

Analisis Kinerja Decision Tree dan Naïve Bayes Pada Klasifikasi Tingkat Kepuasan Masyarakat

Nora Fitaria, Dedi Darwis*

Fakultas Teknik & Ilmu Komputer, Program Studi Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

Email: ¹nora.fitaria@gmail.com, ^{2,*}darwisdedi@teknokrat.ac.id

Email Penulis Korespondensi: darwisdedi@teknokrat.ac.id

Submitted: 31/01/2026; Accepted: 05/03/2026; Published: 06/03/2026

Abstrak—Survei Kepuasan Masyarakat (SKM) merupakan instrumen resmi yang digunakan pemerintah untuk mengevaluasi kinerja pelayanan publik sebagaimana diatur dalam Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (PermenPANRB) Nomor 14 Tahun 2017. Namun, pemanfaatan data SKM di banyak instansi pemerintah masih terbatas pada perhitungan nilai indeks kepuasan tanpa analisis lanjutan yang bersifat prediktif. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kepuasan masyarakat pengguna layanan Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPSTP) Kota Metro menggunakan algoritma Decision Tree dan Naïve Bayes. Data yang digunakan merupakan data SKM tahun 2025 hingga triwulan IV sebanyak 2.760 responden, yang terdiri dari sembilan unsur pelayanan (U1–U9) dengan kategori kepuasan sebagai variabel kelas. Proses penelitian meliputi tahap pra-pemrosesan data, pemodelan klasifikasi menggunakan RapidMiner, serta evaluasi model berdasarkan confusion matrix, accuracy, precision, dan recall. Hasil penelitian menunjukkan bahwa pada dataset dengan jumlah besar, algoritma Naïve Bayes menghasilkan tingkat akurasi sebesar 95,04%, lebih tinggi dibandingkan Decision Tree yang memperoleh akurasi sebesar 84,46% serta memiliki nilai recall yang lebih baik pada kelas dominan (recall kelas Puas sebesar 98.16%). Keunggulan tersebut menunjukkan efisiensi pendekatan probabilistik Naïve Bayes dalam menangani fitur kategorikal pada unsur pelayanan publik. Penelitian ini membuktikan bahwa penerapan teknik data mining pada data SKM dapat mendukung evaluasi pelayanan publik berbasis data.

Kata Kunci : Data Mining; Decision Tree; Naïve Bayes; Pelayanan Publik; Survei Kepuasan Masyarakat

Abstract—The Public Satisfaction Survey (SKM) is an official instrument used by the government to evaluate public service performance as stipulated in Regulation of the Minister of State Apparatus Empowerment and Bureaucratic Reform (PermenPANRB) Number 14 of 2017. However, the use of SKM data in many government agencies is still limited to calculating satisfaction index values without further predictive analysis. This study aims to classify the level of satisfaction of service users of the Metro City Investment and Integrated Services Agency (DPMPSTP) using the Decision Tree and Naïve Bayes algorithms. The data used is SKM data from 2025 to the fourth quarter, consisting of 2,760 respondents, which consists of nine service elements (U1–U9) with satisfaction categories as class variables. The research process includes data pre-processing, classification modeling using RapidMiner, and model evaluation based on confusion matrix, accuracy, precision, and recall. The results showed that the Naïve Bayes algorithm produced an accuracy rate of 95.04%, higher than the Decision Tree, which obtained an accuracy of 84.46%, and had a better recall value in the dominant class (recall of the Satisfied class was 98.16%). These advantages demonstrate the efficiency of the Naïve Bayes probabilistic approach in handling categorical features in public service elements. This study proves that the application of Data mining techniques to SKM data can support data-based public service evaluation.

Keywords: Data Mining; Decision Tree; Naïve Bayes; Public Service; Community Satisfaction Survey

1. PENDAHULUAN

Pelaksanaan Survei Kepuasan Masyarakat (SKM) merupakan instrumen evaluasi yang wajib dilakukan oleh unit penyelenggara pelayanan publik sebagaimana diatur dalam Peraturan Menteri PANRB Nomor 14 Tahun 2017[1]. Namun, dalam praktiknya, hasil SKM pada banyak instansi pemerintah masih terbatas pada pelaporan nilai indeks kepuasan dan penentuan kategori mutu pelayanan, tanpa pemanfaatan lebih lanjut untuk analisis mendalam berbasis data. Kondisi ini menyebabkan data SKM yang berjumlah besar belum optimal digunakan untuk mengidentifikasi pola dan faktor dominan yang memengaruhi tingkat kepuasan masyarakat.

Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPSTP) Kota Metro secara rutin melaksanakan SKM dengan jumlah responden yang relatif besar setiap tahunnya. Berdasarkan laporan internal, nilai SKM hingga triwulan IV tahun 2025 mencapai 93,16 yang tergolong dalam kategori sangat baik[2]. Meskipun menunjukkan capaian kinerja pelayanan yang tinggi, nilai tersebut belum memberikan informasi analitis mengenai bagaimana unsur-unsur pelayanan membentuk tingkat kepuasan masyarakat serta sejauh mana data tersebut dapat digunakan untuk membangun model klasifikasi yang bersifat prediktif[3]. Selain itu, tingginya nilai indeks mengindikasikan dominasi jawaban pada kategori “Sangat Puas”, yang berpotensi menimbulkan ketidakseimbangan distribusi kelas (*class imbalance*) dalam proses klasifikasi. Kondisi ini dapat menyebabkan model lebih mudah memprediksi kelas mayoritas dan kurang sensitif terhadap kelas minoritas seperti “Kurang Puas”, sehingga diperlukan pendekatan klasifikasi yang mampu mengevaluasi performa model secara lebih komprehensif melalui *metrik accuracy, precision, dan recall*[4].

Seiring dengan meningkatnya ketersediaan data SKM, pendekatan *Data mining* menjadi relevan untuk diterapkan dalam analisis kepuasan masyarakat. *Data mining* memungkinkan pengolahan data dalam skala besar untuk menemukan pola tersembunyi serta membangun model klasifikasi yang dapat memprediksi tingkat kepuasan

berdasarkan unsur-unsur pelayanan. Pendekatan ini dinilai lebih informatif dibandingkan analisis deskriptif semata karena mampu memberikan dukungan analitis bagi peningkatan kualitas pelayanan publik secara berkelanjutan.

Dalam konteks klasifikasi data kepuasan masyarakat, algoritma *Decision Tree* dan *Naïve Bayes* merupakan dua metode yang banyak digunakan dalam penelitian sebelumnya. *Decision Tree* membangun model klasifikasi dalam bentuk struktur pohon keputusan yang mudah diinterpretasikan, sehingga aturan keputusan yang dihasilkan dapat dipahami oleh pengambil kebijakan. Keunggulan ini menjadikan *Decision Tree* sering digunakan untuk mengidentifikasi faktor dominan yang memengaruhi tingkat kepuasan masyarakat pada layanan publik. Namun, performa *Decision Tree* sangat dipengaruhi oleh distribusi data dan ketidakseimbangan kelas, yang dapat menyebabkan penurunan akurasi pada dataset dengan karakteristik yang kompleks [5].

Sementara itu, *Naïve Bayes* merupakan algoritma klasifikasi berbasis probabilistik yang bekerja dengan asumsi independensi antar atribut. Algoritma ini dikenal memiliki kinerja yang stabil pada data kategorikal dan dataset berukuran besar, serta relatif tahan terhadap noise dan variasi data [6]. Beberapa penelitian menunjukkan bahwa *Naïve Bayes* mampu menghasilkan nilai *akurasi* dan *precision* yang lebih konsisten dibandingkan *Decision Tree*, khususnya pada data survei kepuasan masyarakat yang memiliki distribusi jawaban tidak seimbang [7].

