tidak |
| Budi | cerah | lambat | panas | ya |
| Heri | berawan | keras | sedang | tidak |
| Irma | hujan | keras | dingin | tidak |
| Diman | cerah | lambat | dingin | ya |

↓ Sample
atribut
↓ Target atribut

Gambar 2. Konsep Data dalam Pohon Keputusan
(Sumber : Basuki dan Syarif, 2003)

Adapun untuk membangun pohon keputusan menggunakan algoritma C4.5 dapat dilakukan melalui langkah-langkah berikut ini:

- 1) Hitung jumlah data, jumlah data berdasarkan anggota atribut hasil dengan syarat tertentu. Untuk proses pertama syaratnya masih kosong.
- 2) Pilih atribut sebagai *node*.
- 3) Buat cabang untuk tiap-tiap anggota dari *node*.
- 4) Periksa apakah nilai *entropy* dari anggota *node* ada yang bernilai nol. Jika ada, tentukan daun yang terbentuk. Jika seluruh nilai *entropy* anggota *node* adalah nol, maka proses pun berhenti.
- 5) Jika ada anggota *node* yang memiliki nilai *entropy* lebih besar dari nol, ulangi lagi

proses dari awal dengan *node* sebagai syarat sampai semua anggota dari *node* bernilai nol.

Keakuratan dalam sebuah pengklasifikasian pada sebuah dataset yang diuji adalah tentang persentase dari sebuah item dalam data yang diuji, yang dikelompokkan secara benar oleh pengelompok. Dalam literature pengenalan pola, juga disebut sebagai presentase keseluruhan dari pengelompoknya, yaitu mencerminkan seberapa baik pengelompok tersebut mengelompokkan item dari berbagai kelas. Misalkan, $Acc(M)$ adalah tingkat keakuratan dari M , maka tingkat kesalahan atau tingkat ketidaktepatan klasifikasi, dari M adalah $1-acc(M)$.

Tabel 1. Tabel *Confusion Matrix*

| | | <i>Predicted class</i> | |
|---------------------|-----------|------------------------|-----------|
| | | Class = 1 | Class = 0 |
| <i>Actual Class</i> | Class = 1 | F11 | F10 |
| | Class = 0 | F10 | F01 |

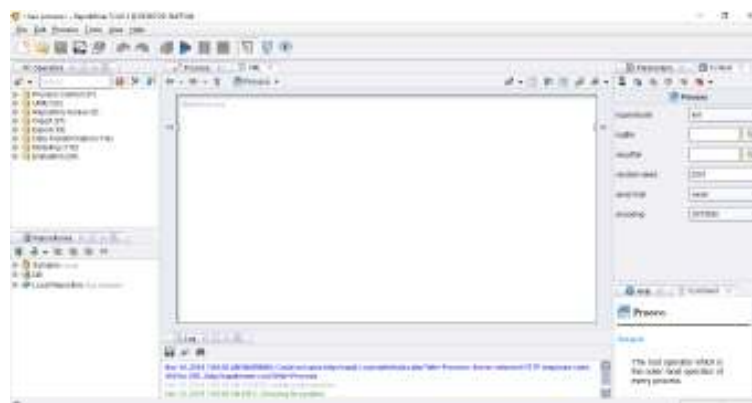
Sumber (Yuliana & Pratomo)

Informasi dari *confusion matrix* diperlukan untuk menentukan kinerja suatu model klasifikasi (*decision tree*). Informasi ini dapat diringkas ke dalam suatu nilai seperti akurasi.

$$Akurasi = \frac{F11 + F01}{F11 + F10 + F01 + F01} \quad (1)$$

1.1. RapidMiner

RapidMiner dapat digunakan untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih 500 operator data mining, termasuk operator untuk input, output, data *preprocessing* dan visualisasi.

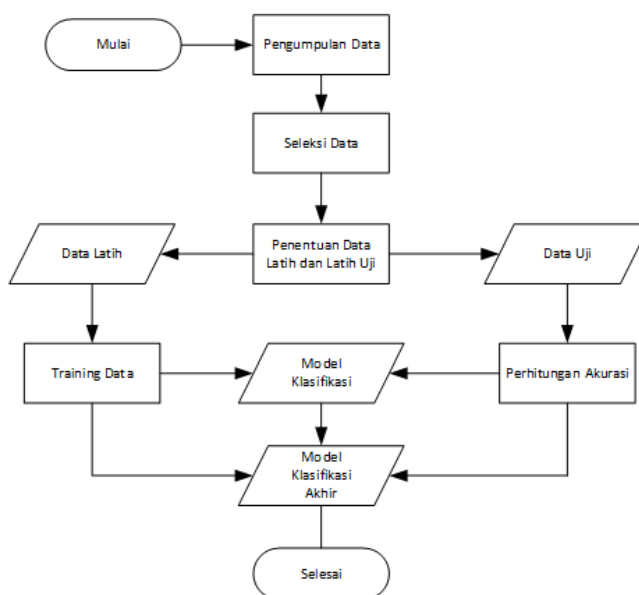


Gambar 3. Tampilan Aplikasi RapidMiner 5.3

2. METODE PENELITIAN

Tahap pertama adalah Pengumpulan Data melalui kuesioner yang akan disebar ke masyarakat yang datang ke Kantor DUKCAPIL Kabupaten Manokwari Selatan. Tahap kedua, Proprocessing untuk pemeriksaan data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Tahap ketiga, Penentuan Data Latih dan Data Uji untuk menyeleksi data. Dari hasil seleksi data

dilakukan pemisahan antara data yang digunakan sebagai data latih dan data yang digunakan untuk data *testing*. Tahap keempat adalah tahap Pembuatan Model Klasifikasi dengan menghitung data *training* dengan menggunakan algoritma C4.5. Dan tahap kelima Pengujian Model Klasifikasi yang telah dihasilkan kemudian diuji menggunakan *confusion matrix*. Penelitian melalui beberapa tahap seperti gambar 4.



