

KOMPARASI ALGORITMA *NAIVE BAYES* DAN *DECISION TREE* UNTUK MENGANALISIS KEPUASAN MASYARAKAT DPMPTSP KABUPATEN BATANG

Putri Rastya Rosmila¹⁾, Risqiati²⁾, Arief Soma Darmawan³⁾

¹⁾³⁾ Sistem Informasi, Fakultas Teknologi Informasi, Institut Widya Pratama

²⁾ Teknik Informatika, Fakultas Teknologi Informasi, Institut Widya Pratama

email: putriputri1331@gmail.com¹⁾, risqiati24@iwima.ac.id²⁾, soma_wp@iwima.ac.id³⁾

INFO ARTIKEL

Riwayat Artikel:

Diterima Oktober, 2025

Revisi November, 2025

Terbit November, 2025

ABSTRAK

Survei Kepuasan Masyarakat (SKM) merupakan instrumen evaluasi yang digunakan pemerintah untuk menilai kualitas pelayanan publik secara berkala. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kepuasan masyarakat pada Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Batang dengan membandingkan dua algoritma klasifikasi, yaitu *Naive Bayes* dan *Decision Tree*. Data penelitian berasal dari 330 responden SKM tahun 2024. Hasil klasifikasi menunjukkan bahwa algoritma *Decision Tree* memiliki performa lebih unggul dengan akurasi sebesar 86,4% dan rata-rata akurasi *cross-validation* sebesar 94,6%, sedangkan algoritma *Naive Bayes* memperoleh akurasi sebesar 63,6% dengan rata-rata *cross-validation* sebesar 77,2%. Berdasarkan uji *Chi-square*, atribut paling berpengaruh terhadap kepuasan masyarakat adalah Kompetensi Pelaksana, Perilaku Pelaksana, dan Waktu Penyelesaian. Temuan ini memberikan kontribusi ilmiah berupa penerapan klasifikasi berbasis *data mining* untuk mendukung pengambilan keputusan dalam peningkatan kualitas layanan publik di masa mendatang.

Kata Kunci :

Survei Kepuasan Masyarakat; Data Mining; Klasifikasi; *Naive Bayes*; *Decision Tree*

ABSTRACT

Community Satisfaction Survey (SKM) is an evaluation instrument used by the government to assess the quality of public services on a regular basis. This study aims to classify the level of community satisfaction at the Investment and One-Stop Integrated Service Office (DPMPTSP) of Batang Regency using the Naive Bayes by comparing two classification algorithms: Naive Bayes and Decision Tree. The research data were obtained from 330 SKM respondents in 2024. The classification results show that the Decision Tree algorithm performs better, achieving an accuracy of 86.4% and an average cross-validation accuracy of 94.6%, while the Naive Bayes algorithm achieves 63.6% accuracy with an average cross-validation accuracy of 77.2%. Based on the Chi-square test, the most influential attributes affecting public satisfaction are Staff Competence, Staff Behavior, and Service Completion Time. These findings contribute scientifically by applying data mining-based classification to support decision-making in improving the quality of public services in the future.

Penulis Korespondensi:

Putri Rastya Rosmila
Sistem Informasi, Fakultas Teknologi
Informasi, Institut Widya Pratama

Email:

putriputri1331@gmail.com

Keywords:

Public Satisfaction Survey; Data Mining; Classification; *Naive Bayes*; *Decision Tree*

1. PENDAHULUAN

Pelayanan publik merupakan indikator penting keberhasilan birokrasi dalam memenuhi kebutuhan masyarakat. Survei Kepuasan Masyarakat (SKM) adalah data informasi mengenai tingkat kepuasan masyarakat yang diperoleh melalui pengukuran kualitatif dan kuantitatif atas pendapat masyarakat terhadap layanan publik dengan membandingkan antara harapan dan kenyataan yang dirasakan sebagai penerima layanan [1]. SKM merupakan instrumen penting yang digunakan instansi pemerintah untuk mengukur tingkat kepuasan masyarakat secara kuantitatif, memberikan umpan balik bagi penyelenggara, dan menjadi dasar perbaikan mutu pelayanan berkelanjutan [2]. Dengan demikian, SKM tidak hanya menjadi alat evaluasi, tetapi juga sarana akuntabilitas dan transparansi.

