JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING
VOL. 04, NO. 3, July 2025: 135-146
Doi: https://doi.org/10.55537/cosie.v4i3.1160

e-ISSN 2827-8356
p-ISSN 2827-9999

Perbandingan Algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) Dalam Memprediksi Kepuasan Pelanggan

Comparison of K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) Algorithms in Predicting Customer Satisfaction

Subhan Rizky Pratama[*1], Ika Nur Fajri[2]
[1,2]Sistem Informasi, Universitas Amikom Yogyakarta
E-mail: [1]subhanrizkypratama@students.amikom.ac.id, [2]fajri@amikom.ac.id

Received: June 6, 2025 | Revised: June 15, 2025 | Accepted: June 16, 2025

Abstrak

Penelitian ini membandingkan algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam memprediksi kepuasan pelanggan pada Warung Makan Indomie (Warmindo). Proses penelitian terdiri dari empat tahap, yaitu: pengumpulan data, pemrosesan data, pembentukan model, dan evaluasi model. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma klasifikasi, yaitu K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM), dalam memprediksi tingkat kepuasan pelanggan berdasarkan data survei. Evaluasi dilakukan dengan menggunakan metrik akurasi dan classification report untuk mengetahui tingkat presisi, recall, dan f1-score dari masing-masing algoritma. Hasil evaluasi menunjukkan bahwa kedua algoritma memiliki akurasi yang sama sebesar 70%. KNN unggul pada f1-score di kelas 2 (0.70), sementara SVM unggul dalam presisi pada kelas 2 (0.79). dengan score rata-rata dari kedua algoritma adalah 0.61. Hasil ini menunjukkan bahwa baik KNN maupun SVM layak digunakan, tergantung pada prioritas performa per kelas.

Kata kunci: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Kepuasan Pelanggan, Klasifikasi, Evaluasi Model

Abstract

This study compares the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms in predicting customer satisfaction at Warung Makan Indomie (Warmindo). The research process consists of four stages, namely: data collection, data processing, model formation, and model evaluation. This study aims to compare the performance of two classification algorithms, namely K-Nearest Neighbor (KNN) and Support Vector Machine (SVM), in predicting customer satisfaction levels based on survey data. The evaluation was carried out using accuracy metrics and classification reports to determine the level of precision, recall, and f1-score of each algorithm. The evaluation results show that both algorithms have the same accuracy of 70%. KNN excels in f1-score in class 2 (0.70), while SVM excels in precision in class 2 (0.79). with an average score of both algorithms being 0.61. These results indicate that both KNN and SVM are feasible to use, depending on the performance priority per class.

Keywords: K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Customer Satisfaction, Classification, Model Evaluation

1. PENDAHULUAN

e-ISSN 2827-8356
p-ISSN 2827-9999

Kepuasan pelanggan merupakan faktor kunci dalam mempertahankan dan meningkatkan kualitas layanan di berbagai sektor, termasuk bisnis kuliner sederhana seperti Warung Makan Indomie (Warmindo). Dalam dunia usaha yang semakin kompetitif, memahami dan memprediksi kepuasan pelanggan menjadi hal penting agar pemilik usaha dapat melakukan perbaikan layanan yang tepat sasaran. Namun, penilaian kepuasan pelanggan seringkali bersifat subjektif dan sulit diukur secara akurat tanpa bantuan data dan teknologi.

Perkembangan teknologi informasi saat ini memungkinkan proses evaluasi kepuasan pelanggan dilakukan secara otomatis dengan bantuan data mining. Data mining merupakan serangkaian proses untuk menggali pola atau informasi yang berguna dari sejumlah data, yang kemudian dapat digunakan untuk pengambilan keputusan yang lebih tepat. Salah satu pendekatan data mining yang populer adalah klasifikasi, yaitu proses mengelompokkan data berdasarkan kategori tertentu, seperti tingkat kepuasan pelanggan. Terdapat beberapa metode algoritma klasifikasi yang umum digunakan dalam data mining, yaitu Naive Bayes, Decision Tree, Support Vector Machine (SVM), dan K-Nearest Neighbors (KNN)[1].

