

Analisis Tingkat Kepuasan Masyarakat Terhadap Pelayanan Administratif di Kantor Kecamatan Cikarang Utara Menggunakan Metode Servqual Berbasis Algoritma C4.5

^{1*}Rafi Eka Candra, ²Khoerul Ahmad Fadillah

¹, Universitas Bina Sarana Informatika, dst

Authors email: ¹raficandra398@gmail.com, ² khoerulahmadfdl@gmail.com dst

Received July 2025 | Revised July 2025 | Accepted Agustus 2025

*Correspondence Author

Abstract

Good public services show that the government is running well, but so far, measuring public satisfaction at Cikarang Utara Sub-District Office has not been done properly and in an organized way. This study aims to classify the level of public satisfaction with administrative services, identify the main factors that influence it based on the SERVQUAL dimensions, and evaluate the accuracy of the model using the C4.5 algorithm. This study uses a quantitative approach by distributing a questionnaire to 105 people who have used the service before. After the data cleaning process was completed, there were 98 valid data entries that were processed using the RapidMiner tool. The data was analyzed using five aspects of SERVQUAL: reliability, responsiveness, assurance, empathy, and tangibles, as a basis for classifying the attributes. The result of the decision tree formation process shows that attribute E1, which is the Empathy dimension from the first question, is in the first position and becomes the root node. This shows that factors like empathy, such as when employees give extra attention and guidance to residents, are the main factors that affect public satisfaction. Evaluating the model's performance using the Accuracy method shows an accuracy rate of 66.33% and a classification error rate of 33.67%. This achievement shows that the C4.5 algorithm model is quite good and suitable for measuring the level of public satisfaction with administrative services at the Cikarang Utara Sub-District Office

Keywords: Kepuasan Masyarakat, Pelayanan Administratif, SERVQUAL, Algoritma C4.5, Decision Tree..

INTRODUCTION

Kualitas pelayanan bisa dilihat dengan beberapa cara, salah satunya adalah metode SERVQUAL yang dirancang oleh Parasuraman, Zeithaml, dan Berry pada tahun 1988. Dalam teori SERVQUAL yang mereka ajukan, kualitas pelayanan diukur melalui lima aspek utama, yaitu keandalan, responsif, kepastian, empati, dan bukti fisik. Keandalan berarti mampu memberikan layanan yang dijanjikan dengan tepat dan bisa diandalkan. Daya tanggap adalah kemampuan untuk siap membantu

pelanggan dan memberikan layanan dengan cepat. Kepastian meliputi pengetahuan dan sikap sopan karyawan, serta kemampuan mereka dalam membangun kepercayaan dan keyakinan. Empati adalah perhatian yang diberikan kepada pelanggan secara pribadi. Sementara itu, bukti fisik mencakup fasilitas fisik, peralatan, dan penampilan anggota staf (Jumawan & Prasetyo, 2024). Kelima dimensi ini bisa digunakan sebagai acuan dalam menilai tingkat kepuasan masyarakat terhadap layanan yang diberikan oleh lembaga pemerintah.

Sehubungan dengan pelayanan publik tersebut, Kecamatan Cikarang Utara sebagai salah satu instansi pemerintah yang melayani masyarakat dengan urusan administratif juga wajib mampu memberikan pelayanan yang baik dan berkualitas.

Namun, dalam pelaksanaannya, pengukuran tingkat kepuasan masyarakat terhadap pelayanan di Kecamatan Cikarang Utara masih belum dilakukan secara optimal dan terstruktur. Evaluasi pelayanan yang dilakukan belum sepenuhnya didukung oleh pengolahan data yang sistematis, sehingga informasi yang diperoleh belum mampu memberikan gambaran yang jelas mengenai tingkat kepuasan masyarakat. Kondisi ini membuat pihak kecamatan kesulitan dalam menentukan faktor-faktor apa saja yang perlu diperbaiki agar pelayanan bisa ditingkatkan.

Sejalan dengan rumusan masalah tersebut Tujuan dari penelitian ini adalah:

1. Mengklasifikasikan tingkat kepuasan masyarakat terhadap pelayanan administratif di Kecamatan Cikarang Utara menggunakan algoritma C4.5.
2. Mengetahui faktor-faktor yang mempengaruhi tingkat kepuasan masyarakat berdasarkan dimensi kualitas pelayanan (SERVQUAL).
3. Mengetahui tingkat akurasi model klasifikasi yang dihasilkan dari penerapan algoritma C4.5 dalam mengukur kepuasan masyarakat.

State Of The Art

Berdasarkan penelitian sebelumnya yang tersaji, Penelitian oleh (Lasarudin et al., 2022) menunjukkan bahwa penerapan algoritma C4.5 mampu menghasilkan tingkat akurasi yang tinggi dalam mengklasifikasikan kepuasan masyarakat. Selanjutnya, penelitian (Ningsih et al., 2025) menyatakan bahwa algoritma C4.5 dapat membantu dalam mengolah data kepuasan yang bersifat subjektif menjadi lebih terstruktur dan sistematis. Selain itu, penelitian (Mukminna et al., 2025)

mengungkapkan bahwa faktor kualitas layanan menjadi salah satu penentu utama dalam tingkat kepuasan pengguna sistem informasi.

Di sisi lain, penelitian (Permatasari et al., 2024) juga menunjukkan bahwa algoritma C4.5 efektif digunakan dalam menganalisis tingkat kepuasan pelanggan pada layanan berbasis teknologi. Sementara itu, penelitian (Nisrina & Kustiyono, 2024) menjelaskan bahwa faktor pelayanan dan fasilitas memiliki pengaruh signifikan terhadap tingkat kepuasan konsumen.