Di Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu (DPMPTSP) Kabupaten Batang, SKM tahun 2024 mencatat indeks kepuasan sebesar 91,42 dengan kategori “Sangat Baik”. Meskipun laporan SKM tersebut memberikan gambaran umum mengenai kualitas pelayanan, diperlukan analisis tambahan untuk melihat pola kepuasan masyarakat secara lebih rinci dan mendukung perumusan strategi peningkatan layanan. Evaluasi pelayanan publik merupakan bagian penting dari manajemen kinerja instansi pemerintah yang bertujuan menilai efektivitas dan efisiensi pelayanan kepada masyarakat sebagai dasar pengambilan kebijakan dalam peningkatan kualitas pelayanan [3]. Evaluasi pelayanan publik juga diwajibkan berdasarkan Peraturan Menteri PANRB Nomor 14 Tahun 2017 dan dilakukan melalui survei kepuasan masyarakat menggunakan metodologi serta indikator yang telah ditentukan [4].

Data mining adalah proses yang menggabungkan teknik statistika, matematika, kecerdasan buatan, dan *machine learning* untuk menemukan informasi berharga serta pengetahuan dari *big data* [5]. Proses ini memungkinkan ekstraksi informasi baru dari basis data berskala besar untuk mendukung pengambilan keputusan [6]. Pendekatan ini sangat relevan dalam konteks SKM karena data survei biasanya memiliki *volume* besar dan variabel beragam. *Data mining* merupakan pendekatan analitis untuk mengekstraksi pengetahuan dan informasi tersembunyi dari kumpulan data dalam *data warehouse* melalui penerapan metode kecerdasan buatan, statistika, dan matematika guna menemukan pola, hubungan, serta tren bermakna, sehingga berperan penting sebagai penghubung antara data kompleks dengan kebutuhan pengguna dalam pengambilan keputusan yang lebih tepat [7]. Melalui pemanfaatan algoritma klasifikasi, hasil SKM tidak hanya ditampilkan secara deskriptif, tetapi juga dapat diprediksi kecenderungannya, sehingga memudahkan instansi pemerintah dalam merumuskan kebijakan. Oleh karena itu, penelitian ini bertujuan memanfaatkan metode klasifikasi dalam *data mining* untuk mengidentifikasi pola kepuasan masyarakat pada DPMPTSP Kabupaten Batang.

Salah satu cara untuk mengetahui kepuasan masyarakat atas mutu pelayanan adalah melalui pendekatan klasifikasi, yaitu proses penyusunan data ke dalam kelompok atau kategori berdasarkan karakteristik tertentu sehingga data lebih terstruktur dan memudahkan pemahaman, pengelolaan, pencarian, serta analisis [8]. *Naïve Bayes Classifier* adalah metode statistik yang digunakan untuk menghitung probabilitas suatu data masuk ke dalam kategori tertentu berdasarkan nilai fitur yang dimilikinya [9]. *Naïve Bayes* merupakan algoritma *machine learning* yang didasarkan pada Teorema Bayes dengan asumsi setiap variabel fitur bersifat saling bebas, dikenal karena kemudahannya implementasi serta efisiensinya dalam memproses data berukuran besar, dan meskipun asumsi “*naïve*” tentang independensi fitur tidak selalu berlaku, metode ini terbukti efektif dengan akurasi yang baik dalam berbagai penerapan seperti klasifikasi teks, deteksi spam, serta analisis opini atau sentimen [10]. *Decision Tree* adalah model klasifikasi berbentuk pohon keputusan, dengan simpul internal sebagai pengujian atribut, cabang sebagai hasil pengujian, dan simpul daun menunjukkan distribusi kelas [11].

Penelitian sebelumnya menunjukkan bahwa *Naïve Bayes* efektif dalam mengelompokkan hasil SKM, sehingga membantu instansi pemerintah dalam mengevaluasi layanan sekaligus mendukung pengambilan keputusan berbasis data [12]. Khoiriyah *et al.* membandingkan kinerja algoritma *C4.5* dan *Naïve Bayes* dalam mengklasifikasikan tingkat kepuasan publik di RUPBASAN Kelas II Blitar, dan menemukan bahwa algoritma *Naïve Bayes* memperoleh akurasi tertinggi sebesar 96,78%, sehingga dianggap paling efektif untuk analisis data kepuasan masyarakat [13]. Temuan lain juga membuktikan bahwa metode klasifikasi ini efektif dalam mendukung proses pengambilan keputusan berbasis data di lingkungan lembaga pemerintahan [1]. Selain *Naïve Bayes*, algoritma *Decision Tree* juga menunjukkan performa yang sangat baik dalam klasifikasi kepuasan masyarakat. Penelitian oleh Windura *et al.* [14] menunjukkan bahwa algoritma *C4.5* mampu membentuk model *Decision Tree* dengan akurasi 100%, sehingga efektif digunakan dalam klasifikasi kepuasan masyarakat. Dalam studi yang dilakukan oleh D. Aditiya dan U. Latifa [15], algoritma *Decision Tree* diidentifikasi sebagai salah satu metode klasifikasi *machine learning* yang paling optimal, dengan tingkat

akurasi tinggi dan performa yang stabil dalam mengolah data survei kepuasan pelanggan. Namun, sebagian besar penelitian terdahulu hanya menekankan akurasi prediksi dan belum mengidentifikasi atribut pelayanan yang paling dominan memengaruhi kepuasan.