Dua algoritma yang cukup populer karena kemampuannya dalam menangani data numerik maupun kategorikal adalah KNN dan SVM. terdapat beberapa penelitian yang membandingkan algoritma KNN dan SVM dalam memprrdiski berbagai hal. pada penelitian yang dilakukan oleh [2] yang dimana penelitian tersebut membahas tentang Perbandingan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Menentukan Prediksi Produk-Produk Terlaris Pada Toko Madura Kecamatan Pondok Aren . Penelitian ini difokuskan pada prediksi produk-produk terlaris yang nantinya dapat dimanfaatkan untuk mengatasi masalah persediaan stok produk. Dengan hasil penelitian Algoritma KNN, dan SVM memiliki nilai akurasi yang sama yaitu sebesar 75%, hingga dapat disimpulkan bahwa algoritma SVM lebih akurat digunakan untuk prediksi produk-produk terlaris pada toko madura Kecamatan Pondok Aren .

Penelitian yang dilakukan oleh [3] membahas tentang Prediksi Produk Penjualan Di Supermarket Dengan Metode Algoritma K-Nearest Neighbor (KNN). Peneliti tersebut menggunakan metode KNN, dengan cara mencari hasil pengujian dengan metode ini yang akan diimplementasikan dengan menggunakan aplikasi Rapidminer yang menghasilkan tingkat akurasi sebesar 85% dalam mengklasifikasikan produk dengan tingkat penjualan tertentu, dengan parameter terbaik diperoleh pada jumlah tetangga (k) sebesar 5. Hasil ini menunjukkan bahwa algoritma KNN dapat menjadi solusi yang efektif untuk membantu manajemen dalam mengambil keputusan terkait strategi penjualan dan pengelolaan stok produk.

Penelitian yang dilakukan oleh [4] membahas tentang komparasi Algoritma K-Nearest Neighbors Dan Support Vector Machines Dalam Prediksi Layanan Produk ICONNET. Penelitian ini mengimplementasikan dan membandingkan algoritma K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM) dalam memprediksi layanan produk Iconnet yang paling diminati oleh pelanggan, sehingga dapat mempermudah manajemen ICON+ dalam perencanaan penyediaan layananan pelanggan, dan penyusunan strategi pasar di masa mendatang. Sebanyak 3.206 dataset (terdiri atas 2.565 data training dan 641 data testing) peminat layanan ICONNET selama 1 tahun yang telah dibersihkan, diuji pada kedua algoritma tersebut, berdasarkan parameter Bandwidth, Request_Date, Status, alamat pelanggan, dan biaya layanan. Akurasi kinerja algoritma diuji menggunakan metode Cross Validation, Confusion Matrix serta ROC curve. Hasil uji akurasi menunjukkan kinerja algoritma K-NN lebih akurat dari algoritma SVM pada berbagai kategori pengujian.

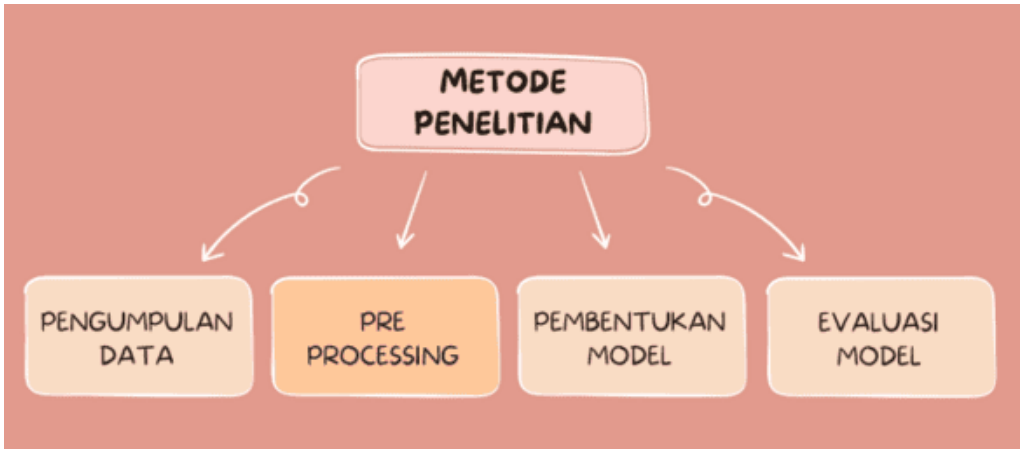
Penelitian yang telah dilakukan oleh [2][3][4] tersebut bertujuan membandingkan performa algoritma KNN dan SVM dalam memprediksi penjualan. Pada penelitian ini memiliki tujuan untuk membandingkan performa algoritma KNN dan SVM dalam memprediksi tingkat kepuasan pelanggan warung makan. Penelitian dilakukan melalui empat tahapan utama, yaitu: (1) pengumpulan data dari kaggle (2) pemrosesan data dengan melakukan preprocessing diantara mencari missing value dll. (3) pembentukan model menggunakan algoritma KNN dan SVM, dan