Gambar 4. Tahapan Penelitian

3. HASIL DAN PEMBAHASAN

Setelah melalui proses pengumpulan data melalui Kuisisioner, dilakukan tabulasi menggunakan aplikasi Microsoft Excel untuk

dilakukan proses *preprocessing*. Nilai dari setiap item pertanyaan di setiap aspek kemudian dirata-ratakan. Adapun hasil *preprocessing* dapat dilihat pada tabel 2.

Tabel 2. Hasil Seleksi Data

| NO | ASPEK | | | | | KUALITAS LAYANAN |
|----|----------------|-----------|--------------|---------|--------|------------------|
| | BUKTI LANGSUNG | KEANDALAN | DAYA TANGGAP | JAMINAN | EMPATI | |
| 1 | 4 | 4.5 | 4 | 4 | 4 | BAIK |
| 2 | 4.5 | 4 | 3.5 | 5 | 4 | BAIK |
| 3 | 3.5 | 5 | 5 | 3 | 4 | BAIK |
| 4 | 4 | 5 | 5 | 3 | 5 | BAIK |
| 5 | 3.75 | 4.5 | 5 | 4 | 5 | BAIK |
| 6 | 4 | 5 | 4.5 | 3 | 5 | BAIK |
| 7 | 2.75 | 3 | 3 | 3 | 3 | BAIK |
| 8 | 3.75 | 4.5 | 4 | 4 | 4 | BAIK |
| 9 | 3.25 | 3 | 3 | 2 | 4 | BAIK |
| 10 | 4 | 2.5 | 2 | 5 | 4 | KURANG BAIK |
| 11 | 4 | 4 | 2 | 5 | 4 | KURANG BAIK |
| 12 | 4.25 | 4 | 2 | 5 | 4 | BAIK |
| 13 | 3 | 3 | 4 | 1 | 4 | KURANG BAIK |
| 14 | 3.5 | 4.5 | 4 | 3 | 5 | BAIK |
| 15 | | | | | | |
| . | ... | ... | ... | ... | ... | ... |
| . | | | | | | |
| . | | | | | | |
| 85 | 2.75 | 1.5 | 1.5 | 2 | 1.5 | KURANG BAIK |

Dari hasil tabulasi, nilai rata-rata dari setiap atribut asal kemudian di bagi kedalam 5 (lima)

kelas. Setiap interval kelas mempunyai jarak nilai seperti pada tabel 3.

Tabel 3. Bobot Interval Atribut

| Nilai Atribut | Bobot | Akronim |
|---------------|--------------------|---------|
| 1 – 1.99 | Sangat Kurang Baik | SKB |
| 2 – 2,00 | Kurang Baik | KB |
| 3 – 3.99 | Cukup | C |
| 4 – 4.99 | Baik | B |
| 5 | Sangat Baik | SB |

Data pada tabel 3 digunakan sebagai data training. Data training terdiri dari 6 (enam) atribut, 5 (lima) menjadi atribut asal yaitu Bukti Langsung, Keandalan, Daya Tanggap, Jaminan, Empati. Sementara Kualitas Layanan nantinya menjadi Atribut target.

Tahap selanjutnya adalah membuat model klasifikasi data mining menggunakan algoritma C4.5.

1. Pilih atribut sebagai akar

Untuk memilih atribut sebagai akar (*node*), didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung *gain* digunakan persamaan (2) dibawah ini:

$$\left[gain(y, A) = entropi(y) - \sum_{c \in nilai(A)} \frac{y_c}{y} * entropi(y_c) \right] \quad (2)$$

Dimana : y : Kasus
 A : Atribut
 n : Jumlah partisi atribut A

y_c : Jumlah kasus pada partisi ke-i
 S : Jumlah kasus

Sehingga didapat *Gain* (Total, *Empati*) = 0.11587328

Untuk menghitung nilai *Entropy* dihitung menggunakan persamaan (3) di bawah ini:

$$Entropi(y) = -p_1 \log_2 p_1 - p_2 \log_2 p_2 \dots - p_n \log_2 p_n \quad (3)$$

Dimana: y : Ruang (data) sample yang digunakan untuk training.

P1 : Jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu.

P2 : Jumlah yang bersolusi negatif (tidak mendukung) pada data sampel untuk kriteria tertentu.

Entropy [Total] =

$$\left(\left(-\frac{9}{68} \log_2 \frac{9}{68} \right) + \left(-\frac{59}{68} \log_2 \frac{59}{68} \right) \right)$$

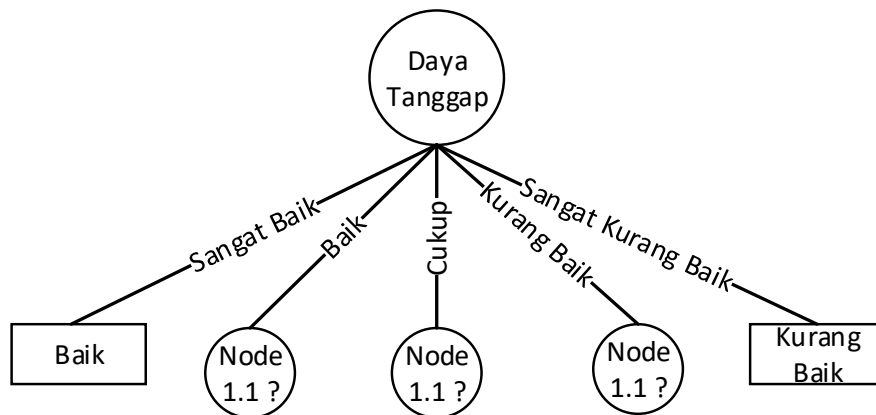
Sehingga didapat hasil Entropy [total] =0.563856
(Tabel 4)

