

PENERAPAN ALGORITMA PENGKLASIFIKASI UNTUK MENGUKUR KEPUASAN PELANGGAN E-COMMERCE (STUDI KASUS : SHOPEE)

Syamsul Bahri¹, Agung Mulyo Widodo²
Teknik Informatika, Universitas Esa Unggul, Jakarta

Abstrak

Penelitian ini bertujuan untuk mengukur tingkat kepuasan pelanggan Shopee dengan menggunakan beberapa algoritma machine learning, yaitu Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, dan Naive Bayes. Tingkat kepuasan pelanggan dikategorikan ke dalam lima tingkat: Sangat Tidak Puas, Tidak Puas, Netral, Puas, dan Sangat Puas. Data survei diperoleh dari 1.000 Responden, mencakup lima variabel utama, yaitu kualitas produk, layanan pengiriman, kualitas pelayanan, harga dan promosi, serta pengalaman berbelanja. Analisis dilakukan dengan heatmap korelasi untuk memahami hubungan antar variabel, serta feature importance menggunakan Random Forest untuk menentukan kontribusi relatif setiap faktor terhadap kepuasan pelanggan. Hasil penelitian menunjukkan bahwa harga dan promosi memiliki pengaruh tertinggi, diikuti oleh kualitas pelayanan dan pengalaman berbelanja. Penelitian ini memberikan wawasan strategis bagi Shopee untuk meningkatkan kualitas layanan berdasarkan analisis data dan memperkuat daya saing di pasar e-commerce.

Kata Kunci Kepuasan Pelanggan, E-Commerce, Shopee, Machine Learning, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Naive Bayes.

Abstract

This study aims to measure the level of customer satisfaction on Shopee using several machine learning algorithms, namely Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, and Naive Bayes. The level of customer satisfaction is categorized into five levels: Very Dissatisfied, Dissatisfied, Neutral, Satisfied, and Very Satisfied. The survey data used includes five main variables, namely product quality, delivery service, service quality, price and promotion, and shopping experience. The analysis was carried out using a correlation heatmap to understand the relationship between variables, and feature importance using Random Forest to determine the relative contribution of each factor to customer satisfaction. The results showed that price and promotion had the highest influence, followed by service quality and shopping experience. This study provides strategic insights for Shopee to improve service quality based on data analysis and strengthen competitiveness in the e-commerce market

Keywords: Customer Satisfaction, E-Commerce, Shopee, Machine Learning, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, Naive Bayes.

Pendahuluan

E-commerce telah menjadi bagian integral dalam kehidupan masyarakat modern, memberikan kemudahan akses bagi konsumen untuk berbelanja berbagai produk tanpa batasan waktu dan tempat. Di Indonesia, platform seperti Shopee telah memimpin pasar e-commerce dengan menawarkan beragam produk dan promosi menarik. Shopee menjadi pilihan utama bagi banyak konsumen karena kemudahan penggunaannya, namun mempertahankan tingkat kepuasan pelanggan tetap menjadi tantangan besar di tengah persaingan ketat industri ini.

Kepuasan pelanggan merupakan indikator penting bagi keberhasilan sebuah platform e-commerce. Faktor-faktor seperti kualitas produk, kecepatan layanan pengiriman, kualitas pelayanan, harga, dan promosi memegang peranan penting dalam membentuk pengalaman berbelanja yang memuaskan. Kegagalan dalam memenuhi ekspektasi pelanggan dapat menyebabkan mereka beralih ke platform lain yang lebih kompetitif.

Penelitian ini bertujuan untuk memahami hubungan antara faktor-faktor tersebut dengan tingkat kepuasan pelanggan Shopee. Dengan menggunakan metode machine learning, penelitian ini menganalisis data survei untuk mengidentifikasi faktor-faktor yang paling memengaruhi kepuasan pelanggan. Pendekatan berbasis data ini memberikan wawasan yang lebih objektif dibandingkan metode tradisional.

Hasil penelitian ini diharapkan dapat memberikan kontribusi yang signifikan bagi perusahaan e-commerce seperti Shopee dalam meningkatkan layanan dan daya saing mereka. Dengan memanfaatkan hasil analisis berbasis machine learning, strategi yang lebih tepat sasaran dapat dirancang untuk mempertahankan loyalitas pelanggan dan menarik konsumen baru di pasar yang semakin kompetitif.