Berdasarkan hasil kajian penelitian sebelumnya, algoritma C4.5 sudah sering digunakan untuk mengklasifikasikan tingkat kepuasan di berbagai bidang, seperti pendidikan, perusahaan, restoran, media, dan layanan berbasis aplikasi. Namun, penelitian yang menggunakan algoritma C4.5 untuk mengklasifikasikan tingkat kepuasan masyarakat terhadap pelayanan administratif di instansi pemerintahan tingkat kecamatan dengan menggunakan dimensi kualitas pelayanan SERVQUAL sebagai atribut utama masih sangat sedikit. Oleh karena itu, penelitian ini menawarkan sesuatu yang baru dengan menerapkan lima dimensi SERVQUAL, yaitu *reliability*, *responsiveness*, *assurance*, *empathy*, dan *tangibles*, sebagai atribut klasifikasi dalam membangun model *decision tree* menggunakan algoritma C4.5 pada pelayanan administratif di Kantor Kecamatan Cikarang Utara.

LITERATURE REVIEW

Metode Kuantitatif

Metode kuantitatif adalah salah satu pendekatan dalam penelitian ilmiah yang sering digunakan oleh para peneliti. Karena dengan metode ini, peneliti mencoba menemukan berbagai jenis pengetahuan yang masih tertutup dan belum diketahui manusia secara sistematis, teratur, objektif, dan dapat dipercaya. Penelitian kuantitatif menggunakan pendekatan primer yang fokus pada hubungan sebab-akibat, hubungan antar variabel, hipotesis, dan pertanyaan yang lebih spesifik. Penelitian

ini juga melibatkan pengukuran, observasi, serta pengujian terhadap sebuah teori. Penelitian dengan pendekatan ini membutuhkan sebuah eksperimen, lalu hasilnya diumpamakan melalui survei, kemudian hasil tersebut dianalisis menggunakan statistika agar bisa mendapatkan hasil yang diharapkan (Syahroni, 2022).

Metode SERVQUAL

Menurut Parasuraman, Zeithaml, dan Berry pada tahun 1988, dalam teori Servqual yang mereka buat, kualitas layanan bisa dilihat dari lima aspek utama yaitu keandalan, responsif, kepastian, empati, dan bukti fisik. Keandalan berarti kemampuan untuk memberikan layanan yang telah dijanjikan dengan tepat dan dapat diandalkan. Daya tanggap adalah kemampuan untuk siap membantu pelanggan dan memberikan layanan dengan cepat. Kepastian mencakup pemahaman dan sikap sopan karyawan, serta kemampuan mereka dalam membangun rasa percaya dan keyakinan. Empati adalah perhatian yang diberikan kepada pelanggan secara pribadi. Sementara itu, bukti fisik mencakup fasilitas fisik, peralatan, dan penampilan anggota staf (Jumawan & Prasetyo, 2024).

Algoritma C4.5

Algoritma C4.5 adalah algoritma yang digunakan untuk mengklasifikasikan data dengan cara membuat keputusan. Algoritma C4.5 adalah versi perluasan dari algoritma ID3 dan menggunakan prinsip pohon keputusan yang hampir sama. Algoritma C4.5 adalah program yang menerima sekumpulan data yang sudah diberi label dan menghasilkan pohon keputusan sebagai hasilnya. Algoritma C4.5 membuat pohon keputusan berdasarkan data latihan yang serupa dengan Algoritma ID3, tetapi menggunakan prinsip entropi informasi sebagai dasarnya. Algoritma ini menerima masukan berupa contoh data latihan yang digunakan untuk membuat pohon keputusan yang sudah diuji kebenarannya, dan data tersebut merupakan

bagian-bagian informasi yang nantinya akan digunakan sebagai acuan dalam mengklasifikasikan data (Nugroho & Kristiana, 2022).

Decision Tree

Untuk membentuk pohon keputusan adalah dengan memilih atribut sebagai *root node* didasarkan pada nilai information gain tertinggi dari atribut-atribut yang ada. Sebelum perhitungan information gain, akan dilakukan perhitungan entropy. Entropy merupakan distribusi probabilitas untuk mengukur tingkat homogenitas distribusi kelas dari sebuah himpunan data. Semakin tinggi tingkat entropy dari sebuah data maka semakin homogen distribusi kelas pada data tersebut. Perhitungan information gain menggunakan rumus 2, sedangkan entropy menggunakan rumus 1 (Tan & Kusnawi, 2022).

Entropy

Entropy mengukur ketidakpastian atau kekacauan dalam data. Entropy (S) untuk himpunan data S dihitung dengan rumus:

$$\text{Entropy}(S) = - \sum_{i=1}^n p_i \cdot \log_2 p_i$$

Dimana: S = Dataset kasus

n = banyaknya label

Pi = probabilitas dari kelas i

(Pradana et al., 2024).

Information Gain

Information gain memeriksa pola data dan menggunakan pola tersebut sebagai pedoman untuk memilih atribut yang akan digunakan dalam pembuatan pohon keputusan. Atribut yang dipilih akan menghasilkan pembagian data yang seragam dan berpotensi menghasilkan pohon keputusan yang sederhana dengan jumlah percabangan yang minimal. Untuk menghitung nilai gain menggunakan rumus:

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

Keterangan:

S: himpunan Kasus

A: Atribut

n: jumlah kasus dalam S

|S_i|: jumlah kasus pada partisi ke-i

|S|: jumlah kasus dalam S

(Xsanal Hakim et al., 2024)

Split Info

Split Info menghitung jumlah informasi yang digunakan untuk melakukan pembagian berdasarkan atribut tertentu, sehingga dapat memperhitungkan pembagian yang kurang berguna. Rumusnya:

$$\text{SplitInfo}(S, A) = - \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S}$$

dimana S_i adalah subset yang dihasilkan oleh nilai i pada atribut A (Pradana et al., 2024).

