



Perbandingan Kinerja Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) dalam Klasifikasi Stunting

Salma Fathiyatur Rizky Munir, Yisti Vita Via, Eka Prakarsa Mandyartha*

Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

Email: ¹salmafathiyatur17@gmail.com, ²yistivita.if@upnjatim.ac.id, ^{3,*}eka_prakarsa.fik@upnjatim.ac.id

Email Penulis Korespondensi: eka_prakarsa.fik@upnjatim.ac.id

Abstrak—Evaluasi dan perbandingan terhadap performansi algoritma K-Nearest Neighbor (KNN) serta Support Vector Machine (SVM) menjadi fokus utama yang ingin dicapai dalam penelitian ini, khususnya dalam mengklasifikasikan kondisi stunting pada kelompok balita. Dataset dalam penelitian merupakan data primer yang diperoleh dari Puskesmas Patianrowo, Kabupaten Nganjuk, dengan jumlah awal 1102 data anak usia 0–60 bulan. Setelah langkah pembersihan data, nilai yang tidak lengkap (missing value) dihapus, jumlah data yang digunakan menjadi 1067. Selanjutnya, data dibagi menjadi 853 data latih dan 214 data uji menggunakan metode split train-test dengan perbandingan 80:20. Tahap praproses meliputi penghapusan data kosong, transformasi label ke bentuk numerik, serta normalisasi menggunakan metode min-max scaling untuk menyamakan rentang nilai fitur. Untuk mengevaluasi model, confusion matrix digunakan dengan metrik akurasi, presisi, recall, dan skor F1-score. Hasil pengujian menunjukkan bahwa algoritma KNN dengan nilai K=5 menghasilkan akurasi sebesar 96,72%, presisi 91,25%, recall 67,73%, dan F1-score 77,52%. Sementara itu, Algoritma SVM polynomial menunjukkan peningkatan kinerja dengan akurasi 97,47%, presisi 90,82%, recall 78,96%, dan F1-score 82,55%. Berdasarkan data tersebut, SVM dinilai lebih efektif dalam mengklasifikasikan stunting pada dataset yang digunakan.

Kata Kunci: Klasifikasi; KNN; Stunting; SVM; Machine Learning

Abstract—The evaluation and comparison of the performance of the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms are the primary objectives of this study, particularly in classifying stunting conditions among toddlers. The dataset in this study consists of primary data obtained from the Patianrowo Community Health Center in Nganjuk Regency, with an initial total of 1,102 data points on children aged 0–60 months. After data cleaning, missing values were removed, reducing the dataset to 1,067 data points. Subsequently, the data was divided into 853 training data points and 214 test data points using the train-test split method with an 80:20 ratio. The preprocessing stage included removing missing data, transforming labels into numerical form, and normalizing the data using the min-max scaling method to standardize the feature value ranges. To evaluate the model, a confusion matrix was used with the metrics of accuracy, precision, recall, and F1-score. The test results showed that the KNN algorithm with K=5 produced an accuracy of 96.72%, precision of 91.25%, recall of 67.73%, and an F1-score of 77.52%. Meanwhile, the polynomial SVM algorithm demonstrated improved performance with an accuracy of 97.47%, precision of 90.82%, recall of 78.96%, and an F1-score of 82.55%. Based on these results, SVM is considered more effective in classifying stunting in the dataset used.

Keywords: Classification; KNN; Stunting; SVM; Machine Learning

1. PENDAHULUAN

Salah satu masalah kekurangan gizi yang terjadi pada balita adalah stunting. Selama 1000 hari pertama kehidupan bayi, kondisi kesehatannya harus diperhatikan. Di Indonesia, stunting pada balita masih merupakan masalah kesehatan yang serius. Stunting adalah kondisi kegagalan tumbuh kembang yang disebabkan oleh kekurangan dan berdampak jangka panjang pada anak (Lonang & Normawati, 2022). Stunting pada balita menyebabkan pertumbuhan fisik yang terhambat. Pendidikan pencegahan stunting diperlukan dari kasus stunting yang ada. Untuk mendapatkan kualitas hidup yang baik, pertumbuhan balita merupakan bagian penting dari kehidupan karena berdampak pada pertumbuhan fisik, mental, dan emosional. Secara umum, stunting diidentifikasi dengan mengukur ukuran antropometri seperti tinggi, berat badan, dan usia, lalu dibandingkan dengan standar pertumbuhan anak. Namun, metode manual yang masih digunakan di fasilitas kesehatan dapat mempersulit deteksi stunting dan menyebabkan kesalahan dalam diagnosis, sehingga diperlukan penggunaan teknologi untuk meningkatkan efisiensi dan ketepatan (Junaidi, Nurdin, & Meiyanti, 2025).

Dengan kemajuan teknologi, machine learning semakin populer dalam klasifikasi dan prediksi kondisi kesehatan. Machine learning memungkinkan sistem untuk membuat prediksi yang sangat akurat dan belajar dari pola data. Untuk memprediksi status gizi secara otomatis dengan memproses data antropometri balita, machine learning dianggap efektif (Hardiani & Putri, 2021). Beberapa faktor, termasuk berat badan, usia, dan tinggi badan, digunakan dalam proses klasifikasi stunting dan tidak stunting. KNN dan SVM adalah dua metode yang umum digunakan dalam proses klasifikasi untuk mengklasifikasikan status gizi, karena keduanya memiliki kemampuan untuk mengolah data yang kompleks dan menghasilkan hasil klasifikasi yang baik (Rahmawati, Sanmas, Nudyawati, & Syaharani, 2024). KNN adalah metode klasifikasi yang bergantung pada kedekatan jarak antar data, sehingga data dengan karakteristik serupa cenderung berada pada kelas yang sama. Namun, pemilihan parameter K dan skala data memengaruhi kinerja KNN, sehingga untuk mendapatkan hasil yang optimal, diperlukan proses normalisasi dan tuning parameter. K-Nearest Neighbor (KNN), yang terbukti memiliki keakuratan tinggi hingga 97% dalam klasifikasi status gizi balita, adalah salah satu metode yang paling umum digunakan (A'la, Homaidi, & Lidimillah, 2025). Sementara itu, SVM berfungsi dengan menemukan hyperplane yang paling efektif untuk memisahkan dua kelompok data dengan ruang yang lebih besar, sehingga mereka dapat mengatasi data berukuran besar dan menjelaskan bagaimana hubungan SVM bekerja dengan

