Journal of Computer Science and Information Systems (JCoInS)
Program Studi Sistem Informasi, Fakultas Sains & Teknologi, Universitas Labuhanbatu
Volume : 6, Nomor : 3, Agustus 2025, Pages. 322-334
e-ISSN: 2747-2221

Analisis Kepuasan Masyarakat Terhadap Kinerja Bupati Labuhanbatu Selatan Periode 2021-2024 Menggunakan Metode *Decision Tree* dan *Naive Bayes*

[1]Ramadhani, [2]Syaiful Zuhri Harahap, [3]Sudi Suryadi, [4]Masrizal

[1,2,3,4]Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Labuhanbatu

Email : [1]ramadhanipria2002@gmail.com, [2]syaifulzuhriharahap@gmail.com, [3]sudisuryadi28@gmail.com, [4]masrizal120405@gmail.com

Corresponding Author : ramadhanipria2002@gmail.com

Abstract

This study was conducted to analyze the level of customer satisfaction with services by comparing the performance of two classification methods, namely Decision Tree and Naive Bayes, so that an accurate model can be obtained to assist decision making. This problem is important because understanding customer satisfaction patterns can be a strategic basis in improving service quality and maintaining loyalty. The theoretical basis used refers to the concept of machine learning classification, where Decision Tree forms a branching rule-based model based on attributes, while Naive Bayes relies on probability calculations based on Bayes' theorem with the assumption of independence between features. The research methodology includes data collection stages, pre-processing to ensure data quality, model training with both methods, and performance evaluation using Test & Score and Confusion Matrix. Based on the classification results, the Decision Tree method produces fairly good accuracy, precision, and recall, but the Naive Bayes method shows higher performance with an accuracy of 91.67%, a precision of the "Satisfied" class of 98.11%, and a recall of 92.86%, which indicates a very good level of prediction accuracy especially for the majority class. Evaluation of both methods shows that Naive Bayes excels in capturing existing data patterns, although Decision Tree still has good interpretability for classification rule analysis. In conclusion, both methods are capable of classifying customer satisfaction data with adequate performance, but Naive Bayes is recommended as the primary model due to its higher and more consistent evaluation results, while Decision Tree can be used as an alternative when model interpretation is a priority.

Keywords *: Decision Tree; Naive Bayes; Classification; Model Evaluation; Naive Bayes.*

1. Pendahuluan

Kinerja kepala daerah merupakan salah satu indikator utama keberhasilan penyelenggaraan pemerintahan di tingkat kabupaten, yang secara langsung berkaitan dengan kesejahteraan masyarakat. Dalam konteks ini, tingkat kepuasan masyarakat menjadi cerminan seberapa efektif kebijakan dan program yang dilaksanakan oleh pemerintah daerah dalam memenuhi kebutuhan dan harapan publik. Kabupaten Labuhanbatu Selatan, sebagai salah satu daerah berkembang di Provinsi Sumatera Utara, memiliki berbagai program pembangunan pada periode kepemimpinan 2021–2024 yang meliputi perbaikan infrastruktur, peningkatan mutu layanan kesehatan, penguatan sektor pendidikan, hingga pengembangan ekonomi berbasis potensi lokal. Periode ini menjadi

momentum penting untuk melakukan evaluasi yang terukur, mengingat banyaknya kebijakan strategis yang diambil dalam upaya mempercepat pembangunan daerah. Penelitian ini berangkat dari kebutuhan untuk memahami sejauh mana kebijakan-kebijakan tersebut mampu memberikan dampak positif yang nyata bagi masyarakat serta bagaimana respons masyarakat terhadap kinerja Bupati selama menjabat. Penelitian ini memusatkan perhatian pada analisis tingkat kepuasan masyarakat terhadap kinerja Bupati Labuhanbatu Selatan dalam kurun waktu 2021–2024, dengan memanfaatkan pendekatan berbasis data mining. Metode yang digunakan adalah Decision Tree dan Naive Bayes, dua algoritma populer dalam klasifikasi yang memiliki karakteristik berbeda namun saling melengkapi. Decision Tree menawarkan struktur visual yang memudahkan interpretasi dan penelusuran faktor-faktor utama yang memengaruhi kepuasan masyarakat. Sementara itu, Naive Bayes dikenal memiliki kecepatan pemrosesan yang tinggi dan tingkat akurasi yang kompetitif untuk data kategorikal. Pendekatan ini memungkinkan penelitian tidak hanya memberikan hasil klasifikasi, tetapi juga pemahaman mendalam mengenai variabel yang paling dominan mempengaruhi tingkat kepuasan, sehingga temuan dapat dijadikan landasan dalam pengambilan kebijakan yang lebih tepat sasaran. Pendekatan yang ditawarkan diharapkan dapat memberikan hasil yang tidak hanya akurat tetapi juga mudah dipahami oleh pihak-pihak terkait, termasuk pemerintah daerah, akademisi, maupun masyarakat umum. Decision Tree akan digunakan untuk memetakan alur logika pengambilan keputusan berdasarkan atribut-atribut kepuasan, yang dapat divisualisasikan secara jelas sehingga memudahkan pemangku kebijakan untuk memahami hasil analisis. Sementara itu, Naive Bayes akan dimanfaatkan untuk menghitung probabilitas tingkat kepuasan dengan mempertimbangkan distribusi variabel yang ada, sehingga dapat memberikan prediksi yang kuat meskipun data bersifat heterogen. Kombinasi kedua metode ini diharapkan mampu memperkaya wawasan terkait evaluasi kinerja kepala daerah, serta menjadi acuan dalam merumuskan kebijakan yang lebih responsif, tepat sasaran, dan berorientasi pada kepentingan masyarakat Labuhanbatu Selatan di masa mendatang.