e-ISSN 2827-8356
p-ISSN 2827-9999

(4) evaluasi model menggunakan metrik akurasi, precision, recall, dan f1-score.. Hasil dari penelitian ini diharapkan dapat memberikan wawasan dalam memilih algoritma yang tepat dalam konteks prediksi kepuasan pelanggan, serta membantu pelaku usaha Warmindo untuk meningkatkan kualitas layanan berdasarkan hasil prediksi tersebut.

2. METODE PENELITIAN

Metode penelitian melibatkan beberapa tahapan diantaranya yaitu, Pra-pemprosesan data, permodelan, evaluasi model, dan prediksi. Penelitian ini dapat dilihat pada Gambar 1. Setiap Langkah dalam alur proses sangat penting untuk dilaksanakan. Penjelasan rinci mengenai masing-masing proses akan disampaikan dalam bab ini, Berikut adalah Gambar 1 yang menunjukan alur metodologi penelitian.



Gambar 1. Metode Penelitian

2.1 Pengumpulan data

Langkah awal dalam penelitian ini adalah pengumpulan data, yang menjadi fondasi utama dalam proses pembentukan model klasifikasi dan evaluasi hasil. Data dalam penelitian ini diperoleh melalui website kaggle [5] (kuesioner online Warung Makan). Dataset tersebut berisi 4 atribut yaitu pelayanan, kebersihan, kualitas makanan, dan tingkat kepuasan secara keseluruhan. Data dikumpulkan dalam format tabel dengan kolom: Timestamp, Nama, Pendapat tentang pelayanan, kebersihan, kualitas makanan, serta kepuasan secara keseluruhan. Informasi ini kemudian diproses untuk membentuk dataset yang digunakan sebagai input dalam pemodelan menggunakan algoritma klasifikasi K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM).

2.2 Pre processing

Tahapan preprocessing dilakukan untuk mempersiapkan data agar layak digunakan dalam proses pelatihan dan pengujian model. Preprocessing bertujuan untuk meningkatkan kualitas data, mengatasi ketidakseimbangan kelas, serta memastikan setiap fitur berada dalam format yang sesuai untuk diproses oleh algoritma klasifikasi. Tahapan ini menjadi krusial karena kualitas data sangat mempengaruhi performa model secara keseluruhan [6][7]. Pertama, dilakukan penghapusan atribut yang tidak relevan. Selanjutnya, dilakukan pengecekan nilai kosong (missing values) pada setiap kolom guna memastikan integritas data. proses berikutnya adalah transformasi data kategorik menjadi numerik menggunakan teknik label encoding. Setelah data bersih diperoleh, dilakukan pemisahan data menjadi data latih dan data uji menggunakan metode train-test split dengan rasio 80:20. Namun, berdasarkan distribusi awal label target, ditemukan bahwa

e-ISSN 2827-8356
p-ISSN 2827-9999

data memiliki ketidakseimbangan kelas (class imbalance). Oleh karena itu, dilakukan penyeimbangan data menggunakan metode SMOTE (Synthetic Minority Over-sampling Technique), yang secara sintetik menghasilkan data pada kelas minoritas sehingga distribusi menjadi lebih merata[8]. Tahap akhir preprocessing adalah standarisasi fitur menggunakan metode Standard Scaler dengan mengubah skala setiap fitur agar memiliki nilai rata-rata nol dan standar deviasi satu. tahapan preprocessing yang sistematis, dataset yang semula berupa data mentah dari hasil survei pelanggan diolah menjadi data bersih, seimbang, dan siap digunakan untuk pembentukan model klasifikasi.

