**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

# NAIVE BAYES DAN DECISION TREE: STUDI KASUS KLASIFIKASI KEPUASAN PELANGGAN E-COMMERCE

**Ofirmince Tulak Bamba[1], Nur Vadila[2], Sri Fitrawati[3], Vilna Wati Tedang[4], Asrawati[5]**
[1]ofirmtb@gmail.com, [2]vadilanur27@gmail.com, [3]fitrawati.1910@gmail.com,
[4]vilnawatitedang28@gmail.com, [5]asrawati026@gmail.com
[1,2,3,4,5]Fakultas Teknik, Program Studi Teknik Informatika, Universitas Papua

**Abstrak**

Peningkatan belanja daring mendorong e-commerce untuk memahami kepuasan pelanggan melalui analisis ulasan otomatis. Studi ini mengevaluasi dan membandingkan kemampuan algoritma *Naive Bayes* dan *Decision Tree* dalam mengklasifikasikan tingkat kepuasan berdasarkan 5.000 ulasan dari platform Olist. Ulasan dikategorikan ke dalam tiga kelas, yaitu Tidak Puas, Netral, dan Puas. Pra-pemrosesan meliputi pembersihan data, ekstraksi fitur dengan TF-IDF, dan pembagian data 80% latih dan 20% uji. Evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan AUC. *Naive Bayes* menunjukkan akurasi lebih tinggi (80,70%) dibanding *Decision Tree* (73,90%) serta performa klasifikasi yang lebih stabil. Dengan demikian, *Naive Bayes* lebih efisien untuk klasifikasi kepuasan pelanggan berbasis teks pada ulasan e-commerce.

**Kata kunci:** *Data Mining*, Klasifikasi, *Naive Bayes, Decision Tree*, E-Commerce

***Abstract***

*The rise in online shopping compels e-commerce platforms to understand customer satisfaction through automated review analysis. This study evaluates and compares the performance of the Naive Bayes and Decision Tree algorithms in classifying customer satisfaction levels based on 5,000 reviews from the Olist platform. The reviews are categorized into three classes: Dissatisfied, Neutral, and Satisfied. Preprocessing involves data cleaning, feature extraction using TF-IDF, and an 80/20 split for training and testing datasets. Evaluation metrics include accuracy, precision, recall, F1-score, and AUC. Naive Bayes achieves higher accuracy (80.70%) compared to Decision Tree (73.90%) and demonstrates more consistent classification performance. Thus, Naive Bayes is more efficient for text-based customer satisfaction classification in e-commerce reviews.*

***Keywords:*** *Data Mining, Classification, Naive Bayes, Decision Tree, E-Commerce*

## 1. Pendahuluan

Peningkatan penggunaan perangkat digital mendorong pelaku bisnis untuk lebih responsif terhadap tingkat kepuasan pelanggan guna mempertahankan loyalitas konsumen dan meningkatkan daya saing. Data ulasan dan respons pelanggan menjadi sumber informasi yang sangat penting dalam mengukur serta memprediksi tingkat kepuasan pelanggan. Menurut [1], kepuasan pelanggan memainkan peran penting dalam menjamin keberlangsungan suatu bisnis, karena pelanggan yang puas cenderung menunjukkan loyalitas yang tinggi dan berkontribusi terhadap stabilitas serta pertumbuhan bisnis dalam jangka panjang.

Salah satu platform yang dianalisis dalam penelitian terkait kepuasan pelanggan berbasis data adalah Olist, sebuah *e-commerce* asal Brazil yang berperan sebagai penghubung antara penjual dan berbagai *marketplace* besar seperti *Mercado Livre*, *Americanas*, dan *platform* lainnya. Olist memfasilitasi penjual toko untuk memasarkan produk mereka secara online melalui sistem yang terintegrasi, sekaligus menghimpun ulasan serta penilaian dari pelanggan terhadap produk yang telah dibeli.