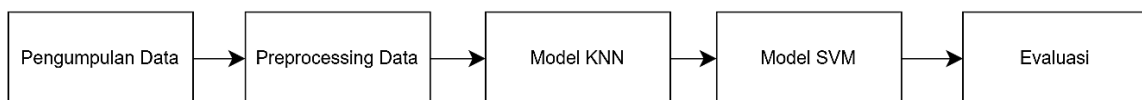
menemukan batas yang paling efektif untuk memisahkan dua kelompok data dengan lebih banyak ruang, sehingga mereka dapat mengatasi data berukuran besar dan juga dapat menggambarkan hubungan yang tidak lurus dengan baik. (Nadroh, Triwibowo, & Sumantri, 2024).

Studi sebelumnya telah membandingkan algoritma machine learning dalam mengklasifikasikan stunting. Studi yang membandingkan SVM, KNN, dan Naive Bayes menunjukkan bahwa SVM paling akurat dalam menangani data yang rumit (Dwinanto, Sandi A, & Ardianto, 2024). Studi tambahan menunjukkan bahwa, tergantung pada pengaturan hyperparameter yang digunakan dan kualitas data yang dimiliki, kinerja KNN dan SVM bias memiliki tingkat akurasi yang sama. Studi juga menekankan betapa pentingnya penerapan parameter penilaian yang sesuai, seperti F1-Score, untuk menilai performa model dengan cara yang lebih adil (Hidayat, Sembiring, Purnomo, & Iriani, 2025). Selain itu, penelitian tambahan yang membandingkan algoritma KNN, Naive Bayes, dan SVM menunjukkan bahwa Naive Bayes memiliki tingkat akurasi terbaik sebesar 99,65%, diikuti oleh SVM dengan 99,96%, dan KNN hanya memiliki 96,36% akurasi (Alamin & Aryani, 2025). Hasil penelitian ini menunjukkan bahwa karakteristik data, persebaran data, preprocessing data, dan parameter yang digunakan dalam proses pelatihan model sangat memengaruhi kinerja algoritma KNN dan SVM. Variasi dalam karakteristik data menyebabkan hasil algoritma KNN dan SVM berbeda. Oleh karena itu, perlu dilakukan pengujian tambahan dengan dataset dengan karakteristik yang berbeda seperti yang dijumpai dalam penelitian ini. Dalam penelitian ini, menggunakan KNN dan SVM karena keduanya memiliki cara yang berbeda untuk mengklasifikasikan data. KNN mengklasifikasikan berdasarkan seberapa dekat jarak antar data, sedangkan SVM menggunakan batasan yang disebut hyperplane untuk membedakan kelas. Sangat menarik untuk membandingkan cara kedua algoritma ini bekerja dalam klasifikasi stunting. Berdasarkan hal tersebut, melalui penelitian ini akan dilakukan penentuan algoritma yang lebih optimal dalam menangani data antropometri balita.

Penelitian ini dirancang untuk mempelajari bagaimana algoritma KNN dan SVM berfungsi saat menggunakan data antropometri balita untuk melakukan klasifikasi stunting. Untuk menghasilkan model klasifikasi terbaik, penelitian ini juga menguji algoritma KNN untuk parameter nilai K dan algoritma SVM untuk beberapa kernel. Fokus penelitian ini adalah memberikan analisis perbandingan kinerja antara algoritma KNN dan SVM menggunakan nilai akurasi, presisi, recall, dan skor F1-score. Oleh karena itu, diharapkan penelitian ini dapat dijadikan referensi dalam pembuatan sistem klasifikasi stunting yang berbasis machine learning.

2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan melalui berbagai tahap yang ditampilkan dalam urutan pada Gambar 1 berikut:



Gambar 1. Metode Penelitian

Dari Gambar 1, dapat dijelaskan sebagai berikut:

a. Pengumpulan Data

Penelitian ini memanfaatkan informasi data yang diperoleh dari puskesmas Patianrowo yang berkaitan dengan stunting pada anak-anak berusia 0 hingga 60 bulan. Total data yang digunakan berjumlah 1102. Informasi dalam dataset mencakup nama, gender, umur, berat badan, tinggi badan, Z Score (TB/U), Z Score (BB/U), Z Score (BB/TB), Status Gizi (TB/U), Status Gizi (BB/U), Status Gizi (BB/TB), dan informasi tentang status gizi yang menunjukkan kondisi gizi anak.

b. Preprocessing Data

Tahapan preprocessing dikenal sebagai proses mengubah data awal menjadi data yang sudah siap digunakan. Penelitian ini melakukan beberapa hal. Agar hasilnya lebih rapi, pertama dilakukan pembersihan data, yaitu menghapus baris dengan data kosong. Pada awalnya, data berjumlah 1102, tetapi kemudian berkurang menjadi 1067 setelah pembersihan. Selanjutnya, label yang menunjukkan tidak stunting diubah menjadi 0 dan label yang menunjukkan stunting menjadi 1. Selain itu, normalisasi dilakukan untuk membuat skala data seimbang, yang berarti bahwa semua data berada dalam rentang antara 0 dan 1. Metode normalisasi min-max digunakan, yang dapat ditemukan dalam rumus berikut:

$$\frac{x - \min}{\max - \min} \quad (1)$$

c. Model KNN

Algoritma K-Nearest Neighbors (KNN) adalah metode klasifikasi supervised learning yang digunakan untuk menentukan kategori baru pada data yang sudah diberi label. Algoritma ini sangat baik dalam memprediksi stunting dan dapat meningkatkan akurasi dengan memilih fitur yang tepat (Drajana & Bode, 2022). Algoritma ini menggunakan nilai K dari semua tetangga terdekat dan kemudian mengklasifikasikannya berdasarkan mayoritas tetangga yang ada untuk menentukan kategori baru. Metode klasifikasi KNN menentukan kelas data berdasarkan jarak terdekat dengan data tetangga yang paling dekat. Untuk mengklasifikasikan status gizi anak balita, metode ini sering digunakan dengan menggunakan ukuran data antropometri (Saleh, Faisal, & Musa, 2019). Dalam proses