Tabel 4. Perhitungann Node 1

| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|----------------|-----|------------------|-----------|------------------|------------|------------|
| 1 | | | 68 | 59 | 9 | 0.563856 | |
| | Bukti Langsung | | | | | | 0.13823521 |
| | | SKB | 2 | 0 | 2 | 0 | |
| | | KB | 5 | 3 | 2 | 0.97095059 | |
| | | C | 29 | 27 | 2 | 0.36205125 | |
| | | B | 27 | 24 | 3 | 0.50325833 | |
| | | SB | 5 | 5 | 0 | 0 | |
| | Keandalan | | | | | | 0.09101972 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 3 | 2 | 1 | 0.91829583 | |
| | | C | 29 | 22 | 7 | 0.79732651 | |
| | | B | 29 | 28 | 1 | 0.21639693 | |
| | | SB | 7 | 7 | 0 | 0 | |
| | Daya Tanggap | | | | | | 0.23384801 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 11 | 5 | 6 | 0.99403021 | |
| | | C | 18 | 17 | 1 | 0.30954343 | |
| | | B | 23 | 22 | 1 | 0.25801867 | |
| | | SB | 15 | 15 | 0 | 0 | |
| | Jaminan | | | | | | 0.22263389 |
| | | SKB | 6 | 3 | 3 | 1 | |
| | | KB | 9 | 5 | 4 | 0.99107606 | |
| | | C | 17 | 17 | 0 | 0 | |
| | | B | 22 | 22 | 0 | 0 | |
| | | SB | 14 | 12 | 2 | 0.59167278 | |
| | Empati | | | | | | 0.11587328 |
| | | SKB | 3 | 1 | 2 | 0.91829583 | |
| | | KB | 5 | 3 | 2 | 0.97095059 | |
| | | C | 13 | 12 | 1 | 0.39124356 | |
| | | B | 34 | 30 | 4 | 0.52255937 | |
| | | SB | 13 | 13 | 0 | 0 | |

Pada perhitungan *Gain* Node 1 dapat diketahui bahwa atribut dengan *gain* tertinggi adalah Daya Tanggap yaitu sebesar 0.23384801

dengan demikian dapat Daya Tanggap menjadi node akar. Sehingga pohon keputusan node 1 dapat digambarkan seperti pada gambar 5.



Gambar 5. Pohon Keputusan Node 1

2. Pilih Atribut Node 1.1

Langkah selanjutnya adalah mencari atribut yang belum diketahui pada node 1.1 yaitu Daya Tanggap Baik, Daya Tanggap Cukup, Daya Tanggap Kurang Baik. Adapun cara yang

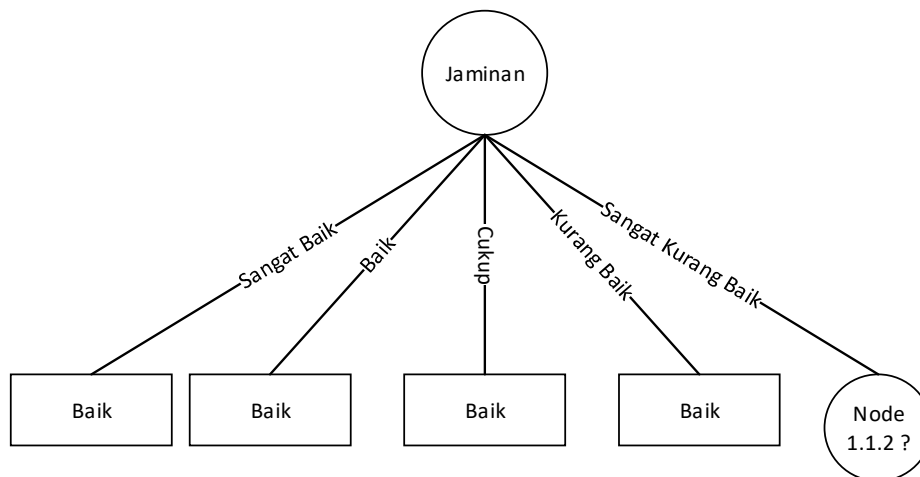
dilakukan adalah dengan memfilter nilai Daya Tanggap = Baik, Daya Tanggap = Cukup dan Daya Tanggap = Kurang Baik. Perhitungan Node 1.1 untuk Atribut Daya Tanggap Baik, dapat dilihat pada tabel 5.

Tabel 5. Perhitungan Node 1.1 Daya Tanggap Baik

| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|---------------------|-----|------------------|-----------|------------------|------------------|------------------|
| 1.1 | Daya Tanggap : Baik | | 23 | 22 | 1 | 0.2580187 | |
| | Bukti Langsung | | | | | | 0.036881 |
| | 1 | SKB | 0 | 0 | 0 | 0 | |
| | 2 | KB | 0 | 0 | 0 | 0 | |
| | 3 | C | 13 | 12 | 1 | 0.3912436 | |
| | 4 | B | 9 | 9 | 0 | 0 | |
| | 5 | SB | 1 | 1 | 0 | 0 | |
| | Keandalan | | | | | | 0.0689528 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 1 | 1 | 0 | 0 | |
| | | C | 8 | 7 | 1 | 0.5435644 | |
| | | B | 13 | 13 | 0 | 0 | |
| | | SB | 1 | 1 | 0 | 0 | |
| | Jaminan | | | | | | 0.1710621 |
| | | SKB | 2 | 1 | 1 | 1 | |
| | | KB | 1 | 1 | 0 | 0 | |
| | | C | 7 | 7 | 0 | 0 | |
| | | B | 12 | 12 | 0 | 0 | |
| | | SB | 1 | 1 | 0 | 0 | |
| | Empati | | | | | | 0.0233821 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 3 | 3 | 0 | 0 | |
| | | B | 16 | 15 | 1 | 0.3372901 | |
| | | SB | 4 | 4 | 0 | 0 | |

Dari perhitungan *Gain* Node 1.1 untuk Daya Tanggap Baik, terlihat bahwa nilai *gain*jaminan merupakan nilai yang tertinggi yaitu sebesar 0.1710621, sehingga atribut untuk node 1.1 daya tanggap baik adalah Atribut Jaminan.