Literature Review

Proses *literature review* dilakukan secara sistematis melalui beberapa langkah, yaitu: merumuskan masalah penelitian, mencari dan mengumpulkan literatur yang relevan, mengevaluasi data yang diperoleh, serta melakukan analisis mendalam terhadap temuan sebelumnya. Berikut ini adalah hasil literature review yang telah dilakukan sebagai dasar untuk mendukung penelitian :

1. Algoritma SVM menunjukkan akurasi 80,31%, dengan precision 73,22% dan recall 95,58%, dalam menganalisis sentimen ulasan pelanggan ShopeeFood dari Twitter (Rismanah *et all*, 2024)
2. Model SVM menghasilkan akurasi 75% dalam mengklasifikasikan sentimen ulasan pelanggan Tokopedia (Milal *et all*, 2023)
3. Naive Bayes lebih unggul dengan akurasi 75,97%, dibandingkan K-NN sebesar 16,69% (Yumarlin *et all*, 2023)
4. Naive Bayes memiliki akurasi 99,06% untuk Shopee, sementara K-NN mencapai 83,29% (Nandaretsa dan Warman, 2023)
5. SVM memiliki akurasi lebih tinggi (79,14%) dibandingkan Naive Bayes (75,61%) (Latifah, 2018).
6. Akurasi model K-NN sebesar 83,33% (Fansyuri, 2020)
7. Penelitian ini membandingkan KNN dan Naive Bayes untuk klasifikasi kepuasan pelanggan menggunakan 800 data dari Kaggle. Hasil menunjukkan KNN lebih unggul dengan akurasi 87,84%, precision, recall, dan F1-score yang tinggi, sehingga direkomendasikan sebagai metode terbaik
8. Penelitian ini menggunakan web scraping untuk mengumpulkan ulasan aplikasi Shopee di Google Play Store. Data diproses dengan preprocessing dan dianalisis menggunakan K-Nearest Neighbor (KNN), menghasilkan akurasi 82%. Penelitian ini menganalisis sentimen masyarakat terhadap e-commerce di Indonesia (Febrian *et all*, 2022)
9. Penelitian ini mengukur kepuasan pelanggan terhadap pelayanan kuliner di Kota Tegal menggunakan analisis sentimen. Algoritma Support Vector Machine (SVM) dengan seleksi fitur diterapkan pada ulasan dari situs TripAdvisor. Evaluasi menggunakan Cross Validation menghasilkan akurasi terbaik sebesar 72,45% dengan model SVM-IG (Kusuma dan Cahyono, 2023)
10. Penelitian ini memprediksi kepuasan pengunjung Taman Tabebuya di Jakarta menggunakan algoritma Support Vector Machine (SVM). Data survei dianalisis dengan SVM dan dievaluasi menggunakan Confusion Matrix dan AUC, menghasilkan akurasi 86% dan AUC 0.947, menunjukkan performa yang sangat baik (Darmawan *et all*, 2018)
11. Penelitian ini mengukur kualitas layanan e-commerce Lejel Home Shopping menggunakan metode E-SERVQUAL pada 100 responden. Hasil menunjukkan kualitas layanan belum memuaskan dengan nilai -1,084, dengan kesenjangan tertinggi pada dimensi pemenuhan (-1,24). Penelitian merekomendasikan perbaikan pada dimensi layanan yang memengaruhi kepuasan pelanggan (Austi dan Salisah, 2016)

12. Penelitian ini menunjukkan bahwa pengalaman dan kepuasan konsumen berpengaruh positif signifikan terhadap niat beli ulang produk skincare di Shopee, dengan t hitung masing-masing 2,920 dan 4,809, serta signifikansi 0,004 dan 0,000 (Agustina dan Julitriarsa, 2021).
13. Penelitian ini menunjukkan bahwa kualitas pelayanan memiliki pengaruh signifikan terhadap kepuasan pelanggan e-commerce Shopee, sementara harga tidak. Secara simultan, kedua variabel berpengaruh signifikan terhadap kepuasan pelanggan dengan nilai signifikansi 0,000 berdasarkan data dari 35 responden menggunakan metode kuantitatif asosiatif (Lonardi *et all*, 2021).
14. Penelitian ini menunjukkan bahwa promosi dan kepercayaan konsumen berpengaruh positif signifikan terhadap kepuasan dan loyalitas pelanggan Shopee. Namun, kepuasan pelanggan tidak memiliki pengaruh signifikan terhadap loyalitas, baik secara langsung maupun tidak langsung melalui promosi dan kepercayaan konsumen. Analisis dilakukan dengan metode Partial Least Square (PLS) pada 100 responden mahasiswa STIESIA Surabaya (Sari dan Lestariningsih, 2021).
15. bahwa kualitas pelayanan, kualitas produk, dan kualitas pengiriman berpengaruh positif signifikan terhadap kepuasan pelanggan toko online Lazada, dengan nilai t-hitung masing-masing 10,952, 5,138, dan 2,996 (lebih besar dari t-tabel 1,661). Uji F menunjukkan pengaruh simultan dengan nilai F-hitung 252,007 dan signifikansi 0,000. Nilai determinasi (R^2) sebesar 88,7% menunjukkan bahwa ketiga variabel memengaruhi kepuasan pelanggan, sementara sisanya 11,3% dipengaruhi oleh faktor lain (Tholok *et all*, 2023)