KNN, ada langkah perhitungan jarak antara data uji dan data latih. Jarak ini dihitung menggunakan Jarak *Eunclidean* dengan rumus berikut:

$$D_{x,y} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{2}$$

Kinerja KNN sangat tergantung pada proses normalisasi data agar jarak antara fitur bisa dihitung dengan baik. Nilai K juga sangat penting dalam menentukan hasil klasifikasi (Muntiari, Hanif, Asma, & Herawati, 2025). Setelah jarak yang dihitung diurutkan dari yang paling dekat, nilai K dipilih berdasarkan jarak yang paling dekat. Nilai K yang dipilih sangat penting untuk hasil klasifikasi karena nilai K sangat penting selama proses klasifikasi. Nilai K terlalu kecil dapat menyebabkan hasil terpengaruh oleh noise. Sebaliknya, nilai K terlalu besar berpotensi menyebabkan penurunan akurasi tingkat hasil prediksi. Nilai K yang optimal dapat ditemukan melalui pengujian *cross validation*.

d. Model SVM

SVM adalah metode yang digunakan untuk mengklasifikasikan data bersifat *supervised learning*, yang bertujuan untuk menemukan garis batas terbaik antara dua kelompok dengan jarak yang paling baik. SVM sangat cocok untuk berbagai jenis data, dan salah satu keunggulan utamanya adalah fungsi kernelnya, yang memungkinkan pemetaan data non-linear ke dalam ruang dengan banyak dimensi sehingga dapat dipisahkan dengan cara yang linear. Untuk klasifikasi status gizi, SVM biasanya memberikan hasil yang akurat (Jalil, Homaidi, & Fatah, 2024). SVM telah terbukti efektif saat digunakan untuk memprediksi masalah stunting (Candra, Erkamim, Muharrom, & Prayitno, 2024). Untuk menentukan kelas dari data yang diuji menggunakan fungsi keputusan berikut :

$$f(x) = w \cdot x + b \tag{3}$$

Pada persamaan (3), *w* adalah vektor bobot yang digunakan untuk menentukan arah hyperplane, *x* adalah data yang nantinya akan diklasifikasikan, dan *b* adalah bias yang berfungsi untuk menggeser posisi hyperplane. Untuk data tidak bisa dipisahkan secara linear (non-linear), maka akan menggunakan fungsi kernel sehingga fungsi keputusan akan berubah menjadi:

$$f(x) = \sum_{i=1}^n \alpha_i \cdot y_i \cdot K(x_i, x) + b \tag{4}$$

Pada persamaan (4), *α_i* adalah parameter Lagrange yang menentukan seberapa besar pengaruh setiap support vector, *y_i* adalah label kelas data latih, *K(x_i, x)* adalah fungsi kernel yang memfasilitasi transformasi data menjadi dimensi yang lebih besar, dan *b* adalah nilai bias dari fungsi keputusan. Penggunaan fungsi kernel membantu algoritma SVM untuk bekerja dengan data non-linear dengan membuat batas keputusan yang lebih baik. Selanjutnya, rumus berikut dapat digunakan untuk menulis hyperplane yang memisahkan dua kelas:

$$w \cdot x_i + b = 0 \tag{5}$$

Pada persamaan (5), *w* merepresentasikan vektor bobot sedangkan *b* merupakan nilai bias. Jarak antara kedua kelas disebut dengan margin, semakin besar nilai margin makasemakin baik hasilnya. Beberapa jenis kernel yang digunakan meliputi *linear*, *polynomial*, *radial basis function (RBF)*, dan *sigmoid*.

$$\hat{y} = (sign)(f(x)) \tag{6}$$

Jika nilai dari *f(x)* positif, maka klasifikasi masuk ke kelas pertama. Sebaliknya, jika nilai *f(x)* negatif, maka data klasifikasi masuk ke kelas kedua.

e. Evaluasi

Evaluasi model adalah proses untuk melihat tingkat kinerja model dalam memberikan hasil yang tepat dalam prediksi. Dalam penelitian ini, algoritma KNN dan SVM dievaluasi untuk mengklasifikasikan stunting. Penilaian kinerja model klasifikasi melihat akurasi, presisi, recall, dan skor F1. Ini memberikan gambaran yang lebih lengkap tentang kemampuan model untuk memprediksi kelas dengan benar (Sulaksono, Pamungkas, & Widodo, 2026). Evaluasi model menggunakan confusion matrix. Confusion matrix adalah tabel yang membandingkan hasil prediksi dengan nilai yang sebenarnya, dan memiliki empat bagian :

Tabel 1. Confusion Matrix

Kelas	Klasifikasi Positif	Klasifikasi Negatif
Positif	TP	FN
Negatif	FP	TN

Berdasarkan Tabel 1 diatas, True Positive (TP) merupakan banyaknya data yang benar diprediksi positif, True Negative (TN) merupakan banyaknya data yang benar diprediksi negatif, False Positive (FP) merupakan banyaknya data yang salah diprediksi positif, dan False Negative (FN) merupakan banyaknya data yang salah diprediksi negatif. Perhitungan kinerja model sebagai berikut:

1. Akurasi

Akurasi untuk mengetahui jumlah prediksi yang tepat dibandingkan dengan melakukan perbandingan terhadap seluruh data yang diuji. Jika hasilnya menunjukkan nilai tinggi, itu menunjukkan bahwa model sangat baik untuk klasifikasi data karena model membuat banyak prediksi yang tepat.