2. Landasan Teori

Data Mining

Data mining adalah proses untuk menggali dan menemukan pola atau informasi penting dari kumpulan data yang besar dan kompleks (Alam, Alana, & Juliane, 2023). Proses ini dilakukan dengan tahapan yang terstruktur, mulai dari pengumpulan data, pembersihan data agar bebas dari kesalahan atau data ganda, transformasi data menjadi format yang sesuai, penerapan algoritma analisis, hingga interpretasi hasil. Tujuan dari data mining adalah mengubah data mentah menjadi informasi yang memiliki nilai guna untuk pengambilan keputusan. Dalam penelitian ini, data mining digunakan untuk mengolah data hasil survei tingkat kepuasan masyarakat sehingga pola hubungan antar faktor kepuasan dapat dianalisis secara objektif dan terukur (Atalya Angelus Leza, Widya Utami, & Anugrah Cahya Dewi, 2024).

Metode *Decision Tree*

Decision Tree adalah metode klasifikasi yang menyajikan proses pengambilan keputusan dalam bentuk pohon bercabang (Barbosa, de Souza, Carneiro, & Farhate, 2021) (Nuraeni, Harliana, & Prabowo, 2024). Setiap cabang merepresentasikan pilihan berdasarkan nilai suatu atribut, dan setiap simpul akhir menunjukkan hasil keputusan atau

kategori tertentu (Zai, Sirait, Nainggolan, Sihombing, & Banjarnahor, 2023). Proses pembentukan pohon dimulai dengan memilih atribut yang paling mampu membedakan data, kemudian data dipisahkan berdasarkan nilai atribut tersebut hingga menghasilkan kelompok yang homogen (Mohammed et al., 2022). Keunggulan Decision Tree adalah kemampuannya menghasilkan model yang mudah dipahami secara visual dan logis, sehingga membantu melihat faktor-faktor utama yang mempengaruhi hasil. Dalam penelitian ini, Decision Tree digunakan untuk melihat jalur keputusan yang mengarah pada kategori kepuasan atau ketidakpuasan masyarakat (Mirbod, Choi, Heinemann, Marini, & He, 2023).

Metode Naïve Bayes

Naive Bayes adalah metode klasifikasi yang menggunakan perhitungan probabilitas untuk menentukan kemungkinan suatu data termasuk ke dalam kategori tertentu (Aulia et al., 2020). Metode ini bekerja dengan mengasumsikan bahwa setiap atribut memiliki pengaruh yang independen terhadap hasil akhir, meskipun pada kenyataannya mungkin terdapat keterkaitan antar atribut (Madjid, Ratnawati, & Rahayudi, 2023).

Model Klasifikasi

Model klasifikasi adalah rancangan atau metode yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut atau variabel yang dimilikinya (Setiawan, Rabi, & Gumilang, 2024). Proses pembentukan model dilakukan melalui tahap pelatihan dengan data berlabel, di mana model belajar mengenali pola dan hubungan antar variabel. Setelah itu, model diuji pada data baru untuk melihat sejauh mana kemampuannya dalam memprediksi kategori yang tepat (Hafizan & Putri, 2020).

3. Metodologi Penelitian

Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode survei untuk mengumpulkan data tingkat kepuasan masyarakat terhadap kinerja Bupati Labuhanbatu Selatan periode 2021–2024. Data diperoleh melalui kuesioner yang disebarkan kepada responden yang dipilih menggunakan teknik *purposive sampling* dengan mempertimbangkan keterwakilan wilayah dan latar belakang masyarakat. Variabel yang diukur mencakup beberapa indikator kepuasan, seperti kualitas pelayanan publik, responsivitas pemerintah, transparansi kebijakan, pemerataan pembangunan, dan realisasi program kerja. Data yang terkumpul kemudian diolah menggunakan teknik data mining dengan dua algoritma klasifikasi, yaitu Decision Tree dan Naive Bayes, untuk melihat pola dan perbedaan hasil antara kedua metode tersebut.

Tahapan penelitian dimulai dari pengumpulan data, dilanjutkan dengan proses *data preprocessing* yang meliputi pembersihan data dari nilai yang hilang, penyamaan format, serta pengkodean data agar sesuai dengan kebutuhan analisis. Selanjutnya, data dibagi menjadi data latih (*training set*) dan data uji (*testing set*) untuk membangun dan menguji model klasifikasi. Algoritma Decision Tree digunakan untuk membuat model berbentuk pohon keputusan yang memetakan jalur penentuan kategori kepuasan, sedangkan Naive Bayes digunakan untuk memprediksi kategori berdasarkan perhitungan probabilitas. Kinerja kedua model dibandingkan berdasarkan nilai akurasi, presisi, dan *recall* untuk menentukan metode yang paling efektif dalam mengklasifikasikan tingkat kepuasan masyarakat.