Berdasarkan jurnal-jurnal yang telah diuraikan, dapat disimpulkan bahwa terdapat potensi besar dalam pengembangan sistem analisis kepuasan pelanggan di *e-commerce*, khususnya melalui pemanfaatan algoritma machine learning. Penelitian baru ini dapat memanfaatkan berbagai metode yang telah diterapkan dalam studi sebelumnya, seperti Support Vector Machine (SVM), K-Nearest Neighbors (KNN), atau Naive Bayes, sesuai dengan kebutuhan dan tujuan penelitian. Sistem yang dihasilkan diharapkan dapat mengatasi berbagai masalah yang telah diidentifikasi, seperti pengolahan data ulasan yang manual, ketidakakuratan prediksi, dan kurangnya pemahaman terhadap faktor-faktor utama yang memengaruhi kepuasan pelanggan.

Dengan memanfaatkan teknologi machine learning dan algoritma yang tepat, sistem ini dapat memberikan wawasan berbasis data, efisiensi dalam analisis, serta kenyamanan bagi perusahaan dalam memahami kebutuhan pelanggan secara real-time. Penelitian ini memiliki potensi untuk meningkatkan efektivitas dan efisiensi analisis kepuasan pelanggan serta mendukung pengambilan keputusan strategis untuk memenuhi kebutuhan pasar e-commerce yang semakin kompetitif

METODE PENELITIAN

Penelitian ini menggunakan pendekatan Berencana untuk mengumpulkan, mengolah, dan menganalisis data terkait kepuasan pelanggan Shopee. Data diperoleh melalui survei berbasis kuesioner yang didistribusikan menggunakan Google Form. Survei mencakup aspek kualitas produk, layanan pengiriman, kualitas layanan pelanggan, harga dan promosi, serta pengalaman berbelanja. Jawaban responden dikategorikan ke dalam lima tingkat kepuasan: sangat tidak puas, tidak puas, netral, puas, dan sangat puas. Data yang terkumpul diproses melalui beberapa tahap, yaitu:

1. Pembersihan Data: Mengidentifikasi dan menghapus data yang tidak lengkap atau tidak relevan.
2. Labeling: Mengubah data kualitatif menjadi data kuantitatif sesuai skala yang telah ditentukan.
3. Pembagian Data: Dataset dibagi menjadi data latih (80%) dan data uji (20%) untuk melatih dan menguji model machine learning.

Analisis dilakukan menggunakan algoritma Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, dan Naive Bayes. Setiap algoritma dievaluasi berdasarkan metrik akurasi, precision, recall, dan F1-score untuk mengidentifikasi model terbaik dalam memprediksi kepuasan pelanggan.

Implementasi Algoritma

Penelitian ini menerapkan beberapa algoritma machine learning melalui tahapan berikut:

1. Preprocessing Data
Data penelitian terdiri dari fitur input (X1 hingga X5) yang merepresentasikan faktor kepuasan pelanggan. Tahapan preprocessing mencakup:
 - Cleaning Data: Menghapus data yang kosong atau tidak relevan.
 - Encoding: Label kepuasan pelanggan (Sangat Tidak Puas hingga Sangat Puas) dikonversi menjadi nilai numerik menggunakan LabelEncoder.
2. Penggunaan Algoritma
Algoritma yang digunakan untuk klasifikasi kepuasan pelanggan meliputi:
 - Support Vector Machine (SVM) dengan kernel linear dan RBF.
 - K-Nearest Neighbor (KNN) dengan nilai $k=5,11,13$
 - Random Forest dengan jumlah pohon $n_estimators=110$ dan 500
 - Naive Bayes untuk klasifikasi berbasis probabilitas.
3. Evaluasi Model
Kinerja algoritma dievaluasi menggunakan confusion matrix dan metrik evaluasi seperti akurasi, precision, recall, dan F1-Score. Evaluasi dilakukan secara macro dan weighted untuk setiap kelas (Sangat Tidak Puas, Tidak Puas, Netral, Puas, Sangat Puas), guna mengukur performa keseluruhan model.

Objek Penelitian

Objek penelitian dalam studi ini adalah kepuasan pelanggan pada platform e-commerce Shopee, yang merupakan salah satu platform e-commerce terbesar dan paling populer di Indonesia. Kepuasan pelanggan menjadi fokus utama penelitian ini, dengan mempertimbangkan berbagai faktor yang memengaruhi pengalaman belanja mereka. Faktor-faktor tersebut meliputi kualitas produk, yang mencakup kesesuaian dengan deskripsi dan kondisi saat diterima. layanan pengiriman, seperti kecepatan dan keakuratan pengiriman kualitas pelayanan, termasuk responsivitas layanan pelanggan dalam menangani keluhan. harga dan promosi, yang mencakup diskon serta penawaran khusus. serta pengalaman berbelanja secara keseluruhan, seperti kemudahan navigasi aplikasi dan fitur pencarian. Penelitian ini bertujuan untuk menganalisis hubungan antara faktor-faktor tersebut dengan tingkat kepuasan pelanggan yang dikategorikan ke dalam lima tingkat (Sangat Tidak Puas hingga Sangat Puas). Dengan menggunakan data survei yang diolah melalui berbagai algoritma machine learning, seperti Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, dan Naive Bayes, penelitian ini bertujuan memberikan wawasan strategis bagi Shopee dalam meningkatkan kualitas layanan dan memenuhi kebutuhan pasar yang semakin kompetitif.