Gain Ratio

Gain Ratio adalah rasio antara Gain dan Split Info. Digunakan untuk menormalkan Gain agar tidak terlalu berpihak pada atribut dengan banyak nilai unik:

$$\text{GainRatio} = \frac{\text{Gain}(A)}{\text{SplitInfo}(S, A)}$$

Setelah atribut terbaik dipilih berdasarkan Gain Ratio tertinggi, algoritma membagi data menjadi cabang-cabang berdasarkan nilai-nilai atribut tersebut. Aturan terbentuk dari setiap cabang menuju ke daun pohon (Pradana et al., 2024).

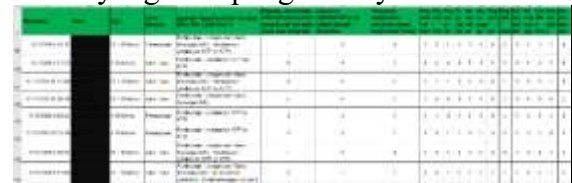
METHODS

Data Responden

Penelitian ini telah mengumpulkan data sebanyak 105 responden yang merupakan masyarakat yang pernah menggunakan pelayanan administratif di Kantor Kecamatan Cikarang Utara. Pemilihan responden ini dilakukan secara *Non-probability sampling* dengan kriteria utama yaitu pernah menggunakan pelayanan administratif di Kantor Kecamatan Cikarang Utara. Pengumpulan data dilakukan dengan menggunakan kuesioner *google form* pada periode januari 2026 hingga april 2026. Total responden 105

tersebut sudah menunjukkan partisipasi yang baik dari masyarakat kecamatan cikarang utara dalam pengisian kuesioner untuk penelitian ini.

Dataset yang akan dipilih untuk diolah menggunakan metode algoritma C4.5 yaitu: data responden yang memiliki beberapa atribut, yaitu keandalan (*reliability*), daya tanggap (*responsiveness*), jaminan (*assurance*), empati (*empathy*), dan bukti fisik (*tangibles*), dengan output berupa tingkat kepuasan masyarakat (Sangat Tidak Puas, Tidak Puas, Cukup Puas, Puas, dan Sangat Puas). Data tersebut digunakan sebagai dasar dalam proses klasifikasi untuk mengetahui tingkat kepuasan masyarakat terhadap pelayanan yang diberikan serta mengidentifikasi faktor-faktor yang mempengaruhinya.



Pembersihan Data



Pada Gambar ini adalah tahapan pembersihan data yang dilakukan bertujuan untuk membersihkan data mentah sebelum masuk ke tahap transformasi data berikut adalah penjelasan tabel hasil pembersihan data yang dilakukan:

Eliminasi Data Tidak Layak (*Noise*): Baris yang ditandai dengan warna merah merupakan data anomali atau data bising (*noise*) yang tidak memenuhi kriteria kelayakan penelitian, seperti adanya ketidakkonsistenan pengisian (contoh: jenis kelamin dan layanan tidak sesuai) atau pengisian nilai yang ekstrem. Pada tahap ini, ditemukan sebanyak 2 (dua) baris data tidak layak, sehingga data tersebut dihapus dari dataset.

Reduksi Data Ganda (*Duplicates*): Baris yang ditandai dengan warna kuning menunjukkan adanya indikasi duplikasi pengisian kuesioner oleh responden yang

sama. Berdasarkan penelusuran data, terdapat 9 (sembilan) baris data yang berasal dari 4 (empat) responden yang mengirimkan jawaban secara berulang. Untuk menjaga objektivitas dan validitas pengujian, data tersebut direduksi dengan cara menyisakan hanya 1 (satu) entri data untuk masing-masing responden tersebut, sehingga total data ganda yang dihapus dari sistem adalah sebanyak 5 (lima) baris.

Melalui tahapan pembersihan data ini, data yang terindikasi *noise* dan duplikat telah berhasil direduksi. Hasil akhir pembersihan data ini menyisakan 98 data bersih (dari total awal 105 data) yang valid, konsisten, dan siap untuk diproses lebih lanjut pada tahapan transformasi data

Uji Validitas



Hasil Pengujian pada Gambar ini dilakukan secara vertikal terhadap 98 sampel data responden. Nilai korelasi (r-hitung) dihitung dengan menggunakan rumus korelasi *Pearson Bivariate* pada *Microsoft Excel*. Caranya adalah dengan menghubungkan skor dari setiap butir pertanyaan dengan skor total dari seluruh atribut yang diisi oleh responden.

hasil uji statistik untuk 15 butir pertanyaan yang mewakili lima dimensi SERVQUAL (Keandalan, Daya Tanggap, Jaminan, Empati, dan Bukti Fisik) memberikan ringkasan analisis sebagai berikut:

1. Dimensi Keandalan (*Reliability*): Tiga butir pertanyaan yaitu P1, P2, dan P3 masing-masing menghasilkan nilai r-hitung sebesar 0,7355, 0,6760, dan 0,6413. Ketiga nilai tersebut jauh lebih besar dari nilai r-tabel acuan yaitu 0,2347, sehingga semua indikator dalam dimensi ini dianggap valid.
2. Dimensi Daya Tanggap (*Responsiveness*): Tiga pertanyaan (P1, P2, P3) memberikan nilai r-hitung masing-masing yaitu 0,7712, 0,7779, dan 0,7422. Karena semua nilai r-hitung lebih besar dari 0,2347, maka indikator pada dimensi ini dianggap valid.