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

2. Presisi

Presisi menampilkan data prediksi positif dari seluruh data yang diprediksi. Jika hasil penelitian ini menunjukkan hasil presisi tinggi, maka model memprediksi "stunting" dengan tepat.

$$\frac{TP}{FP+TP} \quad (8)$$

3. Recall

Recall mengukur model dengan mengidentifikasi seluruh data positif. Recall mengukur model dalam mengenali seluruh data positif dengan benar. Ini digunakan untuk mendeteksi stunting secara keseluruhan tanpa melewatkan apapun.

$$\frac{TP}{FN+TP} \quad (9)$$

4. F1-Score

F1-Score menilai keseimbangan presisi dan recall. Jika nilai keduanya tinggi, maka nilai F1-Score akan tinggi.

$$2 \times \frac{\text{Presisi} \times \text{recall}}{\text{Presisi} + \text{recall}} \quad (10)$$

3. HASIL DAN PEMBAHASAN

Dalam proses klasifikasi stunting pada balita, hasil implementasi algoritma KNN dan dijelaskan pada bagian ini. Setiap tahapan penelitian dimulai dengan pengambilan data, preprocessing data, pelatihan model, pengujian model, dan evaluasi performa dengan confusion matrix. Pembahasan dilakukan untuk mengetahui pengaruh preprocessing terhadap performa model, membandingkan hasil klasifikasi kedua algoritma, serta menganalisis kemampuan masing-masing algoritma dalam mendeteksi kasus stunting berdasarkan data antropometri balita.

3.1 Pengambilan Data

Dataset yang dipakai dalam penelitian ini adalah data asli yang diperoleh melalui Puskesmas Patianrowo, Kabupaten Nganjuk. Data yang dikumpulkan berjumlah 1102 data balita usia 0 hingga 60 bulan. Dalam penelitian ini, variabel yang digunakan termasuk jenis kelamin, umur, berat badan, tinggi badan, Z-Score TB/U, Z-Score BB/U, Z-Score BB/TB, status gizi, dan label klasifikasi stunting. Data antropometri dipilih karena merupakan indikator utama dalam penilaian status gizi balita. Nilai Z-Score digunakan untuk menunjukkan posisi pertumbuhan anak dibandingkan standar pertumbuhan WHO. Variabel tersebut dianggap relevan karena mampu menggambarkan kondisi pertumbuhan fisik balita secara langsung dan sering digunakan dalam proses identifikasi stunting di fasilitas kesehatan.

3.2 Preprocessing Data

Tahapan preprocessing adalah salah satu langkah krusial dalam penelitian machine learning karena kualitas data sangat menentukan kualitas model yang dihasilkan. Proses preprocessing yang digunakan dalam penelitian ini terdiri dari pembersihan data, encoding data, dan normalisasi data. Tujuan utama preprocessing adalah menghasilkan dataset yang bersih, konsisten, dan sesuai untuk diproses oleh algoritma KNN dan SVM. Karena kedua algoritma tersebut sangat sensitif terhadap skala data dan kualitas fitur, sehingga preprocessing diperlukan untuk meningkatkan performa klasifikasi (Daniswara & Nuryana, 2023).

3.2.1 Pembersihan Data

Tahapan pertama preprocessing adalah pembersihan data, juga dikenal sebagai data cleaning. Pada tahap ini data yang memiliki missing value atau data kosong. Data kosong dapat menyebabkan kesalahan dalam proses pelatihan model karena beberapa algoritma machine learning tidak mendukung pemrosesan nilai kosong secara langsung. Penghapusan data kosong dilakukan untuk menjaga kualitas dataset agar model dapat mempelajari pola data dengan lebih baik. Walaupun jumlah data menjadi berkurang, data yang digunakan menjadi lebih konsisten dan valid. Setelah proses cleaning selesai, seluruh kolom diperiksa kembali dan tidak ditemukan missing value pada dataset.

```
... Jumlah data awal: 1102
    Jumlah data setelah menghapus NaN: 1067
```

Gambar 2. Penghapusan Data Kosong

Hasil pembersihan data ditunjukkan pada Gambar 2. Data awal terdiri dari 1102 baris, tetapi setelah dibersihkan, menjadi 1067 baris. Itu berarti 35 baris data dihapus. Setelah itu, setiap kolom diperiksa dan ditemukan bahwa tidak ada nilai kosong di seluruh kolom. Selanjutnya, bagikan data dengan proporsi 80:20. Akurasi model klasifikasi sangat dipengaruhi oleh pemilihan perbandingan untuk pelatihan dan pengujian pembagian data. Oleh karena itu, mendapatkan rasio yang tepat sangat penting agar model dapat berfungsi dengan baik (Oktafian, Hermawan, & Avianto, 2023).

3.2.2 Encoding Data

Encoding Data merupakan tahapan yang mengubah data kategorikal (teks) menjadi data berbentuk numerik. Proses encoding dilakukan agar label target dapat diproses oleh algoritma klasifikasi yang berbasis numerik. Data kategorikal perlu diubah menjadi angka menggunakan teknik encoding agar algoritma machine learning bisa memprosesnya (Karim, Prasetyo, & Saputro, 2023).

NO	NAMA	Gender (L=1,P=2)	Umur (Bln)	BB (Kg)	TB (Cm)	\
0	1	AGNES	2	59	16.0	105.0
1	2	AISYAH	2	57	13.7	101.0
2	3	AHMAD GIBRAN HASBY	1	50	12.2	94.5
3	4	ICHA	2	56	14.3	102.0
4	5	ALMEERA FATIMATUS	2	28	9.5	78.5

	Z Score (TB/U)	Z Score (BB/U)	Z Score (BB/TB)	STATUS GIZI (TB/U)	\
0	-0.857660	-0.836	-0.515	Normal	
1	-1.496258	-1.775	-1.368	Normal	
2	-2.403131	-2.477	-1.660	Pendek	
3	-1.193434	-1.429	-1.104	Normal	
4	-3.121872	-2.099	-0.495	S.Pendek	

STATUS GIZI (BB/U)	STATUS GIZI (BB/TB)	KETERANGAN
0	BB Normal	Normal 0
1	BB Normal	Normal 0
2	BB Krg	Normal 1
3	BB Normal	Normal 0
4	BB Krg	Normal 1

Gambar 3. Setelah Encoding

Gambar 3 menampilkan data setelah dilakukan encoding data pada label. Data yang awalnya tertulis tidak dan stunting. Keterangan tersebut menandakan data masih berupa data kategorikal, sehingga diperlukan proses encoding data untuk mengonversi data menjadi format numerik dengan menampilkan data tidak stunting menjadi 0 dan stunting menjadi 1.