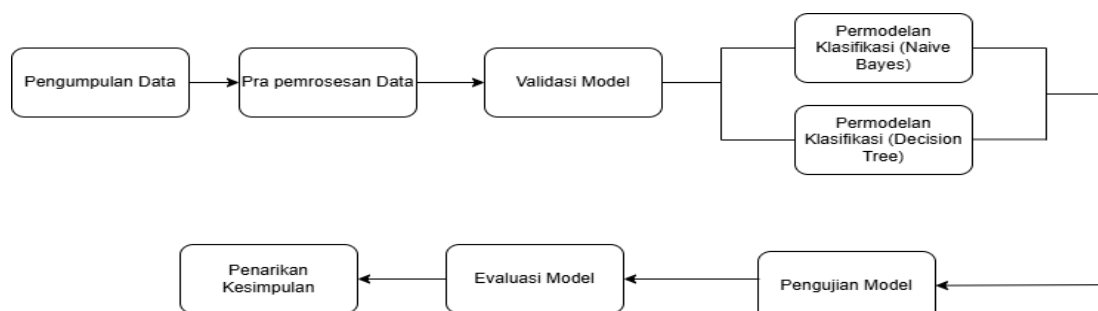
Penelitian-penelitian terdahulu umumnya menggunakan jumlah data yang relatif terbatas yaitu data yang digunakan oleh penelitian sebelumnya kurang dari 1000 data dan distribusi kelas yang lebih seimbang, sehingga menghasilkan nilai akurasi yang tinggi. Sebagai contoh, Khoiriyah et al. (2024) menggunakan 195 data responden dalam analisis kepuasan publik di RUPBASAN Blitar dan memperoleh akurasi rata-rata sebesar 99,31% untuk algoritma C4.5 serta 96,78% untuk *Naïve Bayes* [8]. Demikian pula penelitian Dafwen (2023) dan Sari et al. (2023) yang menggunakan dataset dalam skala ratusan responden dengan distribusi kelas yang relatif merata, sehingga menghasilkan akurasi di atas 90% [6]. Meskipun menunjukkan performa yang sangat baik, penggunaan dataset berukuran kecil dengan distribusi kelas yang seimbang cenderung belum sepenuhnya merepresentasikan kompleksitas data kepuasan masyarakat pada kondisi nyata yang memiliki jumlah responden besar dan kecenderungan distribusi kelas yang tidak merata. Sebaliknya, data SKM DPMPTSP Kota Metro memiliki jumlah responden yang besar yaitu 2.760 data dengan distribusi kelas yang tidak seimbang, di mana kategori "Sangat Puas" mendominasi secara signifikan. Kondisi extreme *class imbalance* ini dapat menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas dan mengabaikan sensitivitas terhadap kelas minoritas. Oleh karena itu, pengujian pada dataset yang lebih besar dan tidak seimbang menjadi penting untuk mengevaluasi robustnes algoritma dalam konteks implementasi SKM yang lebih realistis, sehingga menimbulkan tantangan tersendiri dalam proses klasifikasi [9]. Kondisi ini membuka peluang penelitian untuk mengkaji kinerja algoritma klasifikasi pada data SKM yang lebih realistis dan kompleks [8].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun dan membandingkan model klasifikasi tingkat kepuasan masyarakat pengguna layanan DPMPTSP Kota Metro menggunakan algoritma *Decision Tree* dan *Naïve Bayes*. Penelitian ini menggunakan data riil Survei Kepuasan Masyarakat (SKM) DPMPTSP Kota Metro yang memiliki karakteristik distribusi kelas sangat tidak seimbang. Selain itu, evaluasi model dilakukan menggunakan 10-Fold *Cross Validation* dengan *stratified sampling* untuk memperoleh hasil yang lebih representatif terhadap kondisi dataset. Penelitian ini tidak menerapkan teknik resampling seperti SMOTE, sehingga hasil yang diperoleh mencerminkan performa algoritma pada kondisi distribusi data asli. Penelitian ini diharapkan dapat memberikan gambaran performa kedua algoritma pada data SKM yang mengacu pada PermenPANRB Nomor 14 Tahun 2017 serta menjadi dasar pemanfaatan data SKM secara lebih analitis dalam peningkatan kualitas pelayanan publik.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan dataset survei kepuasan masyarakat yang dihimpun melalui pemindaian barcode sebagai sumber data utama. Tahapan penelitian meliputi pengumpulan dan pra-pemrosesan data, validasi model, serta pemodelan klasifikasi menggunakan algoritma *Decision Tree* dan *Naïve Bayes*. Kedua algoritma diuji untuk mengukur kemampuan prediksi dan dievaluasi guna menentukan tingkat akurasi terbaik sebelum penarikan kesimpulan. Alur penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan alur tahapan penelitian yang dilakukan dalam proses klasifikasi tingkat kepuasan masyarakat, mulai dari pengumpulan data hingga evaluasi model menggunakan algoritma *Decision Tree* dan *Naive Bayes*. Setiap tahapan pada gambar tersebut dijelaskan secara rinci pada bagian berikut:

2.1.1 Pengumpulan Data

Tahap pengumpulan data merupakan tahap awal dalam penelitian ini. Data yang digunakan adalah data Survei Kepuasan Masyarakat (SKM) yang diperoleh dari Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kota Metro selama periode tahun 2025. Data SKM dikumpulkan dari masyarakat pengguna layanan perizinan dan nonperizinan yang telah menerima pelayanan secara langsung yang menjadi responden dalam pengisian survey kepuasan masyarakat tersebut.

Instrumen SKM yang digunakan mengacu pada Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi (PermenPANRB) Nomor 14 Tahun 2017, yang menetapkan sembilan unsur pelayanan sebagai indikator pengukuran kepuasan masyarakat. Kesembilan unsur tersebut terdiri dari persyaratan (U1), sistem, mekanisme, dan prosedur (U2), waktu penyelesaian (U3), biaya atau tarif (U4), produk spesifikasi jenis pelayanan (U5), kompetensi pelaksana (U6), perilaku pelaksana (U7), sarana dan prasarana (U8), serta penanganan pengaduan, saran, dan masukan (U9)[1]. Selain atribut unsur pelayanan, dataset juga memiliki atribut kategori tingkat kepuasan masyarakat yang digunakan sebagai variabel kelas (label) dalam proses klasifikasi.

2.1.2 Pra-Pemrosesan Data

Setelah data SKM dikumpulkan, tahap selanjutnya adalah pra-pemrosesan data (*data preprocessing*). Tahap ini bertujuan untuk memastikan kualitas data agar dapat digunakan secara optimal dalam proses pemodelan klasifikasi. Pra-pemrosesan data meliputi beberapa aktivitas, yaitu seleksi atribut, penghapusan atribut yang tidak relevan, serta pengecekan dan penanganan data kosong (*missing value*)[8]. Atribut yang tidak berkaitan langsung dengan analisis kepuasan masyarakat, seperti nomor urut responden, dan identitas pribadi responden, dihilangkan dari dataset. Selanjutnya, atribut tanggal pengisian survey, unsur pelayanan U1 sampai U9 dipertahankan sebagai variabel prediktor.

Pada tahap ini juga dilakukan penetapan atribut kategori kepuasan masyarakat sebagai label kelas (*class label*), dimana label kelas tersebut di bagi menjadi 3 kategori yaitu, Kurang Puas, Puas, dan Sangat Puas. Berdasarkan hasil pelabelan, distribusi kelas pada dataset menunjukkan ketidakseimbangan yang cukup signifikan. Dari total 2.760 data responden, sebanyak 1.546 data (55,99%) termasuk kategori Sangat Puas, 1.193 data (43,22%) kategori Puas, dan hanya 21 data (0,76%) kategori Kurang Puas. Proporsi ini mengonfirmasi adanya kondisi *class imbalance* sebagaimana telah dijelaskan pada bagian pendahuluan, di mana kelas mayoritas mendominasi distribusi data. Proses pra-pemrosesan dilakukan menggunakan perangkat lunak untuk memastikan data siap digunakan pada tahap pemodelan *RapidMine*.

2.1.3 Validasi Model (10-Fold Cross Validation)

Data yang telah melalui tahap pra-pemrosesan selanjutnya dipersiapkan untuk proses pembentukan dan evaluasi model klasifikasi. Secara konseptual, Pembagian data dilakukan menggunakan teknik *split data* dengan rasio tertentu, misalnya 80:20 antara data latih dan data uji, untuk menguji kemampuan generalisasi model. Namun demikian, distribusi kelas pada dataset menunjukkan kondisi tidak seimbang (*imbalanced*), sehingga penggunaan split tunggal berpotensi menghasilkan estimasi kinerja yang kurang stabil[10].