3. Dimensi Jaminan (*Assurance*): Tiga pertanyaan yaitu P1, P2, dan P3 memberikan nilai r-hitung masing-masing sebesar 0,7960, 0,6507, dan 0,7230. Seluruh pertanyaan dalam dimensi jaminan terbukti melebihi batas r-tabel sebesar 0,2347, sehingga dianggap valid.
4. Dimensi Empati (*Empathy*): Tiga pertanyaan, yaitu P1, P2, dan P3, memberikan skor r-hitung yang tertinggi dibandingkan dimensi lainnya, masing-masing sebesar 0,8017, 0,8351, dan 0,8216. Dengan nilai yang lebih besar dari r tabel 0,2347, semua indikator empati dianggap valid. Hal ini juga sesuai dengan hasil analisis pohon keputusan yang menunjukkan bahwa variabel ini mempunyai pengaruh yang paling besar.
5. Dimensi bukti fisik (*tangibles*): Tiga pertanyaan yaitu P1, P2, dan P3 masing-masing menghasilkan nilai r-hitung sebesar 0,8014, 0,8330, dan 0,7644. Semua indikator fisik tersebut secara pasti berada di atas nilai r-tabel 0,2347 dan dinyatakan valid.

Berdasarkan hasil analisis korelasi Pearson, dapat disimpulkan bahwa semua butir pertanyaan (15 indikator) memiliki nilai r-hitung yang jauh lebih besar dibandingkan nilai r-tabel (r-hitung > 0,2347). Nilai 0,2347 tersebut merupakan standar dalam perhitungan uji validitas. Oleh karena itu, seluruh instrumen kuesioner dianggap valid dan memenuhi syarat formal untuk digunakan sebagai data primer dalam pemodelan klasifikasi menggunakan algoritma C4.5 pada tahap penambangan data (*data mining*).

Uji Reliabilitas



Setelah semua pertanyaan dalam instrumen dianggap valid yang ditunjukkan pada Gambar hasil uji reliabilitas, langkah berikutnya adalah melakukan uji reliabilitas untuk mengetahui seberapa terpercaya dan konsisten kuesioner kualitas pelayanan

(SERVQUAL) di Kantor Kecamatan Cikarang Utara. Pengujian ini menggunakan metode *Cronbach's Alpha* yang dihitung dengan bantuan perangkat lunak *Microsoft Excel*.

Berdasarkan perhitungan statistik, proses dijalankan menggunakan rumus array yang menjumlahkan varians dari setiap pertanyaan kemudian dibagi dengan varians skor total per dimensi.

Hasil perhitungan koefisien *Cronbach's Alpha* untuk setiap dimensi yang diuji pada 98 data jawaban disajikan dalam tabel pembahasan, dengan penjelasan analisis sebagai berikut:

1. Dimensi Keandalan (*Reliability*): Menghasilkan nilai *Cronbach's Alpha* sebesar 0,7896. Nilai ini jauh di atas batas standar keandalan (0,7), sehingga kuesioner untuk dimensi ini dianggap Sangat Reliabel.
2. Dimensi Daya Tanggap (*Responsiveness*): Menghasilkan koefisien sebesar 0,8684. Karena nilainya lebih besar dari 0,7, maka alat ukur untuk mengukur dimensi responsif dianggap sangat dapat dipercaya.
3. Dimensi Jaminan (*Assurance*): Menghasilkan koefisien sebesar 0,7394. Angka ini melebihi batas minimal 0,7, sehingga alat pengukurannya terbukti sangat dapat diandalkan.
4. Dimensi Empati (*Empathy*): Menghasilkan koefisien sebesar 0,8668. Alat ukur yang digunakan untuk mengukur dimensi empati - yang kemudian dikenal sebagai faktor yang paling signifikan - telah terbukti sangat konsisten secara statistik dan dianggap sangat dapat dipercaya.
5. Dimensi bukti fisik (*tangibles*) menghasilkan nilai koefisien tertinggi dibandingkan dimensi lainnya, yaitu sebesar 0,8635. Dengan nilai mendekati angka 1, instrumen fisik ini dianggap sangat terpercaya dan memiliki konsistensi dalam pengukuran yang sangat baik.

Berdasarkan hasil analisis tersebut, terlihat bahwa semua aspek dari variabel SERVQUAL memiliki nilai *Cronbach's Alpha* yang cukup tinggi, yaitu di atas 0,7. Menurut Uyanto (2009), skala pengukuran yang reliabel sebaiknya memiliki *cronbach alpha* $\geq 0,7$. Oleh karena itu, kuesioner penelitian ini terbukti cukup baik, konsisten, dan tidak terpengaruh oleh bias pengukuran yang berulang, sehingga data primer yang berhasil dikumpulkan sangat memadai dan dapat dipercaya untuk dianalisis lebih lanjut dalam tahapan data mining dengan menggunakan algoritma C4.5.

Hasil Transformasi Data

Tahap transformasi data dilakukan untuk mengubah dan menyederhanakan struktur data mentah menjadi format yang lebih seragam, sehingga data tersebut dapat siap untuk diproses serta dianalisis menggunakan algoritma C4.5 *Decision tree*.



