

## PERBANDINGAN ALGORITMA KNN, NAIVE BAYES DAN SVM UNTUK KLASIFIKASI STATUS STUNTING ANAK BALITA DI PUSKESMAS SEPATAN

<sup>1</sup>Pandu Bashir Alamin, <sup>2</sup>Diah Aryani

<sup>1,2</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul

<sup>1</sup>pandubashir@gmail.com, <sup>2</sup>diah.aryani@esaunggul.ac.id

### INFO ARTIKEL

#### Riwayat Artikel :

Diterima : 13 Oktober 2025

Disetujui : 16 Oktober 2025

#### Kata Kunci :

klasifikasi, KNN, naive bayes, machine learning, SVM, status Stunting.

### ABSTRAK

Penelitian ini membandingkan kinerja tiga algoritma Machine Learning, yaitu K-Nearest Neighbors (KNN), Naive Bayes (NB), dan Support Vector Machine (SVM), dalam mengklasifikasikan status stunting pada anak balita di Puskesmas Sepatan. Status stunting ditentukan berdasarkan indikator antropometri seperti berat badan, tinggi badan, dan usia. Penelitian ini menggunakan metode GridSearchCV untuk melakukan penyetelan hyperparameter dan mengevaluasi performa model berdasarkan metrik Accuracy, Precision, Recall, F1-score, Specificity, dan ROC AUC. Validasi model dilakukan menggunakan teknik K-Fold Cross Validation untuk memastikan konsistensi hasil pada berbagai subset data. Hasil penelitian menunjukkan bahwa Naive Bayes memberikan performa terbaik dengan Accuracy 99,65%, Precision 99,67%, Recall 99,65%, F1-score 99,64%, Specificity 99,94%, dan ROC AUC 0,9999. Algoritma SVM dengan kernel linear berada di posisi kedua dengan akurasi 98,96% dan nilai ROC AUC sempurna (1,0000), sementara KNN memperoleh akurasi 96,36% dengan performa stabil pada kelas mayoritas namun menurun pada kelas minoritas. Secara keseluruhan, Naive Bayes terbukti sebagai algoritma paling optimal untuk klasifikasi status gizi balita pada dataset ini. Meskipun demikian, penelitian selanjutnya disarankan untuk mengeksplorasi algoritma lain yang lebih kompleks seperti Random Forest atau XGBoost, serta mempertimbangkan penerapan teknik validasi tambahan seperti stratified K-Fold atau repeated K-Fold guna meningkatkan reliabilitas dan generalisasi model terhadap variasi data yang lebih.

### ARTICLE INFO

#### Article History :

Received : Oct 13, 2025

Accepted : Oct 16, 2025

#### Keywords:

classification, KNN, naive bayes, machine learning, SVM, stunting Status.

### ABSTRACT

*This study compares the performance of three Machine Learning Algorithms K Nearest Neighbors (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) in classifying stunting status among toddlers at Sepatan Health Center. Stunting status was determined using anthropometric indicators such as weight, height, and age. GridSearchCV was applied for hyperparameter tuning, and the models were evaluated using Accuracy, Precision, Recall, F1-score, Specificity, and ROC AUC. The results show that Naive Bayes achieved the best performance with an Accuracy of 99.65%, Precision of 99.67%, Recall of 99.65%, F1-score of 99.64%, Specificity of 99.94%, and ROC AUC of 0.9999. The SVM with a linear kernel ranked second with an accuracy of 98.96% and a perfect ROC AUC score (1.0000),*

---

*while KNN obtained 96.36% accuracy with stable performance on majority classes but lower performance on minority classes. Overall, Naive Bayes proved to be the most optimal algorithm in this study, followed by SVM and KNN. These findings indicate that the choice of machine learning algorithm strongly depends on data characteristics, particularly class distribution in the nutritional status dataset of toddlers.*