Penelitian yang dilakukan oleh [2] menggunakan algoritma C4.5, yang juga dikenal sebagai algoritma *decision tree*, untuk menganalisis tingkat kepuasan pelanggan pada aplikasi TikTok Shop. Hasil penelitian menunjukkan bahwa tingkat kepuasan pelanggan berhasil diprediksi dengan akurasi sebesar

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

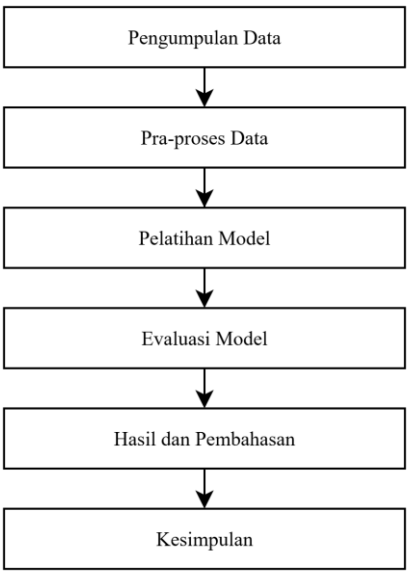
87,27%. Penelitian lain yang ditulis oleh [3] berjudul Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma *Decision Tree*, *Naive Bayes* dan *K-Nearest Neighbor*, ditemukan bahwa jalur yang ditemukan pada *Decision Tree* dapat digunakan untuk mengklasifikasikan data baru. Hasil penelitian menunjukkan bahwa algoritma *Decision Tree* memiliki tingkat akurasi 92,31% dan nilai AUC yang mencapai 93,8%.

Studi yang ditulis oleh [4] dengan judul "Analisis Sentimen pada Ulasan Produk di E-Commerce Menggunakan Metode Naive Bayes", menemukan bahwa algoritma *Naive Bayes* dapat secara otomatis memprediksi perasaan yang diberikan oleh ulasan produk, dengan nilai akurasi yang cukup memuaskan. Studi tambahan yang dilakukan oleh [5] dengan judul *"Analisis Klasifikasi Ulasan Sentimen pada E-Commerce Shopee Menggunakan Word Cloud berbasis Naive Bayes dan K-Nearest Neighbor"* terbukti memberikan hasil yang lebih baik berdasarkan evaluasi *akurasi*, *precision*, *recall*, dan *F1-score*. Sebaliknya, *Naive Bayes* menunjukkan keunggukan pada aspek *confusion matrix* dalam memprediksi ulang sentimen negatif, tetapi tidak dalam memprediksi sentimen ulasan positif.

Berdasarkan studi terdahulu menunjukkan bahwa algoritma *Decision Tree* dan *Naive Bayes* keduanya memiliki keunggulan dalam tugas klasifikasi. Karena kedua algoritma ini digunakan untuk menangani dataset '*order review'* Olist. Penelitian tambahan diperlukan untuk membandingkan kedua algoritma ini dalam konteks yang lebih khusus. Penelitian ini bertujuan membandingkan kinerja algoritma *Naive Bayes* dan *Decision* Tree dalam mengklasifikasikan tingkat kepuasan pelanggan pada platform e-commerce Olist asal Brasil. Proses pengklasifikasian dilakukan menggunakan algoritma ini dengan metrik evaluasi yang meliputi *accuracy, precision, recall,* dan *F1-score*.

## 2. Metode Penelitian

Studi ini menerapkan pendekatan kuantitatif komparatif dengan tujuan membandingkan performa dua algoritma klasifikasi, yaitu *Naive Bayes* dan *Decision Tree* dalam mengklasifikasikan tingkat kepuasan pelanggan berdasarkan metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score*. Rangkaian tahapan penelitian disajikan pada Gambar 1.




Gambar 1. Tahapan Penelitian

### 2.1. Pengumpulan Data

Data dalam studi ini merupakan data sekunder yang bersumber dari e-commerce Olist dan diakses melalui platform Kaggle, Olist sendiri merupakan sebuah platform *e-commerce* asal Brazil. Dari 99.224 entri data, sebanyak 5.000 entri dianalisis dan telah diklasifikasikan dalam tiga kelas: skor 1 dan 2 sebagai pelanggan tidak puas, skor 3 sebagai pelanggan netral, skor 4 dan 5 sebagai pelanggan puas.

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

## 2.2. Pra-proses Data

Untuk meningkatkan kualitas data, proses pra-pemrosesan menjadi langkah awal yang esensial dalam pengelolaan data dan kebersihan data, sehingga hasil evaluasi menjadi lebih optimal. Proses pembersihan data dilakukan melalui beberapa tahapan berikut.