Attribute	Value	Attribute	Value	Attribute	Value
Keandalan	1	Keandalan	2	Keandalan	3
Keandalan	4	Keandalan	5	Keandalan	6
Keandalan	7	Keandalan	8	Keandalan	9
Keandalan	10	Keandalan	11	Keandalan	12
Keandalan	13	Keandalan	14	Keandalan	15
Keandalan	16	Keandalan	17	Keandalan	18
Keandalan	19	Keandalan	20	Keandalan	21
Keandalan	22	Keandalan	23	Keandalan	24
Keandalan	25	Keandalan	26	Keandalan	27
Keandalan	28	Keandalan	29	Keandalan	30
Keandalan	31	Keandalan	32	Keandalan	33
Keandalan	34	Keandalan	35	Keandalan	36
Keandalan	37	Keandalan	38	Keandalan	39
Keandalan	40	Keandalan	41	Keandalan	42
Keandalan	43	Keandalan	44	Keandalan	45
Keandalan	46	Keandalan	47	Keandalan	48
Keandalan	49	Keandalan	50	Keandalan	51
Keandalan	52	Keandalan	53	Keandalan	54
Keandalan	55	Keandalan	56	Keandalan	57
Keandalan	58	Keandalan	59	Keandalan	60
Keandalan	61	Keandalan	62	Keandalan	63
Keandalan	64	Keandalan	65	Keandalan	66
Keandalan	67	Keandalan	68	Keandalan	69
Keandalan	70	Keandalan	71	Keandalan	72
Keandalan	73	Keandalan	74	Keandalan	75
Keandalan	76	Keandalan	77	Keandalan	78
Keandalan	79	Keandalan	80	Keandalan	81
Keandalan	82	Keandalan	83	Keandalan	84
Keandalan	85	Keandalan	86	Keandalan	87
Keandalan	88	Keandalan	89	Keandalan	90
Keandalan	91	Keandalan	92	Keandalan	93
Keandalan	94	Keandalan	95	Keandalan	96
Keandalan	97	Keandalan	98	Keandalan	99

Pada Gambar ini Pada tahap sebelum transformasi data pertanyaan yang berisi masing - masing dari variabel yaitu keandalan (*reliability*), daya tanggap (*responsiveness*), jaminan (*assurance*), empati (*empathy*), dan bukti fisik (*tangibles*), dengan soal yang masih berbentuk berupa pertanyaan panjang menunjukkan format data sebelum proses transformasi. Dalam tahap ini, judul kolom atribut masih menggunakan kalimat pertanyaan yang panjang dari kuesioner yang mencerminkan lima dimensi kualitas pelayanan, yaitu keandalan (*reliability*), daya tanggap (*responsiveness*), jaminan (*assurance*), empati (*empathy*), dan bukti fisik (*tangibles*). Pertanyaan yang terlalu panjang ini dapat menyulitkan pembacaan

pola saat algoritma C4.5 melakukan pemrosesan.



Pada tahap setelah transformasi pertanyaan yang sebelumnya masih berupa pertanyaan panjang diubah kedalam bentuk kode atribut variabel yang dimasukkan untuk dijadikan pertanyaan, karena itu penyederhanaan telah dilakukan sebagaimana ditunjukkan Pada Gambar ini. Dalam data yang telah ditransformasi, seluruh kalimat pertanyaan panjang dari kuesioner telah diubah dan disingkat menjadi kode atribut variabel yang lebih ringkas, padat, dan terorganisir.

Hasil Penentuan Atribut dan Label

No	Item	Skala	Dimensi
1	Tampilan	5	Visual
2	Rasa	5	Alat
3	Ukuran	5	Alat
4	Warna	5	Alat
5	Kecepatan	5	Alat
6	Kecepatan	5	Alat
7	Kecepatan	5	Alat
8	Kecepatan	5	Alat
9	Kecepatan	5	Alat
10	Kecepatan	5	Alat
11	Kecepatan	5	Alat
12	Kecepatan	5	Alat
13	Kecepatan	5	Alat
14	Kecepatan	5	Alat
15	Kecepatan	5	Alat
16	Kecepatan	5	Alat
17	Kecepatan	5	Alat
18	Kecepatan	5	Alat
19	Kecepatan	5	Alat
20	Kecepatan	5	Alat
21	Kecepatan	5	Alat
22	Kecepatan	5	Alat
23	Kecepatan	5	Alat
24	Kecepatan	5	Alat

Setelah tahap pembersihan dan transformasi data selesai, kemudian ditentukan atribut dan label yang digunakan dalam proses klasifikasi dengan algoritma C4.5. Tahap ini bertujuan untuk menentukan variabel-variabel yang digunakan sebagai data masukan serta variabel yang akan menjadi acuan dalam proses klasifikasi.

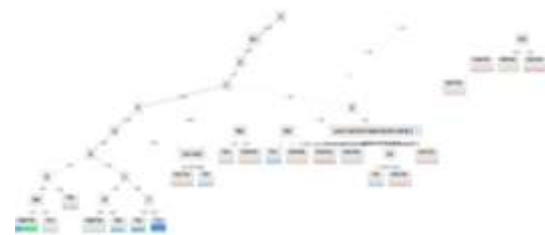
Berdasarkan Tabel hasil penentuan atribut dan label, atribut penelitian mencakup atribut karakteristik responden, atribut pelayanan yang diukur melalui lima dimensi SERVQUAL, serta atribut kepuasan keseluruhan (FF). Atribut pelayanan terdiri dari R1 hingga R3 yang mengacu pada Reliability, RES1 hingga RES3 yang berkaitan dengan Responsiveness, A1 hingga A3 yang

menggambarkan Assurance, E1 hingga E3 yang menunjukkan Empathy, dan T1 hingga T3 yang mengacu pada Tangibles. Sementara itu, indeks ditetapkan sebagai label atau target klasifikasi yang digunakan untuk mengetahui tingkat kepuasan masyarakat.