Oleh karena itu, evaluasi model dilakukan menggunakan metode *Stratified 10-Fold Cross Validation*. Pada metode ini, dataset dibagi menjadi sepuluh subset (*fold*) dengan proporsi kelas yang tetap terjaga pada setiap fold. Proses pelatihan dan pengujian dilakukan secara bergantian sebanyak 10 iterasi, di mana 9 *fold* digunakan sebagai data latih dan satu *fold* sebagai data uji. Pendekatan ini menghasilkan estimasi kinerja model yang lebih representatif dan reliabel dalam mengukur kemampuan klasifikasi terhadap data yang belum pernah diproses sebelumnya.

2.1.4 Pemodelan Klasifikasi

Tahap pemodelan klasifikasi dilakukan dengan menerapkan dua algoritma *data mining*, yaitu *Decision Tree* dan *Naive Bayes*, untuk mengklasifikasikan tingkat kepuasan masyarakat berdasarkan sembilan unsur pelayanan SKM (U1–U9). Pemilihan kedua algoritma ini didasarkan pada perbedaan karakteristik dan pendekatan klasifikasi yang digunakan, sehingga memungkinkan dilakukan analisis perbandingan kinerja model secara komprehensif.

a. Pemodelan Menggunakan *Decision Tree*

Algoritma *Decision Tree* membangun model klasifikasi dalam bentuk struktur pohon keputusan yang terdiri atas *root node*, *internal node*, dan *leaf node*. Setiap node merepresentasikan atribut unsur pelayanan, sedangkan cabang menggambarkan nilai atau kategori dari atribut tersebut. Proses pembentukan pohon keputusan diawali dengan pemilihan atribut terbaik sebagai *root node* berdasarkan kriteria pemisahan tertentu, seperti *information gain* atau *gain ratio*[11], yang bertujuan untuk memaksimalkan pemisahan kelas kepuasan masyarakat.

Pada penelitian ini, atribut unsur pelayanan U1 hingga U9 dianalisis untuk menentukan atribut yang paling berpengaruh dalam membedakan kategori tingkat kepuasan masyarakat. Atribut dengan kemampuan pemisahan tertinggi akan ditempatkan pada node teratas, kemudian proses pemisahan dilanjutkan secara rekursif hingga

seluruh data terklasifikasi atau memenuhi kondisi penghentian tertentu. Hasil dari proses ini berupa aturan keputusan (*if-then rules*) yang menggambarkan hubungan antara unsur pelayanan dan kategori kepuasan masyarakat[12].

Keunggulan utama *Decision Tree* terletak pada kemampuannya menghasilkan model yang mudah dipahami dan diinterpretasikan oleh pengambil kebijakan. Dalam konteks pelayanan publik, aturan keputusan yang dihasilkan dapat digunakan untuk mengidentifikasi unsur pelayanan yang paling dominan memengaruhi tingkat kepuasan masyarakat. Pada penelitian ini, pemodelan *Decision Tree* dilakukan menggunakan perangkat lunak *RapidMiner* dengan memanfaatkan data latih untuk membangun struktur pohon keputusan.

b. Pemodelan Menggunakan *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan metode klasifikasi berbasis probabilitas yang bekerja berdasarkan *Teorema Bayes* dengan asumsi independensi antar atribut. Dalam konteks penelitian ini, *Naïve Bayes* menghitung probabilitas suatu data SKM termasuk ke dalam kategori tingkat kepuasan tertentu (misalnya Baik, Sangat Baik, atau Kurang Baik) berdasarkan nilai-nilai unsur pelayanan U1–U9 yang dimiliki data tersebut.

Proses klasifikasi dimulai dengan menghitung probabilitas awal (*prior probability*) dari setiap kelas kepuasan berdasarkan proporsi data pada data latih. Selanjutnya, dihitung probabilitas bersyarat (*conditional probability*) dari setiap atribut unsur pelayanan terhadap masing-masing kelas. Nilai probabilitas ini kemudian dikombinasikan untuk memperoleh probabilitas posterior, yaitu peluang suatu data termasuk ke dalam kelas kepuasan tertentu. Kelas dengan nilai probabilitas tertinggi akan ditetapkan sebagai hasil klasifikasi[13]

Meskipun menggunakan asumsi independensi antar atribut, yang dalam praktiknya sering tidak sepenuhnya terpenuhi, *Naïve Bayes* dikenal memiliki performa yang baik pada data kategorikal dan dataset berukuran besar. Algoritma ini juga relatif sederhana dan efisien secara komputasi, sehingga mampu memberikan hasil klasifikasi yang stabil pada data SKM dengan jumlah responden yang besar dan distribusi kelas yang tidak seimbang. Pada penelitian ini, pemodelan *Naïve Bayes* dilakukan menggunakan *RapidMiner* dengan memanfaatkan data latih yang sama seperti pada pemodelan *Decision Tree* untuk menjaga konsistensi perbandingan.

2.1.5 Evaluasi Kinerja Model

Tahap akhir dalam penelitian ini adalah evaluasi kinerja model klasifikasi yang telah dibangun menggunakan algoritma *Decision Tree* dan *Naïve Bayes*. Evaluasi kinerja dilakukan untuk mengetahui sejauh mana model mampu mengklasifikasikan tingkat kepuasan masyarakat secara tepat berdasarkan data Survei Kepuasan Masyarakat (SKM). Metode evaluasi yang digunakan adalah *confusion matrix*, yang merupakan pendekatan standar dalam penilaian performa algoritma klasifikasi.

Confusion matrix menyajikan perbandingan antara kelas aktual dan kelas hasil prediksi yang dihasilkan oleh model. Matriks ini terdiri dari empat komponen utama, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Berdasarkan nilai-nilai tersebut, beberapa metrik evaluasi dapat dihitung untuk mengukur performa model secara kuantitatif [14].

a. Akurasi (*Accuracy*)

Akurasi digunakan untuk mengukur proporsi data yang berhasil diklasifikasikan dengan benar oleh model terhadap keseluruhan data yang diuji[15]. Nilai akurasi dihitung menggunakan persamaan(1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Metrik ini memberikan gambaran umum tentang kemampuan model dalam melakukan klasifikasi, namun pada dataset dengan distribusi kelas yang tidak seimbang, akurasi saja belum cukup merepresentasikan performa model secara menyeluruh.

b. Presisi (*Precision*)

Presisi menunjukkan tingkat ketepatan model dalam memprediksi suatu kelas tertentu, yaitu perbandingan antara jumlah data yang diprediksi benar pada suatu kelas dengan seluruh data yang diprediksi sebagai kelas tersebut[15]. Nilai presisi dihitung dengan persamaan(2) :

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Presisi penting untuk melihat sejauh mana model mampu meminimalkan kesalahan prediksi positif, terutama pada kelas kepuasan tertentu seperti Sangat Baik atau Baik.

c. *Recall* (*Sensitivity*)

Recall mengukur kemampuan model dalam mengenali seluruh data yang benar-benar termasuk dalam suatu kelas. Nilai *recall* dihitung dengan persamaan(3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. *F1-Score*

F1-Score digunakan untuk mengukur keseimbangan antara *precision* dan *recall*, terutama pada dataset dengan distribusi kelas yang tidak seimbang. Metrik ini lebih representatif dibandingkan akurasi dalam mengevaluasi performa model pada kelas minoritas. Nilai *F1-Score* dihitung menggunakan persamaan (4):

$$F1 = \frac{2x(Precision \times Recall)}{Precision+Recall} \tag{4}$$

Keterangan: TP (*True Positive*) merupakan data yang diprediksi benar sesuai kelas aktual, TN (*True Negative*) adalah data yang diprediksi benar sebagai bukan kelas tertentu FP (*False Positive*) menunjukkan data yang diprediksi masuk kelas tertentu tetapi sebenarnya termasuk kelas lain, FN (*False Negative*) adalah data yang seharusnya termasuk kelas tertentu tetapi diprediksi ke kelas lain

Nilai akurasi, *precision*, dan *recall* pada penelitian ini diperoleh secara otomatis melalui proses evaluasi model menggunakan perangkat lunak *RapidMiner* berdasarkan data uji yang digunakan. Hasil perhitungan akurasi, *precision*, dan *recall* dari masing-masing algoritma digunakan untuk membandingkan performa *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan tingkat kepuasan masyarakat pengguna layanan DPMPTSP Kota Metro. Evaluasi berbasis *confusion matrix* memungkinkan analisis performa yang lebih mendalam pada setiap kelas kepuasan, sehingga tidak hanya bergantung pada nilai akurasi secara keseluruhan[16].