### 2.2.1. Inisialisasi dan *Import Library*

Langkah pertama pra-proses data dilakukan dengan menginisialisasi *Jupyter Notebook* menggunakan *%matplotlib inline*, sehingga visualisasi grafik dapat langsung ditampilkan. Selanjutnya, dilakukan proses *import library* penting seperti *pandas, numpy*, *Matplotlib* dan *Seaborn* digunakan sebagai alat untuk manipulasi dan visualisasi data. *Pandas* berfungsi dalam pengelolaan dan menganalisis data terstruktur dalam bentuk *Data Frame*, sementara *numpy* menyediakan dukungan untuk komputasi *numerik* dan manipulasi *array* multidimensi. Untuk visualisasi data, *matplotlib* digunakan dalam pembuatan grafik dasar seperti garis dan batang, sedangkan *seaborn* mempermudah pembuatan visualisasi statistik yang lebih informatif dan estetis.

### 2.2.2. Pemuatan dan Eksplorasi Dataset

Dalam penelitian ini, *dataset Order Reviews* digunakan sebagai sumber data, yang berisi ulasan pelanggan pada platform e-commerce. Untuk memahami struktur dan isi awal dataset, lima baris pertama data ditampilkan, dengan fokus utama pada kolom *review_score* dan *review_comment_message* yang menjadi objek analisis.

### 2.2.3. Pemeriksaan dan Pembersihan Data

Data diperiksa terlebih dahulu untuk mengidentifikasi adanya data duplikat serta komentar kosong. Komentar kosong didefinisikan sebagai entri yang bernilai *NaN* atau hanya berisi spasi kosong. Seluruh data duplikat kemudian dihapus menggunakan *drop_duplicates()*, dan hanya dua kolom yang dipertahankan, yaitu *review_score* dan *review_comment_message*. Selanjutnya, semua baris yang mengandung komentar kosong juga dihapus dari dataset. Proses ini dilakukan untuk memastikan data yang akan dianalisis dalam kondisi bersih dan relevan, sehingga dapat meningkatkan akurasi dan validitas model klasifikasi yang dibangun.

### 2.2.4. Pengambilan Sampel Data

Dari data yang telah dibersihkan, diambil sampel acak sebanyak 5000 baris menggunakan fungsi *sample()*. Langkah ini bertujuan untuk mempercepat proses pelatihan model serta mempermudah pengolahan data dalam skala menengah, tanpa mengorbankan representasi distribusi data asli.

### 2.2.5. Kategorisasi Sentimen

Kategorisasi sentimen ini dilakukan dengan menerapkan fungsi *apply()* pada kolom *review_score*. Skor ulasan pelanggan kemudian diklasifikasikan ke dalam tiga kelas sentimen, yaitu: skor 1 dan 2 dikategorikan sebagai tidak puas, skor 3 sebagai netral, skor 4 dan 5 sebagai puas.

### 2.2.6. Ekstraksi Fitur Teks (TF-IDF)

Dalam proses ini, teks ulasan pelanggan direpresentasikan secara numerik dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF *Vectorizer* dari *library scikit-learn* digunakan untuk menghasilkan vektor yang menunjukkan tingkat relevansi setiap kata dalam kumpulan dokumen.

### 2.2.7. Pembagian Data

Data yang telah dikonversi kemudian dipisahkan menjadi data latih dan data uji dengan rasio 80:20 menggunakan fungsi *train_test_split()* dari pustaka *scikit-learn*. Model klasifikasi dilatih menggunakan data latih, sedangkan data uji dimanfaatkan untuk menilai kemampuan model dalam mengklasifikasikan data yang tidak termasuk dalam proses pelatihan.

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

### 2.3. Pelatihan Model

Dalam penelitian ini, diterapkan dua model klasifikasi, yaitu *Naive Bayes* dan *Decision Tree*. *Naive Bayes* adalah metode klasifikasi yang didasarkan pada asumsi bahwa setiap atribut bersifat saling bebas secara kondisional jika nilai kelas (*output*) diketahui, sehingga probabilitas gabungan dapat dihitung dari perkalian probailitas masing-masing atribut [6]. Algoritma *Naive Bayes* dipilih karena karakteristiknya yang sesuai untuk data teks serta kemampuannya dalam memberikan performa yang baik pada tugas klasifikasi dokumen. Metode *Decision Tree* mengklasifikasikan data melalui struktur pohon, di mana setiap *node* menggambarkan sebuah atribut, cabang menunjukkan nilai atribut tersebut, dan daun menandakan kelas hasil klasifikasi [7]. *Decision Tree* digunakan untuk membangun struktur pohon keputusan berdasarkan fitur yang tersedia, dengan tujuan menghasilkan aturan klasifikasi yang mudah diinterpresentasikan.