Teknik Pengumpulan Data

Penelitian ini menggunakan dua jenis data, yaitu data primer dan data sekunder, dengan teknik pengumpulan data sebagai berikut:

1. Data Primer

Data primer diperoleh melalui survei online yang didistribusikan kepada pelanggan Shopee menggunakan Google Form. Survei ini dirancang untuk mengumpulkan informasi tentang kepuasan pelanggan berdasarkan pengalaman mereka dalam berbelanja di Shopee, khususnya untuk kategori produk tertentu, seperti handphone. Pertanyaan survei mencakup lima aspek utama, yaitu kualitas produk, layanan pengiriman, kualitas pelayanan, harga dan promosi, serta pengalaman berbelanja. Setiap pertanyaan dikategorikan dalam skala Likert dengan lima tingkatan, mulai dari Sangat Tidak Puas hingga Sangat Puas. Teknik ini memungkinkan peneliti untuk mendapatkan data langsung dari responden yang relevan dengan topik penelitian.

2. Data Sekunder

Data sekunder diperoleh dari berbagai literatur, seperti jurnal ilmiah, buku, laporan penelitian sebelumnya, dan artikel yang relevan. Data ini digunakan untuk mendukung landasan teori, memahami metode yang digunakan dalam penelitian terdahulu, serta memberikan konteks tambahan dalam analisis hasil. Sumber data sekunder meliputi studi yang membahas faktor-faktor kepuasan pelanggan, implementasi algoritma machine learning, dan perkembangan e-commerce di Indonesia.

Library Python

Penelitian ini memanfaatkan beberapa library Python utama untuk preprocessing data, pembangunan model, evaluasi, dan visualisasi hasil:

1. Pandas

Digunakan untuk membaca, mengelola, dan memproses data berbasis tabel (DataFrame), seperti membaca file CSV, membersihkan data, dan transformasi data. (Rismanah *et al*, 2024)

2. Numpy

Library untuk komputasi numerik, digunakan dalam operasi array multidimensi, aljabar linier, dan manipulasi data (Milai *et al*, 2023)

3. Scikit-Learn(sklearn)

Library utama untuk machine learning yang menyediakan algoritma, preprocessing, evaluasi, dan analisis model. Fungsi yang digunakan meliputi (Yumarlin *et al*, 2023):

- `train_test_split`: Membagi dataset menjadi data pelatihan dan pengujian.
- `LabelEncoder`: Mengubah nilai kategorikal menjadi numerik.

Algoritma Klasifikasi:

- `SVC` (Support Vector Classifier): Implementasi SVM untuk klasifikasi dengan berbagai kernel.
- `KNeighborsClassifier`: Klasifikasi berbasis k-Nearest Neighbors.
- `RandomForestClassifier`: Model ensemble berbasis pohon keputusan.
- `GaussianNB`: Implementasi Naive Bayes berbasis distribusi Gaussian.

Fungsi Evaluasi: Menggunakan metrik seperti akurasi, precision, recall, F1-score, dan confusion matrix.

4. Matplotlib

Digunakan untuk membuat visualisasi data dan grafik, termasuk confusion matrix (Nandaresta dan warman, 2023)

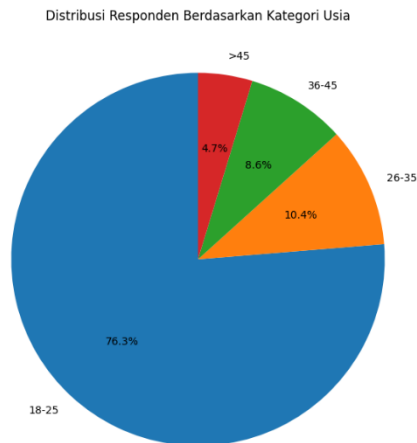
5. Seaborn

Library berbasis Matplotlib untuk membuat grafik estetik, seperti heatmap untuk visualisasi korelasi dan confusion matrix (Latifah, 2018).