2.3 Pembentukan Model

Pada penelitian ini, digunakan dua algoritma supervised learning, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Pemilihan kedua algoritma ini didasarkan pada kemampuannya dalam menangani masalah klasifikasi multi kelas serta telah banyak digunakan dalam studi prediksi kepuasan pelanggan.

a. Algoritma K-Nearest Neighbor (KNN)

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi yang bekerja berdasarkan prinsip kedekatan jarak antar data. Suatu data baru akan diklasifikasikan berdasarkan mayoritas kelas dari k=5 tetangga terdekatnya. Dalam penelitian ini, nilai k ditentukan melalui proses tuning parameter untuk memperoleh akurasi terbaik[9]. KNN memiliki keunggulan dalam kesederhanaan implementasi serta tidak memerlukan proses pelatihan eksplisit, karena termasuk algoritma lazy learning. KNN sangat bergantung pada normalisasi data dan sensitif terhadap fitur yang memiliki skala berbeda, sehingga tahap standarisasi sebelumnya menjadi krusial untuk menjaga performa algoritma.

b. Algoritma Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang mencari hyperplane optimal yang memisahkan data ke dalam kelas-kelas yang berbeda dengan margin maksimum. SVM bekerja dengan menggunakan fungsi kernel, dan dalam penelitian ini digunakan kernel linier sebagai pendekatan awal, mengingat data bersifat kategorikal yang telah di encoding menjadi numerik. Model SVM dibangun menggunakan parameter default, namun dapat dilakukan optimasi parameter seperti C dan gamma untuk meningkatkan performa[10].

2.4 Evalusi Model

Evaluasi model dilakukan dengan menggunakan data uji yang telah dipisahkan pada tahap sebelumnya. bertujuan untuk mengukur sejauh mana model mampu menggeneralisasi terhadap data baru yang belum pernah dilatihkan sebelumnya. dilakukan menggunakan beberapa metrik performa meliputi akurasi, precision, recall, dan F1-score. Selain itu, digunakan pula Confusion Matrix atau alat bantu evaluasi berdasarkan kategori aktual untuk menganalisis hasil klasifikasi secara lebih rinci[11]. Tabel berikut menjelaskan struktur dasar confusion matrix:

Tabel 1. Struktur Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Kelas Positif	TP (True Positive)	FN (False Negative)
kelas Negatif	FP (False Positive)	TN (True Negative)

e-ISSN 2827-8356
p-ISSN 2827-9999

Berdasarkan nilai-nilai dalam confusion matrix tersebut, beberapa metrik evaluasi utama dapat dihitung sebagai berikut:

a. Akurasi

Akurasi mengukur proporsi total prediksi yang benar terhadap seluruh jumlah data. Metrik ini menunjukkan sejauh mana model mampu mengklasifikasi data dengan benar secara keseluruhan[12].

$$\text{Akurasi} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \times 100\% \qquad (1)$$

b. Presisi

Precision mengukur ketepatan prediksi positif yang dihasilkan model, yaitu seberapa banyak dari prediksi positif yang benar-benar relevan[12].

$$\text{Presisi} = \frac{\text{TP}}{\text{TP}+\text{FP}} \qquad (2)$$

c. Recall

Recall menggambarkan sensitivitas model, yaitu kemampuan model dalam menangkap semua data yang sebenarnya positif[12].

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

d. F1-Score

F1-Score merupakan rata-rata harmonis antara precision dan recall. Metrik ini digunakan untuk memberikan penilaian yang seimbang antara ketepatan dan sensitivitas model[12].

$$\text{F1-Score} = 2\,\frac{(\text{Presisi} \times \text{Recall})}{\text{Presisi}+\text{Recall}} \qquad (4)$$

Evaluasi ini penting dilakukan untuk memastikan bahwa model yang dibangun tidak hanya memiliki akurasi tinggi secara keseluruhan, tetapi juga mampu mengenali distribusi data dari setiap kelas dengan baik, khususnya dalam kasus ketidakseimbangan kelas.

3. HASIL DAN PEMBAHASAN

3.1 Pembentukan Model

Tahap awal dalam penelitian ini adalah pengumpulan data, yang menjadi dasar dalam proses analisis dan pembentukan model prediksi. Data yang digunakan dalam penelitian ini diperoleh melalui kuesioner Google Form yang disebarkan kepada responden yang pernah berkunjung dan menggunakan layanan Warung Makan Prima. Dataset yang dihasilkan berisi tanggapan pelanggan mengenai beberapa aspek pelayanan dan kualitas, yang nantinya digunakan untuk memprediksi tingkat kepuasan secara keseluruhan. Dataset ini terdiri dari 4 atribut utama

e-ISSN 2827-8356
p-ISSN 2827-9999

yaitu pelayanan , kebersihan, kualitas makanan dan Tingkat kepuasan. Dataset tersebut dapat dilihat pada tabel 2.