4. Hasil Dan Pembahasan

Pengumpulan Data

Pada tahap pengumpulan data, seluruh informasi yang diperoleh dari hasil survei tingkat kepuasan masyarakat terhadap kinerja Bupati Labuhanbatu Selatan periode 2021–2024 dikumpulkan dan disusun secara terstruktur. Total data yang berhasil dihimpun berjumlah 105 record, yang masing-masing berisi nilai pada indikator kepuasan seperti pendidikan, kesehatan, dan infrastruktur jalan. Setelah data terkumpul, proses pembagian dataset dilakukan secara langsung menjadi dua bagian, yaitu data *training* dan data *testing*. Data *training* digunakan untuk membangun dan melatih model klasifikasi agar dapat mengenali pola hubungan antar variabel, sedangkan data *testing* digunakan untuk menguji performa model dalam memprediksi tingkat kepuasan pada data yang belum pernah dilihat sebelumnya. Pembagian ini bertujuan untuk memastikan bahwa model yang dihasilkan mampu melakukan prediksi secara akurat dan tidak hanya menghafal pola dari data pelatihan. Data training yang digunakan dalam penelitian untuk membangun model klasifikasi tingkat kepuasan masyarakat terhadap kinerja Bupati Labuhanbatu Selatan periode 2021–2024. Data training ini berjumlah sebanyak 45 record, di mana setiap record berisi informasi mengenai nama responden, penilaian terhadap tiga indikator utama yaitu pendidikan, kesehatan, dan infrastruktur jalan, serta kategori kepuasan yang menjadi target klasifikasi (label). Kategori kepuasan terbagi menjadi dua kelas, yaitu "puas" dan "tidak puas", yang nantinya akan digunakan sebagai acuan bagi algoritma untuk mempelajari pola hubungan antara nilai pada variabel indikator dengan kelas yang dihasilkan. Secara umum, data training adalah sekumpulan data berlabel yang digunakan untuk melatih sebuah model dalam mengenali pola dan hubungan antar variabel, sehingga model mampu melakukan prediksi yang akurat pada data yang belum pernah dilihat sebelumnya. Proses pelatihan menggunakan data ini bertujuan agar algoritma seperti Decision Tree dan Naive Bayes dapat membentuk aturan atau perhitungan probabilitas yang sesuai dengan karakteristik data, sehingga siap untuk diuji pada tahap testing. Data testing yang digunakan dalam penelitian, dengan jumlah total sebanyak 60 data. Setiap data memuat informasi nama responden beserta penilaian terhadap tiga indikator utama, yaitu pendidikan, kesehatan, dan infrastruktur jalan. Data testing adalah sekumpulan data berlabel yang tidak digunakan dalam proses pelatihan model, melainkan dipakai untuk mengukur dan mengevaluasi kinerja model yang telah dibangun menggunakan data training. Melalui data ini, dapat diketahui sejauh mana model mampu memprediksi kategori kepuasan masyarakat secara akurat pada data yang belum pernah dilihat sebelumnya. Dengan demikian, penggunaan data testing membantu memastikan hasil evaluasi model bersifat objektif dan mencerminkan kemampuan prediksi di dunia nyata.

Pra-Pemrosesan Data

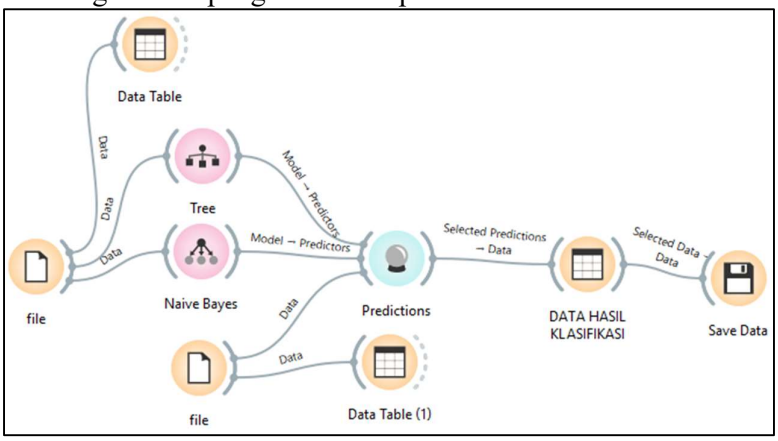
Pra-pemrosesan data adalah tahap penting dalam siklus analisis data yang bertujuan untuk memastikan bahwa data yang akan digunakan berada dalam kondisi terbaik sebelum diproses oleh algoritma. Proses ini biasanya mencakup pembersihan data untuk menghilangkan kesalahan penulisan, duplikasi, atau nilai yang tidak logis, penanganan data hilang, serta standarisasi format agar seragam. Selain itu, pra-pemrosesan juga dapat melibatkan proses transformasi seperti konversi data kategorikal menjadi bentuk numerik, normalisasi nilai numerik, dan pengelompokan (*binning*) agar sesuai dengan kebutuhan metode analisis. Kualitas data sangat menentukan performa

model yang dibangun, sehingga pra-pemrosesan berfungsi sebagai pondasi yang memastikan hasil analisis akurat, konsisten, dan dapat dipertanggungjawabkan. Tanpa tahap ini, model berisiko menghasilkan kesimpulan yang keliru akibat data yang kotor atau tidak seragam.

Pada penelitian ini, data yang diperoleh dari responden sudah berada dalam kondisi baik sehingga tidak memerlukan proses pembersihan atau transformasi yang rumit. Seluruh variabel telah terisi lengkap tanpa ada nilai yang hilang, format penulisan seragam, dan kategori setiap jawaban konsisten sesuai desain kuesioner. Data juga telah memenuhi prinsip keterwakilan sehingga dapat langsung digunakan untuk proses pembagian menjadi *training set* dan *testing set*. Kelayakan data ini mempercepat alur penelitian karena tahap pra-pemrosesan dapat dilewati secara cepat, sehingga peneliti dapat segera masuk ke tahap pemodelan dan analisis menggunakan metode yang dipilih, seperti *Decision Tree* atau *Naive Bayes*. Kondisi ini menunjukkan bahwa proses pengumpulan data telah dilakukan secara baik dan terstruktur, yang pada akhirnya berkontribusi pada ketepatan hasil klasifikasi yang dihasilkan.

Perancangan Model Klasifikasi

Perancangan model klasifikasi merupakan tahap di mana peneliti menentukan rancangan atau kerangka kerja algoritma yang akan digunakan untuk memisahkan data ke dalam kategori tertentu berdasarkan pola yang ada pada data pelatihan (*training set*). Tahap ini mencakup pemilihan metode klasifikasi yang sesuai, penentuan variabel atau fitur yang akan digunakan, serta pengaturan parameter model agar dapat bekerja secara optimal. Model klasifikasi dirancang sedemikian rupa untuk mengenali pola dari data *training* dan kemudian mampu memprediksi label atau kategori pada data baru yang belum pernah dilihat sebelumnya, yaitu data *testing*. Perancangan yang tepat akan memastikan bahwa model tidak hanya akurat pada data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik pada data uji, sehingga hasil analisis dapat dipercaya dan digunakan sebagai dasar pengambilan keputusan.