Sehingga pohon keputusan node 1.1 Daya Tanggap Baik dapat digambarkan seperti pada gambar 6. Selanjutnya untuk Hasil perhitungan node 1.1 untuk Daya Tanggap Cukup dapat dilihat pada tabel 4.6 di bawah ini:



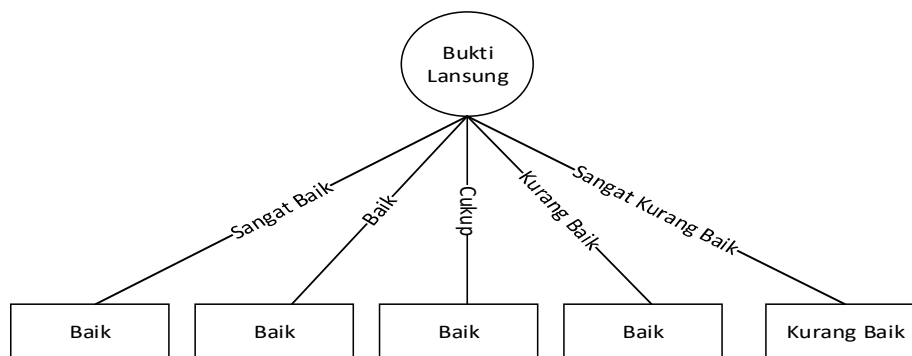
Gambar 6. Pohon Keputusan Node 1.1 Daya Tanggap Baik

Tabel 6. Perhitungan Node 1.1 Daya Tanggap Cukup

| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|----------------------|-----|------------------|-----------|------------------|------------------|------------------|
| 1.1 | Daya Tanggap : Cukup | | 18 | 17 | 1 | 0.3095434 | |
| | Bukti Langsung | | | | | | 0.3095434 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 1 | 1 | 0 | 0 | |
| | | C | 8 | 8 | 0 | 0 | |
| | | B | 6 | 6 | 0 | 0 | |
| | | SB | 2 | 2 | 0 | 0 | |
| | Keandalan | | | | | | 0.0409619 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 1 | 1 | 0 | 0 | |
| | | C | 11 | 10 | 1 | 0.439497 | |
| | | B | 5 | 5 | 0 | 0 | |
| | | SB | 1 | 1 | 0 | 0 | |
| | Jaminan | | | | | | 0.1292594 |
| | | SKB | 1 | 1 | 0 | 0 | |
| | | KB | 4 | 3 | 1 | 0.8112781 | |
| | | C | 3 | 3 | 0 | 0 | |
| | | B | 5 | 5 | 0 | 0 | |
| | | SB | 5 | 5 | 0 | 0 | |
| | Empati | | | | | | 0.1984323 |
| | | SKB | 2 | 1 | 1 | 1 | |
| | | KB | 1 | 1 | 0 | 0 | |
| | | C | 4 | 4 | 0 | 0 | |
| | | B | 10 | 10 | 0 | 0 | |
| | | SB | 1 | 1 | 0 | 0 | |

Dari perhitungan *Gain* Node 1.1 untuk Daya Tanggap : Cukup, terlihat bahwa nilai *gain*Bukti Langsung merupakan nilai yang tertinggi yaitu sebesar 0.3095434, sehingga

atribut untuk node 1.1 daya tanggap Cukup adalah Atribut Bukti Langsung. Pohon keputusan node 1.1 Daya Tanggap Cukup dapat digambarkan seperti pada gambar 7.



Gambar 7. Pohon Keputusan Node 1.1 Daya Tanggap Cukup

Selanjutnya dilakukan perhitungan untuk node 1.1 pada Daya Tanggap Kurang Baik. Untuk Hasil perhitungan node 1.1 untuk Daya

Tanggap Kurang Baik dapat dilihat pada tabel 7.

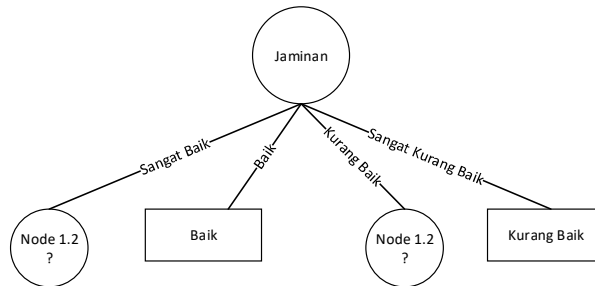
Tabel 7. Perhitungan Node 1.1 Daya Tanggap Kurang Baik

| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|----------------------------|-----|------------------|-----------|------------------|------------------|------------------|
| 1.1 | Daya Tanggap : Kurang Baik | | 11 | 5 | 6 | 0.9940302 | |
| | Bukti Langsung | | | | | | 0.1204265 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 3 | 1 | 2 | 0.9182958 | |
| | | C | 2 | 1 | 1 | 1 | |
| | | B | 5 | 3 | 2 | 0.9709506 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Keandalan | | | | | | 0.1981313 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 1 | 0 | 1 | 0 | |
| | | C | 6 | 2 | 4 | 0.9182958 | |
| | | B | 4 | 3 | 1 | 0.8112781 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Jaminan | | | | | | 0.4485757 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 4 | 3 | 1 | 0.8112781 | |
| | | C | 0 | 0 | 0 | 0 | |
| | | B | 3 | 3 | 0 | 0 | |
| | | SB | 3 | 1 | 2 | 0.9182958 | |
| | Empati | | | | | | 0.1911336 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 3 | 2 | 1 | 0.9182958 | |
| | | C | 2 | 1 | 1 | 1 | |
| | | B | 5 | 2 | 3 | 0.9709506 | |
| | | SB | 0 | 0 | 0 | 0 | |