3.2.3 Normalisasi Data

Normalisasi data merupakan tahapan yang bertujuan untuk menyeragamkan skala antar fitur, sehingga tidak terjadi kondisi adanya variabel dengan nilai lebih besar. Data sebelum dilakukan proses normalisasi data masih memiliki nilai rentang skala yang berbeda-beda, sehingga menyebabkan perbedaan skala yang cukup besar antar fitur. Proses preprocessing pada bagian normalisasi data harus dilakukan sebelum implementasi KNN dan SVM. Normalisasi memiliki peran yang sangat penting terutama untuk algoritma KNN yang berbasis perhitungan jarak. Tanpa normalisasi, fitur dengan skala besar seperti tinggi badan dapat mendominasi perhitungan jarak dibandingkan fitur lain. Pada SVM, normalisasi membantu dalam menghasilkan margin yang lebih optimal. Dengan menyamakan rentang nilai fitur, model dapat bekerja secara lebih adil dalam mempertimbangkan setiap variabel.

```

...
Data uji setelah normalisasi:
[[[1. 0.03333333 0.08661417 0.12130178 0.59318731 0.59162207
0.35868861]
[1. 0.96666667 0.38188976 0.71005917 0.36493936 0.40966968
0.17701179]
[1. 0.26666667 0.17322835 0.34171598 0.26901874 0.35523159
0.02669431]
[1. 0.7 0.42125984 0.64349112 0.4382315 0.56034324
0.40248801]
[0. 0.55 0.37795276 0.61538462 0.51326568 0.55019376
0.29493326]]

```

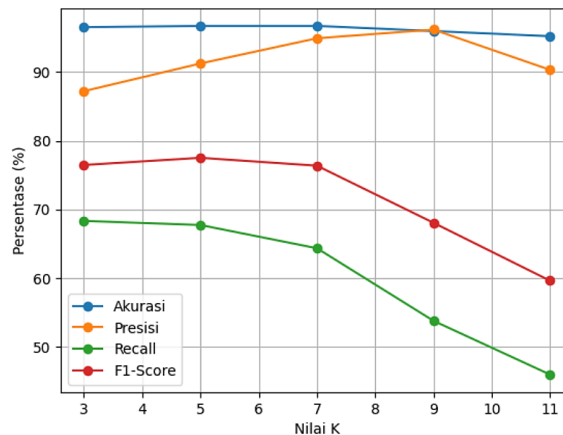
Gambar 4. Setelah Normalisasi

Gambar 4 menunjukkan hasil setelah langkah normalisasi dilakukan dengan menggunakan normalisasi min-max yang telah mengubah informasi data menjadi berada dalam rentang antara 0 sampai 1.

3.3 Pelatihan Model KNN

Dalam penelitian ini, penerapan algoritma KNN dilakukan dengan beberapa langkah penting. Pertama, dimulai dengan mengatur model dan menentukan K, yang adalah jumlah tetangga terdekat yang akan dipakai saat klasifikasi. Data yang digunakan untuk latihan disimpan tanpa perlu dilatih secara khusus, karena KNN adalah metode yang langsung menggunakan data saat memprediksi. Klasifikasi dilakukan dengan menggunakan metode jarak antar data untuk menghitung jarak antara semua data latihan dan data yang ingin diuji, lalu dipilih K data yang paling dekat sebagai

tetangga. Untuk menentukan kelas, dilakukan voting mayoritas dari tetangga terdekat tersebut. Untuk mengevaluasi seberapa baik model ini bekerja, dipakai confusion matrix sehingga memperoleh ukuran seperti akurasi, presisi, recall, dan F1-score. Metrik F1-Score ini penting karena menilai keseimbangan antara presisi dan recall, yang relevan untuk kasus 'Normal' dan 'Stunting' (Banurea, Hutagaol, & Sihombing, 2023). Uji coba model dilakukan dengan metode K-Fold Cross Validation, yang melakukan pembagian data ke dalam lima bagian. Setiap bagian selanjutnya diproses dengan melakukan normalisasi menggunakan Min-Max Normalization berdasarkan data latih. Hasil evaluasi dari setiap bagian ini kemudian dihitung rata-ratanya untuk mendapatkan performa akhir model.



Gambar 5. Pengaruh nilai K

Gambar 5 menunjukkan pengaruh nilai K terhadap performa algoritma KNN. Berdasarkan grafik tersebut, terlihat bahwa nilai akurasi cenderung stabil, namun nilai recall dan F1-score mengalami penurunan pada nilai K yang lebih besar. Hal ini menunjukkan bahwa pemilihan nilai K sangat mempengaruhi kemampuan model dalam mendeteksi kasus stunting. Nilai K = 5 dipilih sebagai parameter terbaik karena memberikan keseimbangan antara akurasi, presisi, recall, dan F1-score.

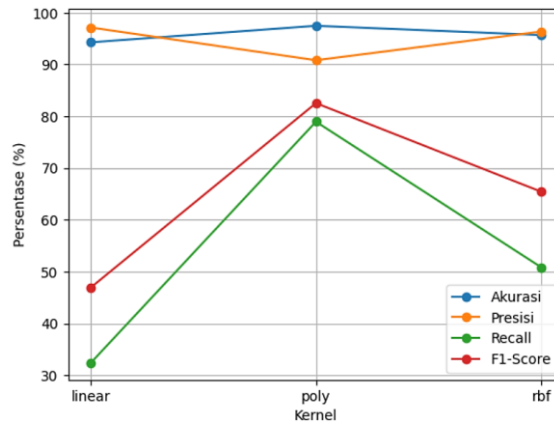


Gambar 6. Hasil KNN

Berdasarkan perolehan dari pengujian pada Gambar 6, model KNN dengan K = 5 menunjukkan kinerja yang baik, terutama di bagian akurasi dan presisi. Hal ini menunjukkan model ini cukup efektif dalam mengklasifikasikan data secara umum dan bisa mengurangi kesalahan dalam memprediksi kelas positif. Namun, nilai recall yang lebih rendah menunjukkan bahwa masih ada beberapa kasus stunting yang tidak terdeteksi dengan baik oleh model ini. Hal ini dapat dijelaskan karena sifat KNN yang sangat bergantung pada distribusi data di sekitar titik uji. Jika data stunting memiliki distribusi yang tidak merata atau jumlahnya lebih sedikit dibandingkan kelas lain, maka kemungkinan besar data tersebut akan kalah dalam proses voting mayoritas.