Perbandingan ini menjadi dasar dalam menentukan algoritma yang paling sesuai untuk analisis data SKM yang memiliki jumlah data besar serta distribusi kelas yang tidak seimbang, sehingga dapat memberikan hasil klasifikasi yang lebih representatif dan relevan dalam konteks pelayanan publik. Hasil evaluasi kinerja model kemudian digunakan untuk membandingkan performa algoritma *Decision Tree* dan *Naïve Bayes* dalam mengklasifikasikan tingkat kepuasan masyarakat pengguna layanan DPMPTSP Kota Metro. Perbandingan ini menjadi dasar dalam menentukan algoritma yang paling sesuai untuk analisis data SKM dalam konteks pelayanan publik dengan jumlah data besar dan distribusi kelas yang tidak seimbang. Pada implementasinya di *RapidMiner*, *Decision Tree* dibangun menggunakan kriteria pemisahan *Gain Ratio* untuk menentukan atribut terbaik pada setiap node pohon keputusan. Pemilihan *Gain Ratio* bertujuan untuk mengurangi bias terhadap atribut dengan jumlah kategori yang banyak sehingga proses pemisahan kelas menjadi lebih proporsional, sedangkan *Naïve Bayes* diterapkan dengan *Laplace Correction* untuk menghindari permasalahan probabilitas nol (*zero probability*) pada perhitungan probabilitas bersyarat, terutama pada kelas minoritas dengan jumlah data yang sangat sedikit[17]. Evaluasi dilakukan menggunakan *10-Fold Cross Validation* dengan *stratified sampling* untuk memperoleh hasil yang lebih stabil dan representatif dalam konteks pelayanan publik.

3. HASIL DAN PEMBAHASAN

Pada bab ini menyajikan hasil penelitian yang diperoleh berdasarkan tahapan penelitian yang telah dilakukan, dimulai dari hasil pengumpulan data Survei Kepuasan Masyarakat (SKM), pra-pemrosesan data, pembagian data, pemodelan klasifikasi menggunakan algoritma *Decision Tree* dan *Naïve Bayes*, hingga evaluasi dan perbandingan kinerja model. Pembahasan dilakukan dengan mengaitkan hasil penelitian dengan tujuan penelitian serta ketentuan PermenPANRB Nomor 14 Tahun 2017.

3.1 Hasil

3.1.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data Survei Kepuasan Masyarakat (SKM) DPMPTSP Kota Metro yang dikumpulkan selama tahun 2025. Proses pengumpulan data dilakukan terhadap masyarakat yang menjadi responden langsung dimana masyarakat tersebut merupakan pengguna layanan perizinan dan nonperizinan yang telah menerima pelayanan dari DPMPTSP Kota Metro, responden tersebut mengisi survey pada *website* survey dengan melakukan *scan barcode* yang bisa di akses.

Jumlah data yang berhasil dikumpulkan sebanyak 2.760 responden. Setiap responden memberikan penilaian terhadap sembilan unsur pelayanan (U1–U9) yang mengacu pada PermenPANRB Nomor 14 Tahun 2017. Adapun unsur penilaian tersebut dapat dilihat pada Tabel 1.

Tabel 1. Unsur Pelayanan

Kode Pelayanan	Unsur Pelayanan
U1	Persyaratan
U2	Prosedur
U3	Waktu Pelayanan
U4	Biaya/Tarif
U5	Produk Layanan
U6	Kompetensi Pelaksana
U7	Perilaku Pelaksana
U8	Penanganan Pengaduan, saran dan masukan
U9	Sarana & Prasarana

Setiap unsur pelayanan dinilai menggunakan skala 1 sampai 4, kemudian diolah untuk menghasilkan Nilai Indeks Kepuasan Masyarakat (IKM). Nilai IKM selanjutnya diklasifikasikan ke dalam kategori mutu pelayanan dan kinerja unit pelayanan sesuai ketentuan PermenPANRB No. 14 Tahun 2017. Kategori tersebut digunakan sebagai

dasar penentuan kategori tingkat kepuasan masyarakat, yang dalam penelitian ini berperan sebagai variabel kelas (*class label*) pada proses klasifikasi. Tabel 2 menunjukkan kategori mutu pelayanan dan kinerja unit pelayanan berdasarkan PermenPANRB No. 14 Tahun 2017.

Tabel 2. Nilai Persepsi, Nilai Interval, Nilai Interval Konversi, Mutu Pelayanan dan Kinerja Unit Pelayanan

Nilai Persepsi	Nilai Interval (NI)	Nilai Interval Konversi (NIK)	Mutu Pelayanan	Kinerja Unit Pelayanan
1	1,00 - 2.5996	25,00 - 64,99	D	Tidak Baik
2	2,60 - 3.064	65,00 - 76.60	C	Kurang Baik
3	3.0644 - 3,532	76,61 - 88,30	B	Baik
4	3,5324 - 4,00	88,31 - 100,00	A	Sangat Baik

Berdasarkan hasil pengolahan data awal, nilai Indeks Kepuasan Masyarakat (IKM) DPMPTSP Kota Metro hingga triwulan IV tahun 2025 mencapai 93,16, yang termasuk dalam kategori Sangat Baik. Hasil ini menunjukkan bahwa secara umum masyarakat merasa puas terhadap kualitas pelayanan yang diberikan. Namun demikian, nilai indeks tersebut masih bersifat agregat dan belum menggambarkan pola hubungan antar unsur pelayanan secara rinci.

Oleh karena itu, data SKM yang telah dikumpulkan selanjutnya digunakan sebagai dasar untuk membangun model klasifikasi tingkat kepuasan masyarakat guna memperoleh pemahaman yang lebih mendalam mengenai pola kepuasan berdasarkan unsur pelayanan. Hasil pengumpulan data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengumpulan Data

U1	U2	U3	U4	U5	U6	U7	U8	U9	NI	Kategori
3	4	3	4	3	4	3	3	4	86,11	Puas
3	4	3	4	3	4	3	2	4	83.33	Puas
3	4	3	4	3	4	3	3	3	86.11	Puas
3	4	3	4	3	4	3	3	4	83.33	Puas
3	4	3	4	3	4	3	4	3	80.56	Puas
4	3	4	4	4	4	3	3	3	88.89	Sangat Puas

3.1.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan terhadap data hasil pengumpulan Survei Kepuasan Masyarakat (SKM) sebelum digunakan dalam proses pemodelan klasifikasi. Tahapan ini bertujuan untuk memastikan bahwa data yang digunakan telah sesuai dengan kebutuhan algoritma klasifikasi dan dapat diproses secara optimal.

Pada tahap awal pra-pemrosesan, dilakukan seleksi atribut dengan mempertahankan atribut unsur pelayanan U1 hingga U9 sebagai variabel prediktor. Atribut lain yang tidak relevan dengan proses klasifikasi, seperti nomor urut responden, tanggal pengisian survei, dan nilai indeks kepuasan (IKM), dihapus dari dataset karena tidak memiliki pengaruh langsung terhadap pembentukan model klasifikasi.

Selanjutnya, dilakukan proses pelabelan (*labeling*) data [18]. Pada tahap ini, atribut kategori tingkat kepuasan masyarakat ditetapkan sebagai label kelas menggunakan operator *Set Role* pada perangkat lunak *RapidMiner*. Proses pelabelan ini bertujuan untuk menentukan variabel target yang akan diprediksi oleh algoritma klasifikasi [19]. Dengan penetapan label tersebut, setiap data responden memiliki kelas kepuasan yang jelas, seperti Baik dan Sangat Baik, sesuai dengan ketentuan PermenPANRB Nomor 14 Tahun 2017.