Gambar 1. Perancangan Model Klasifikasi

Pada gambar di atas merupakan rancangan model klasifikasi yang digunakan dalam penelitian, di mana prosesnya melibatkan beberapa node penting. Node File berfungsi untuk memanggil dan membaca dataset yang digunakan, kemudian diteruskan ke Data Table untuk menampilkan isi data secara lengkap sebelum diproses lebih lanjut. Selanjutnya, data tersebut dianalisis menggunakan dua metode klasifikasi, yaitu Decision Tree melalui node Tree dan Naive Bayes melalui node Naive Bayes, yang masing-masing

memiliki cara berbeda dalam membangun model prediksi. Setelah model terbentuk, hasilnya dihubungkan ke node Prediction yang berfungsi memprediksi kategori pada data uji (*testing*) dan mengevaluasi tingkat akurasi dari kedua metode tersebut. Susunan node ini dirancang secara sistematis agar proses klasifikasi berjalan optimal dan hasil analisis dapat dibandingkan dengan jelas.

Pelatihan dan Pengujian Model Klasifikasi

Pelatihan dan pengujian model klasifikasi merupakan tahap penting dalam memastikan bahwa model yang dibangun mampu melakukan prediksi secara akurat. Pada tahap pelatihan (*training*), data training digunakan untuk membentuk model berdasarkan pola dan hubungan antar variabel yang ada dalam dataset. Selanjutnya, pada tahap pengujian (*testing*), model yang telah dilatih diuji menggunakan data testing yang belum pernah dilihat oleh model sebelumnya, sehingga dapat mengukur sejauh mana model mampu menggeneralisasi dan memberikan hasil prediksi yang tepat. Proses ini memungkinkan peneliti untuk mengevaluasi performa model serta membandingkan efektivitas metode yang digunakan, seperti *Decision Tree* dan *Naive Bayes*, dalam mengklasifikasikan data secara optimal.

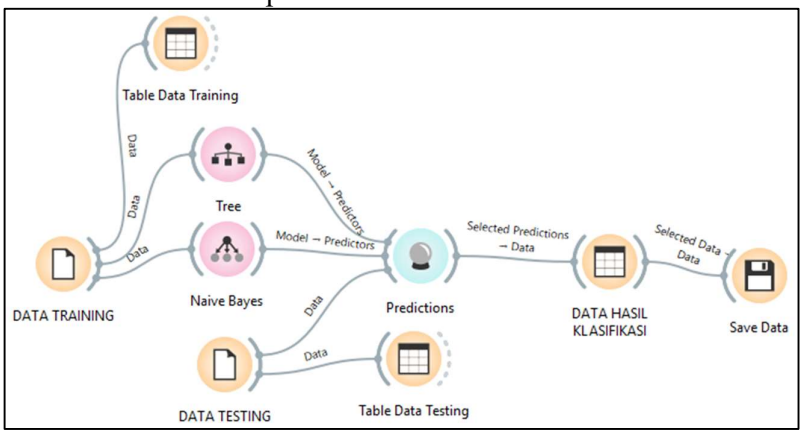


Figure 2. Model Pelatihan dan Pengujian

Pada gambar di atas terlihat rancangan alur kerja yang digunakan untuk proses pelatihan dan pengujian model klasifikasi. Alur tersebut diawali dengan node file yang berisi dua jenis dataset, yaitu data training dan data testing. Data training digunakan untuk membangun model dengan mempelajari pola dari data yang ada, sedangkan data testing digunakan untuk menguji sejauh mana model dapat melakukan prediksi terhadap data baru. Data dari node file ini kemudian dihubungkan ke beberapa node metode klasifikasi, seperti Decision Tree dan Naive Bayes, yang masing-masing memiliki algoritma berbeda dalam mengolah data dan menghasilkan model prediksi.

Metode Decision Tree berfungsi untuk membentuk pohon keputusan yang memisahkan data ke dalam kelas-kelas tertentu berdasarkan atribut yang paling berpengaruh, sementara Naive Bayes menggunakan pendekatan probabilitas untuk mengklasifikasikan data berdasarkan peluang terbesar dari suatu kategori. Dalam alur tersebut juga terdapat node prediction yang bertugas menggabungkan model dengan data testing untuk menghasilkan hasil prediksi. Rangkaian ini memungkinkan proses klasifikasi dilakukan secara sistematis, dimulai dari pembacaan data, pembentukan model, hingga menghasilkan prediksi akhir yang dapat dievaluasi tingkat akurasinya.

Setelah dilakukan pelatihan dat pengujian model, kemudian diperoleh hasil klasifikasi yang akan dipaparkan dalam bentu tabel yaitu sebagai berikut.