### 2.4. Evaluasi Model

Kinerja dari kedua model dievaluasi menggunakan berbagai metrik evaluasi. Metrik *accuracy* digunakan untuk mengukur proporsi prediksi yang benar secara keseluruhan. Selain itu, metrik *precision*, *recall*, dan F1-score diterapkan untuk mengevaluasi performa model dalam konteks klasifikasi multikelas, terutama dalam menilai keseimbangan antara prediksi positif yang benar dan kesalahan klasifikasi. *Confusion Matrix* digunakan untuk mengidentifikasi distribusi prediksi yang benar dan salah pada setiap kelas. Sementara itu, *Receiver Operating Characteristic* (ROC) *dan Area Under the Curve* (AUC) digunakan sebagai metrik untuk menilai kemampuan model dalam membedakan antar kelas.

## 3. Hasil dan Pembahasan

### 3.1. Hasil Tahap Pra-proses Data

#### 3.1.1. Distribusi Frekuensi Skor dan Kelas

Tabel 1 menyajikan distribusi frekuensi skor ulasan (review_skor) yang diberikan oleh pelanggan terhadap suatu produk atau layanan dalam Skala *likert* dengan rentang nilai 1 sampai 5, dimana skor 1 merepresentasikan tingkat ketidakpuasan yang sangat tinggi, sementara skor 5 menunjukkan tingkat kepuasan yang sangat tinggi.

Tabel 1. Distribusi Frekuensi Skor dan Kelas

| review_score | Skor Review (Awal) |
|---|---|
| 1 | 11424 |
| 2 | 3151 |
| 3 | 8179 |
| 4 | 19142 |
| 5 | 57328 |

Berdasarkan Tabel 1, skor ulasan dengan frekuensi tertinggi adalah skor 5, yaitu sebesar 57.328 ulasan atau sekitar 55% dari total keseluruhan. Skor 4 menempati urutan kedua dengan jumlah 19.142 ulasan. Jika digabungkan, skor 4 dan 5 mencakup lebih dari 70% total ulasan, yang mengindikasikan tingkat kepuasan pengguna yang relatif tinggi terhadap produk atau layanan yang ditinjau.

Sebaliknya, skor 2 menyumbang sebesar 3151 ulasan. Dan skor 1 dengan frekuensi ketidakpuasan sangat tinggi menyumbang 11424 ulasan. Ini menunjukkan bahwa hanya sebagian kecil pengguna yang memberikan penilaian negatif. Sedangkan skor 3 merupakan frekuensi netral yang menyumbang 8179 ulasan.

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

Tingginya frekuensi skor 4 dan 5 menunjukkan bahwa mayoritas pengguna memiliki persepsi yang positif terhadap produk atau layanan yang dievaluasi. Temuan ini dapat diinterpretasikan sebagai indikasi bahwa produk atau layanan tersebut mampu memenuhi, bahkan melampaui, ekspektasi pengguna. Sebaliknya, rendahnya frekuensi skor 1 dan 2 mencerminkan tingkat ketidakpuasan yang relatif rendah.

### 3.1.2. Pembersihan Data

Tabel 2 menyajikan tahapan praproses data yang mencakup kegiatan pembersihan dan seleksi data ulasan sebelum dilakukan analisis lebih lanjut.

Tabel 2. Dataset

| Uraian | Jumlah |
|---|---|
| Jumlah data awal | 99224 |
| Duplikat | 0 |
| Komentar kosong/null | 58274 |
| Jumlah data setelah bersih | 5000 |

Berdasarkan Tabel 2, jumlah data ulasan awal yang berhasil dikumpulkan adalah sebanyak 99.224 entri. Dari jumlah tersebut, tidak ditemukan data duplikat, sehingga tidak diperlukan tindakan penghapusan entri ganda. Namun demikian, terdapat sejumlah besar entri berupa komentar kosong atau bernilai *null*, yaitu sebanyak 58.274 entri, yang kemudian dieliminasi dari dataset karena tidak memuat informasi yang dapat dianalisis.