Proses pengecekan data selanjutnya menunjukkan bahwa dataset tidak mengandung nilai kosong (*missing value*) [20]. Seluruh atribut unsur pelayanan dan label kelas memiliki nilai yang lengkap, sehingga data dapat langsung digunakan tanpa perlu dilakukan proses imputasi data.

Hasil dari tahap pra-pemrosesan ini adalah dataset yang telah bersih, terstruktur, dan berlabel, sehingga siap digunakan pada tahap pemodelan klasifikasi menggunakan algoritma *Decision Tree* dan *Naïve Bayes*, hasilnya dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pra-pemrosesan Data

No	Kategori	U1	U2	U3	U4	U5	U6	U7	U8	U9
1	Puas	3	4	3	4	3	4	3	3	4
2	Puas	3	4	3	4	3	4	3	2	4
3	Puas	3	4	3	4	3	4	3	3	3
4	Puas	3	4	3	4	3	4	3	3	4
5	Puas	3	4	3	4	3	4	3	4	3
6	Sangat Puas	4	3	4	4	4	4	3	3	3

3.1.3 Validasi Model (10-Fold Cross Validation)

Evaluasi model dilakukan menggunakan metode *10-Fold Cross Validation* untuk memperoleh estimasi kinerja yang lebih representatif terhadap keseluruhan data. Hasil pengujian menunjukkan bahwa algoritma *Decision Tree*

memperoleh nilai akurasi sebesar $84,46\% \pm 1,89\%$, sedangkan algoritma *Naïve Bayes* memperoleh akurasi sebesar $95,04\% \pm 1,13\%$. Nilai rata-rata akurasi yang lebih tinggi serta standar deviasi yang lebih kecil pada *Naïve Bayes* menunjukkan bahwa model tersebut memiliki kinerja yang lebih baik dan lebih stabil dibandingkan *Decision Tree* dalam mengklasifikasikan data survei kepuasan masyarakat.

3.1.4 Klasifikasi Menggunakan Algoritma *Decision Tree*

Hasil pengujian, model menghasilkan nilai akurasi sebesar $84,46\% \pm 1,89\%$ dengan *micro average* $84,46\%$, yang menunjukkan bahwa algoritma *Decision Tree* memiliki kemampuan yang cukup baik dalam melakukan klasifikasi tingkat kepuasan masyarakat. Evaluasi performa model dilakukan menggunakan *confusion matrix* sebagaimana ditunjukkan pada Tabel 5.

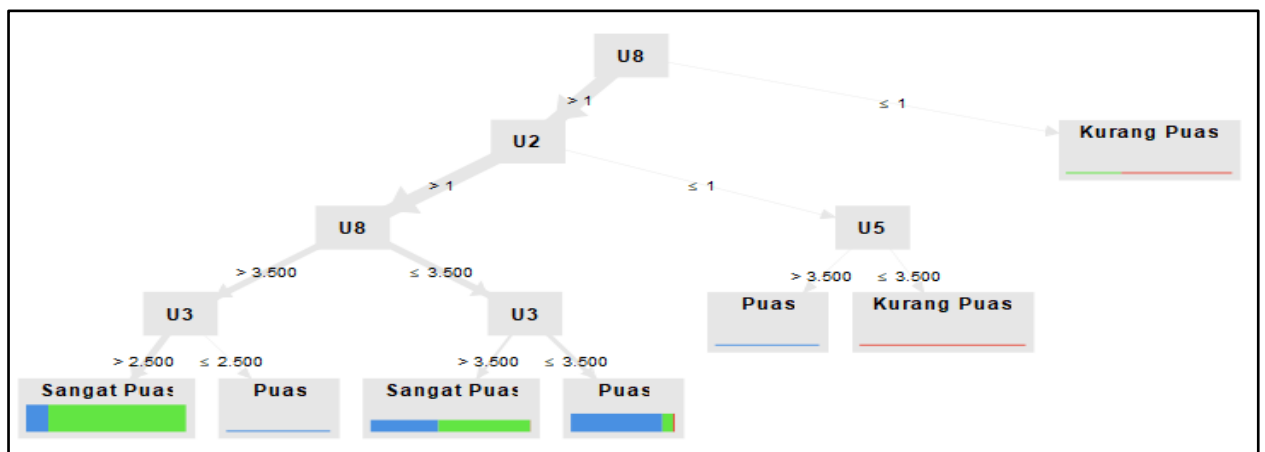
Tabel 5. Tabel *Confusion matrix Decision Tree*

	True Puas	True Sangat Puas	True Kurang Puas	Class Precision
Pred Puas	862	77	17	90,17%
Pred Sangat Puas	329	1.468	3	81,56%
Pred Kurang Puas	2	1	1	25,00%
Class Recall	72,25%	94,95%	4,76%	

Berdasarkan tabel *confusion matrix* di atas, model menunjukkan kinerja yang sangat baik pada kelas Sangat Puas, dengan nilai *recall* sebesar $94,95\%$, yang menunjukkan hampir seluruh data pada kelas tersebut berhasil teridentifikasi oleh model. Sementara itu, kelas Puas memiliki *precision* tinggi ($90,17\%$), namun *recall* lebih rendah ($72,25\%$), yang menunjukkan adanya kecenderungan model mengklasifikasikan sebagian data Puas ke dalam kelas Sangat Puas.

Sebaliknya, pada kelas Kurang Puas, model hanya mampu mengenali sebagian kecil data, terbukti dengan *precision* $25,00\%$ dan *recall* $4,76\%$. Kondisi ini mengindikasikan bahwa setelah proses pembersihan data, jumlah maupun pola data pada kelas Kurang Puas sangat terbatas sehingga tidak cukup representatif untuk dipelajari model.

Hal ini menegaskan bahwa tingginya nilai akurasi keseluruhan model ($84,46\%$) tidak selalu mencerminkan kemampuan model dalam mendeteksi seluruh kelas, terutama pada dataset survei kepuasan masyarakat yang bersifat tidak seimbang. Selanjutnya untuk memahami pola klasifikasi yang dihasilkan model, dilakukan analisis terhadap struktur pohon keputusan (*Decision Tree*) sebagaimana ditampilkan pada Gambar 4.



Gambar 4. Struktur Pohon *Decision Tree*

Pohon keputusan yang ditampilkan pada Gambar 4 tersebut memberikan representasi visual mengenai aturan klasifikasi yang terbentuk berdasarkan unsur-unsur Survei Kepuasan Masyarakat (SKM) DPMPSTSP Kota Metro. Berdasarkan struktur pohon keputusan, unsur Penanganan Pengaduan, Saran, dan Masukan (U8) muncul sebagai simpul akar[21], yang menunjukkan bahwa unsur tersebut merupakan faktor paling dominan dalam menentukan tingkat kepuasan masyarakat. Nilai U8 yang rendah secara langsung mengarah pada klasifikasi Kurang Puas, tanpa melibatkan unsur pelayanan lainnya. Namun, sejalan dengan hasil *confusion matrix*, jumlah data pada kondisi ini sangat terbatas, sehingga klasifikasi Kurang Puas tidak muncul secara signifikan pada hasil prediksi model.

Pada kondisi nilai U8 yang lebih baik, proses klasifikasi selanjutnya dipengaruhi oleh unsur Prosedur Pelayanan (U2). Apabila prosedur pelayanan dinilai kurang jelas, maka kepuasan masyarakat cenderung berhenti pada kategori Puas atau bahkan Kurang Puas, meskipun unsur pelayanan lainnya telah terpenuhi. Sebaliknya, prosedur pelayanan yang jelas dan mudah dipahami menjadi prasyarat utama untuk mencapai tingkat kepuasan yang lebih tinggi.

Selanjutnya, unsur Waktu Pelayanan (U3) berperan sebagai faktor pembeda antara kategori Puas dan Sangat Puas. Kepuasan tertinggi terbentuk ketika penanganan pengaduan berjalan efektif, prosedur pelayanan jelas, serta waktu pelayanan dinilai cepat dan tepat. Pola ini menunjukkan bahwa kepuasan masyarakat DPMPSTSP Kota Metro

bersifat akumulatif, di mana tingkat Sangat Puas hanya tercapai apabila beberapa unsur pelayanan terpenuhi secara simultan.