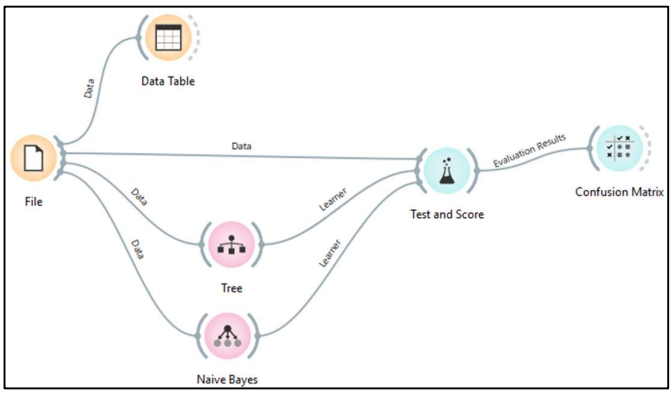
Hasil klasifikasi tingkat kepuasan responden terhadap layanan publik yang diukur berdasarkan tiga variabel, yaitu Pendidikan, Kesehatan, dan Infrastruktur Jalan. Dari total 60 responden, mayoritas atau sebanyak 56 responden dikategorikan puas, sedangkan hanya 4 responden yang termasuk dalam kategori tidak puas. Hal ini menunjukkan bahwa tingkat kepuasan masyarakat terhadap ketiga aspek tersebut secara umum sangat tinggi. Kategori "puas" didominasi oleh responden yang memberikan jawaban Setuju atau Sangat Setuju pada ketiga indikator, sedangkan kategori "tidak puas" lebih banyak muncul pada responden yang memberikan jawaban Netral atau Tidak Setuju pada salah satu atau lebih indikator, terutama pada aspek infrastruktur jalan.

Jika dilihat lebih detail, responden yang tergolong puas umumnya menunjukkan konsistensi jawaban positif pada semua indikator, misalnya kombinasi Setuju-Setuju-Setuju atau Sangat Setuju-Setuju-Setuju. Sementara itu, responden yang masuk kategori tidak puas memiliki pola penilaian yang cenderung menurun pada salah satu indikator, seperti Setuju-Setuju-Tidak Setuju atau Setuju-Netral-Sangat Setuju namun dengan persepsi keseluruhan yang negatif. Pola ini mengindikasikan bahwa infrastruktur jalan menjadi faktor yang paling berpengaruh terhadap penurunan tingkat kepuasan, karena sebagian responden yang puas pada pendidikan dan kesehatan tetap memberikan penilaian kurang baik pada aspek jalan, sehingga memengaruhi klasifikasi akhir.

Secara keseluruhan, hasil klasifikasi ini memperlihatkan bahwa layanan publik di bidang pendidikan dan kesehatan sudah cukup memuaskan bagi mayoritas masyarakat, sementara perbaikan masih diperlukan pada infrastruktur jalan untuk mengurangi ketidakpuasan. Nilai proporsi 93,33% puas (56 dari 60) menjadi indikator positif bahwa pelayanan yang diberikan telah sesuai harapan masyarakat, namun keberadaan 4 responden tidak puas menjadi masukan penting bagi pihak terkait. Analisis pola jawaban ini juga dapat menjadi dasar pengambilan kebijakan, di mana pemerintah dapat mempertahankan kualitas layanan yang sudah baik pada sektor pendidikan dan kesehatan, sekaligus fokus melakukan perbaikan signifikan pada infrastruktur jalan agar tingkat kepuasan masyarakat dapat meningkat secara merata di semua aspek.

Evaluasi Metode

Evaluasi model merupakan tahap akhir dalam proses klasifikasi yang bertujuan untuk mengukur kinerja model yang telah dibangun. Pada tahap ini, hasil prediksi yang dihasilkan oleh model dibandingkan dengan data sebenarnya pada dataset testing untuk mengetahui tingkat akurasi, presisi, recall, dan metrik evaluasi lainnya. Proses evaluasi ini penting agar dapat memastikan bahwa model tidak hanya bekerja baik pada data training, tetapi juga mampu memberikan prediksi yang akurat pada data baru. Dengan evaluasi yang tepat, peneliti dapat menentukan apakah model yang digunakan sudah optimal atau perlu dilakukan perbaikan dan penyesuaian lebih lanjut.



Gambar 3. Perancangan Model Evaluasi

Gambar model yang dirancang untuk melakukan evaluasi data pada penelitian ini menunjukkan alur kerja yang menghubungkan hasil pelatihan dan pengujian model dengan metode evaluasi yang digunakan. Pada model tersebut, terdapat dua komponen utama evaluasi yaitu Test and Score dan Confusion Matrix. Node Test and Score berfungsi untuk menghitung nilai akurasi, presisi, recall, F1-score, dan metrik evaluasi lainnya berdasarkan hasil prediksi model terhadap data testing. Sedangkan node Confusion Matrix memberikan gambaran visual berupa matriks yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas, sehingga memudahkan peneliti dalam menganalisis performa model secara detail. Kombinasi kedua metode evaluasi ini memungkinkan penilaian yang lebih menyeluruh terhadap efektivitas model yang digunakan.

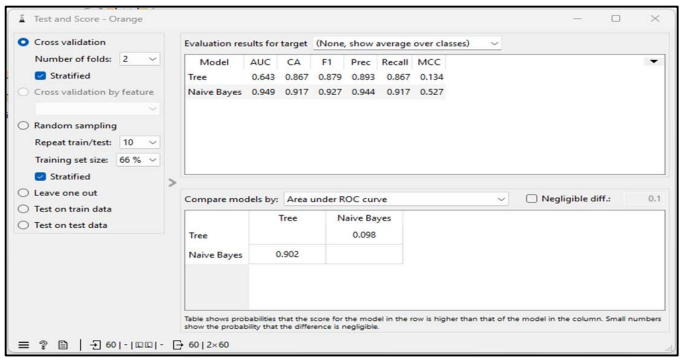
Melalui rancangan model evaluasi tersebut, peneliti dapat memastikan bahwa algoritma klasifikasi seperti Decision Tree dan Naive Bayes tidak hanya memiliki nilai akurasi yang tinggi, tetapi juga seimbang dalam mengklasifikasikan setiap kategori. Test and Score akan memberikan ringkasan metrik kuantitatif, sementara Confusion Matrix akan menyoroti pola kesalahan prediksi, misalnya jika model cenderung salah mengklasifikasikan kelas tertentu. Informasi ini sangat penting untuk mengambil keputusan apakah model sudah layak digunakan pada data baru atau memerlukan perbaikan, seperti penyesuaian parameter atau penambahan data training. Dengan demikian, evaluasi yang ditampilkan pada gambar tersebut menjadi langkah krusial untuk menjamin keandalan dan kualitas model klasifikasi yang dibangun.