3.4 Pelatihan Model SVM

Penerapan algoritma SVM pada penelitian ini dilakukan dengan beberapa langkah. Prosesnya dimulai dengan inisialisasi model yang meliputi parameter kernel, gamma, dan derajat degree yang digunakan untuk memilih jenis fungsi kernel seperti linear, polynomial, atau RBF. Fungsi kernel digunakan untuk menghitung tingkat kemiripan antara data uji dan data latih, sehingga membantu model dalam membentuk batas keputusan (decision boundary) yang terbaik, terutama ketika data tidak bisa dipisahkan dengan cara linear. Data latih kemudian disimpan ke dalam model, dengan label kelas dikonversi menjadi nilai -1 dan 1 sesuai dengan konsep klasifikasi biner pada SVM. Proses prediksi dilakukan dengan menghitung nilai kernel antara data uji dan seluruh data latih, kemudian hasilnya dikalikan dengan label dan dijumlahkan untuk menentukan kelas menggunakan fungsi sign. Pengujian model dilakukan dengan cara K-Fold Cross Validation yang memiliki 5 fold. Di setiap fold, proses normalisasi diterapkan menggunakan Min-Max Normalization untuk menyamakan skala fitur. Setelah itu, model dilatih dengan algoritma SVM dari library scikit-learn dan digunakan untuk memprediksi data uji. Kinerja model dinilai dengan memanfaatkan confusion matrix yang mendapatkan nilai metrik akurasi, presisi, recall, dan F1-score, yang kemudian dirata-rata dari semua fold untuk mendapatkan hasil akhir.



Gambar 7. Perbandingan performa algoritma

Gambar 7 menunjukkan perbandingan performa algoritma SVM dengan berbagai jenis kernel. Berdasarkan grafik, kernel polynomial memberikan nilai akurasi, recall, dan F1-score yang lebih baik daripada kernel lainnya. Ini berarti bahwa kernel polynomial lebih optimal untuk mendeteksi pola non-linear dalam data. Oleh karena itu, kernel polynomial dipilih sebagai model terbaik dalam penelitian ini.

```

...
Hasil 5 fold SVM
Kernel : poly
Akurasi : 97.47%
Presisi : 90.82%
Recall : 78.96%
F1-Score : 82.55%

```

Gambar 8. Hasil SVM

Dari hasil pengujian yang terlihat pada Gambar 8, SVM dengan kernel polynomial menunjukkan kinerja yang luar biasa untuk mengklasifikasikan data stunting. Seperti yang terbukti dari akurasi dan keseimbangan metrik evaluasi yang lebih tinggi dibanding kernel lain. Model SVM memperlihatkan hasil yang lebih konsisten dan seimbang dibandingkan KNN, terutama pada metrik recall dan F1-score. Ini memperlihatkan performa SVM yang lebih optimal dalam mengenali pola kompleks dalam data, termasuk dalam mendeteksi kasus stunting yang sebelumnya sulit diklasifikasikan oleh KNN. Keunggulan lain dari SVM adalah kemampuannya dalam membentuk margin yang optimal antara dua kelas, sehingga model tidak hanya fokus pada data terdekat seperti KNN, tetapi juga mempertimbangkan batas pemisah.

3.5 Hasil Klasifikasi

Setelah model KNN menemukan nilai K terbaik dan SVM menemukan nilai kernel terbaik, maka dievaluasi pada data uji untuk mendapatkan hasil kinerja kedua model dalam melakukan prediksi. Hasil klasifikasi yang dilakukan pada data uji menampilkan dua kelas seperti yang ditampilkan pada gambar 7, yaitu kelas stunting dan tidak stunting.

```

...
Kelas_Asl  Prediksi_KNN  Prediksi_SVM
0  Tidak Stunting  Tidak Stunting  Tidak Stunting
1  Stunting  Tidak Stunting  Stunting
2  Stunting  Tidak Stunting  Tidak Stunting
3  Tidak Stunting  Tidak Stunting  Tidak Stunting
4  Tidak Stunting  Tidak Stunting  Tidak Stunting
5  Tidak Stunting  Tidak Stunting  Tidak Stunting
6  Tidak Stunting  Tidak Stunting  Tidak Stunting
7  Tidak Stunting  Tidak Stunting  Tidak Stunting
8  Tidak Stunting  Tidak Stunting  Tidak Stunting
9  Tidak Stunting  Tidak Stunting  Tidak Stunting
10 Tidak Stunting  Tidak Stunting  Tidak Stunting
11 Tidak Stunting  Tidak Stunting  Tidak Stunting
12 Tidak Stunting  Tidak Stunting  Tidak Stunting
13 Tidak Stunting  Tidak Stunting  Tidak Stunting
14 Stunting  Tidak Stunting  Tidak Stunting
15 Tidak Stunting  Tidak Stunting  Tidak Stunting
16 Tidak Stunting  Tidak Stunting  Tidak Stunting
17 Tidak Stunting  Tidak Stunting  Tidak Stunting
18 Tidak Stunting  Tidak Stunting  Tidak Stunting
19 Tidak Stunting  Tidak Stunting  Tidak Stunting

```

Gambar 9. Hasil Klasifikasi

Pada Gambar 9, hasil klasifikasi menunjukkan bahwa kedua algoritma mampu memisahkan kelas stunting dan tidak stunting dengan cukup baik. Namun, terdapat beberapa data yang masih mengalami kesalahan klasifikasi. Pada algoritma KNN, kesalahan klasifikasi lebih banyak terjadi pada data stunting yang memiliki karakteristik mirip dengan



data tidak stunting. Karena KNN hanya mempertimbangkan kedekatan jarak antar data, maka data minoritas dapat kalah voting apabila berada di lingkungan yang didominasi oleh kelas mayoritas. Sementara itu, algoritma SVM mampu menghasilkan batas pemisah yang lebih jelas antar kelas. Hal ini menyebabkan jumlah kesalahan klasifikasi menjadi lebih sedikit dibandingkan KNN. Perbedaan performa kedua algoritma menunjukkan bahwa karakteristik dataset sangat mempengaruhi hasil klasifikasi. Dataset stunting memiliki pola hubungan antar variabel yang cukup kompleks sehingga algoritma yang mampu menangani pola non-linear seperti SVM cenderung memberikan hasil yang lebih baik.