Penentuan atribut dan label itu merupakan langkah persiapan sebelum data digunakan dalam proses pemodelan C4.5. Selanjutnya, atribut-atribut yang tidak penting dalam proses pengklasifikasian dibuang agar mendapatkan dataset yang sudah siap digunakan untuk membuat pohon keputusan.

RESULT

Dari hasil pengolahan data pada *rapidminer* yang sudah dilakukan di penelitian ini menghasilkan pohon keputusan atau *decision tree* berikut adalah detailnya:



Pada Gambar ini menampilkan bentuk pohon keputusan dari hasil pengolahan algoritma C4.5 dengan operator *decision tree*. Hasil tersebut menampilkan tingkat kepuasan dari masing - masing variabel skala *likert* yang berupa singkatan kode agar mempermudah algoritma C4.5 dalam mengolah datanya. Sistem secara otomatis menghasilkan *decision tree* yang menunjukkan tingkat pengolahan data matematis mesin terhadap setiap attribute. Keluaran *decision tree* yang menjadikan E1 sebagai *root node* merupakan tolak ukur dalam melakukan penerapan model algoritma C4.5 yang sudah dibuat pada fase *training* ke sekumpulan data baru yang belum pernah dilakukan apapun dengan operator *apply model* serta akan dikalkulasi oleh operator *performance* untuk menghasilkan *confusion matrix* nilai akurasi dan nilai klasifikasi error.



Pada Gambar Ini Adalah Deskripsi yang dihasilkan oleh *Decision Tree* Yang di olah Menggunakan *RapidMiner*.

Hasil Perhitungan Entropy

Entropy digunakan untuk mengukur tingkat ketidakpastian atau variasi kelas dalam sebuah dataset. Nilai entropy awal dihitung dengan melihat cara kelompok kepuasan masyarakat terdistribusi pada 98 data responden. Distribusi kelas pada dataset adalah sebagai berikut.

Kelas Kepuasan	Jumlah
Sangat Puas	19
Puas	61
Cukup Puas	16
Tidak Puas	2
Sangat Tidak Puas	0
Total	98

Berdasarkan distribusi tersebut, diperoleh nilai entropy keseluruhan sebesar:

$$\text{Entropy}(S) = 1,426081269$$

Perhitungan entropy dilakukan dengan menggunakan persamaan:

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

dengan p_i menunjukkan perbandingan jumlah anggota dari setiap kelas dalam dataset. Setelah mendapatkan nilai entropy awal, selanjutnya dihitung nilai entropy untuk setiap atribut berdasarkan pengelompokan nilai menjadi dua bagian, yaitu kurang dari atau sama dengan 2,5 dan lebih dari 2,5. Hasil perhitungannya ditunjukkan pada tabel berikut.

No.	Atribut	Entropy ≤ 2,5	Entropy > 2,5
1	R1	1,459147917	1,350819793
2	R2	1,530493057	1,267946122
3	R3	1,584962501	1,372724899
4	RES1	1,251629167	1,326348820
5	RES2	0,000000000	1,348638854
6	RES3	1,370950594	1,343139727
7	A1	0,970950594	1,263983363
8	A2	1,448815636	1,358572258
9	A3	0,918295834	1,351028564
10	E1	1,000000000	1,280242370
11	E2	1,251629167	1,326348820
12	E3	1,500000000	1,358556645
13	T1	0,970950594	1,263983363
14	T2	0,811278124	1,335532291
15	T3	0,918295834	1,351028564

Berdasarkan Tabel 3.X, nilai entropy pada setiap atribut berbeda tergantung pada distribusi kelas di setiap cabangnya. Nilai entropy tersebut kemudian digunakan untuk mendapatkan nilai Information Gain di tahap berikutnya.

Hasil Perhitungan Information Gain

Setelah nilai entropy dari setiap cabang sudah diketahui, langkah berikutnya adalah menghitung Information Gain. Information Gain digunakan untuk mengetahui seberapa besar ketidakpastian berkurang setelah dataset dibagi berdasarkan suatu atribut. Persamaan Information Gain yang digunakan adalah:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v=1}^n \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

Dengan:

- S = dataset awal,
- A = atribut,
- S_v = subset data berdasarkan atribut,
- $|S|$ = jumlah seluruh data,
- $|S_v|$ = jumlah data pada subset.

Hasil perhitungan *Information Gain* dari masing-masing atribut adalah sebagai berikut.

No.	Atribut	Information Gain
1	R1	0,068629142
2	R2	0,134023694
3	R3	0,046859301
4	RES1	0,104307122
5	RES2	0,083785697
6	RES3	0,081522620
7	A1	0,177048558
8	A2	0,061063056
9	A3	0,088299626
10	E1	0,157277363
11	E2	0,104307122
12	E3	0,061751426
13	T1	0,177048558
14	T2	0,111947107
15	T3	0,088299626

Berdasarkan tabel tersebut, nilai Information Gain yang paling besar dalam perhitungan yang digunakan terdapat pada atribut A1 dan T1, yaitu sebesar 0,177048558, sedangkan atribut E1 memiliki nilai Information Gain sebesar 0,157277363.