Hasil Evaluasi

Berdasarkan hasil evaluasi menggunakan *node Test and Score*, model klasifikasi menunjukkan tingkat akurasi yang tinggi dalam membedakan kategori *puas* dan *tidak puas*. Sementara itu, evaluasi menggunakan Confusion Matrix memperlihatkan bahwa sebagian besar data berhasil diprediksi dengan benar sesuai kategori aslinya, dengan kesalahan prediksi yang sangat minimal.

Hasil Test and Score

Node Test and Score digunakan untuk mengukur kinerja model klasifikasi dengan menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Hasil dari node ini memberikan gambaran seberapa baik model mampu memprediksi data baru berdasarkan pola yang telah dipelajari dari data training.



Gambar 4. Evaluasi Model *Test and Score*

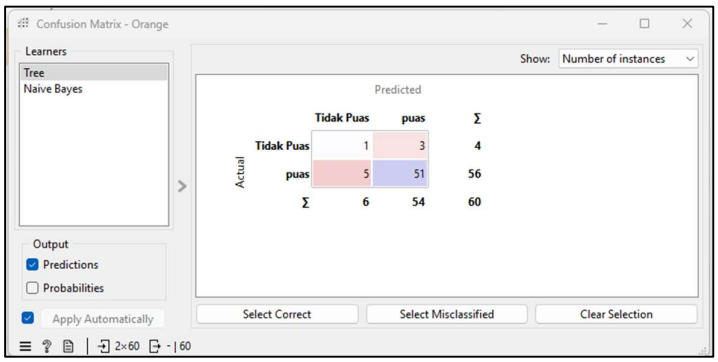
Pada gambar di atas merupakan hasil evaluasi model klasifikasi menggunakan node Test and Score di aplikasi Orange. Dua algoritma yang dibandingkan adalah *Decision Tree* (*Tree*) dan Naive Bayes, dengan metrik evaluasi yang ditampilkan meliputi AUC (Area Under Curve), CA (*Classification Accuracy*), F1 Score, Precision, Recall, dan MCC (*Matthews Correlation Coefficient*). Berdasarkan hasil yang diperoleh, Naive Bayes menunjukkan performa yang lebih tinggi dibandingkan Decision Tree pada semua metrik evaluasi. Misalnya, nilai AUC untuk Naive Bayes mencapai 0.949 sedangkan Decision Tree hanya 0.643, yang menunjukkan bahwa Naive Bayes memiliki kemampuan pemisahan kelas yang lebih baik. Begitu juga pada CA (akurasi), Naive Bayes mencapai 0.917 sementara Decision Tree hanya 0.867.

Dari metrik F1 Score, Naive Bayes mendapatkan nilai 0.927, sedikit lebih tinggi dibandingkan Decision Tree yang memperoleh 0.879, menunjukkan bahwa Naive Bayes mampu menyeimbangkan antara Precision dan Recall dengan lebih baik. Nilai Precision pada Naive Bayes (0.944) juga lebih unggul dibandingkan Decision Tree (0.893), yang berarti model ini lebih tepat dalam memprediksi kelas positif. Sementara itu, Recall untuk Naive Bayes mencapai 0.917, lebih tinggi dibandingkan Decision Tree dengan nilai 0.867, yang menandakan bahwa Naive Bayes lebih baik dalam mendeteksi seluruh data positif. Nilai MCC pada Naive Bayes (0.527) juga jauh lebih besar dibandingkan Decision Tree (0.134), mengindikasikan bahwa korelasi antara prediksi dan label sebenarnya lebih kuat pada model Naive Bayes.

Bagian bawah tabel menunjukkan hasil perbandingan model berdasarkan Area Under ROC Curve. Nilai perbandingan ini menunjukkan bahwa probabilitas model Naive Bayes memiliki AUC lebih tinggi dibandingkan Decision Tree adalah sebesar 0.902, sedangkan probabilitas Decision Tree mengungguli Naive Bayes hanya 0.098. Hasil ini mengonfirmasi bahwa Naive Bayes tidak hanya unggul pada sebagian besar metrik evaluasi, tetapi juga secara keseluruhan memiliki performa yang lebih konsisten dalam memisahkan kelas pada dataset yang digunakan. Dengan demikian, berdasarkan hasil evaluasi ini, Naive Bayes dapat dipertimbangkan sebagai model yang lebih optimal dibandingkan Decision Tree untuk data dan skenario klasifikasi yang sedang diteliti.

Hasil Confusion Matrix

Confusion Matrix adalah tabel yang digunakan untuk menggambarkan kinerja model klasifikasi dengan menampilkan jumlah prediksi benar dan salah pada setiap kelas. Melalui matriks ini, kita dapat melihat secara detail jumlah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), sehingga memudahkan dalam menghitung metrik evaluasi seperti akurasi, presisi, dan recall.