3.6 Evaluasi

Evaluasi model dilakukan dengan memanfaatkan confusion matrix untuk memahami secara lebih mendalam kinerja klasifikasi. Dari hasil confusion matrix, diperoleh nilai true positive (TP), true negative (TN), false positive (FP), dan false negative (FN) dari setiap algoritma seperti pada Tabel 2 berikut :

Tabel 2. Hasil Confusion Matrix

Algoritma	TP	TN	FP	FN
KNN	13	191	3	6
SVM	17	191	3	2

Pada algoritma KNN diperoleh nilai TP sebanyak 13, TN sebanyak 191, FP sebanyak 3, dan FN sebanyak 6. Sedangkan pada algoritma SVM diperoleh nilai TP sebanyak 17, TN sebanyak 191, FP sebanyak 3, dan FN sebanyak 2. Hasil tersebut menunjukkan bahwa SVM memiliki jumlah false negative yang lebih rendah dibandingkan KNN. Hal ini sangat penting dalam kasus deteksi stunting karena kesalahan mendeteksi balita stunting sebagai normal dapat berdampak terhadap keterlambatan penanganan kesehatan. Jumlah true positive pada SVM juga lebih tinggi dibandingkan KNN. Artinya, SVM mampu mengenali lebih banyak balita stunting secara benar. Walaupun kedua algoritma memiliki jumlah false positive yang sama, kemampuan SVM dalam menekan false negative menjadikan performanya lebih baik..

Pada Tabel 3 ditampilkan hasil perhitungan evaluasi metrik dengan menampilkan hasil akurasi, presisi, recall, dan f1-score dari metode KNN dan SVM.

Tabel 3. Hasil Perhitungan

Algoritma	Akurasi	Presisi	Recall	F1-Score
KNN	96,72%	91,25%	67,73%	77,52%
SVM	97,47%	90,82%	78,96%	82,55%

Berdasarkan hasil evaluasi, algoritma KNN memperoleh akurasi sebesar 96,72%, presisi 91,25%, recall 67,73%, dan F1-score 77,52%. Sementara algoritma SVM memperoleh akurasi sebesar 97,47%, presisi 90,82%, recall 78,96%, dan F1-score 82,55%. Perbedaan nilai akurasi kedua model memang tidak terlalu besar. Namun, jika dilihat dari recall dan F1-score, SVM memberikan hasil yang lebih memuaskan. Hal ini menunjukkan bahwa evaluasi model tidak cukup hanya melihat akurasi, tetapi juga perlu mempertimbangkan kemampuan model dalam mendeteksi kelas positif. Dalam penelitian kesehatan seperti klasifikasi stunting, recall menjadi metrik yang sangat penting karena berkaitan langsung dengan kemampuan model dalam mengenali seluruh kasus stunting. Model dengan recall rendah berpotensi melewatkan banyak kasus stunting sehingga kurang efektif digunakan dalam proses deteksi dini. Secara keseluruhan, hasil penelitian mengindikasikan bahwa algoritma SVM mampu memberikan tingkat efektivitas yang lebih tinggi dibandingkan KNN dalam pengklasifikasian stunting pada dataset yang digunakan. Kemampuan SVM dalam membentuk hyperplane optimal dan menangani pola non-linear membuat model lebih mampu mengenali karakteristik data stunting dengan baik.

4. KESIMPULAN

Penelitian ini bertujuan untuk menilai efektivitas algoritma KNN dan SVM dalam klasifikasi stunting pada anak-anak. Dalam pelaksanaan, penelitian ini melibatkan sejumlah tahap praproses data yang ditujukan untuk memastikan kualitas data yang digunakan. Proses-proses ini termasuk pembersihan data untuk menghilangkan nilai yang tidak lengkap, proses pengkodean untuk mentransformasikan label kategori ke dalam format angka, dan normalisasi dengan menggunakan metode normalisasi min-max untuk memastikan bahwa semua fitur berada dalam rentang nilai yang sama. Hal ini adalah prosedur yang dilakukan untuk meningkatkan akurasi model agar dapat memastikan bahwa seluruh fitur berada dalam skala nilai yang serupa dalam mengidentifikasi pola pada data. Hasil pengujian menunjukkan bahwa algoritma SVM lebih baik daripada KNN dalam mengklasifikasikan kondisi stunting. Hasil tersebut menunjukkan bahwa SVM dapat mengenali pola data yang lebih kompleks dan membuat batas pemisah antar kelas yang ideal. Meskipun algoritma KNN memiliki nilai akurasi sebesar 96,72%, presisi sebesar 91,25%, recall sebesar 67,73%, dan nilai F1 sebesar 77,52%, SVM mendapatkan nilai akurasi sebesar 97,47%, presisi sebesar 90,82%, dan nilai F1 sebesar 82%. Meskipun nilai presisi KNN sedikit lebih tinggi, tetapi nilai recall dan skor F1-score lebih rendah menunjukkan bahwa KNN masih kurang efisien dalam mendeteksi seluruh kasus stunting. Dapat disimpulkan bahwa dataset dalam penelitian ini lebih baik menerapkan algoritma SVM untuk klasifikasi stunting.