Tabel 2. Tabel Dataset

Timestamp	Nama	Bagaimana pendapat anda mengenai pelayanan di warung makan Prima?	Bagaimana pendapat anda tentang kebersihan di warung makan Prima?	Bagaimana pendapat anda tentang kualitas makanan di warung makan Prima	Bagaimana tingkat kepuasan anda terhadap di warung makan Prima?
11/12/2023 21:50:00	Nelti	Cukup ramah	Cukup bersih	Cukup berkualitas	Sangat puas
11/12/2023 22:00:19	Nur haziza	Cukup ramah	Cukup bersih	Cukup berkualitas	Cukup puas
11/12/2023 22:15:31	Nelfi Juliani Safera	Sangat ramah	Cukup bersih	Cukup berkualitas	Cukup puas
12/12/2023 4:53:06	Mursawal	Cukup ramah	Cukup bersih	Cukup berkualitas	Cukup puas
12/12/2023 4:55:55	Afdal	Cukup ramah	Cukup bersih	Kurang berkualitas	Cukup puas

3.2 Pre Processing

Pada tahap awal, dilakukan penghapusan atribut yang tidak relevan, yaitu kolom Timestamp dan Nama, karena tidak memberikan kontribusi terhadap proses prediksi. Selanjutnya,

e-ISSN 2827-8356
p-ISSN 2827-9999

dilakukan pengecekan nilai kosong (missing values) pada setiap kolom guna memastikan integritas data. Proses tersebut dapat dilihat pada gambar 2.

```
# Menghapus kolom yang tidak relevan
df = df.drop(columns=['Timestamp', 'Nama'])
```

Gambar 2. Penghapusan Kolom

Proses berikutnya adalah teknik *label encoding* mengubah ke dalam bentuk numerik untuk memungkinkan algoritma machine learning memproses data tersebut secara optimal.

```
# Encoding categorical features jika ada
label_encoders = {}
for column in df.columns:
    le = LabelEncoder()
    df[column] = le.fit_transform(df[column])
    label_encoders[column] = le
```

Gambar 3. Encoding Data

Setelah data bersih diperoleh, dilakukan pemisahan data menjadi data latih dan data uji menggunakan metode *train-test split* dengan rasio 80:20.

```
# Membagi data menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 4. Pembagian Data

Sebagai tahap akhir preprocessing, dilakukan standarisasi fitur menggunakan metode *Standard Scaler*. Proses ini mengubah skala setiap fitur agar memiliki nilai rata-rata nol dan standar deviasi satu.

```
# Standarisasi fitur
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Gambar 5. Standarisasi Fitur

3.3 Pembentukan Model

Setelah melalui tahap preprocessing, data siap untuk digunakan dalam proses pembentukan model klasifikasi. Pada penelitian ini, digunakan dua algoritma supervised learning, yaitu Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). Pemilihan kedua algoritma ini didasarkan pada efektivitasnya dalam menyelesaikan permasalaha

a. Algoritma K-Nearest Neighbors (KNN)

Algoritma KNN bekerja berdasarkan prinsip kedekatan jarak (distance-based) antara data baru dan data pelatihan. Dalam implementasi ini, digunakan nilai k = 5, sebagaimana diatur dalam parameter model KNeighborsClassifier(n_neighbors=5). Pemilihan nilai ini mempertimbangkan keseimbangan antara kompleksitas model dan akurasi prediksi. Proses pembelajaran KNN dilakukan dengan metode lazy learning, di mana model tidak melakukan generalisasi selama

e-ISSN 2827-8356
p-ISSN 2827-9999

pelatihan, melainkan hanya menyimpan data pelatihan dan melakukan klasifikasi saat proses inferensi. Oleh karena itu, proses standarisasi data sebelumnya menjadi penting untuk mencegah dominasi fitur yang memiliki skala besar dalam perhitungan jarak.

```
# Membuat model KNN
knn_model = KNeighborsClassifier(n_neighbors=5)
knn_model.fit(X_train, y_train)
```

```
KNeighborsClassifier
KNeighborsClassifier()
```

```
# Prediksi
y_pred = knn_model.predict(X_test)
```

Gambar 6. Pembentukan Model KNN

b. Algoritma Support Vector Machine (SVM)

Algoritma Support Vector Machine (SVM) bertujuan untuk menemukan hyperplane optimal yang dapat memisahkan data ke dalam kelas yang berbeda dengan margin maksimum. Dalam penelitian ini, digunakan fungsi kernel Radial Basis Function (RBF) untuk menangani data non-linear, sebagaimana didefinisikan dalam parameter kernel='rbf'. Model SVM dibangun dengan parameter default, yaitu C=1.0 dan gamma='scale'.

```
# Membuat model SVM
svm_model = SVC(kernel='rbf', C=1.0, gamma='scale')
svm_model.fit(X_train, y_train)
```

```
SVC
SVC()
```

```
# Prediksi
y_pred = svm_model.predict(X_test)
```