Setelah proses pembersihan dilakukan, khususnya melalui penghapusan komentar kosong atau *null*, jumlah data yang dinyatakan valid dan layak untuk dianalisis menyusut menjadi 5.000 entri. Tujuan dari proses ini adalah memastikan bahwa hanya data yang signifikan dan bermakna yang digunakan dalam analisis lanjutan.

Langkah pembersihan data merupakan bagian krusial dalam memastikan kualitas hasil analisis. Penghapusan entri yang tidak lengkap atau tidak relevan membantu meningkatkan *validitas*, *reliabilitas*, serta akurasi interpretasi terhadap ulasan pengguna, sehingga kesimpulan yang dihasilkan dapat lebih dipercaya.

### 3.1.3. Klasifikasi Data

Setelah melalui proses pembersihan, data ulasan kemudian diklasifikasikan ke dalam tiga kategori sentimen utama, yaitu *tidak puas*, *netral*, dan *puas*, sebagaimana ditampilkan pada Tabel 3.

Tabel 3. Statement

| Statement | Jumlah |
|---|---|
| tidak_puas | 1348 |
| netral | 442 |
| puas | 3209 |

Berdasarkan Tabel 3, kelas *puas* merupakan kategori dengan jumlah tertinggi, yaitu sebanyak 3.209 entri. Hal ini mencerminkan bahwa mayoritas pengguna memberikan ulasan dengan nada positif terhadap produk atau layanan yang ditinjau. Kelas *tidak puas* menempati urutan kedua dengan jumlah 1.348 entri, yang mengindikasikan adanya sebagian pengguna yang menyampaikan ketidakpuasan. Sementara itu, kelas *netral* memiliki jumlah paling sedikit, yakni hanya 442 entri, yang mempresentasikan ulasan tanpa kecenderungan emosional yang kuat, baik ke arah positif maupun negatif.

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

### 3.2. Hasil Klasifikasi

Tabel 4 menyajikan evaluasi performa model klasifikasi terhadap tiga kategori kepuasan, yaitu Tidak Puas, Netral, dan Puas, berdasarkan metrik *precision*, *recall*, *F1-score*, serta jumlah data (*support*) pada masing-masing kelas. Jumlah kemunculan kelas yang sebenarnya dalam dataset disebut sebagai *support*. Informasi ini sangat penting untuk memahami distribusi kelas secara keseluruhan, terutama dalam dataset yang tidak seimbang di mana ketidakseimbangan jumlah kelas dapat memengaruhi kinerja model klasifikasi [8]. Model yang diajukan berhasil mencapai *precision* sebesar 0.85, *recall* sebesar 0.80, dan *F1-score* sebesar 0.825, yang mengindikasikan kinerja model yang seimbang pada ketiga metrik evaluasi tersebut [9]. Secara keseluruhan, model mencapai tingkat akurasi sebesar 73.90% dari total 1000 data yang diuji. Meskipun demikian, performa model menunjukkan variasi yang cukup signifikan di antara ketiga kelas tersebut.

Tabel 4. Hasil Klasifikasi Algoritma *Decision Tree*

| Kelas | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Tidak Puas | 0.075472 | 0.045455 | 0.056738 | 88.0 |
| Netral | 0.833085 | 0.870717 | 0.851485 | 642.0 |
| Puas | 0.637681 | 0.651852 | 0.644689 | 270.0 |
| Akurasi | | | 0.7390 | 1000.0 |

Berdasarkan Tabel 4, pada kelas Tidak Puas, model memperlihatkan kinerja yang sangat rendah dengan tingkat *precision* sebesar, *recall*, dan *F1*-score yang diperoleh masing-masing adalah 0,075, 0,045, dan 0,057. Temuan ini mengindikasikan bahwa model jarang mengklasifikasikan data ke dalam kelas Tidak Puas, dan ketika klasifikasi tersebut dilakukukan, tingkat ketepatannya sangat rendah. Rendahnya nilai *recall* juga menunjukkan bahwa model gagal mengenali sebagian besar data yang sebenarnya termasuk dalam kategori ini, yang kemungkinan besar disebabkan oleh ketidakseimbangan jumlah data antar kelas.