---

## 1. PENDAHULUAN

Stunting Stunting didefinisikan sebagai kondisi gagal tumbuh pada anak di bawah lima tahun (balita), yang ditandai dengan tinggi badan atau panjang tubuh yang lebih rendah dari standar usia anak kondisi ini diukur berdasarkan z-score panjang atau tinggi badan menurut usia, dengan batasan lebih dari minus dua standar deviasi dari median pertumbuhan anak sesuai standar WHO (Trisnawati et al., 2022). Menurut Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2022 tentang *Standar Antropometri Anak*, stunting didefinisikan sebagai kondisi tubuh anak di bawah usia lima tahun yang pendek atau sangat pendek berdasarkan indikator antropometri, yaitu panjang badan atau tinggi badan menurut umur (PB/U atau TB/U) anak usia 0–60 bulan dengan ambang batas (*z-score*) antara -3 Standar Deviasi (SD) dan < -2 SD (Lestari et al., 2024). Penyebab utama stunting adalah kekurangan gizi kronis yang terjadi secara terus-menerus, terutama selama 1000 hari pertama kehidupan (sejak konsepsi hingga anak berusia dua tahun), masa ini merupakan periode kritis karena organ-organ vital, termasuk otak, berkembang pesat selama waktu tersebut kekurangan nutrisi yang cukup selama periode ini dapat mengakibatkan gangguan pertumbuhan yang bersifat permanen (Laksono et al., 2022). Faktor yang bisa menyebabkan stunting pada balita antara lain kekurangan energi protein (KEP), kekurangan vitamin A (KVA), anemia gizi besi (AGB), gangguan akibat kekurangan yodium (GAKY), dan gizi lebih. Salah satu masalah gizi lain yang signifikan adalah stunting, yang merupakan gangguan pertumbuhan pada fisik anak di bawah usia lima tahun yang ditandai dengan penurunan kecepatan pertumbuhan akibat ketidakseimbangan gizi (Laksono et al., 2022). Menurut Standar Antropometri Anak di Indonesia yang mengacu pada World Health Organization (WHO) Child Growth Standards untuk anak usia 0-5 tahun, pertumbuhan anak

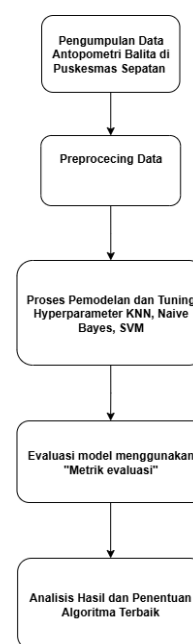
dapat dicapai apabila memenuhi syarat-syarat tertentu. Standar Antropometri Anak didasarkan pada parameter berat badan dan panjang/tinggi badan yang terdiri atas empat indeks, meliputi Berat Badan menurut Umur (BB/U), Panjang/Tinggi Badan menurut Umur (TB/U), Berat Badan menurut Panjang/Tinggi Badan (BB/PB atau BB/TB), dan Indeks Massa Tubuh menurut Umur (IMT/U) (Permatasari and Chadirin, 2022). Di Indonesia, kekurangan gizi pada balita menjadi masalah serius, salah satunya adalah stunting, Meskipun Indonesia pernah mencatatkan penurunan angka stunting hingga 27,67% (Pramulia Fitri et al., 2024), angka tersebut masih belum memenuhi standar dari World Health Organization (WHO), yang menetapkan angka stunting harus di bawah 20%. Di Provinsi Banten sendiri, angka kejadian stunting saat ini mencapai 23% (Rizal et al., 2024), yang menunjukkan bahwa masalah ini masih menjadi tantangan besar bagi pemerintah daerah maupun masyarakat. Kondisi ini termasuk di wilayah Puskesmas Sepatan, Kabupaten Tangerang, yang juga memiliki kasus stunting yang signifikan. Oleh karena itu, diperlukan langkah strategis untuk mendeteksi dan mengklasifikasikan status stunting sebagai upaya pencegahan dan penanganan dini. Klasifikasi status gizi termasuk stunting pada anak balita umumnya didasarkan pada indikator seperti berat badan, tinggi badan, dan usia anak. Indikator ini diolah untuk mendapatkan Z-score yang kemudian digunakan sebagai dasar penentuan status gizi. Namun, pengolahan data secara manual seringkali tidak efisien dan berpotensi menghasilkan kesalahan. Untuk itu, diperlukan pendekatan berbasis teknologi yang mampu mengolah data secara cepat dan akurat (Fitrylia Azis et al., 2024). Beberapa algoritma pembelajaran mesin populer untuk klasifikasi adalah K-Nearest Neighbors (KNN) (Wulandari and Avianto, 2024), Naive Bayes, dan Support Vector Machine (SVM). KNN dikenal karena kesederhanaannya, bekerja dengan menyimpan