Hasil dan Pembahasan

Deskripsi Data

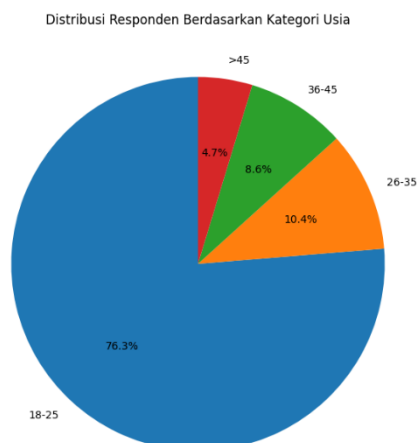
a. Distribusi Responden Berdasarkan Kategori Usia



Gambar 1 Grafik Usia Responden

Berdasarkan survei, mayoritas responden berasal dari kelompok usia 18–25 tahun dengan persentase sebesar 76,3%, menunjukkan bahwa Shopee sangat populer di kalangan generasi muda. Kelompok usia 26–35 tahun berada di urutan kedua dengan persentase 10,4%, diikuti oleh kelompok usia 36–45 tahun sebesar 8,6%, dan kelompok usia di atas 45 tahun hanya sebesar 4,7%. Hasil ini menunjukkan bahwa mayoritas pengguna aktif Shopee adalah generasi muda yang terbiasa dengan teknologi dan transaksi online.

b. Distribusi Responden Berdasarkan Jenis Kelamin



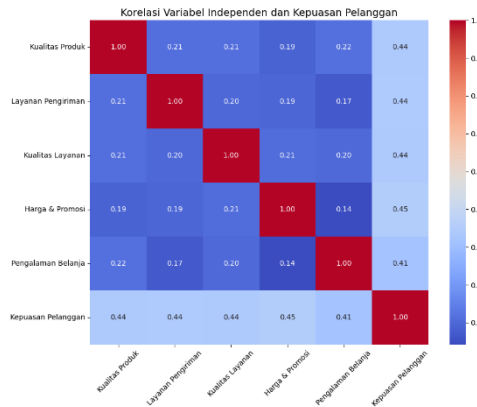
Gambar 2 Grafik Jenis Kelamin Responden

Distribusi responden berdasarkan jenis kelamin menunjukkan pembagian yang hampir merata antara laki-laki dan perempuan. Responden laki-laki mencakup 50,6%, sedangkan perempuan mencakup 49,4%. Hal ini mengindikasikan bahwa Shopee berhasil menarik

minat pengguna dari kedua gender secara hampir seimbang, yang mencerminkan inklusivitas platform dalam menyediakan berbagai produk yang sesuai untuk berbagai kebutuhan pelanggan.

Korelasi Variabel Independen dan Kepuasan Pelanggan

Korelasi adalah visualisasi hubungan antara variabel, menunjukkan kekuatan dan arah hubungan antar variabel, seperti faktor-faktor yang memengaruhi kepuasan pelanggan. Nilai korelasi berkisar antara -1 hingga 1, dengan nilai positif menunjukkan hubungan searah dan nilai negatif menunjukkan hubungan berlawanan.

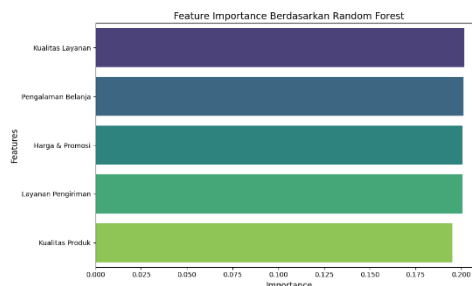


Gambar 3 Korelasi Variabel Independen dan Kepuasan Pelanggan

Gambar 4 Korelasi menunjukkan hubungan antara variabel independen dan kepuasan pelanggan. Harga dan Promosi memiliki korelasi tertinggi (0,45), menjadikannya faktor utama yang memengaruhi kepuasan pelanggan. Kualitas Produk, Layanan Pengiriman, dan Kualitas Pelayanan memiliki korelasi yang sama (0,44), sementara Pengalaman Berbelanja memiliki korelasi lebih rendah (0,41). Semua variabel menunjukkan hubungan positif dengan kepuasan pelanggan, dengan fokus utama pada Harga dan Promosi sebagai prioritas dalam meningkatkan kepuasan pelanggan Shopee.

Feature Importance Random Forest

Feature importance menunjukkan seberapa besar pengaruh tiap variabel terhadap hasil prediksi dalam model machine learning, membantu mengidentifikasi faktor yang paling signifikan.