Dari perhitungan Gain Node 1.1 untuk Daya Tanggap : Kurang Baik, terlihat bahwa nilai *gain* Atribut Jaminan merupakan nilai yang tertinggi yaitu sebesar 0.4484757,

sehingga atribut untuk node 1.1 Daya Tanggap Kurang Baik adalah Atribut Jaminan. Pohon keputusan node 1.1 Daya Tanggap

Kurang Baik dapat digambarkan seperti pada gambar 8.



Gambar 8. Pohon Keputusan Node 1.1 Daya Tanggap Kurang Baik

3. Pilih Atribut Node 1.1.2

Pada perhitungan Node 1.1 masih ada atribut yang belum diketahui yaitu atribut Daya Tanggap Baik dan Jaminan Sangat Kurang Baik, Daya Tanggap Kurang Baik dan Jaminan Kurang Baik serta Daya Tanggap Kurang Baik dan Jaminan Sangat Baik, sehingga perlu dilanjutkan menghitung

kembali *Entropy* dan *Gain* untuk Node 1.1.2 dengan atribut yang belum diketahui tersebut. Untuk node 1.1.2 yang pertama adalah Atribut Daya Tanggap Baik dan Jaminan Sangat Kurang Baik. Adapun hasil perhitungan *entropy* dan *gain* Atribut Daya Tanggap Baik dan Jaminan Sangat Kurang Baik dapat dilihat pada tabel 8.

Tabel 8. Hasil Perhitungan Node 1.1.2 Daya Tanggap Baik dan Jaminan Sangat Kurang Baik

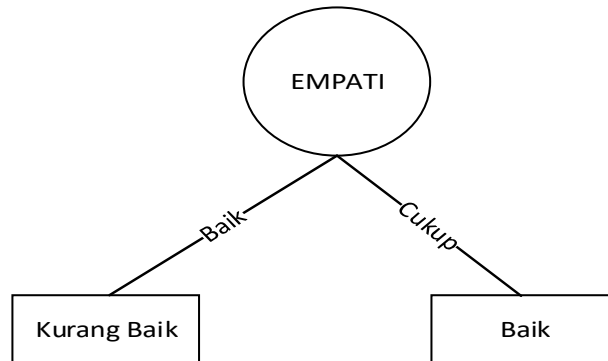
| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|---|-----|------------------|-----------|------------------|---------|------|
| 1.2 | Daya Tanggap Baik, Jaminan : Sangat Kurang Baik | | 2 | 1 | 1 | 1 | |
| | Bukti Langsung | | | | | | 0 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 2 | 1 | 1 | 1 | |
| | | B | 0 | 0 | 0 | 0 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Keandalan | | | | | | 0 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 2 | 1 | 1 | 1 | |
| | | B | 0 | 0 | 0 | 0 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Empati | | | | | | 1 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 1 | 1 | 0 | 0 | |
| | | B | 1 | 0 | 1 | 0 | |
| | | SB | 0 | 0 | 0 | 0 | |

Dari perhitungan Gain Node 1.1.2 untuk Daya Tanggap Baik, Jaminan Sangat Kurang Baik, terlihat bahwa nilai *gain* Atribut Empati

merupakan nilai yang tertinggi yaitu sebesar 1, sehingga atribut untuk node 1.1.2 Daya Tanggap Baik, Jaminan Sangat Kurang Baik

adalah Atribut Empati. Sehingga pohon keputusan node 1.1.2 Daya Tanggap Baik

dan Jaminan Sangat Kurang Baik dapat digambarkan seperti pada gambar 9.



Gambar 9. Pohon Keputusan Node 1.1.2 Daya Tanggap Baik & Jaminan Sangat Kurang Baik

Tahapan selanjutnya adalah mencari nilai *entropy* dan *gain* Untuk node 1.1.2 Atribut Daya Tanggap Kurang Baik dan Jaminan

Kurang Baik. Adapun hasil perhitungan dapat dilihat pada tabel 9.

Tabel 9. Hasil Perhitungan Node 1.1.2 Daya Tanggap Kurang Baik dan Jaminan Kurang Baik