Secara keseluruhan, struktur pohon keputusan menjelaskan dominasi kelas Puas dan Sangat Puas pada hasil klasifikasi, serta mengonfirmasi temuan pada *confusion matrix* bahwa model cenderung bias terhadap kelas mayoritas[22]. Ketidakhadiran prediksi kelas Kurang Puas pada hasil akhir bukan disebabkan oleh kegagalan algoritma, melainkan oleh keterbatasan representasi data kepuasan rendah setelah proses pembersihan data. Temuan ini menunjukkan bahwa *Decision Tree* efektif sebagai alat interpretatif untuk mengidentifikasi faktor kunci kepuasan masyarakat, namun perlu dikombinasikan dengan pendekatan lain untuk analisis ketidakpuasan secara lebih mendalam.

3.1.5 Klasifikasi Menggunakan Algoritma *Naïve Bayes*

Hasil pemodelan klasifikasi tingkat kepuasan masyarakat menggunakan algoritma *Naïve Bayes* menunjukkan nilai akurasi sebesar akurasi 95,04% ± 1,13% dengan *micro average* 95,04% yang menunjukkan peningkatan kinerja yang signifikan dibandingkan dengan model *Decision Tree* sebelumnya. Hasil permodelan tersebut dapat dilihat pada Tabel 6.

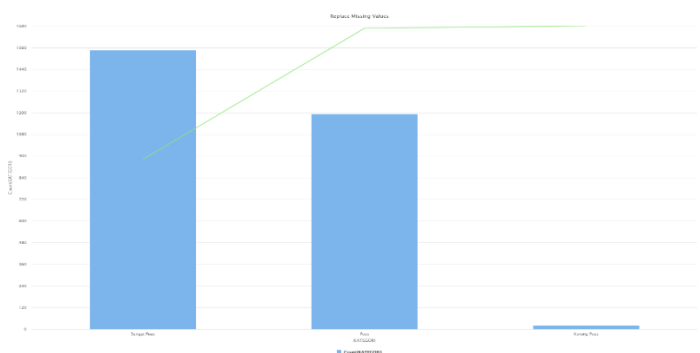
Tabel 6. *Confusion Matrix Naïve Bayes*

	True Puas	True Sangat Puas	True Kurang Puas	Class Precision
Pred Puas	1171	96	18	91.13%
Pred Sangat Puas	18	1449	0	98.77%
Pred Kurang Puas	4	1	3	37.50%
Class Recall	98.16%	93.73%	14.29%	

Berdasarkan tabel *confusion matrix* tersebut menunjukkan performa klasifikasi yang sangat baik dalam mengelompokkan tingkat kepuasan masyarakat DPMPTSP Kota Metro. Pada kelas Puas, model mencapai nilai *recall* sebesar 98,16% dengan *precision* 91,13% yang berarti hampir seluruh data pada kelas tersebut berhasil dikenali, sementara pada kelas Sangat Puas diperoleh *recall* 93,73% dan *precision* 98,77%. Hasil ini menunjukkan sebagian besar prediksi Sangat Puas benar, meskipun terdapat sedikit kesalahan klasifikasi dari kelas Puas.

Namun demikian, kelas Kurang Puas masih menjadi kelas yang paling sulit dikenali, terbukti dengan *recall* hanya 14,29% dan *precision* 37,50%. Kondisi ini menunjukkan keterbatasan jumlah data pada kelas minoritas sehingga model kesulitan mempelajarinya secara representatif. Meskipun akurasi keseluruhan tinggi, evaluasi per kelas menegaskan bahwa nilai akurasi 95,04% tidak sepenuhnya mencerminkan kemampuan model dalam mendeteksi seluruh kelas, khususnya kelas minoritas. Untuk penelitian mendatang, teknik seperti SMOTE atau *Cost-Sensitive Learning* dapat dipertimbangkan untuk meningkatkan kemampuan model dalam mengenali kelas Kurang Puas.

Pada Gambar 5 menunjukkan distribusi hasil prediksi tingkat kepuasan masyarakat menggunakan algoritma *Naïve Bayes*. Terlihat bahwa mayoritas data hasil prediksi berada pada kategori Sangat Puas dan Puas, sedangkan kategori Kurang Puas memiliki jumlah yang sangat kecil.



Gambar 5. Distribusi Hasil Prediksi Tingkat Kepuasan Masyarakat Menggunakan Algoritma *Naïve Bayes*

Evaluasi kinerja kedua metode klasifikasi dapat dilihat pada Tabel 7. Tabel ini menyajikan nilai *precision*, *recall*, rata-rata makro, dan akurasi mikro untuk ketiga kelas kepuasan (Puas, Sangat Puas, dan Kurang Puas), sehingga memberikan gambaran keseluruhan performa masing-masing model.

Tabel 7. Perbandingan kinerja *Decision Tree* & *Naïve Bayes*

Metode	Kelas	Precision	Recall	Macro Precision	Macro Recall	Micro Avg / Accuracy
<i>Decision Tree</i>	Puas	90,17%	72,25%			
	Sangat Puas	81,56%	94,95%	69,74%	63,16%	84,46%
	Kurang Puas	37,50%	14,29%			
<i>Naïve Bayes</i>	Puas	91,13%	98,16%			

Metode	Kelas	Precision	Recall	Macro Precision	Macro Recall	Micro Avg / Accuracy
	Sangat Puas	98,77%	93,73%	75,80%	68,06%	95,04%
	Kurang Puas	37,50%	14,29%			

Dari Tabel 7, terlihat bahwa *Naïve Bayes* menunjukkan performa yang lebih baik dibanding *Decision Tree*. Hal ini ditunjukkan oleh nilai rata-rata makro dan akurasi mikro yang lebih tinggi, meskipun kedua metode sama-sama memiliki kesulitan dalam mengklasifikasikan kelas Kurang Puas yang jumlah datanya paling sedikit. Secara keseluruhan, *Naïve Bayes* lebih konsisten dalam mengenali pola kepuasan pelanggan di dataset ini.

3.2 Pembahasan

Berdasarkan hasil pemodelan dan evaluasi yang telah dilakukan, algoritma *Decision Tree* dan *Naïve Bayes* menunjukkan perbedaan kinerja yang cukup signifikan dalam mengklasifikasikan tingkat kepuasan masyarakat pengguna layanan DPMPSTSP Kota Metro. Evaluasi model dilakukan menggunakan *confusion matrix* dengan indikator akurasi, *precision*, dan *recall* untuk menilai kemampuan prediksi masing-masing algoritma terhadap data uji.

Distribusi data responden dalam penelitian ini menunjukkan ketidakseimbangan kelas yang cukup signifikan. Dari total 2.760 data Survei Kepuasan Masyarakat (SKM) DPMPSTSP Kota Metro, kategori Sangat Puas mendominasi dengan jumlah 1.546 responden, diikuti oleh kategori Puas sebanyak 1.193 responden, sedangkan kategori Kurang Puas hanya berjumlah 21 responden. Kondisi ini mencerminkan persepsi mayoritas masyarakat terhadap kualitas pelayanan publik yang berada pada tingkat kepuasan tinggi, sekaligus menunjukkan karakteristik data SKM yang secara alami tidak seimbang.

Ketidakseimbangan distribusi data tersebut berpengaruh langsung terhadap kinerja algoritma klasifikasi yang digunakan[23]. Berdasarkan hasil evaluasi menggunakan *confusion matrix*, baik algoritma *Decision Tree* maupun *Naïve Bayes* mampu mengklasifikasikan kelas Puas dan Sangat Puas dengan performa yang relatif baik. Namun, kedua model menunjukkan keterbatasan dalam mengidentifikasi kelas Kurang Puas, yang ditunjukkan oleh rendahnya nilai *recall* pada kelas tersebut. Hal ini mengindikasikan bahwa sebagian besar data dengan tingkat kepuasan rendah cenderung salah diklasifikasikan ke dalam kelas mayoritas.