Sebaliknya, pada kelas Puas, model menunjukkan *precision* 0.0638, *recall* 0.652, serta *F1-score* 0.645 dalam evaluasi performanya. Meskipun performanya tidak sebaik pada kelas Netral, ketiga metrik tersebut menunjukkan kinerja yang cukup baik dan relatif seimbang antara kemampuan model dalam mengenali data dan menghasilkan prediksi yang akurat untuk kelas ini.

Tabel 5 menyajikan hasil evaluasi performa model klasifikasi terhadap tiga kategori kepuasan, yaitu Tidak Puas, Netral, dan Puas, menggunakan metrik *precision*, *recall*, dan *F1-score*, serta jumlah data (*support*) pada masing-masing kelas. Secara keseluruhan, model mencapai tingkat akurasi sebesar 80.70% dari total 1000 data yang diuji. Meskipun akurasi tergolong tinggi, terdapat ketimpangan performa yang cukup signifikan di antara kelas yang diuji.

Tabel 5. Hasil Klasifikasi Algoritma *Naive Bayes*

| Kelas | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| Tidak Puas | 0.0 | 0.0 | 0.0 | 88.0 |
| Netral | 0.810209 | 0.964174 | 0.880512 | 642.0 |
| Puas | 0.79661 | 0.696296 | 0.743083 | 270.0 |
| Akurasi | | | 0.8070 | 1000.0 |

Berdasarkan Tabel 5, pada kelas Tidak Puas, model sama sekali tidak mampu mengenali atau memprediksi data yang termasuk dalam kategori ini. Hal tersebut tercermin dari nilai *precision*, *recall*, dan *F1-score* yang semuanya bernilai 0.0. Artinya, model tidak mengklasifikasikan satu pun data ke

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

dalam kelas Tidak Puas, sehingga tidak terdapat prediksi yang benar untuk kategori tersebut. Kegagalan ini sangat mungkin disebabkan oleh distribusi data yang tidak seimbang, dimana jumlah data dalam kelas Tidak Puas (88 data) jauh lebih sedikit dibandingkan kelas lainnya, sehingga model cenderung mengabaikannya selama proses pelatihan.

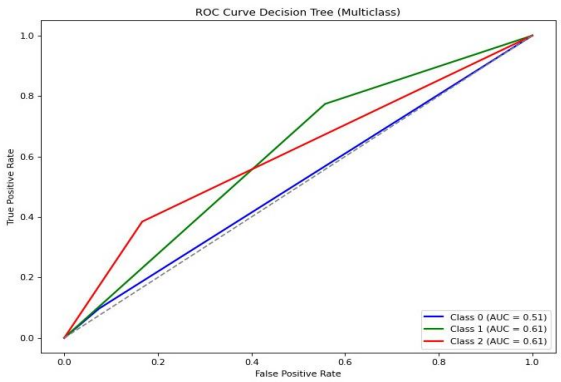
Sebaliknya, performa model pada kelas Netral sangat baik, yang memperoleh nilai *precision* 0.810, *recall* 0.964, dan *F1-score* 0.881. Nilai *recall* yang tinggi menunjukkan bahwa hampir seluruh data yang sebenarnya termasuk dalam kelas Netral berhasil dikenali oleh model, sedangkan nilai *precision* yang tinggi mengindikasikan bahwa sebagian besar prediksi untuk kelas ini benar. Kinerja yang baik ini didukung oleh jumlah data yang besar dalam kategori ini, yaitu sebanyak 642 data.

Pada kelas Puas, model menunjukkan performa yang juga baik, meskipun tidak setinggi kelas Netral. Dengan *precision* 0.787 dan *recall* 0.696, model mencapai *F1-score* sebesar 0.743. Temuan ini mengindikasikan bahwa model mampu mengidentifikasi dan memprediksi data dengan tingkat efektivitas yang memadai. Secara keseluruhan, hasil ini mengindikasikan bahwa meskipun model bekerja sangat baik pada dua kelas utama, yaitu Netral dan Puas, model masih mengalami kegagalan dalam menangani kelas minoritas seperti Tidak Puas.