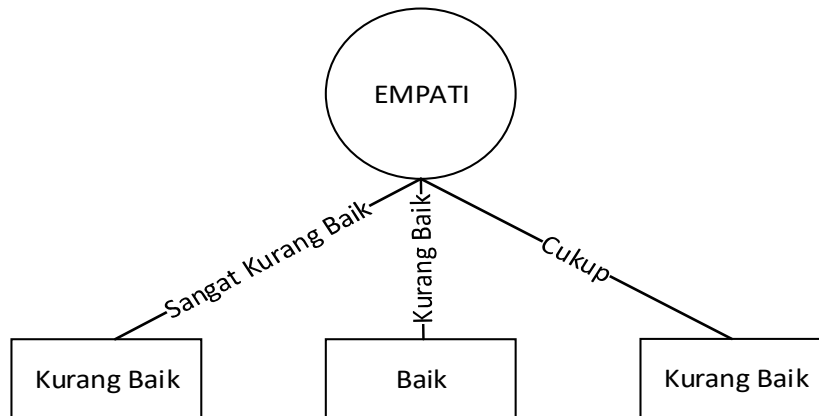
| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|---|-----|------------------|-----------|------------------|-----------|-----------|
| 1.2 | Daya Tanggap: Kurang Baik, Jaminan: Kurang Baik | | 4 | 1 | 3 | 0.8112781 | |
| | Bukti Langsung | | | | | | 0.1225562 |
| | 1 | SKB | 1 | 0 | 1 | 0 | |
| | 2 | KB | 3 | 1 | 2 | 0.9182958 | |
| | 3 | C | 0 | 0 | 0 | 0 | |
| | 4 | B | 0 | 0 | 0 | 0 | |
| | 5 | SB | 0 | 0 | 0 | 0 | |
| | Keandalan | | | | | | 0 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 4 | 1 | 3 | 0.8112781 | |
| | | B | 0 | 0 | 0 | 0 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Empati | | | | | | 0.3112781 |
| | | SKB | 1 | 0 | 1 | 0 | |
| | | KB | 2 | 1 | 1 | 1 | |
| | | C | 1 | 0 | 1 | 0 | |
| | | B | 0 | 0 | 0 | 0 | |
| | | SB | 0 | 0 | 0 | 0 | |

Dari perhitungan Gain Node 1.1.2 untuk Daya Tanggap: Kurang Baik, Jaminan:

Kurang Baik, terlihat bahwa nilai *gain* Atribut Empati merupakan nilai yang tertinggi yaitu

sebesar 0.3112781, sehingga atribut untuk node 1.1.2 Daya Tanggap Kurang Baik dan Jaminan Kurang Baik adalah Atribut Empati. Pohon keputusan yang terbentuk dari

perhitungan node 1.1.2 Daya Tanggap Kurang Baik dan Jaminan Kurang Baik seperti pada gambar 10.



Gambar 10. Pohon Keputusan Node 1.1.2 Daya Tanggap Kurang Baik & Jaminan Kurang Baik

Tahapan selanjutnya adalah mencari nilai *entropy* dan *gain* Untuk node 1.1.2 Atribut Daya Tanggap Kurang Baik dan Jaminan

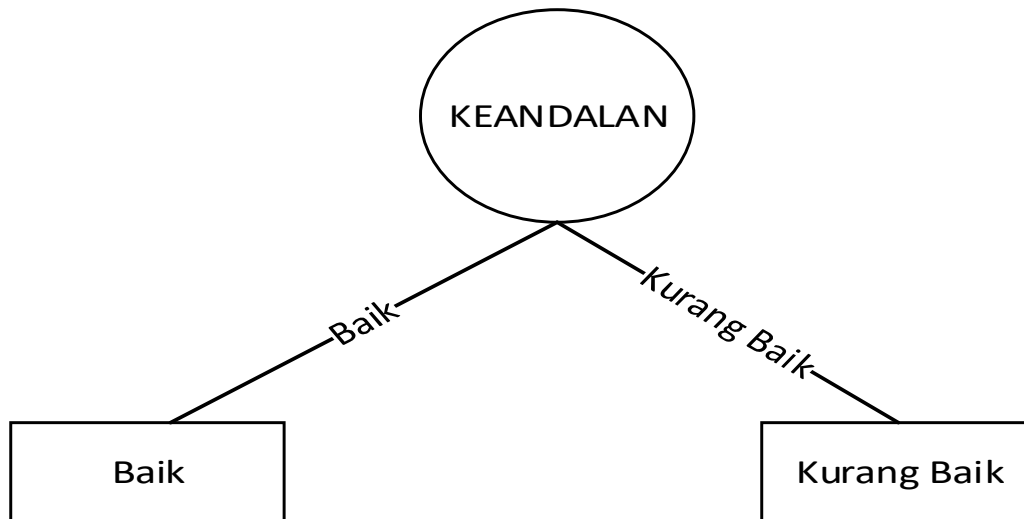
Sangat Baik. Adapun hasil perhitungan dapat dilihat pada tabel 10.

Tabel 10. Hasil Perhitungan Node 1.1.2 Daya Tanggap Kurang Baik dan Jaminan Sangat Baik

| Node | Total | | Jumlah Kasus (S) | Baik (S1) | Kurang Baik (S2) | Entropy | Gain |
|------|---|-----|------------------|-----------|------------------|------------------|------------------|
| 1.2 | Daya Tanggap: Kurang Baik, Jaminan: Sangat Baik | | 3 | 1 | 2 | 0.9182958 | |
| | Bukti Langsung | | | | | | 0 |
| | 1 | SKB | 0 | 0 | 0 | 0 | |
| | 2 | KB | 0 | 0 | 0 | 0 | |
| | 3 | C | 0 | 0 | 0 | 0 | |
| | 4 | B | 3 | 1 | 2 | 0.9182958 | |
| | 5 | SB | 0 | 0 | 0 | 0 | |
| | Keandalan | | | | | | 0.2516292 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 1 | 0 | 1 | 0 | |
| | | C | 0 | 0 | 0 | 0 | |
| | | B | 2 | 1 | 1 | 1 | |
| | | SB | 0 | 0 | 0 | 0 | |
| | Empati | | | | | | 0.1225562 |
| | | SKB | 0 | 0 | 0 | 0 | |
| | | KB | 0 | 0 | 0 | 0 | |
| | | C | 0 | 0 | 0 | 0 | |
| | | B | 3 | 1 | 2 | 0.9182958 | |
| | | SB | 0 | 0 | 0 | 0 | |