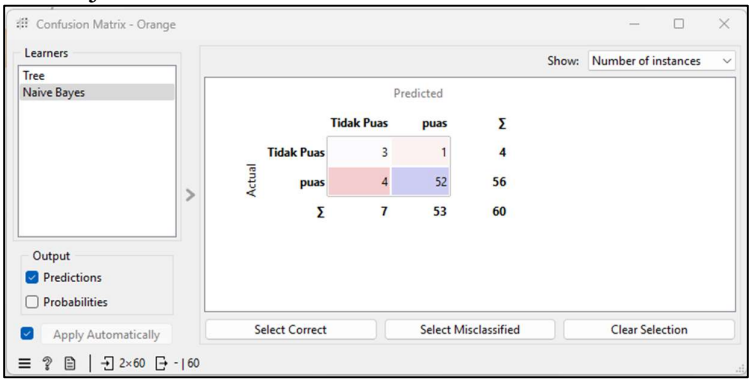


Gambar 5. *Confusion Matrix Metode Decision Tree*

Pada gambar di atas merupakan hasil evaluasi Confusion Matrix untuk metode Decision Tree yang digunakan dalam mengklasifikasikan data kepuasan pelanggan menjadi dua kategori, yaitu "Puas" dan "Tidak Puas". Dari total 60 data uji, model memprediksi dengan benar 1 data untuk kelas "Tidak Puas" (*True Negative*) dan 51 data untuk kelas "Puas" (*True Positive*). Sementara itu, kesalahan prediksi terjadi pada 3 data "Tidak Puas" yang diprediksi sebagai "Puas" (*False Positive*) dan 5 data "Puas" yang diprediksi sebagai "Tidak Puas" (*False Negative*). Hasil ini memberikan gambaran umum tentang kemampuan model dalam membedakan kedua kelas tersebut.

Berdasarkan hasil evaluasi, akurasi model mencapai 86,67%, yang berarti sebagian besar prediksi yang dihasilkan model sesuai dengan kondisi sebenarnya. Nilai presisi pada kelas "Puas" sebesar 94,44%, menunjukkan bahwa dari semua data yang diprediksi sebagai "Puas", sebagian besar benar-benar merupakan data "Puas". Hal ini menandakan bahwa model memiliki kemampuan yang sangat baik dalam meminimalkan kesalahan positif palsu pada kelas ini.

Nilai *Recall* untuk kelas "Puas" sebesar 91,07%, yang berarti model mampu mengidentifikasi mayoritas data "Puas" yang sebenarnya ada di dataset uji. Meskipun demikian, masih terdapat beberapa data yang terlewat dan diklasifikasikan ke kelas "Tidak Puas". Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa metode Decision Tree bekerja dengan baik, namun performanya dapat lebih ditingkatkan, terutama pada kemampuan model dalam mengenali kelas minoritas "Tidak Puas" yang cenderung memiliki jumlah data lebih sedikit.



Gambar 6. *Confusion Matrix Metode Naive Bayes*

Berdasarkan gambar *Confusion Matrix* di atas, metode Naive Bayes digunakan untuk mengklasifikasikan data kepuasan pelanggan ke dalam dua kategori, yaitu “Tidak Puas” dan “Puas”. Dari total 60 data uji, model berhasil memprediksi dengan benar 3 data

"Tidak Puas" (*True Negative*) dan 52 data "Puas" (True Positive). Kesalahan prediksi terjadi pada 1 data "Tidak Puas" yang diprediksi sebagai "Puas" (False Positive) dan 4 data "Puas" yang diprediksi sebagai "Tidak Puas" (False Negative). Distribusi hasil ini menunjukkan bahwa model mampu mengenali pola data dengan cukup baik, meskipun masih terdapat kesalahan pada kedua kelas.

Hasil evaluasi menunjukkan bahwa akurasi model mencapai 91,67%, yang berarti sebagian besar prediksi yang dihasilkan sesuai dengan kondisi aktual pada data uji. Nilai presisi untuk kelas "Puas" adalah 98,11%, yang menandakan bahwa hampir semua data yang diprediksi sebagai "Puas" benar-benar termasuk dalam kelas tersebut, sehingga kesalahan positif palsu pada kelas "Puas" sangat kecil. Sementara itu, presisi untuk kelas "Tidak Puas" adalah 42,86%, yang menunjukkan bahwa prediksi pada kelas minoritas ini masih cenderung rawan kesalahan karena jumlah datanya yang jauh lebih sedikit dibandingkan kelas mayoritas.

Dari sisi *Recall*, nilai untuk kelas "Puas" adalah 92,86%, yang berarti sebagian besar data "Puas" berhasil teridentifikasi oleh model, dengan hanya sedikit data yang terlewat. Namun, nilai Recall untuk kelas "Tidak Puas" adalah 75%, menunjukkan bahwa meskipun model cukup baik dalam mengenali sebagian besar data "Tidak Puas", masih ada peluang perbaikan agar tidak terjadi kesalahan prediksi yang menganggap data tersebut sebagai "Puas". Secara keseluruhan, performa Naive Bayes pada data ini tergolong sangat baik untuk kelas mayoritas, tetapi memerlukan optimasi lebih lanjut agar dapat meningkatkan kinerja pada kelas minoritas.

5. Kesimpulan

Berdasarkan hasil klasifikasi menggunakan metode ***Naive Bayes***, model mampu memprediksi tingkat kepuasan pelanggan dengan performa yang sangat baik. Dari total 60 data uji, sebanyak 55 data berhasil diklasifikasikan dengan benar, sedangkan 5 data lainnya salah klasifikasi. Mayoritas prediksi tepat berada pada kelas "Puas" dengan jumlah 52 data benar dan hanya 4 yang salah prediksi. Sementara itu, pada kelas "Tidak Puas" terdapat 3 data benar dan 1 data yang salah prediksi menjadi "Puas". Hasil ini menunjukkan bahwa Naive Bayes mampu mengidentifikasi pola mayoritas data dengan baik, meskipun masih terdapat sedikit kesalahan pada kelas minoritas.