Perbedaan kinerja antara algoritma *Decision Tree* dan *Naïve Bayes* juga dipengaruhi oleh pendekatan pemodelan yang digunakan. *Decision Tree* membangun struktur pohon keputusan berdasarkan pemisahan data yang paling dominan, sehingga model cenderung bias terhadap kelas mayoritas ketika distribusi data tidak seimbang. Sebaliknya, *Naïve Bayes* menggunakan pendekatan probabilistik yang mempertimbangkan distribusi data secara keseluruhan, sehingga menghasilkan performa klasifikasi yang lebih stabil dan tingkat akurasi yang lebih tinggi pada kelas mayoritas, meskipun masih mengalami keterbatasan dalam mendeteksi kelas minoritas[24].

Selain faktor algoritma, ukuran dataset yang digunakan dalam penelitian ini turut memengaruhi hasil klasifikasi. Dengan jumlah data yang relatif besar, model dihadapkan pada variasi persepsi kepuasan masyarakat yang lebih kompleks dibandingkan dengan penelitian terdahulu yang umumnya menggunakan jumlah data kurang dari 500 responden. Pada dataset dengan ukuran yang lebih kecil, model cenderung menghasilkan nilai akurasi yang tinggi karena variasi data yang terbatas dan pola klasifikasi yang lebih sederhana[10]. Sebaliknya, penggunaan dataset yang lebih besar dalam penelitian ini menghasilkan nilai akurasi yang lebih representatif terhadap kondisi nyata pelayanan publik.

Temuan ini menunjukkan bahwa tingginya nilai akurasi pada penelitian dengan jumlah data terbatas tidak selalu mencerminkan kemampuan generalisasi model yang sesungguhnya. Pada penelitian ini, ketidakseimbangan distribusi kelas dan kompleksitas data SKM menyebabkan model klasifikasi lebih realistis dalam menggambarkan kondisi lapangan, meskipun performa pada kelas minoritas masih terbatas. Meskipun ketidakseimbangan kelas berpotensi memengaruhi nilai *recall* pada kategori Kurang Puas, penelitian ini tidak menerapkan teknik penyeimbangan data seperti *oversampling* (misalnya SMOTE) atau *undersampling*. Hal ini dikarenakan tujuan penelitian adalah mengevaluasi performa algoritma terhadap data Survei Kepuasan Masyarakat dalam kondisi aslinya (*real-world* data). Distribusi kelas yang tidak seimbang mencerminkan kondisi faktual tingkat kepuasan masyarakat terhadap layanan DPMPSTSP Kota Metro, sehingga data dipertahankan dalam bentuk aslinya agar model mampu merepresentasikan situasi pelayanan secara lebih realistis. Namun demikian, penerapan teknik penanganan ketidakseimbangan kelas dapat menjadi rekomendasi untuk penelitian selanjutnya guna meningkatkan sensitivitas model terhadap kelas minoritas[25].

Dalam konteks Survei Kepuasan Masyarakat DPMPSTSP Kota Metro yang mengacu pada PermenPANRB Nomor 14 Tahun 2017, dominasi kategori Puas dan Sangat Puas sejalan dengan nilai SKM sebesar 90,93 hingga triwulan IV tahun 2025, yang berada pada kategori sangat baik. Namun demikian, keterbatasan model dalam mengidentifikasi responden dengan tingkat kepuasan rendah menjadi perhatian penting, karena kelompok Kurang Puas merepresentasikan aspek pelayanan yang perlu menjadi fokus evaluasi dan perbaikan berkelanjutan.

Dari sisi pemanfaatan model, *Naïve Bayes* lebih sesuai digunakan sebagai model utama dalam klasifikasi tingkat kepuasan masyarakat karena memiliki kinerja prediksi yang lebih optimal pada data dengan distribusi kelas tidak seimbang. Sementara itu, *Decision Tree* memiliki keunggulan dalam menghasilkan aturan keputusan yang mudah dipahami dan dapat dimanfaatkan sebagai alat analisis eksploratif untuk mengidentifikasi pola hubungan antar unsur pelayanan (U1–U9) terhadap tingkat kepuasan masyarakat. Berdasarkan struktur pohon keputusan yang

dihasilkan, unsur U8 (Penanganan Pengaduan, Saran, dan Masukan) muncul sebagai atribut yang paling dominan dalam menentukan klasifikasi tingkat kepuasan masyarakat. Temuan ini menunjukkan bahwa aspek responsivitas pelayanan memiliki pengaruh signifikan terhadap persepsi kepuasan publik. Secara konseptual, responsivitas merupakan salah satu dimensi penting dalam kualitas pelayanan publik dan sejalan dengan indikator penilaian dalam PermenPANRB Nomor 14 Tahun 2017 tentang Pedoman Survei Kepuasan Masyarakat. Implikasi kebijakan dari temuan ini adalah perlunya penguatan sistem penanganan pengaduan yang cepat, transparan, dan mudah diakses sebagai strategi peningkatan mutu pelayanan secara berkelanjutan di DPMPTSP Kota Metro.

4 KESIMPULAN

Penelitian ini bertujuan untuk mengklasifikasikan tingkat kepuasan masyarakat pengguna layanan DPMPTSP Kota Metro berdasarkan data Survei Kepuasan Masyarakat (SKM) yang mengacu pada PermenPANRB Nomor 14 Tahun 2017 dengan menggunakan algoritma *Decision Tree* dan *Naïve Bayes*. Hasil penelitian, menunjukkan bahwa penerapan teknik *data mining* pada data SKM mampu memberikan analisis yang lebih komprehensif dibandingkan pengolahan SKM secara konvensional. Berdasarkan hasil evaluasi model menggunakan *confusion matrix* dan *cross validation*, algoritma *Naïve Bayes* memperoleh nilai akurasi sebesar $95,04\% \pm 1,13\%$ (*micro average* 95,04%), lebih tinggi dibandingkan *Decision Tree* yang memperoleh akurasi sebesar $84,46\% \pm 1,89\%$ (*micro average* 84,46%), sehingga algoritma *Naïve Bayes* memiliki performa yang lebih optimal dalam mengklasifikasikan tingkat kepuasan masyarakat pada data dengan distribusi kelas tidak seimbang. Meskipun demikian, *Decision Tree* memiliki keunggulan dalam aspek interpretabilitas karena menghasilkan aturan keputusan yang mudah dipahami. Disisi lain *Naïve Bayes* mampu mengklasifikasikan data kepuasan masyarakat secara lebih stabil, karena pendekatan probabilistik yang digunakan mampu menangani dominasi kelas mayoritas serta mengurangi pengaruh variasi dan ketidakseimbangan data, terutama pada kelas dengan jumlah data yang dominan. Berdasarkan struktur pohon keputusan yang terbentuk, unsur U8 (Penanganan Pengaduan, Saran, dan Masukan) menjadi faktor paling dominan dalam menentukan klasifikasi tingkat kepuasan masyarakat, sehingga aspek responsivitas pelayanan perlu menjadi perhatian utama dalam upaya peningkatan kualitas layanan. Penelitian ini juga menemukan adanya ketidakseimbangan distribusi kelas, di mana kategori Kurang Puas memiliki jumlah data yang sangat kecil sehingga memengaruhi kemampuan model dalam mendeteksi kelas minoritas. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penanganan ketidakseimbangan data seperti SMOTE (Synthetic Minority Over-sampling Technique) atau metode ensemble seperti Random Forest guna meningkatkan sensitivitas model terhadap kelas minoritas tanpa mengurangi performa keseluruhan. Secara keseluruhan, integrasi *data mining* dalam pengolahan SKM terbukti dapat mendukung evaluasi dan pengambilan kebijakan pelayanan publik secara lebih berbasis data dan terukur.