Tabel 4 dan Tabel 5 masing-masing menyajikan hasil evaluasi performa algoritma *Decision Tree* dan *Naive Bayes* untuk mengklasifikasikan data kepuasan pelanggan. Berdasarkan evaluasi yang dilakukan, algoritma *Naive Bayes* berhasil meraih akurasi sebesar. 0.807, lebih tinggi dibandingkan dengan algoritma *Decision Tree* yang mencatatkan akurasi sebesar 0.739. Perbedaan tersebut mengindikasikan bahwa algoritma *Naive Bayes* memperlihatkan performa yang lebih baik secara keseluruhan dalam mengklasifikasikan data kepuasan pelanggan dibandingkan *Decision Tree*.

### 3.3. Analisis ROC dan AUC

Gambar 2 memvisualisasikan kurva ROC (*Receiver Operating Characteristic*) dari model klasifikasi *Decision Tree* dalam skenario *multiclass classification*, yang berfungsi untuk menilai kemampuan model dalam membedakan tiga kelas berbeda yang berbeda, *Class* 0, *Class* 1, dan *Class* 2. Dalam penelitian ini, *Area Under the Curve* (AUC) berperan dalam menilai sejauh mana model mampu membedakan kelas positif dan negatif. Nilai AUC diklasifikasikan ke dalam kategori sangat baik (0.9–1), baik (0.8–0.9), cukup (0.7–0.8), buruk (0.6–0.7), serta gagal (0.5–0.6) [10] .



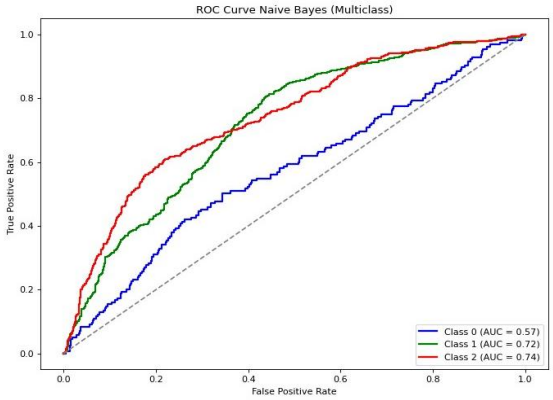

Gambar 2. Kurva ROC *Decision Tree*

Data yang ditampilkan pada Gambar 2 menunjukkan bahwa performa model dalam membedakan antar kelas berada pada tingkat yang relatif rendah. Untuk *Class* 0, model menghasilkan nilai AUC sebesar 0.51 yang mendekati garis diagonal atau *baseline* (AUC = 0.5), sehingga mengindikasikan bahwa model hampir tidak memiliki kemampuan untuk membedakan kelas ini dari yang

**Jurnal Sistem Informasi dan Sistem Komputer,** Vol. 10, No. 2, Juli 2025
**ISSN:** 2715-906X (Online)
10.51717/simkom.v10i2.897

kelas lainnya. Sementara itu, *Class* 1 dan *Class* 2 masing-masing memiliki nilai AUC sebesar 0.61, yang menunjukkan performa yang sedikit lebih baik, namun tetap tergolong lemah dalam konteks klasifikasi. Luas di bawah kurva ROC yang berkisar antara 0.5 dan 1.0 menunjukkan kemampuan model untuk membedakan antar subjek yang mengalami kejadian yang menjadi fokus penelitian dengan subjek yang tidak mengalaminya [11].

Gambar 3 menampilkan kurva ROC (*Receiver Operating Characteristic*) yang merepresentasikan model klasifikasi. *Naive Bayes* dalam skenario *multiclass classification*, yang digunakan untuk mengevaluasi kemampuan model dalam membedakan antar tiga kelas yang berbeda, Class 0, Class 1, dan Class 2.




Gambar 3. Kurva ROC *Naive Bayes*

Visualisasi pada Gambar 3 memperlihatkan bahwa performa model bervariasi antar kelas. Untuk *Class* 0, model menghasilkan nilai AUC sebesar 0.57, yang mengindikasikan bahwa kemampuan model dalam membedakan kelas ini dari yang lain sangat terbatas dan mendekati performa acak. Hal ini menunjukkan bahwa model mengalami kesulitan dalam mengenali karakteristik secara khusus yang membedakan *Class* 0 dari kelas lainnya. Sebaliknya, *Class* 1 menampilkan performa yang lebih optimal dengan skor AUC 0.72, yang menunjukkan kemampuan model yang cukup baik dalam mengklasifikasikan data ke dalam kelas tersebut. Performa terbaik ditunjukkan pada *Class* 2, dengan nilai AUC 0.74 yang mengindikasikan bahwa model mampu membedakan *Class 2* dengan tingkat ketepatan yang memadai dibandingkan dua kelas lainnya