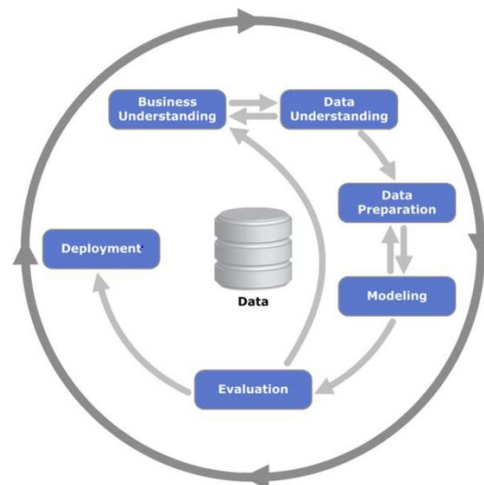
Penelitian ini bertujuan untuk mengklasifikasikan tingkat kepuasan masyarakat pada DPMPTSP Kabupaten Batang menggunakan algoritma *Naïve Bayes* dan *Decision Tree*, sekaligus mengidentifikasi faktor atau indikator pelayanan yang paling berpengaruh terhadap tingkat kepuasan tersebut. Penelitian dibatasi pada data SKM tahun 2024 DPMPTSP Kabupaten Batang. Kontribusi utama penelitian ini adalah melengkapi laporan SKM yang bersifat deskriptif dengan analisis prediktif dan identifikasi atribut dominan, sehingga memberikan dasar pengambilan keputusan yang lebih terukur bagi peningkatan kualitas pelayanan publik.

Dengan demikian, integrasi antara data survei dan metode kecerdasan buatan seperti *Naïve Bayes* dan *Decision Tree* menjadi langkah strategis dalam menciptakan sistem pelayanan publik yang lebih adaptif, efisien, dan transparan.

2. METODOLOGI PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif melalui metode *data mining classification* dengan algoritma *Naïve Bayes* dan *Decision Tree*. Kedua metode ini dipilih karena mampu menangani data kategorikal seperti hasil Survei Kepuasan Masyarakat (SKM) dengan tingkat akurasi tinggi serta komputasi yang efisien. Untuk memastikan alur penelitian sistematis, digunakan kerangka kerja *CRISP-DM* (*Cross Industry Standard Process for Data Mining*).

CRISP-DM adalah kerangka kerja standar lintas industri dalam *data mining* yang mencakup enam tahap berulang, mulai dari pemahaman tujuan bisnis hingga penerapan model, dan dipandang sebagai pendekatan yang komprehensif, sistematis, serta fleksibel, sehingga sering dipilih dalam berbagai proyek analisis data [16]. *CRISP-DM* juga merupakan pendekatan terstandarisasi yang digunakan untuk mengidentifikasi pola-pola penting dalam data [17]. Menurut Sulianta [18], terdapat enam tahapan pokok dalam metode *CRISP-DM*, sebagaimana ditampilkan pada Gambar 1.



Gambar 1. Tahapan dalam metode *CRISP-DM*.

2.1 Business Understanding

Tahap awal ini menekankan pemahaman menyeluruh terhadap permasalahan yang ingin diselesaikan menggunakan *data mining*. Proses ini melibatkan diskusi dengan pemangku kepentingan guna merumuskan tujuan bisnis, target yang ingin dicapai, serta batasan yang berlaku.

2.2 Data Understanding

Pada fase ini, data yang relevan dihimpun dan dieksplorasi guna mengidentifikasi karakteristiknya. Aktivitas mencakup identifikasi sumber data, eksplorasi awal, serta analisis kualitas data.

2.3 Data Preparation

Data yang telah diperoleh kemudian diproses melalui tahap pembersihan, transformasi, dan penyusunan agar siap digunakan dalam pembangunan model. Proses ini mencakup penghapusan data yang tidak diperlukan, penanganan data kosong, serta pelabelan tingkat kepuasan masyarakat berdasarkan hasil penilaian responden. Tahap ini memastikan data berada dalam kondisi optimal untuk membangun model klasifikasi yang akurat dan dapat diandalkan.