Namun, nilai Information Gain tidak langsung digunakan untuk menentukan node akar dalam penelitian ini, karena proses C4.5 berikutnya menggunakan Split Information dan Gain Ratio. Oleh karena itu, hasil Information Gain digunakan sebagai salah satu bagian dalam menghitung Gain Ratio.

Hasil Perhitungan Split Information

Setelah mendapatkan Information Gain, langkah selanjutnya adalah menghitung Split Information. Split Information digunakan untuk mengukur informasi yang dihasilkan dari membagi dataset berdasarkan suatu atribut.

Persamaan Split Information yang digunakan adalah:

$$SplitInfo(S, A) = - \sum_{v=1}^n \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

Hasil perhitungan *Split Information* dari setiap atribut adalah sebagai berikut.

No.	Atribut	Split Information
1	R1	0,332286630
2	R2	0,442571065
3	R3	0,197453004
4	RES1	0,332286630
5	RES2	0,197453004
6	RES3	0,290715865
7	A1	0,290715865
8	A2	0,371232327
9	A3	0,197453004
10	E1	0,246022578
11	E2	0,332286630
12	E3	0,246022578
13	T1	0,290715865
14	T2	0,246022578
15	T3	0,197453004

Nilai Split Information diambil dari bagaimana data terdistribusi pada setiap pembagian atribut, seperti yang terlihat pada Tabel hasil split information. Nilai tersebut kemudian digunakan sebagai pembagi dalam menghitung Gain Ratio. Sehingga, tahap Split Information menjadi bagian yang menghubungkan hasil Information Gain dengan pemilihan atribut terbaik berdasarkan Gain Ratio.

Hasil Perhitungan Gain Ratio

Tahap selanjutnya adalah menghitung Gain Ratio. Dalam penelitian ini, Gain Ratio digunakan untuk menentukan atribut yang paling efektif dalam membagi data saat membangun pohon keputusan.

Persamaan yang digunakan adalah:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

Hasil perhitungan *Gain Ratio* dari seluruh atribut adalah sebagai berikut.

No.	Atribut	Information Gain	Split Information	Gain Ratio
1	R1	0,068629142	0,332286630	0,206535972
2	R2	0,134023694	0,442571065	0,302829770
3	R3	0,046859301	0,197453004	0,237318750
4	RES1	0,104307122	0,332286630	0,313907067
5	RES2	0,083785697	0,197453004	0,424323246
6	RES3	0,081522620	0,290715865	0,280402266
7	A1	0,177048558	0,290715865	0,609008928
8	A2	0,061063056	0,371232327	0,164488742
9	A3	0,088299626	0,197453004	0,447193124
10	E1	0,157277363	0,246022578	0,639280200
11	E2	0,104307122	0,332286630	0,313907067
12	E3	0,061751426	0,246022578	0,250969102
13	T1	0,177048558	0,290715865	0,609008928
14	T2	0,111947107	0,246022578	0,455027778
15	T3	0,088299626	0,197453004	0,447193124

Berdasarkan Tabel hasil perhitungan gain ratio, ditemukan nilai Gain Ratio yang paling tinggi pada atribut E1, yaitu mencapai 0,639280200. Nilai tersebut lebih

besar dibandingkan atribut lainnya, misalnya A1 dan T1, yang masing-masing memiliki nilai Gain Ratio sebesar 0,609008928.

Oleh karena itu, E1 dipilih sebagai node akar pada awal pembentukan pohon keputusan menggunakan algoritma C4.5. Atribut E1 adalah indikator dalam dimensi Empati, sehingga hasil tersebut menunjukkan bahwa pada tahap awal klasifikasi, indikator E1 memiliki nilai Gain Ratio tertinggi dalam membagi data berdasarkan tingkat kepuasan masyarakat. Misalnya, perhitungan untuk atribut E1 dilakukan dengan memisahkan data menjadi dua bagian, yaitu E1 kurang dari atau sama dengan 2,5 dan E1 lebih besar dari 2,5.

Misalnya, perhitungan untuk atribut E1 dilakukan dengan memisahkan data menjadi dua bagian, yaitu E1 kurang dari atau sama dengan 2,5 dan E1 lebih besar dari 2,5. Di cabang E1 kurang dari atau sama dengan 2,5 terdapat 4 data dengan nilai entropy 1,000000000, sedangkan di cabang E1 lebih besar dari 2,5 terdapat 94 data dengan nilai entropy 1,280242370.

Nilai *Information Gain* E1 dihitung dengan cara sebagai berikut:

$$Gain(S, E1) = 1,420081269 - \left[\frac{4}{98} (1,000000000) + \frac{94}{98} (1,280242370) \right]$$

Sehingga diperoleh:

$$Gain(S, E1) = 0,157277363$$

Selanjutnya diperoleh Split Information:

$$SplitInfo(S, E1) = 0,246022578$$

Maka Gain Ratio E1 adalah:

$$GainRatio(S, E1) = \frac{0,157277363}{0,246022578}$$

$$GainRatio(S, E1) = 0,639289200$$

Karena nilai tersebut merupakan *gain ratio* tertinggi, maka E1 ditetapkan sebagai node akar / *root node*.