Gambar 7. Pembentukan Model SVM

3.4 Evaluasi Model

Evaluasi kinerja model dalam penelitian ini dilakukan menggunakan confusion matrix, dengan mengukur metrik akurasi, presisi, recall, dan F1-score. Visualisasi hasil evaluasi model disajikan pada Gambar 8 untuk algoritma KNN dan gambar 9 untuk algoritma SVM.

e-ISSN 2827-8356
p-ISSN 2827-9999

```
# Confusion Matrix
conf_matrix = confusion_matrix(y_test, y_pred)
conf_matrix
```

```
array([[53,  5,  4,  0],
       [12,  6,  0,  0],
       [ 7,  0, 13,  0],
       [ 1,  2,  0,  2]])
```

Gambar 8. Hasil Algoritma KNN

```
# Confusion Matrix
conf_matrix = confusion_matrix(y_test, y_pred)
conf_matrix
```

```
array([[53,  6,  3,  0],
       [11,  7,  0,  0],
       [ 9,  0, 11,  0],
       [ 1,  2,  0,  2]])
```

Gambar 9. Hasil Algoritma SVM

Pada Gambar 9 ditampilkan hasil confusion matrix yang dihasilkan dari proses evaluasi model klasifikasi terhadap empat kelas tingkat kepuasan pelanggan. Matriks kebingungan ini dihitung menggunakan fungsi confusion_matrix() dari pustaka scikit-learn yang berfungsi untuk membandingkan antara label aktual (y_test) dengan label prediksi (y_pred) yang dihasilkan oleh model klasifikasi. Matriks berukuran 4×4 ini merepresentasikan empat kelas yang berbeda. Baris menunjukkan kelas aktual, sedangkan kolom menunjukkan prediksi model. Sebagai contoh, pada kelas 0, sebanyak 53 data berhasil diklasifikasikan dengan benar, sementara 5 data salah diklasifikasikan sebagai kelas 1 dan 4 data diklasifikasikan sebagai kelas 2. Untuk kelas 1, hanya 6 data yang diklasifikasikan dengan benar, sementara 12 lainnya salah diklasifikasikan sebagai kelas 0. Pada kelas 2, terdapat 13 prediksi yang benar dan 7 yang salah diklasifikasikan ke kelas 0. Sedangkan pada kelas 3, hanya 2 data yang diklasifikasikan dengan benar, dengan kesalahan klasifikasi sebanyak 3 data (1 ke kelas 0 dan 2 ke kelas 1).

Sedangkan Pada Gambar 10 ditampilkan hasil *confusion matrix* yang dihasilkan dari proses evaluasi model klasifikasi menggunakan fungsi confusion_matrix() dari pustaka *scikit-learn*. Matriks ini menyajikan perbandingan antara nilai aktual (y_test) dengan hasil prediksi (y_pred) dari model yang diuji. Dalam penelitian ini, *confusion matrix* berukuran 4×4 karena terdapat empat kelas kategori kepuasan pelanggan yang digunakan sebagai target klasifikasi.

Baris dalam matriks menunjukkan kelas aktual, sedangkan kolom menunjukkan kelas hasil prediksi model. Misalnya, pada kelas 0 (kelas dominan), model berhasil mengklasifikasikan 53 data dengan benar sebagai kelas 0. Namun, terdapat 6 data dari kelas 0 yang salah diklasifikasikan sebagai kelas 1, dan 3 data lainnya sebagai kelas 2. Untuk kelas 1, model mengklasifikasikan 7 data dengan benar, namun terdapat kesalahan klasifikasi sebanyak 11 data yang diprediksi sebagai kelas 0. Hal serupa juga terjadi pada kelas 2, dengan 11 prediksi benar dan 9 salah prediksi sebagai kelas 0. Sementara itu, kelas 3 hanya diklasifikasikan dengan benar sebanyak 2 data, dan 3 data lainnya salah diklasifikasikan ke kelas 0 dan kelas 1.