2.4 Modeling

Pada tahap ini, pemodelan data dilakukan dengan menerapkan algoritma *Naïve Bayes* dan *Decision Tree* untuk mengklasifikasikan tingkat kepuasan masyarakat berdasarkan hasil Survei Kepuasan Masyarakat (SKM) di DPMPTSP Kabupaten Batang. Kedua algoritma dipilih karena mampu menangani data kategorikal dan memberikan hasil klasifikasi yang interpretatif. Algoritma *Naïve Bayes* bekerja dengan menghitung probabilitas setiap kelas berdasarkan nilai atribut sehingga dapat menentukan kategori dengan probabilitas tertinggi. Sedangkan *Decision Tree* membentuk struktur pohon keputusan untuk menentukan kategori kepuasan dengan tingkat kejelasan yang tinggi. Model dibangun menggunakan bahasa pemrograman *Python* dengan pustaka *scikit-learn* untuk klasifikasi, *pandas* untuk pengolahan data, dan *matplotlib* untuk visualisasi.

2.5 Evaluation

Setelah model *Naïve Bayes* dan *Decision Tree* dibangun, tahap selanjutnya adalah melakukan evaluasi performa untuk membandingkan tingkat akurasi dan keandalan kedua model dalam mengklasifikasikan tingkat kepuasan masyarakat. Evaluasi model dilakukan berdasarkan metrik kinerja seperti *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix*. Selain itu, digunakan juga uji *cross-validation* untuk mengukur konsistensi performa model terhadap data yang berbeda. Tujuan dari tahap ini adalah menjamin bahwa model klasifikasi yang dibangun menghasilkan *output* yang *valid* dan akurat, sehingga dapat menjadi dasar bagi evaluasi serta pengambilan keputusan berbasis data dalam peningkatan kualitas pelayanan publik.

2.6 Deployment

Fase terakhir adalah penerapan model ke dalam sistem nyata di lingkungan organisasi atau perusahaan. Tahap ini dapat berupa penerapan hasil pemodelan *Naïve Bayes* dan *Decision Tree* dalam bentuk visualisasi dan analisis hasil klasifikasi.

3. HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan hasil penelitian yang meliputi proses penerapan metode *CRISP-DM* pada data Survei Kepuasan Masyarakat (SKM) DPMPTSP Kabupaten Batang. Tahapan analisis dilakukan secara sistematis untuk memperoleh model klasifikasi yang dapat mengidentifikasi tingkat kepuasan masyarakat terhadap pelayanan publik. Pembahasan disusun berdasarkan setiap fase dalam metode *CRISP-DM*, mulai dari *Business Understanding* hingga *Evaluation*.

3.1 Business Understanding

Penelitian ini berfokus pada pengembangan model klasifikasi kepuasan masyarakat terhadap pelayanan publik pada DPMPTSP Kabupaten Batang berdasarkan data Survei Kepuasan Masyarakat (SKM). Analisis yang dilakukan selama ini masih deskriptif, sehingga diperlukan pengembangan untuk memberikan *insight* yang lebih strategis. Dengan pendekatan *data mining* berbasis algoritma *Naïve Bayes* dan *Decision Tree*, penelitian ini bertujuan menghasilkan model prediktif yang mampu mengelompokkan tingkat kepuasan masyarakat serta mengidentifikasi faktor layanan dominan yang paling berpengaruh terhadap kepuasan publik. Komparasi kedua algoritma tersebut juga diharapkan memberikan informasi mengenai metode klasifikasi mana yang lebih efektif dan akurat dalam konteks SKM DPMPTSP Kabupaten Batang.

3.2 Data Understanding

Dataset penelitian diperoleh dari Survei Kepuasan Masyarakat (SKM) DPMPTSP Kabupaten Batang tahun 2024 dengan total 330 responden. Pada pelaksanaannya, terdapat perbedaan jumlah indikator pelayanan yang dinilai oleh responden di tiap triwulan. Pada Triwulan 1, responden hanya menilai 9 unsur utama pelayanan, yaitu: Persyaratan, Sistem, Mekanisme dan Prosedur, Waktu Penyelesaian, Biaya/Tarif, Produk Layanan, Kompetensi Pelaksana, Perilaku Pelaksana, Sarana Prasarana, Serta Penanganan Pengaduan, Saran dan Masukan. Mulai Triwulan 2 hingga 4, kuesioner diperluas menjadi 11 unsur dengan penambahan dua indikator baru, yaitu Transparansi Pelayanan dan Integritas Pelaksana. Perbedaan jumlah unsur ini diatasi dengan standarisasi variabel sebelum pemodelan, sehingga semua data dapat dianalisis secara konsisten.

Pemodelan klasifikasi difokuskan pada 11 unsur agar mencerminkan seluruh aspek layanan yang dievaluasi oleh responden di sepanjang tahun.

Eksplorasi awal menunjukkan mayoritas responden memberikan skor 3-4 dari skala 1-4. Dimensi yang relatif tinggi nilainya adalah Persyaratan dan Penanganan Pengaduan Saran Masukan, sedangkan variasi skor lebih besar muncul pada *Sarana Prasarana* dan *Waktu Penyelesaian*.