Hasil Performance Vector



Pada Gambar ini Berdasarkan tingkat akurasi yang dihasilkan dari *confusion matrix* yang sebesar 66,33% dan klasifikasi error sebesar 33,67% menunjukkan bahwa presisi / ketepatan prediksi yang mengukur seberapa presisi atau tepat prediksi yang dibuat oleh mesin pengolahan data menggunakan algoritma C4.5 dengan hasil *decision tree* cukup tinggi dikarenakan tingkat akurasi yang sudah mencukupi untuk mengukur seberapa tanggap mesin dalam menemukan seluruh data yang relevan juga sudah sesuai dengan data primer bisa dilihat dari data yang sudah kita olah bahwa kategori skala *likert* yang dominan adalah 3, 4, dan 5.

CONCLUSSION

Berdasarkan hasil penelitian dan pembahasan mengenai kepuasan masyarakat terhadap pelayanan administratif di kantor kecamatan Cikarang Utara menggunakan algoritma C4.5, penulis dapat mengambil kesimpulan yaitu: Penerapan algoritma C4.5 dalam penelitian ini berhasil membangun model klasifikasi tingkat kepuasan masyarakat terhadap pelayanan administratif di Kantor Kecamatan Cikarang Utara. Dalam penelitian ini, sebanyak 105 data dari hasil kuesioner diproses dan dibersihkan, sehingga menghasilkan 98 data valid yang

memenuhi kriteria. Data tersebut kemudian digunakan untuk membuat model klasifikasi dengan algoritma C4.5 yang menargetkan indeks pertanyaan final sebagai label kelasnya.

Selanjutnya, berdasarkan hasil pemrosesan data yang membentuk *decision tree*, atribut E1 (dimensi *Empathy* pada pertanyaan 1) terbukti menempati urutan pertama sebagai *root node*. Oleh karena itu, dapat disimpulkan bahwa faktor empati, yakni ketika pegawai memberikan perhatian dan pendampingan kepada warga, merupakan faktor yang paling dominan dalam menentukan tingkat kepuasan masyarakat. Sebagai pendukung, hasil evaluasi performa model menggunakan pengujian *Accuracy* juga menunjukkan tingkat akurasi yang tergolong baik, yaitu dengan persentase sebesar 66,32% serta tingkat *classification error* sebesar 33,68%. Pada akhirnya, pencapaian ini membuktikan bahwa model *decision tree* yang dihasilkan cukup kuat dan layak digunakan sebagai alat ukur tingkat kepuasan masyarakat terhadap pelayanan administratif di Kantor Kecamatan Cikarang Utara.

REFERENCES

Jumawan, & Prasetyo, A. I. (2024). Analisis Pengaruh Kualitas Pelayanan, Harga, dan Fasilitas Terhadap Kepuasan Konsumen di Wuling Arista Bekasi. *Jurnal Penelitian Ilmu-Ilmu Sosial*, 2(1), 129–143. <https://doi.org/10.30871/jaba.v2i1.703>

Lasarudin, A., Hasyim, W., & Dumako, R. (2022). *Analisis Kepuasan Masyarakat Terhadap Layanan Siaran RRI Gorontalo Menggunakan Algoritma C4.5*. 4(4), 483–490.

Mukminna, H., Kurniasari, I., & Legowo, G. (2025). IMPLEMENTASI DATA MINING ALGORITMA C4.5 UNTUK KLASIFIKASI TINGKAT KEPUASAN MAHASISWA TERHADAP LAYANAN SISTEM INFORMASI. *JURNAL*

INFORMATIKA TEKNOLOGI DAN SAINS (JINTEKS), 7(4), 1837–1847.

- Ningsih, L. N., Septiani, R., Pramadjaya, A., & Nuralisah, S. (2025). Implementing the C4 . 5 Algorithm for Customer Satisfaction Classification Implementasi Algoritma C4 . 5 untuk Klasifikasi Kepuasan Pelanggan. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(October), 1470–1480.
- Nisrina, D., & Kustiyono. (2024). Analisis Kepuasan Konsumen Menggunakan Metode Algoritma C4.5 Berbasis RapidMiner pada PT. Adeaksa Indo Jayatama. *MEANS (Media Informasi Analisa dan Sistem)*, 9(1).
- Nugroho, C. R. A., & Kristiana, T. (2022). Penerapan Algoritma C4.5 Untuk Kepuasan Pelanggan Toko Online Parfume Chantik. *Jurnal Algoritme*, 3(1), 10–21.
- Permatasari, V. N., Aula, R. F., Akbar, Y., & Hidayat, A. Z. (2024). Analisis Tingkat Kepuasan Pelanggan Terhadap Pengguna Jasa Layanan Grab Menggunakan Metode C4 . 5. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 5(3), 3189–3198.
- Pradana, M. A., Putra, R. S., Misbachuddin, M., & Puspaningrum, E. Y. (2024). Penerapan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung. *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, 5(3), 1–10. <https://doi.org/10.33005/jifosi.v5i3.471>
- Syahroni, M. I. (2022). *PROSEDUR PENELITIAN KUANTITATIF*. 2(3), 43–56.
- Tan, N. F. W., & Kusnawi. (2022). Klasifikasi Menggunakan Algoritma C4.5 dalam Memprediksi Status

Keorganisasian Mahasiswa
Universitas Amikom Yogyakarta.
Jurnal Sistem Informasi, 11(3), 696–
712. <http://sistemasi.ftik.unisi.ac.id>

Xsanal Hakim, R., Putrawansyah, F., &
Syahri, R. (2024). Penerapan
Algoritma C4.5 Untuk Prediksi Anak
Stunting Di Kota Pagar Alam. *JATI*
(*Jurnal Mahasiswa Teknik*
Informatika), 8(2), 2469–2478.
<https://doi.org/10.36040/jati.v8i2.930>

1