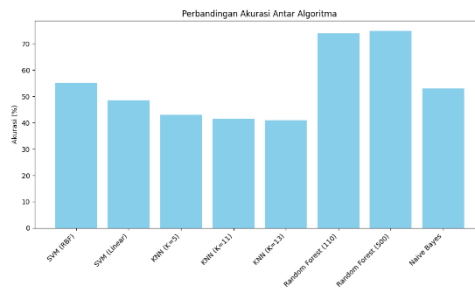


Gambar 4 Feature Importance Random Forest

Gambar di atas menunjukkan Feature Importance berdasarkan algoritma Random Forest. Faktor Kualitas Layanan dan Pengalaman Berbelanja memiliki tingkat kepentingan

tertinggi (20%), menunjukkan bahwa kedua aspek ini berperan dominan dalam memengaruhi kepuasan pelanggan. Harga dan Promosi, Layanan Pengiriman, dan Kualitas Produk memiliki kontribusi yang hampir setara, dengan nilai sedikit lebih rendah. Hasil ini menegaskan bahwa setiap faktor memiliki pengaruh signifikan, namun prioritas utama dapat diberikan pada Kualitas Layanan dan Pengalaman Berbelanja untuk meningkatkan kepuasan pelanggan.

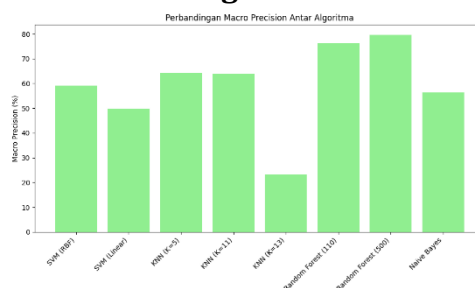
Perbandingan Akurasi antar Algoritma



Gambar 5 Perbandingan Akurasi antar Algoritma

Gambar 6 di atas menunjukkan perbandingan akurasi antar algoritma machine learning. Algoritma Random Forest dengan 500 estimators memiliki akurasi tertinggi, yaitu 70%, diikuti oleh Random Forest dengan 110 estimators. Support Vector Machine (SVM) dengan kernel RBF dan linear mencapai akurasi yang moderat, masing-masing sekitar 60% dan 50%. K-Nearest Neighbors (KNN) menunjukkan hasil yang lebih rendah, dengan akurasi berkisar antara 40%-45% tergantung pada nilai K. Algoritma Naive Bayes menghasilkan akurasi yang lebih rendah dibandingkan Random Forest dan SVM. Hasil ini menunjukkan bahwa Random Forest adalah algoritma terbaik untuk dataset ini, memberikan performa paling unggul dalam prediksi kepuasan pelanggan.

Perbandingan macro Precision antar Algoritma Machine Learning.

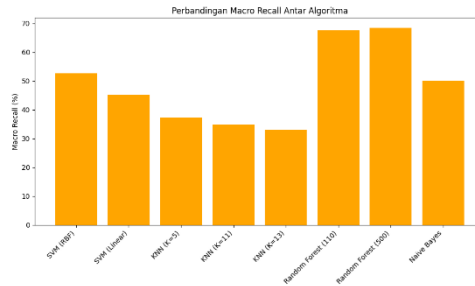


Gambar 6 perbandingan macro precision antar algoritma machine learning.

Gambar 7 di atas menunjukkan perbandingan macro precision antar algoritma machine learning. Algoritma Random Forest dengan 500 estimators memiliki nilai macro precision tertinggi, yaitu sekitar 78%, diikuti oleh Random Forest dengan 110 estimators. K-Nearest Neighbors (KNN) dengan K=5 dan K=11 juga menunjukkan performa yang cukup baik dengan nilai precision di atas 60%. Support Vector Machine (SVM) dengan kernel RBF mencapai sekitar 60%, sedangkan SVM dengan kernel linear lebih rendah. Naive

Bayes menghasilkan macro precision yang moderat, lebih rendah dari Random Forest namun masih kompetitif. Hasil ini menunjukkan bahwa Random Forest tidak hanya unggul dalam akurasi, tetapi juga dalam precision untuk dataset ini.

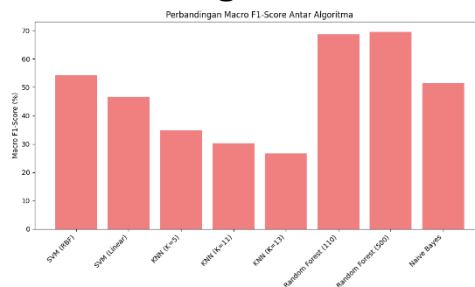
Perbandingan macro Recall antar Algoritma Machine Learning.



Gambar 7 Perbandingan macro Recall antar Algoritma Machine Learning.

Gambar 8 di atas menunjukkan perbandingan macro recall antar algoritma machine learning. Algoritma Random Forest dengan 500 estimators memiliki macro recall tertinggi, yaitu sekitar 70%, diikuti oleh Random Forest dengan 110 estimators. Support Vector Machine (SVM) dengan kernel RBF menunjukkan performa yang baik dengan nilai sekitar 50%, sementara kernel linear berada di angka yang sedikit lebih rendah. K-Nearest Neighbors (KNN) menunjukkan hasil yang lebih bervariasi, dengan nilai recall tertinggi pada K=5. Naive Bayes memiliki nilai recall yang lebih rendah dibandingkan Random Forest tetapi masih kompetitif. Hasil ini menegaskan bahwa Random Forest unggul dalam mengidentifikasi kelas positif secara konsisten dibandingkan algoritma lainnya.