3.3. Data Preparation

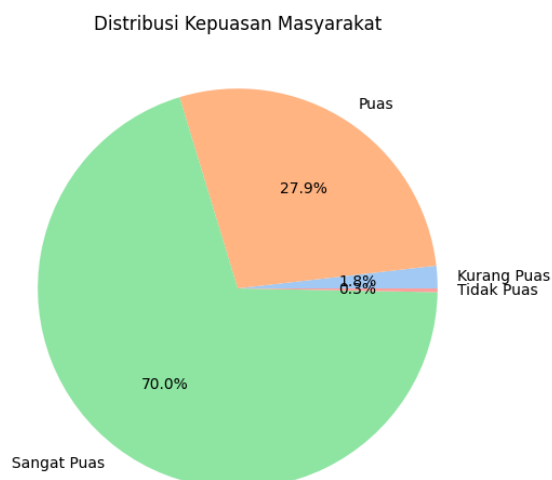
Proses persiapan data bertujuan menjamin bahwa *dataset* yang digunakan dalam pemodelan memiliki kualitas yang baik dan konsisten. Proses ini diawali dengan *data cleaning*, yaitu menghapus duplikasi data dan menangani nilai yang kosong (*missing values*) menggunakan metode imputasi modus agar data menjadi konsisten. Selain itu, sejumlah atribut yang tidak relevan terhadap analisis dikeluarkan, seperti data demografi (nama, nomor telepon, usia, jenis kelamin, pendidikan, dan pekerjaan) karena tidak berpengaruh langsung terhadap hasil penilaian kepuasan masyarakat.

Selanjutnya, dilakukan transformasi data. Skor Survei Kepuasan Masyarakat (SKM) dari sebelas indikator utama pelayanan dihitung rata-ratanya untuk setiap responden. Skala penilaian yang semula menggunakan rentang 1–4 kemudian dikonversi menjadi Indeks Kepuasan Masyarakat (IKM) dengan rentang 25–100. Setelah itu, nilai IKM tersebut dikategorikan ke dalam empat kategori kepuasan, yaitu Tidak Puas, Kurang Puas, Puas, dan Sangat Puas seperti ditunjukkan pada Tabel 1. Tabel ini diadaptasi dari Peraturan Menteri PANRB Nomor 14 Tahun 2017 dengan penyesuaian kategori dan rentang nilai untuk klasifikasi kepuasan per responden.

Tabel 1. Interval kategori IKM.

Nilai Persepsi	Nilai Interval (NI)	Nilai Interval Konversi (NIK)	Kategori Kepuasan (y)
1	1,00 - 2,5996	25,00 - 64,99	Tidak Puas
2	2,60 - 3,064	65,00 - 76,60	Kurang Puas
3	3,0644 - 3,532	76,61 - 88,30	Puas
4	3,5324 - 4,00	88,31 - 100,00	Sangat Puas

Distribusi hasil pengelompokan kategori dapat dilihat pada Gambar 2. Mayoritas responden berada pada kategori Sangat Puas (70,0%), diikuti Puas (27,9%), sedangkan kategori Kurang Puas (1,8%) dan Tidak Puas (0,3%) memiliki proporsi yang sangat kecil. Kondisi ini menandakan adanya ketidakseimbangan kelas (*class imbalance*) yang cukup signifikan, sehingga perlu dilakukan penanganan khusus sebelum tahap pemodelan.



Gambar 2. Distribusi kepuasan masyarakat.

Tahap akhir dari persiapan data adalah pembagian dataset, di mana 80% data (264 responden) digunakan sebagai data latih, sedangkan 20% data (66 responden) digunakan sebagai data uji. Setelah pembagian dilakukan, diterapkan proses *oversampling* pada data latih untuk menyeimbangkan distribusi kelas agar model *Naïve Bayes* tidak bias terhadap kategori mayoritas. Pendekatan ini bertujuan meningkatkan kemampuan model dalam mengenali kelas minoritas seperti Kurang Puas dan Tidak Puas, sehingga menghasilkan prediksi yang lebih akurat dan representatif.

3.4. Modeling

Pada tahap pemodelan, dilakukan proses pembangunan dan komparasi antara dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Decision Tree*, guna memperoleh model terbaik dalam mengklasifikasikan tingkat kepuasan masyarakat terhadap pelayanan publik di DPMPTSP Kabupaten Batang. Kedua model menggunakan sebelas indikator pelayanan publik dari hasil Survei Kepuasan Masyarakat (SKM) sebagai variabel independen (X), sedangkan kategori kepuasan masyarakat (Tidak Puas, Kurang Puas, Puas, dan Sangat Puas) sebagai variabel dependen (Y). Sebelum dilakukan pelatihan model, dataset dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% dengan metode *stratified split* agar proporsi setiap kelas tetap seimbang. Selanjutnya, digunakan teknik *Random Over Sampling* (ROS) pada data latih guna menyeimbangkan distribusi kelas, mengingat kategori Sangat Puas memiliki proporsi responden yang jauh lebih besar dibandingkan kategori lainnya.