## 4. Kesimpulan dan Saran

### 4.1. Kesimpulan

Penelitian ini berhasil melakukan perbandingan dengan performa antara algoritma *Naive Bayes* dan *Decision Tree* yang digunakan untuk mengklasifikasikan tingkat kepuasan pelanggan berdasarkan data ulasan dari platform *e-commerce* Olist. Hasil evaluasi mengindikasikan bahwa algoritma *Naive Bayes* memperoleh tingkat akurasi yang lebih baik, yaitu sebesar 80.70%, dibandingkan dengan algoritma *Decision Tree* yang hanya mencapai 73.90%. *Naive Bayes* juga memperlihatkan kestabilan performa yang lebih baik dan seimbang dalam metrik *precision*, *recall*, dan *F1-score*, khususnya dalam mengklasifikasikan kelas Netran dan Puas. Namun, kedua algoritma sama-sama menunjukkan kelemahan dalam mengenali kelas Tidak Puas, yang disebabkan oleh distribusi data yang tidak seimbang.

Dengan demikian, dapat disimpulkan bahwa algoritma *Naive Bayes* lebih direkomendasikan untuk digunakan dalam tugas klasifikasi kepuasan pelanggan berbasis teks.

### 4.1. Saran

Untuk penelitian selanjutnya, disarankan untuk menerapkan teknik penyeimbangan data atau eksplorasi algoritma lain yang lebih adaptif terhadap kelas minoritas, guna meningkatkan akurasi klasifikasi secara keseluruhan.

## Daftar Pustaka

[1] M. Siregar, "Analisis Kepuasan Pelanggan Ompu Gende Coffee Medan," *J. Divers.*, vol. 7, no. 1, pp. 114–120, Jun. 2021, doi: 10.31289/diversita.v7i1.5190.

[2] N. Farhana, H. Okprana, and K. R. Sorim, "Analisis Tingkat Kepuasan Pelanggan Pada Aplikasi Tiktok Shop Dengan Metode Algoritma C4.5," *SmartEDU J.*, vol. 1, pp. 101–111, 2022.

[3] H. Priatmojo, F. Saputra, M. H. Prasetiyo, D. Puspitasari, and D. Nurlaela, "Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket Dengan Pendekatan Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor," 2023. [Online]. Available: http://jurnal.bsi.ac.id/index.php/jinsan

[4] A. Muzaki *et al.*, "Analisis Sentimen Pada Ulasan Produk di E-Commerce Dengan Metode Naive Bayes," *J. Ris. dan Apl. Mhs. Inform.*, vol. 05, 2024.

[5] J. J. A. Limbong, I. Sembiring, K. D. Hartomo, U. Kristen, S. Wacana, and P. Korespondensi, "Analisis Klasifikasi Sentimen Ulasan Pada E-Commerce Shopee Berbasis Word Cloud Dengan Metode Naive Bayes dan K-Nearest Analysis Of Review Sentiment Classification On E-Commerce Shopee Word Cloud Based With Naïve Bayes And K-Nearest Neighbor Methods," vol. 9, no. 2, pp. 347–356, 2022, doi: 10.25126/jtiik.202294960.

[6] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," vol. 7, no. 1, pp. 59–64, 2013.

[7] M. Zarlis, R. W. Sembiring, T. B. Pematangsiantar, U. Methodist, and P. N. Medan, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," no. 2, pp. 180–185, 2017.

[8] D. Chicco and G. Jurman, "The Advantages Of The Matthews Correlation Coefficient (Mcc) Over F1 Score And Accuracy In Binary Classification Evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.

[9] H. P. Newquist, *Artificial Intelligence: Technology In Transition*, vol. 36. 1992.

[10] N. White, R. Parsons, G. Collins, and A. Barnett, "Evidence Of Questionable Research Practices In Clinical Prediction Models," *BMC Med.*, vol. 21, no. 1, pp. 1–10, 2023, doi: 10.1186/s12916-023-03048-6.

[11] Hosmer, *et al.*, *Applied Logistic Regression*, 3rd ed. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey, 2013.