Perbandingan macro F-1 Score antar Algoritma Machine Learning.



Gambar 8 Perbandingan macro F-1 Score antar Algoritma Machine Learning.

Gambar 9 di atas menunjukkan perbandingan macro F1-score antar algoritma machine learning. Algoritma Random Forest dengan 500 estimators mencatat F1-score tertinggi, yaitu sekitar 70%, diikuti oleh Random Forest dengan 110 estimators. Support Vector Machine (SVM) dengan kernel RBF memiliki performa baik dengan nilai sekitar 50%, sedangkan kernel linear sedikit lebih rendah. K-Nearest Neighbors (KNN) menunjukkan hasil yang lebih bervariasi, dengan performa terbaik pada K=5. Naive Bayes menghasilkan F1-score yang lebih rendah dibandingkan Random Forest, tetapi tetap

kompetitif. Kesimpulannya, Random Forest unggul dalam keseimbangan antara precision dan recall, menjadikannya algoritma terbaik untuk dataset ini.

Hasil Penelitian

Hasil penelitian menunjukkan bahwa mayoritas responden berasal dari kelompok usia 18–25 tahun (76,3%), yang mencerminkan popularitas Shopee di kalangan generasi muda. Dari segi jenis kelamin, distribusi responden hampir merata, dengan 50,6% laki-laki dan 49,4% perempuan. Analisis korelasi menunjukkan bahwa Harga dan Promosi memiliki hubungan paling signifikan dengan kepuasan pelanggan (nilai korelasi 0,45), diikuti oleh Kualitas Layanan, Kualitas Produk, dan Layanan Pengiriman (masing-masing 0,44). Pengalaman Berbelanja memiliki nilai korelasi lebih rendah (0,41), namun tetap memberikan kontribusi penting terhadap kepuasan pelanggan.

Berdasarkan analisis feature importance menggunakan algoritma Random Forest, Kualitas Layanan dan Pengalaman Berbelanja memiliki tingkat kepentingan tertinggi (20%), diikuti oleh Harga dan Promosi, Layanan Pengiriman, dan Kualitas Produk. Dalam perbandingan kinerja algoritma machine learning, Random Forest dengan 500 estimators memberikan performa terbaik dengan akurasi sebesar 70%, macro precision 78%, macro recall 70%, dan macro F1-score 70%. Algoritma SVM dengan kernel RBF menunjukkan performa moderat dengan akurasi sekitar 60%, sementara KNN menunjukkan hasil yang bervariasi tergantung pada nilai KKK, dengan performa terbaik pada $K=5$. Naive Bayes menghasilkan performa yang cukup baik tetapi lebih rendah dibandingkan Random Forest.

Secara keseluruhan, penelitian ini menegaskan bahwa algoritma Random Forest adalah yang paling unggul untuk menganalisis kepuasan pelanggan Shopee. Faktor Harga dan Promosi menjadi elemen paling penting yang perlu diprioritaskan oleh Shopee, diikuti oleh Kualitas Layanan dan Pengalaman Berbelanja, untuk meningkatkan kepuasan dan loyalitas pelanggan. Hasil penelitian ini memberikan wawasan strategis bagi Shopee dalam merancang pengembangan layanan berbasis data yang lebih efektif.

Kesimpulan

Kepuasan pelanggan Shopee dipengaruhi oleh lima faktor utama, yaitu Harga dan Promosi, Kualitas Layanan, Kualitas Produk, Layanan Pengiriman, dan Pengalaman Berbelanja. Analisis korelasi menunjukkan bahwa Harga dan Promosi memiliki hubungan paling signifikan terhadap kepuasan pelanggan, menjadikannya faktor prioritas untuk ditingkatkan. Berdasarkan analisis feature importance, Kualitas Layanan dan Pengalaman Berbelanja juga berkontribusi besar dalam menentukan tingkat kepuasan pelanggan.

Dalam perbandingan kinerja algoritma machine learning, Random Forest dengan 500 estimators terbukti menjadi model terbaik, dengan akurasi, macro precision, macro recall, dan macro F1-score tertinggi dibandingkan algoritma lainnya, seperti SVM, KNN, dan Naive Bayes. Hasil ini menunjukkan bahwa Random Forest mampu memberikan prediksi yang paling akurat dan konsisten dalam menganalisis data kepuasan pelanggan.