Kedua model kemudian dilatih menggunakan data latih hasil *oversampling* untuk mempelajari pola hubungan antara indikator pelayanan dengan tingkat kepuasan masyarakat. Proses pengujian dilakukan menggunakan data uji untuk menilai performa model berdasarkan metrik akurasi, presisi, *recall*, dan *F1-score*. Hasil pengujian dari kedua algoritma selanjutnya dianalisis dan dibandingkan untuk menentukan model dengan performa paling optimal dalam mengklasifikasikan tingkat kepuasan masyarakat terhadap pelayanan publik di DPMPTSP Kabupaten Batang. Perbandingan kinerja algoritma ini bertujuan untuk memberikan pemahaman yang lebih menyeluruh terkait efektivitas masing-masing algoritma dalam mendukung evaluasi dan peningkatan kualitas pelayanan publik berbasis data.

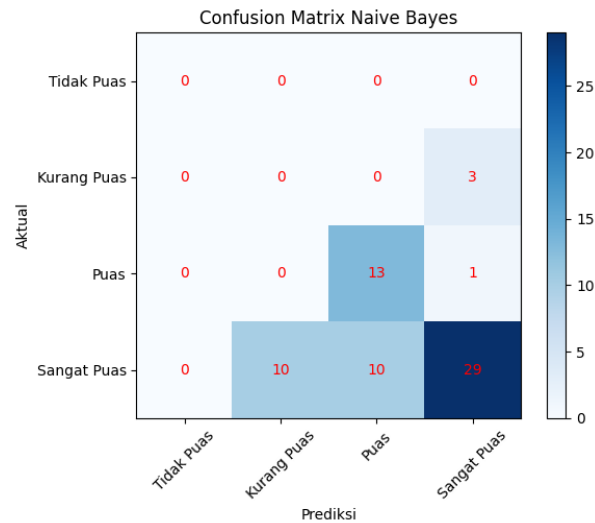
3.5. Evaluation

Tahap evaluasi dilakukan untuk menilai performa model *Naïve Bayes* dan *Decision Tree* dalam mengklasifikasikan tingkat kepuasan masyarakat berdasarkan hasil Survei Kepuasan Masyarakat (SKM) di DPMPTSP Kabupaten Batang. Evaluasi dilakukan melalui dua tahap, yaitu pengujian menggunakan data uji (*test set*) dan validasi silang (*cross validation*) untuk memastikan konsistensi performa model.

Pada model *Naïve Bayes*, hasil pengujian awal setelah dilakukan proses *oversampling* menghasilkan akurasi sebesar 63,6%. Nilai ini menunjukkan bahwa model cukup mampu mengenali pola data, meskipun masih terdapat kesalahan klasifikasi pada beberapa kategori kepuasan. Berdasarkan pada Gambar 3., kategori Puas dan Sangat Puas menunjukkan performa terbaik dengan nilai *recall* masing-masing sebesar 0,93 dan 0,59, sedangkan kategori Kurang Puas memiliki nilai *precision*, *recall*, dan *F1-score* sebesar 0. Hal ini memperlihatkan bahwa model *Naïve Bayes* cenderung bias terhadap kelas mayoritas. Visualisasi yang ditampilkan sebagaimana Gambar 4., juga memperkuat temuan tersebut, di mana sebagian besar data aktual dari kategori Kurang Puas dan Puas terprediksi sebagai Sangat Puas. Meski demikian, hasil *5-Fold Cross Validation* menunjukkan rata-rata akurasi sebesar 77,2%, menandakan bahwa model masih cukup stabil dan memiliki kemampuan generalisasi yang baik.

=== Classification Report ===				
	precision	recall	f1-score	support
Kurang Puas	0.00	0.00	0.00	3
Puas	0.57	0.93	0.70	14
Sangat Puas	0.88	0.59	0.71	49
accuracy			0.64	66
macro avg	0.48	0.51	0.47	66
weighted avg	0.77	0.64	0.67	66

Gambar 3. Classification report Naïve Bayes.