REFERENCES

- A'la, F. R., Homaidi, A., & Lidimillah, L. F. (2025). Implementasi Algoritma K-Nearest Neighbor (KNN) Untuk Klasifikasi Status Gizi Balita. *Prosiding Seminar Nasional KARSA NUSANTARA*, 2, 18-23.
- Alamin, P. B., & Aryani, D. (2025). Perbandingan Algoritma Knn, Naive Bayes Dan Svm Untuk Klasifikasistatus Stunting Anak Balita Di Puskesmas Sepatan. *Jurnal Ilmiah Teknik dan Ilmu Komputer*, 4(4), 360-369, DOI: 10.55123.
- Banurea, M., Hutagaol, D. B., & Sihombing, O. (2023). Klasifikasi Penyakit Stunting Dengan Menggunakan Algoritma Support Vector Machine Dan Random Forest. *Jurnal TEKINKOM*, 6(2), 540-549, DOI : 10.37600/tekinkom.v6i2.927.
- Candra, A., Erkamim, M., Muharrom, M., & Prayitno, E. (2024). Klasifikasi Stunting Pada Balita Berdasarkan Status Gizi Menggunakan Pendekatan Support Vector Machine (SVM). *Jurnal Ilmiah FIFO*, 16(2), 171-181, DOI: <http://dx.doi.org/10.22441/fifo.2024.v16i2.007>.
- Daniswara, A. A., & Nuryana, I. D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 5(1), 97-100, DOI: <https://doi.org/10.26740/jinacs.v5n01.p97-100>.
- Drajana, I. C., & Bode, A. (2022). Prediksi Status Penderita Stunting Pada Balita Provinsi Gorontalo Menggunakan K-Nearest Neighbor Berbasis Seleksi Fitur Chi Square. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 5(2), 309-316, DOI: 10.32672/jnkti.v5i2.4205.
- Dwinanto, R. W., Sandi A, A. S., & Ardianto, R. (2024). Klasifikasi Berisiko Stunting pada Balita: Perbandingan K-Nearest Neighbor, Naïve Bayes, Support Vector Machine. *Jurnal Manajemen Informatika & Komputersisasi Akuntansi*, 8(2), 264-273, DOI: <https://doi.org/10.46880/jmika.Vol8No2.pp264-273>.
- Hardiani, T., & Putri, R. N. (2021). Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Stunting. *Digital Transformation Technology (Digitech)*, 4(1), 54-59, DOI: <https://doi.org/10.47709/digitech.v4i1.4481>.
- Hidayat, T., Sembiring, I., Purnomo, H. D., & Iriani, A. (2025). Prediksi Prevalensi Stunting Balita dengan Pendekatan Algoritma Support Vector Machine dan Synthetic Minority Oversampling Technique (SMOTE). *Jurnal_Pekommas*, 10(1), 9-16, doi: 10.56873/jpkm.v9i1.5389.
- Jalil, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Algoritma Support Vector Machine Untuk Klasifikasi Status Stunting Pada Balita. *G-Tech : Jurnal Teknologi Terapan*, 8(3), 2070-2079, DOI: <https://doi.org/10.33379/gtech.v8i3.4811>.
- Junaidi, A., Nurdin, & Meiyanti, R. (2025). Klasifikasi Status Anak Stunting Menggunakan Metode Knearest Neighbor. *RABIT : Jurnal Teknologi dan Sistem Informasi Univrab*, 10(2), 1435-1445, DOI : <https://doi.org/10.36341/rabit.v10i2.6536>.
- Karim, A. A., Prasetyo, M. A., & Saputro, M. R. (2023). Perbandingan Metode Random Forest, K-Nearest Neighbor, dan SVM Dalam Prediksi Akurasi Pertandingan Liga Italia. *Prosiding Seminar Nasional Teknologi Dan Sains*, 2(1), 377-382, DOI: <https://doi.org/10.29407/stains.v2i1.2877>.
- Lonang, S., & Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, 6(1), 49-59, DOI: <https://doi.org/10.30865/mib.v6i1.3312>.
- Muntiari, N. R., Hanif, K. H., Asma, & Herawati, L. (2025). Deteksi Stunting Pada Balita Dengan Menggunakan Perbandingan Algoritma Machine Learning. *Jurnal Khatulistiwa Informatika*, 13(1), 17-23, DOI: <https://doi.org/10.31294/jki.v13i1.25132>.
- Nadroh, A., Triwibowo, D. N., & Sumantri, R. B. (2024). Klasifikasi Status Gizi Balita Menggunakan Algoritma Support Vector Machine Dengan Optimasi Grid Search Cross-Validation. *METHOMIKA: Jurnal Manajemen Informatika & Komputersisasi Akuntansi*, 8(2), 250-257, DOI: <https://doi.org/10.46880/jmika.Vol8No2.pp250-257>.
- Oktafian, R., Hermawan, A., & Avianto, D. (2023). Pengaruh Komposisi Split Data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning. *Jurnali Sainsi dan iInformatika*, 9(1), 19-28, DOI: 10.34128/jsi.v9i1.622.
- Rahmawati, G., Sanmas, S. A., Nudyawati, E., & Syaharani, N. D. (2024). Studi Perbandingan Performa: Prediksi Status Stunting Pada Anak Berdasarkan Data Antropometri Menggunakan Algoritma Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN). *Prosiding Seminar Nasional Sains Data*, 4(1), 782-790, DOI:10.33005/senada.v4i1.330.
- Sahamony, N. F., Terttiaavini, & Rianto, H. (2024). Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 413-422, DOI: <https://doi.org/10.57152/malcom.v4i2.1210>.
- Saleh, H., Faisal, M., & Musa, R. I. (2019). klasifikasi status gizi balita menggunakan metode k-nearest neighbor. *jurnal sistem informasi dan teknik komputer*, 4(2), 120-126, DOI: <https://doi.org/10.51876/simtek.v4i2.60>.
- Sulaksono, J., Pamungkas, D. P., & Widodo, D. W. (2026). Analisis Perbandingan Model Machine Learning untuk Prediksi Penyakit Jantung. *Prosiding Seminar Nasional Teknologi Dan Sains*, 5(1), 707-712, DOI: <https://doi.org/10.29407/b4yzdb65>.