Hasil evaluasi melalui *Confusion Matrix* menunjukkan bahwa akurasi model mencapai **91,67%**, yang berarti sebagian besar prediksi sesuai dengan data aktual. Presisi pada kelas "Puas" mencapai **98,11%**, menunjukkan bahwa hampir semua prediksi "Puas" memang benar-benar berasal dari pelanggan yang puas. Sementara itu, *recall* pada kelas "Puas" sebesar **92,86%** memperlihatkan kemampuan model dalam menangkap sebagian besar data yang benar-benar puas dari seluruh data aktual pada kelas tersebut. Dengan nilai evaluasi yang tinggi ini, dapat disimpulkan bahwa metode Naive Bayes sangat layak digunakan untuk memprediksi tingkat kepuasan pelanggan, terutama jika fokusnya adalah pada identifikasi pelanggan yang puas dengan tingkat kesalahan yang sangat rendah.

6. Dafta Pustaka

Alam, A., Alana, D. A. F., & Juliane, C. (2023). Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes. *Sinkron*, *8*(4), 2641–2650. https://doi.org/10.33395/sinkron.v8i4.12998

Atalya Angelus Leza, M., Widya Utami, N., & Anugrah Cahya Dewi, P. (2024). Prediksi

Prestasi Siswa Smas Katolik Santo Yoseph Denpasar Berdasarkan Kedisiplinan Dan Tingkat Ekonomi Orang Tua Menggunakan Metode Knowledge Discovery in Database Dan Algoritma Regresi Linier Berganda. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 373–379. https://doi.org/10.36040/jati.v8i1.8754

Aulia, R., Putri, I., Pudjiantoro, T. H., Informatika, T., Informatika, S., Jenderal, U., & Yani, A. (2020). *Prediksi Perguruan Tinggi Negeri dengan Menggunakan Metode Naive Bayes*. 106–111.

Barbosa, R. S., de Souza, Z. M., Carneiro, M. P., & Farhate, C. V. V. (2021). Root system and its relations with soil physical and chemical attributes in orange culture. *Applied Sciences (Switzerland)*, *11*(4), 1–14. https://doi.org/10.3390/app11041790

Farid Naufal, M., Fernando Susanto, A., Nathaneil Kansil, C., Huda, S., & kunci, K. (2023). Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Potensi Hilangnya Nasabah Bank Application of Machine Learning to Predict Potential Loss of Bank Customer. *Februari*, *22*(1), 1–11.

Hafizan, H., & Putri, A. N. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, *1*(2), 68–72. https://doi.org/10.30645/kesatria.v1i2.23

Madjid, F. M., Ratnawati, D. E., & Rahayudi, B. (2023). Sentiment Analysis on App Reviews Using Support Vector Machine and Naïve Bayes Classification. *Jurnal Dan Penelitian Teknik Informatika*, *8*(1), 556–562. Retrieved from https://doi.org/10.33395/sinkron.v8i1.12161

Mirbod, O., Choi, D., Heinemann, P. H., Marini, R. P., & He, L. (2023). On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling. *Biosystems Engineering*, *226*, 27–42. https://doi.org/10.1016/j.biosystemseng.2022.12.008

Mohammed, S., Elbeltagi, A., Bashir, B., Alsafadi, K., Alsilibe, F., Alsalman, A., … Harsányi, E. (2022). A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean. *Computers and Electronics in Agriculture*, *197*(March). https://doi.org/10.1016/j.compag.2022.106925

Naufal, M. F., Arifin, T., & Wirjawan, H. (2023). Analisis Perbandingan Tingkat Performa Algoritma SVM, Random Forest, dan Naïve Bayes untuk Klasifikasi Cyberbullying pada Media Sosial. *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK)*, *8*, 82. Retrieved from https://tunasbangsa.ac.id/ejurnal/index.php/jurasik

Ninditama, I. P., Ninditama, I. P., Cholil, W., Akbar, M., & Antoni, D. (2020). *Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang*. *15*(2), 37–49.

Nuraeni, S., Harliana, H., & Prabowo, T. (2024). Analisis Akurasi Naïve Bayes Dan Knn Dalam Penentuan Penerima Pkh Di Lombok Utara. *Journal of Information System Management (JOISM)*, *5*(2), 121–126. https://doi.org/10.24076/joism.2024v5i2.1205

Pratama, H. A., Yanris, G. J., Nirmala, M., & Hasibuan, S. (2023). *Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels*. *8*(3), 1832–1851.

Setiawan, A., Rabi, A., & Gumilang, Y. S. A. (2024). Pengolahan Citra untuk Sortir Buah

Stroberi Berdasarkan Kematangan Menggunakan Algoritma K-Nearst Neighbors (KNN). *Blend Sains Jurnal Teknik*, *2*(4), 322–328. https://doi.org/10.56211/blendsains.v2i4.551

Siahaan, T., Laia, Y., Silitonga, M., Pasaribu, C., Sains, F., & Teknologi, D. (2023). Penerapan Data Mining Classification Untuk Data Pasien Covid-19 Menggunakan Metode Naïve Bayes. *AJurnal TEKINKOM*, *6*(1), 245–250. https://doi.org/10.37600/tekinkom.v6i1.879

Simanjuntak, A. Y., Simatupang, I. S. S., & Anita. (2022). Implementasi Data Mining Menggunakan Metode Naïve Bayes Classifier Untuk Data Kenaikan Pangkat Dinas. *Journal of Science and Social Research*, *4307*(1), 85–91.

Sulaiman, A. P., Liliana, L., & Santoso, L. W. (2021). Kecerdasan Buatan dengan Metode ID3 Finite State Machine dalam Turn-Based Tactics Game. *Jurnal Infra*, (031). Retrieved from http://publication.petra.ac.id/index.php/teknik-informatika/article/viewFile/11421/10031

Zai, F., Sirait, J., Nainggolan, D. W., Sihombing, N. G. D., & Banjarnahor, J. (2023). Comparison Analysis of C4.5 Algorithm and KNN Algorithm for Predicting Data of Non-Active Students at Prima Indonesia University. *Sinkron*, *8*(4), 2027–2035. https://doi.org/10.33395/sinkron.v8i4.12879