Gambar 4. *Confusion Matrix Naïve Bayes.*

Sementara itu, model *Decision Tree* menunjukkan performa yang jauh lebih baik dibandingkan *Naïve Bayes*. Berdasarkan pengujian awal, *Decision Tree* memperoleh akurasi sebesar 86,4%, dengan nilai *F1-score* tertinggi pada kategori Kurang Puas (1,00) dan Sangat Puas (0,90) sebagaimana ditunjukkan pada Gambar 5. Visualisasi pada Gambar 6., memperlihatkan bahwa sebagian besar data aktual berhasil diklasifikasikan dengan benar, hanya terdapat sedikit kesalahan pada kategori Puas. Hasil ini juga konsisten dengan *5-Fold Cross Validation*, yang menghasilkan akurasi rata-rata sebesar 94,6%, menandakan bahwa model *Decision Tree* memiliki kestabilan dan kemampuan prediksi yang sangat baik terhadap seluruh kelas kepuasan masyarakat.

```

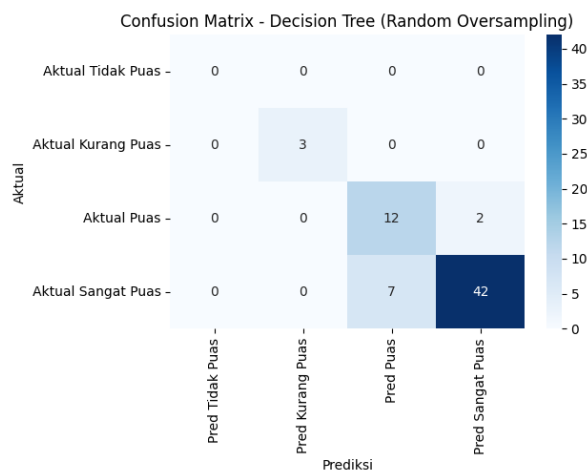
=== Evaluasi Decision Tree (Random Oversampling) ===
Akurasi: 0.864

Classification Report:
              precision    recall  f1-score   support

   Kurang Puas         1.00        1.00        1.00         3
      Puas            0.63        0.86        0.73        14
   Sangat Puas         0.95        0.86        0.90        49

 accuracy              0.86              0.86         66
  macro avg           0.86              0.90        0.88         66
 weighted avg         0.89              0.86        0.87         66
  
```

Gambar 5. *Classification report Decision Tree.*



Gambar 6. *Confusion Matrix Decision Tree.*

serta dasar pengambilan keputusan strategis dalam upaya peningkatan kualitas pelayanan publik di DPMPTSP Kabupaten Batang.

4. KESIMPULAN

Penelitian ini berhasil mengklasifikasikan tingkat kepuasan masyarakat berdasarkan data Survei Kepuasan Masyarakat (SKM) di DPMPTSP Kabupaten Batang menggunakan dua algoritma klasifikasi, yaitu *Naïve Bayes* dan *Decision Tree*. Hasil pengujian menunjukkan bahwa model *Decision Tree* memperoleh akurasi tertinggi sebesar 86,4% dengan rata-rata akurasi *cross-validation* sebesar 94,6%, sedangkan model *Naïve Bayes* menghasilkan akurasi sebesar 63,6% dengan rata-rata akurasi *cross-validation* sebesar 77,2%. Perbandingan tersebut menunjukkan bahwa *Decision Tree* memiliki performa yang lebih stabil dan unggul dalam mengklasifikasikan tingkat kepuasan masyarakat. Secara umum, mayoritas responden menilai layanan pada kategori Sangat Puas dan Puas, menandakan bahwa kinerja pelayanan publik DPMPTSP Kabupaten Batang dinilai sangat baik oleh masyarakat.

Berdasarkan hasil uji *Chi-square*, atribut yang paling dominan memengaruhi tingkat kepuasan masyarakat meliputi Kompetensi Pelaksana, Perilaku Pelaksana, dan Waktu Penyelesaian, disusul oleh Sistem Mekanisme Prosedur, Produk Layanan, serta Biaya Tarif. Temuan ini memperlihatkan bahwa aspek sumber daya manusia dan efisiensi proses layanan berperan penting dalam membentuk persepsi kepuasan masyarakat. Kontribusi utama penelitian ini terletak pada penerapan pendekatan data mining berbasis algoritma klasifikasi untuk mendukung evaluasi kinerja pelayanan publik serta memperkaya laporan SKM yang selama ini bersifat deskriptif. Dengan adanya model prediktif ini, instansi dapat mengidentifikasi area pelayanan yang perlu ditingkatkan serta menetapkan kebijakan berbasis data. Untuk penelitian selanjutnya, disarankan melakukan pengembangan model menggunakan algoritma lain seperti *k-Nearest Neighbors*, *Random Forest* atau *Gradient Boosting*, serta menerapkan teknik praproses dan penyeimbangan data yang lebih optimal agar model mampu memberikan hasil klasifikasi yang lebih akurat dan representatif terhadap seluruh kategori kepuasan masyarakat.