Secara keseluruhan, penelitian ini memberikan wawasan strategis bagi Shopee untuk fokus pada penguatan program promosi, peningkatan kualitas layanan, dan penyempurnaan pengalaman berbelanja guna meningkatkan loyalitas pelanggan. Dengan memanfaatkan pendekatan berbasis data ini, Shopee dapat mengoptimalkan layanan mereka untuk memenuhi kebutuhan pasar yang semakin kompetitif.

Daftar Referensi

- Agustina, A., & Julitriarsa, D. (2021). *Pengaruh Pengalaman Konsumen dan Kepuasan Konsumen Terhadap Niat Beli Ulang pada Pembelian Produk Skincare di E-Commerce Shopee*. CAKRAWANGSA BISNIS, 2(2), 287-300.
- Astuti, D., & Salisah, F. N. (2016). *Analisis Kualitas Layanan E-commerce Terhadap Kepuasan Pelanggan Menggunakan Metode E-Servqual (Studi Kasus: Lejel Home Shopping Pekanbaru)*. Jurnal Rekayasa dan Manajemen Sistem Informasi, 2(1), 44-46. e-ISSN 2502-8995
- Darmawan, A., Kustian, N., & Rahayu, W. (2018). *Implementasi Data Mining Menggunakan Model SVM untuk Prediksi Kepuasan Pengunjung Taman Tabebuaya*. Jurnal String, 2(3), 299-300.
- Fansyuri, M. (2020). *Analisa Algoritma Klasifikasi K-Nearest Neighbor dalam Menentukan Nilai Akurasi Terhadap Kepuasan Pelanggan (Studi Kasus PT. Trigatra Komunikatama)*. Humanika: Jurnal Ilmu Sosial, Pendidikan, dan Humaniora, 3(1), 29-30.
- Febrian, A. K., Chrisnanto, Y. H., & Sabrina, P. N. (2022). *Studi Komparasi Metode Klasifikasi K-Nearest Neighbor dan Naïve Bayes dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk*. SNESTIK Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika, Institut Teknologi Adhi Tama Surabaya.
- Kusuma, I. H., & Cahyono, N. (2023). *Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor*. Jurnal Informatika: Jurnal Pengembangan IT (JPIT), 8(3), 302-303.
- Latifah, E. F. U. (2018). *Perbandingan Kinerja Machine Learning Berbasis Algoritma Support Vector Machine dan Naive Bayes (Studi Kasus: Data Tanggapan Mengenai Traveloka Melalui Media Sosial Twitter)*. Skripsi, Universitas Islam Indonesia, Yogyakarta.
- Lonardi, S., Lan, W. P., Hutabarat, F. A. M., Nugroho, N., & Supriyanto. (2021). *Peranan Kualitas Pelayanan dan Harga Terhadap Kepuasan Konsumen Pengguna E-commerce Shopee Di Politeknik Cendana*. Journal of Business and Economics Research (JBE), 2(3), 80-85.
- Milal, I. S., Hasanudin, M., Azhari, M. A. N., Nugraha, R. A., & Agustina, N. (2023). *Klasifikasi Teks Review pada E-Commerce Tokopedia Menggunakan Algoritma SVM*. NARATIF: Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika, 5(1), 34-36.
- Nandaresta, S. C., & Warman, C. (2023). *Analisis Sentimen Tanggapan Masyarakat Terhadap TikTok Shop dan Shopee di Twitter Menggunakan Metode Naïve Bayes dan KNN*. SISMATIK Seminar Nasional Sistem Informasi dan Manajemen Informatika, Universitas Nusa Putra, Sukabumi.
- Rismanah, S. N., Astuti, R., & Basysyar, F. M. (2024). *Penerapan Algoritma Support Vector Machine dalam Menganalisis Sentimen Ulasan Pelanggan ShopeeFood Berdasarkan Twitter*. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 406-408.

- Sari, A. C. M., & Lestariningsih, M. (2021). *Pengaruh Promosi dan Kepercayaan Konsumen Terhadap Loyalitas Pelanggan Melalui Kepuasan Pelanggan Aplikasi Shopee (Studi Pada Mahasiswa STIESIA Surabaya)*. Jurnal Ilmu dan Riset Manajemen, 10(5).
- Tholok, F. W., Sugandha, S., Janamarta, S., & Parameswari, R. (2023)**. Analisis kualitas pelayanan, kualitas produk dan kualitas pengiriman terhadap kepuasan pelanggan toko online Lazada (Studi kasus mahasiswa Universitas Buddhi Dharma Fakultas Bisnis). *Primanomics: Jurnal Ekonomi & Bisnis*, 21(2), 26–34. <https://doi.org/10.31253/pe.v21i2.1825>
- Yumarlin, M. Z., Bororing, J. E., Rahayu, S., & Putra, J. A. (2023). *Analisis Sentimen Pengguna Aplikasi Shopee Menggunakan Metode Naive Bayes Classifier dan K-Nearest Neighbor (K-NN)*. *Smart Comp*, 12(3), 745-746.