REFERENCES

- [1] {Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia}, "Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik," Kementerian PANRB, Jakarta, 2017. [Online]. Available: <https://peraturan.bpk.go.id/Details/132600/permen-pan-rb-no-14-tahun-2017>
- [2] DINAS PENANAMAN MODAL DAN PELAYANAN TERPADU SATU PINTU, "Laporan Pelaksanaan Survei Kepuasan Masyarakat (Skm) Tahun 2025 Pemerintah Kota Metro Dinas Penanaman Modal Dan Pelayanan Terpadu Satu Pintu (Dpmptsp) Tahun 2025," Metro, 2025. [Online]. Available: <https://mpp.metrokota.go.id/v1/layanan.php?id=12>
- [3] A. Setiawan, A. Nugraha Saputra, Z. A. Dewi, Y. Nuryamin, and Ade Priyatna, "Perbandingan Akurasi Algoritma Decision Tree dan Naive Bayes untuk Prediksi Gagal Jantung," *J. Teknol. Komput. dan Sist. Inf. (SINTA 5)*, vol. 10, no. 2, pp. 134–142, 2025, doi: Prefix 10.23960/jtii.
- [4] A. P. Pristiawati, I. Permana, Z. Zarnelly, and F. Muttakin, "Klasifikasi Penerima Bantuan Beras Miskin Menggunakan Algoritma K-NN, NBC dan C4.5," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, pp. 336–344, 2023, doi: 10.47065/bits.v5i1.3617.
- [5] K. Aprilia, U. Khaira, and B. Ferdian Hutabarat, "Komparasi Algoritma Naive Bayes Dan Decision Tree Pada Klasifikasi Penerimaan Peserta Didik Baru," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 10, no. 2, pp. 547–553, 2025, doi: 10.51876/simtek.v10i2.1694.
- [6] L. Winda Sari Siburian, D. Saripurna, S. Kusnasari, P. Studi Sistem Informasi, and S. Triguna Dharma, "Analisis Tingkat Kepuasan Masyarakat Terhadap Pelayanan Kantor Desa Dengan Menggunakan Algoritma C4.5," vol. 3, no. 2, pp. 263–273, 2024, doi: <https://doi.org/10.53513/jursi.v3i2.6023>.
- [7] I. A. Wisky, E. Rianti, and A. Syahira, "Analisis Kepuasan Pelayanan Publik Menggunakan Metode Naïve Bayes Pada Dinas Kependudukan Dan Pencatatan Sipil Kabupaten Agam," *J. Sains Inform. Terap.*, vol. 4, no. 3, pp. 487–492, Oct. 2025, doi: 10.62357/joseamb.v2i2.
- [8] M. W. Khoiriyah, I. H. Santi, and R. D. Romadhona, "Analisis Algoritma C4.5 Dan Naïve Bayes Dalam Menentukan Tingkat Kepuasan Publik Di Rupbasan Kelas 2 Blitar," *JIP (Jurnal Inform. Polinema)*, vol. 11, Nov. 2024, doi: <https://doi.org/10.33795/jip.v11i1.5831>.
- [9] T. A. Putra and M. Permana, Inggih Zarnelly, "Klasifikasi Penerima Bantuan Program Indonesia Pintar (PIP) Pada Siswa SMK Menggunakan Algoritma KNN, NBC dan C4.5," *Technol. Sci.*, vol. 6, no. 4, pp. 2131–2138, 2025, doi: 10.47065/bits.v6i4.6395.
- [10] P. A. M. Zidane, R. W. P. P. Zer, and I. Gunawan, "Penerapan Data Mining Naïve Bayes Dalam Klasifikasi Kepuasan



- Mahasiswa Berlangganan WiFi Indihome,” *J. MEDIA Inform. [JUMIN]*, vol. 3, pp. 112–118, 2022, doi: <https://doi.org/10.55338/jumin.v3i2.488>.
- [11] T. S. Lestari, I. Ismaniah, and W. Priatna, “Particle Swarm Optimization for Optimizing Public Service Satisfaction Level Classification,” *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 1, Mar. 2024, doi: 10.23887/janapati.v13i1.69612.
- [12] P. R. Rosmila, R. Risqiati, and A. S. Darmawan, “Komparasi Algoritma Naïve Bayes Dan Decision Tree Untuk Menganalisis Kepuasan Masyarakat Dpmptsp Kabupaten Batang,” *Inf. Syst. J.*, vol. 8, no. 02, pp. 109–118, Dec. 2025, doi: 10.24076/infosjournal.2025v8i02.2364.
- [13] Y. Umaidah and U. Enri, “Prediction Of Public Service Satisfaction Using C4.5 And Naïve Bayes Algorithm,” *J. PILAR Nusa Mandiri*, vol. 17, no. 2, 2021, doi: <https://doi.org/10.33480/pilar.v17i2.2403>.
- [14] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. Jambi: PT Sonpedia Publishing Indonesia, 2024. [Online]. Available: www.buku.sonpedia.com
- [15] A. A. Purnama and Y. R. Sipayung, “Sentiment Analysis of Public Service Using Naïve Bayes Classifier,” *J. Inf. Syst. Informatics*, vol. 7, no. 3, pp. 2439–2457, Sep. 2025, doi: 10.51519/journalisi.v7i3.1207.
- [16] Suparyanto, “Klasifikasi Kepuasan Layanan Akademik Di STMIK El Rahma Menggunakan Metode Algoritma Naive Bayes,” *J. Inform. Komputer, Bisnis dan Manaj.*, vol. 20, no. 2, p. 100, 2022, doi: <https://doi.org/10.61805/fahma.v20i2.37>.
- [17] W. D. Septiani and M. Marlina, “Comparison of Decision Tree, Naïve Bayes, and Neural Network Algorithm for Early Detection of Diabetes,” *J. Pilar Nusa Mandiri*, vol. 17, no. 1, pp. 73–78, 2021, doi: 10.33480/pilar.v17i1.2213.
- [18] D. A. Eryana, Filius Deivivi, M. Ilham Baehaqi, Arya Wicaksono, Zalfa Alykha Puspita, and Herwis Gultom, “STUDI KOMPARATIF NAIVE BAYES DAN DECISION TREE PADA DATASET BUNGA IRIS: EVALUASI AKURASI DAN EFISIENSI,” *J. Chem. Inf. Model.*, vol. 53, p. 160, 1989, doi: <https://doi.org/10.56486/jris.vol5no2.898>.
- [19] D. Fatmawati, W. Trisnawati, Y. Jumaryadi, and G. Triyono, “Kajian Ilmiah Informatika dan Komputer Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, pp. 1056–1062, 2023, doi: 10.30865/klik.v3i6.803.
- [20] A. Natuzzuhriyyah, N. Nafisah, and R. Mayasari, “Klasifikasi Tingkat Kepuasan Mahasiswa Terhadap Pembelajaran Secara Daring Menggunakan Algoritma Naïve Bayes,” *J. Inform. Sunan Kalijaga*, vol. 6, no. 3, pp. 161–170, 2021, doi: <https://doi.org/10.14421/jiska.2021.6.3.161-170>.
- [21] O. Tulak Bamba, Nur Vadila, Sri Fitrawati, V. W. Tedang, and Asrawati, “Naive Bayes dan Decision Tree: Studi Kasus Klasifikasi Kepuasan Pelanggan E-Commerce,” *SIMKOM*, vol. 10, no. 2, pp. 254–262, Jul. 2025, doi: 10.51717/simkom.v10i2.897.
- [22] D. Amalia Lestari, Sugiyono, Y. Akbar, and A. Z. Hidayat, “Klasifikasi Tingkat Kepuasan Masyarakat terhadap Pelayanan Pembuatan KTP Elektronik di Dinas Dukcapil Semper Barat Menggunakan Metode Naïve Bayes,” Jakarta, 2025. doi: <https://doi.org/10.63447/jimik.v6i3.1598>.
- [23] Y. A. Singgalen, “Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT),” *J. MEDIA Inform. BUDIDARMA*, vol. 7, no. 3, p. 1551, Jul. 2023, doi: 10.30865/mib.v7i3.6461.
- [24] M. Abdurrohman, R. Husna, I. Ali, G. Dwilestari, and N. Rahaningsih, “Penerapan Model Klasifikasi Dalam Tingkat Kepuasan Layanan Publik Kelurahan Karyamulya Menggunakan Algoritma Decision Tree,” *Inf. Manag. Educ. Prof.*, vol. 6, no. 1, pp. 81–90, 2021, doi: <https://doi.org/10.51211/imbi.v6i1.1678>.
- [25] Tasya Nadina Firyal, Hamdani Hamdani, and Anindita Septiarini, “Penerapan Metode Naïve Bayes Untuk Klasifikasi Kepuasan Masyarakat Terhadap Kualitas Pelayanan,” *J. Rekayasa Teknol. Inf.*, vol. Vol.8 No.2, Dec. 2024, doi: <http://dx.doi.org/10.30872/jurti.v8i2>.