Dari perhitungan Gain Node 1.1.2 untuk Daya Tanggap : Kurang Baik, Jaminan : Sangat Baik, terlihat bahwa nilai *gain* Atribut keandalan merupakan nilai yang tertinggi yaitu sebesar 0.2516292, sehingga atribut untuk node 1.1.2 Daya Tanggap : Kurang

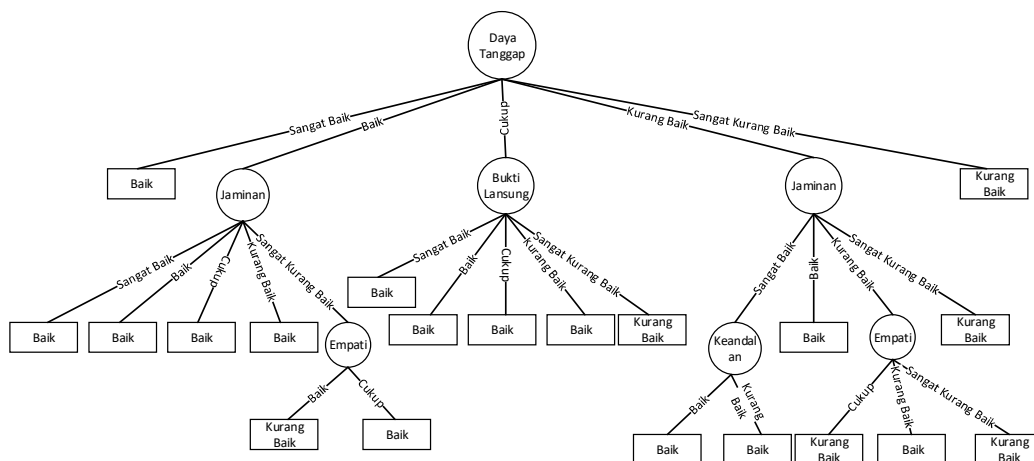
Baik, Jaminan : Sangat Baik adalah Atribut Keandalan. pohon keputusan hasil perhitungan node 1.1.2 Daya Tanggap : Kurang Baik, Jaminan : Sangat Baik yang terbentuk seperti pada gambar 11.



Gambar 11. Pohon Keputusan Node 1.2 Daya Tanggap Kurang Baik & Jaminan Sangat Baik

Sehingga dari hasil *entropy* dan *gain* pada algoritma C4.5 secara keseluruhan pohon

keputusan yang dibentuk seperti pada gambar 12.



Gambar 12. Pohon Keputusan Klasifikasi Kualitas Layanan Publik

Dari pohon keputusan diatas maka didapatkan rule model klasifikasi sebagai berikut:

1. Jika Daya Tanggap= "Sangat Baik" maka Kualitas Layanan = "Baik"
2. Jika Daya Tanggap = "Baik" dan Jaminan = "Sangat Baik" maka Kualitas Layanan = "Baik"
3. Jika Daya Tanggap = "Baik" dan Jaminan = "Baik" maka Kualitas Layanan = "Baik"
4. Jika Daya Tanggap = "Baik" dan Jaminan = "Cukup" maka Kualitas Layanan = "Baik"
5. Jika Daya Tanggap = "Baik" dan Jaminan = "Kurang Baik" maka Kualitas Layanan = "Baik"
6. Jika Daya Tanggap = "Baik" dan Jaminan = "Sangat Kurang Baik" maka Kualitas Layanan = "Kurang Baik"
7. Jika Daya Tanggap = "Cukup" dan Bukti Langsung = "Sangat Baik" maka Kualitas Layanan = "Baik"
8. Jika Daya Tanggap = "Cukup" dan Bukti Langsung = "Baik" maka Kualitas Layanan = "Baik"
9. Jika Daya Tanggap = "Cukup" dan Bukti Langsung = "Cukup" maka Kualitas Layanan = "Baik"
10. Jika Daya Tanggap = "Cukup" dan Bukti Langsung = "Kurang Baik" maka Kualitas Layanan = "Baik"
11. Jika Daya Tanggap = "Cukup" dan Bukti Langsung = "Sangat Kurang Baik" maka Kualitas Layanan = "Kurang Baik"

12. Jika Daya Tanggap = "Kurang Baik" dan Jaminan = "Sangat Baik" dan Keandalan = "Baik" maka Kualitas Layanan = "Baik"
13. Jika Daya Tanggap = "Kurang Baik" dan Jaminan = "Sangat Baik" dan Keandalan = "Kurang Baik" maka Kualitas Layanan = "Kurang Baik"
14. Jika Daya Tanggap = "Kurang Baik" dan Jaminan = "Kurang Baik" dan Empati= "Cukup" maka Kualitas Layanan = "Kurang Baik"
15. Jika Daya Tanggap = "Kurang Baik" dan Jaminan = "Kurang Baik" dan Empati= "Kurang Baik" maka Kualitas Layanan = "Baik"
16. Jika Daya Tanggap = "Kurang Baik" dan Jaminan = "Kurang Baik" dan Empati= "Sangat Kurang Baik" maka Kualitas Layanan = "Kurang Baik"
17. Jika Daya Tanggap= "Sangat Kurang Baik" maka Kualitas Layanan = "Kurang Baik"

4. Pengujian Model

Setelah melakukan tranning terhadap data latih untuk membuat model klasifikasi dengan algoritma C4.5, kemudian dilakukan pengujian terhadap model yang telah dihasilkan dari *training* data. Pengujian model dilakukan terhadap 85 data testing. Untuk menghitung model digunakan nilai *confusion matrix*.