Tabel 3. Perbandingan Performa Algoritma KNN dan SVM

e-ISSN 2827-8356
p-ISSN 2827-9999

Kelas	Presisi KNN	Recall KNN	F1-Score KNN	Presisi SVM	Recall SVM	F1-Skore SVM
0	0,716	0,855	0,780	0,720	0,850	0,780
1	0,465	0,389	0,420	0,470	0,390	0,420
2	0,789	0,550	0,650	0,790	0,550	0,650
3	1,000	0,400	0,571	1,000	0,400	0,570

4. KESIMPULAN

Berdasarkan hasil pengujian terhadap dua algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM), dapat disimpulkan bahwa keduanya menunjukkan performa yang cukup baik dalam mengklasifikasikan tingkat kepuasan pelanggan, dengan nilai akurasi keseluruhan sebesar 70%. Kedua algoritma layak digunakan sebagai pendekatan awal dalam membangun sistem prediksi kepuasan pelanggan berbasis data, khususnya pada kelas mayoritas. Secara khusus, algoritma KNN menunjukkan kinerja terbaik pada kelas 0 (pelanggan puas), dengan nilai precision sebesar 71,6% dan recall sebesar 85,5%. Hal ini mengindikasikan bahwa KNN mampu mengklasifikasikan pelanggan puas dengan baik. Namun, performa KNN menurun pada kelas minoritas seperti kelas 1 (kurang puas), yang ditunjukkan oleh nilai precision dan recall yang lebih rendah. Pola yang sama juga ditunjukkan oleh algoritma SVM, yang mencatat precision sebesar 72% dan recall 85% pada kelas 0, tetapi menunjukkan penurunan kinerja pada kelas lainnya, khususnya kelas 1 dan kelas 3. Nilai F1-score makro rata-rata untuk SVM sebesar 0,61 dan untuk KNN sebesar 0,59, menunjukkan bahwa SVM memiliki sedikit keunggulan dalam keseimbangan antara precision dan recall di semua kelas. Kelemahan utama yang diidentifikasi dari kedua algoritma adalah kurang optimalnya performa dalam memprediksi kelas minoritas. Hal ini disebabkan oleh ketidakseimbangan distribusi data antar kelas, yang mengakibatkan bias terhadap kelas mayoritas. Oleh karena itu, disarankan untuk melakukan peningkatan performa model melalui penerapan teknik balancing data seperti oversampling (misalnya SMOTE) atau undersampling, serta tuning parameter yang lebih optimal baik untuk algoritma KNN (penentuan nilai k) maupun SVM (penyesuaian parameter kernel, C, dan gamma). Sebagai batasan penelitian ini, penggunaan dataset yang relatif kecil dan tidak seimbang menjadi salah satu faktor yang mempengaruhi akurasi model dalam memprediksi kelas-kelas tertentu. Selain itu, belum dilakukan validasi silang (cross-validation) serta optimasi parameter secara menyeluruh. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi algoritma lain seperti Random Forest, Gradient Boosting, atau pendekatan berbasis Deep Learning, serta menerapkan teknik evaluasi tambahan guna meningkatkan kemampuan generalisasi model dalam konteks klasifikasi kepuasan pelanggan.

DAFTAR PUSTAKA

[1] Y. Maulina, A. Gunaryati, and R. T. Aldisa, “Sistem Pakar Diagnosis Awal Penyakit Anemia Menggunakan Metode Naïve Bayes dan Certainty Factor,” STRING (Satuan Tulisan Ris. dan Inov. Teknol., vol. 8, no. 1, p. 110, 2023, doi:10.30998/string.v8i1.16468.

[2] I. H. Pratama and U. Salamah, “Perbandingan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Menentukan Prediksi Produk-Produk Terlaris Pada Toko Madura Kecamatan Pondok Aren,” *J. Tek. Inform. Kaputama*, vol. 6, no. 2, pp. 846–858, 2022, [Online]. Available: https://jurnal-backup.kaputama.ac.id/index.php/JTIK/article/view/1163

[3] A. M. W. S. A and Z. Fatah, “Prediksi Produk Penjualan Di Supermarket Dengan Metode Algoritma K-Nearest Neighbors (KNN),” vol. 3, no. 1, 2025.

e-ISSN 2827-8356
p-ISSN 2827-9999

[4] I. W. A. Purnawibawa, I. N. Purnama, and I. N. Y. A. Wijaya, "Komparasi Algoritme K-Nearest Neighbors Dan Support Vector Machines Dalam Prediksi Layanan Produk ICONNET," *Progresif J. Ilm. Komput.*, vol. 18, no. 2, p. 271, 2022, doi: 10.35889/progresif.v18i2.894.