DAFTAR PUSTAKA

- [1] S. Wirma, "Data Mining Dengan Metode Naïves Bayes Classifier dalam Memprediksi Tingkat Kepuasan Pelayanan Dokumen Kependudukan," *J. Inform. Ekon. Bisnis*, vol. 4, no. 3, pp. 156–160, 2022, doi: 10.37034/infneb.v4i3.155.
- [2] Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi, "Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 14 Tahun 2017 tentang Pedoman Penyusunan Survei Kepuasan Masyarakat Unit Penyelenggara Pelayanan Publik," Jakarta, 2017. [Online]. Available: <https://peraturan.bpk.go.id/Details/132600/permen-pan-rb-no-14-tahun-2017>
- [3] M. S. Aprilia and R. C. Puspitarini, "Analisis Indeks Kepuasan Masyarakat Mal Pelayanan Publik Kota Probolinggo," *Publicio J. Ilm. Polit. Kebijak. dan Sos.*, vol. 3, no. 1, pp. 56–61, 2021, doi: 10.51747/publicio.v3i1.497.
- [4] R. Amrulloh, "Survei Kepuasan Masyarakat Terhadap Kualitas Pelayanan Pdam Intan Banjar," *Al-KALAM J. KOMUNIKASI, BISNIS DAN Manaj.*, vol. 9, no. 2, p. 128, 2022, doi: 10.31602/al-kalam.v9i2.6526.
- [5] Sutisna and N. M. Yuniar, "Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes," *J. Sains dan Teknol.*, vol. 5, no. 1, pp. 243–246, 2023, [Online]. Available: <https://doi.org/10.55338/saintek.v5i1.1383>
- [6] R. N. Ainum, N. Hidayat, and A. A. Soebroto, "Klasifikasi Buku Perpustakaan menggunakan Metode Naive Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 8, pp. 3726–3732, 2022.
- [7] D. Jollyta, W. Ramdhan, and M. Zarlis, *KONSEP DATA MINING DAN PENERAPAN*. Yogyakarta: DEEPUBLISH, 2020.
- [8] A. L. Nasution and R. N. S. Fathonah, *Klasifikasi Kondisi Peralatan Elektronik Metode Gaussian Naïve Bayes*. Bandung: Penerbit Buku Pedia, 2023.
- [9] J. Siswoyo et al., "Klasifikasi Tingkat Kepuasan Pelanggan Facebook Marketplace," vol. 9, no. 1, pp. 237–245, 2025.
- [10] S. A. S. Mola, S. N. R. Djawa, and A. Y. Mauko, *Text Mining Analisis Sentimen dengan Naïve Bayes*. Bandung: Kaizen Media Publishing, 2025.
- [11] B. Zaman, L. margareta Huizen, and M. B. Ardima, "Prediksi Kepuasan Mahasiswa Terhadap Pelayanan Akademik Menggunakan Model Decision Tree," *J. Transform.*, vol. 21, no. 2, pp. 46–55, 2024, doi: 10.26623/transformatika.v21i2.8214.
- [12] T. Nadina Firyal, H. Hamdani, and A. Septiarini, "Penerapan Metode Naïve Bayes Untuk Klasifikasi Kepuasan Masyarakat Terhadap Kualitas Pelayanan," *Jurti*, vol. 8, no. 2, pp. 2579–8790, 2024.
- [13] 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," *J. Inform. Polinema*, vol. 11, no. 1, pp. 13–18, 2024, doi: 10.33795/jip.v11i1.5831.

- [14] R. A. Windura, I. O. Kirana, and M. Safii, "Analisis Kepuasan Masyarakat Terhadap Pelayanan Kantor Camat Siantar Marimbun Menggunakan Algoritma C4.5," *Student Res. J.*, vol. 2, no. 1, pp. 312–320, 2023, doi: 10.55606/srjyappi.v2i1.1035.
- [15] D. Aditiya and U. Latifa, "Uji Efektivitas Penerapan Machine Learning Classification Untuk Survey Kepuasan Pelanggan Maskapai Penerbangan X," *Barometer*, vol. 8, no. 1, pp. 9–18, 2023, doi: 10.35261/barometer.v8i1.6566.
- [16] D. Kurniawan and M. Yasir, "Optimization Sentimen Analysis using CRISP-DM and Naive Bayes Methods Implemented on Social Media," *Cybersp. J. Pendidik. Teknol. Inf.*, vol. 6, no. 2, p. 74, 2022, doi: 10.22373/cj.v6i2.12793.
- [17] Y. Yuliani, H. Sulaiman, Yanto, I. Al Farobi, and K. Panggalih, "Penerapan Model CRISP-DM Untuk Klasifikasi Tumor Otak Menggunakan Algoritma CNN," vol. 8, no. 2, 2025.
- [18] F. Sulianta, *Buku Dasar Data Mining from A to Z*, no. January. 2023