Untuk hasil perhitungan akurasi dengan *confusion matrix* dapat dilihat pada tabel 11

Tabel 11. Hasil Perhitungan *Confusion Matrix*

| | | Kelas Prediksi | |
|--------------|---------------------|----------------|---------------------|
| | | Kelas "Baik" | Kelas "Kurang Baik" |
| Kelas Aktual | Kelas "Baik" | 72 | 0 |
| | Kelas "Kurang Baik" | 5 | 8 |

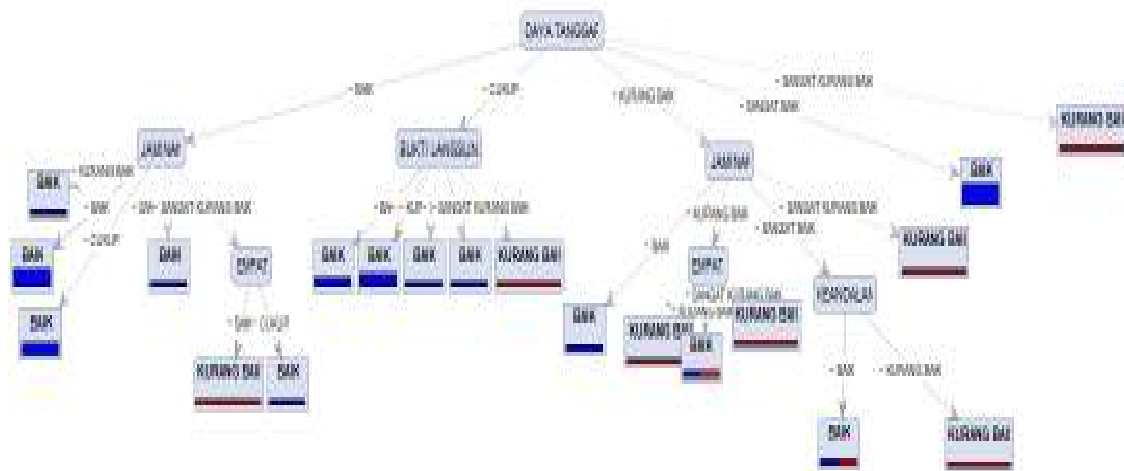
$$\text{Akurasi} = \frac{72+8}{72+5+0+8}$$

$$\text{Akurasi} = \frac{80}{85}$$

Akurasi = 0.9412 atau 94, 12%

pada Rapid Miner didapatkan *Rule* yang sama sesuai dengan perhitungan algoritma C4.5 secara manual. Adapun hasil pembentukan pohon keputusan dengan aplikasi Rapid Miner 5.3 dapat dilihat pada gambar 13.

Langkah selanjutnya ada melihat hasil hasil pemodelan Algoritma C4.5 yang telah dihitung secara manual dengan aplikasi Rapid Miner. Dari pembentukan pohon keputusan



Gambar 13. Pohon Keputusan Algoritma C4.5 pada Rapid Miner

| accuracy: 94.12% | | | |
|-------------------|-----------|------------------|-----------------|
| | true BAIK | true KURANG BAIK | class precision |
| pred. BAIK | 72 | 5 | 93.51% |
| pred. KURANG BAIK | 0 | 8 | 100.00% |
| class recall | 100.00% | 61.54% | |

Gambar 14. Hasil Pengujian Model Klasifikasi pada RapidMiner

Hasil perhitungan akurasi dengan menggunakan aplikasi RapidMiner 5.3 dapat dilihat pada gambar 14 di atas.

Dari hasil perhitungan dengan Aplikasi RapidMiner 5.3 didapatkan hasil yang sama dengan perhitungan manual, sehingga model klasifikasi dapat dikatakan sudah baik.

4. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan, maka dapat diambil kesimpulan, telah didapatkan model klasifikasi algoritma C4.5 untuk menklasifikasikan kualitas layanan publik yang ada di Kabupaten Manokwari Selatan. Atribut Daya Tanggap dan Jaminan merupakan indikator yang memiliki pengaruh paling besar dalam penelitian ini, hal ini terlihat dari hasil akhir

pohon keputusan yang memunculkan Daya Tanggap sebagai node tertinggi dan Atribut Jaminan merupakan Atribut yang paling sering muncul. Hasil pembentukan pohon keputusan klasifikasi algoritma C4.5 secara manual sudah sama dengan pohon keputusan yang dibentuk oleh Aplikasi Rapid Miner. Dari Pengujian yang dilakukan terhadap 85 data *testing*, didapatkan tingkat akurasi sebesar 94,12%

5. DAFTAR PUSTAKA

- [1] H. Dr, Kualitas Layanan Publik ; Konsep, Dimensi, Indikator dan Implementasinya, Yogyakarta: Gava Media, 2011.
- [2] A. P. Sujana, "MEMANFAATKAN BIG DATA UNTUK MENDETEKSI EMOSI," *Jurnal Teknik Komputer Unikom* –

- Komputika*, vol. Volume 2, no. No. 2, pp. 1-4, 2013.
- [3] F. S., "Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer," p. 63–70, 2013.
- [4] L. A. G and H. G. P, "Optimization of C4.5 Decision Tree Algorithm for Data Mining Application," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 3, no. 3, p. 341–345, 2013.
- [5] S. A, "Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree," *Studi Kasus : Yayasan Perguruan vol. V*, p. 7–12, 2015.
- [6] U. D, "Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5," 2010.
- [7] P. Destarianto, W. . K. Dewanto and H. A. Putranto, "ANALISIS PERBANDINGAN IMPLEMENTASI KERNEL PADA LIBRARY LibSVM UNTUK KLASIFIKASI SENTIMEN MENGGUNAKAN WEKA," in *Seminar Nasional APTIKOM (SEMNASTIKOM)*, Jayapura, 2017.
- [8] http://cs.calstatela.edu/wiki/index.php/Cources/CS_461/Fall_2011/Data_Mining_and_Weka 2011 [Online]. [Accessed Oktober 2018]