[5] "Kaggle: Your Machine Learning and Data Science Community," *Kaggle.com*, 2025. https://www.kaggle.com/ (accessed Jun. 03, 2025).

[6] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[7] J. Muliawan and E. Dazki, "Sentiment Analysis of Indonesia'S Capital City Relocation Using Three Algorithms: Naïve Bayes, Knn, and Random Forest," *J. Tek. Inform.*, vol. 4, no. 5, pp. 1227–1236, 2023, doi: 10.52436/1.jutif.2023.4.5.1436.

[8] W. A. Ridho, T. Wuryandari, and A. R. Hakim, "Perbandingan Kinerja Metode Klasifikasi K-Nearest Neighbor Dan Support Vector Machines Pada Dataset Parkinson," *J. Gaussian*, vol. 12, no. 3, pp. 372–381, 2024, doi: 10.14710/j.gauss.12.3.372-381.

[9] W. A. Ridho, T. Wuryandari, and A. R. Hakim, "Perbandingan Kinerja Metode Klasifikasi K-Nearest Neighbor Dan Support Vector Machines Pada Dataset Parkinson," *J. Gaussian*, vol. 12, no. 3, pp. 372–381, 2024, doi: 10.14710/j.gauss.12.3.372-381.

[10] M. Utami and V. Ayumi, "Prediksi Penjualan Produk Terlaris Menggunakan Algoritma K-Nearest Neighboard (KNN)," pp. 43–47, 2024.

[11] G. Rahmawati, S. A. Sanmas, E. Nudyawati, and N. D. Syaharani, "Studi Perbandingan Performa : Prediksi Status Stunting Pada Anak Berdasarkan Data Antropometri Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN)," vol. 2024, no. Senada, pp. 782–790, 2024.

[12] S. A. Lashari, M. M. Khan, A. Khan, and S. Salahuddin, "Comparative Evaluation of Machine Learning Models for Mobile Phone Price Prediction : Assessing Accuracy , Robustness , and Generalization Performance," vol. 3, no. 3, 2024.

[13] M. Dava, R. Fajar, D. Puspitasari, and Q. N. Azizah, "K-Nearest Neighbor and Naive Bayes Algorithm Approach for Online Sales Level Classification Optimization In South Tangerang," vol. 12, no. 3, pp. 663–668, 2024.

[14] F. Pritama, E. Rueh, D. Leluni, and J. Parhusip, "Analisis Distribusi Kinerja SVM dan KNN Berdasarkan Rata Rata Simpangan Baku dan Stabilitas Analysis of SVM and KNN Performance Distribution Based on Average Standard Deviation and Stability," vol. 1, 2024.

[15] M. Muchtar and R. A. Muchtar, "Perbandingan Metode Knn Dan Svm Dalam Klasifikasi Kematangan Buah Mangga Berdasarkan Citra Hsv Dan Fitur Statistik," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 876–884, 2024, doi: 10.23960/jitet.v12i2.4010.

[16] M. Fadawkas Oemarki, M. Dimas Mufti Baskara, and I. Ernawati, "Perbandingan Akurasi Metode Support Vector Machine Dan K-Nearest Neighbour Dalam Prediksi Curah Hujan Potensi Banjir," no. April, pp. 160–167, 2024, [Online]. Available: https://dataonline.bmkg.go.id/home.

[17] A. Sharma, M. Dutta, and R. Prakash, "Comparative Performance Analysis of Machine Learning Algorithms for COVID-19 Cases in India," *Commun. Comput. Inf. Sci.*, vol. 1929, no. 11, pp. 243–257, 2024, doi: 10.1007/978-3-031-48774-3_17.

[18] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.

[19] Amanda Pratiwi, Ananto Tri Sasongko, and D. K. Pramudito, "Analisis Prediksi Gilingan Plastik Terlaris Menggunakan Algoritma K-Nearest Neighbor Di Cv Menembus Batas,"

e-ISSN 2827-8356
p-ISSN 2827-9999

J. Inform. Teknol. dan Sains, vol. 5, no. 3, pp. 437–445, 2023, doi: 10.51401/jinteks.v5i3.3323.

[20] V. Sariayu and P. Sugiartawan, "Analisis Prediksi Penjualan Lampu Dengan Metode Svm Pada PT. Terang Abadi Raya," *J. Sist. Inf. dan Komput. Terap. Indones.*, vol. 5, no. 1, pp. 1–10, 2022, doi: 10.33173/jsikti.172.