semua data latih dan mengklasifikasikan data baru berdasarkan kedekatan dengan data sebelumnya. Naive Bayes, algoritma berbasis probabilitas, sering digunakan karena kemampuannya menangani data kategoris dengan baik (Hardiani and Putri, 2024). Sedangkan SVM, yang diperkenalkan oleh Vapnik pada tahun 1992, adalah algoritma berbasis konsep pattern recognition. Secara sederhana, SVM mencari hyperplane-hyperplane terbaik yang berfungsi sebagai pemisah dua kelas pada input space, kemampuan SVM untuk memisahkan data secara linear maupun non-linear membuatnya menjadi pilihan yang kuat dalam berbagai aplikasi klasifikasi (Muh. Faried Muchtar et al., 2024). Algoritma-algoritma ini banyak digunakan dalam berbagai penelitian klasifikasi modern, terutama ketika berhadapan dengan data berukuran besar dan kompleks (El-Sofany, 2024). Penelitian terdahulu juga telah melakukan perbandingan beberapa algoritma machine learning dalam bidang kesehatan, seperti pada deteksi penyakit jantung (Fawwaz et al., 2025), untuk mengetahui model klasifikasi dengan performa terbaik. Hasil penelitian tersebut menunjukkan bahwa algoritma KNN memiliki tingkat akurasi yang cukup tinggi dibandingkan algoritma lainnya, sementara SVM dikenal unggul dalam menangani data berdimensi tinggi dan pola yang kompleks (Wahyuni and Kusumodestoni, 2024). Oleh karena itu, penelitian ini berfokus pada perbandingan kinerja algoritma KNN, Naive Bayes, dan SVM dalam klasifikasi status stunting balita di Puskesmas Sepatan. Tujuannya adalah untuk mengevaluasi dan menentukan algoritma terbaik dalam hal akurasi, efisiensi, dan ketepatan pengolahan data.

## 2. METODE

Metode Penelitian ini membandingkan kinerja tiga algoritma Machine Learning K-Nearest Neighbors (KNN) (Homepage et al., 2023), Naive Bayes (NB) (Lestari et al., 2024) dan Support Vector Machine (SVM) (Dwinanto et al., 2024), Dalam mengklasifikasikan status gizi balita berdasarkan data yang diperoleh dari Puskesmas Sepatan. Metodologi penelitian ini terdiri dari enam tahapan, yaitu: Pengumpulan data yang melibatkan data antropometri anak balita seperti berat badan, tinggi badan, dan usia. Preprocessing data, meliputi pembersihan data,

penanganan nilai kosong, dan normalisasi agar data siap digunakan untuk pemodelan. (Wahyu Dwinanto et al., 2024) Pemodelan tiga algoritma yaitu KNN, Naive Bayes, dan SVM, di mana tiga model *Machine Learning* diterapkan untuk klasifikasi status stunting. *Tuning hyperparameter*, untuk mengoptimalkan parameter model menggunakan *GridSearchCV* agar diperoleh performa terbaik. Evaluasi model, menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-score*, *Specificity*, dan *ROC AUC*. (Widyawati et al., 2025) dan Analisis hasil. Diagram alur metodologi penelitian dapat dilihat pada Gambar 1:



Gambar 1 Alur Metodologi Penelitian

### 2.1. Pengumpulan Data

Data dalam penelitian ini diperoleh dari data sekunder yang disediakan oleh Puskesmas Sepatan Tangerang. Data yang digunakan dalam penelitian ini merupakan hasil penimbangan massal balita yang dilakukan pada tiga periode berbeda, yaitu bulan April (2023), Oktober (2024), dan Februari (2025). Data yang dikumpulkan mencakup beberapa atribut penting, antara lain nama balita, jenis kelamin laki-laki atau perempuan (L/P), tanggal lahir, berat badan dalam kilogram, tinggi badan dalam sentimeter (cm), usia balita, nilai z-score TB/U yang menunjukkan status gizi balita.

## 2.2. Load Data

Proses pemuatan data status gizi balita dalam format Excel dilakukan di lingkungan Google Colab menggunakan fungsi `read_excel` dari library `pandas`. File yang digunakan berjudul `Dataset Balita Sepatan.xlsx`, yang merupakan data sekunder hasil pengukuran antropometri balita di Kecamatan Sepatan. Setelah berhasil dimuat, data tersebut disimpan dalam struktur `DataFrame` dengan nama `df` untuk memudahkan proses analisis dan pemodelan lebih lanjut.

## 2.3. Preprocessing Data

Setelah data berhasil dimuat, tahap selanjutnya adalah `preprocessing` data untuk memastikan bahwa dataset yang digunakan dalam pelatihan model memiliki kualitas yang baik. Proses `preprocessing` meliputi beberapa langkah penting, antara lain pemeriksaan data hilang menggunakan `df.isnull()` dan pengecekan duplikasi data dengan `df.duplicate()`. Data yang hilang ditangani dengan metode imputasi atau penghapusan baris, sedangkan data duplikat dihapus agar tidak memengaruhi kualitas model. Dilakukan `feature selection` untuk memilih atribut yang relevan terhadap klasifikasi status gizi TB/U. Pada tahap ini, hanya kolom-kolom yang memiliki hubungan langsung dengan penentuan status gizi yang dipertahankan, yaitu Desa/Kelurahan, usia saat ukur, berat badan, tinggi badan, TB/U, dan z-score TB/U. Atribut lain yang tidak relevan seperti nama orang tua, alamat, dan tanggal lahir dihapus agar tidak menambah `noise` pada data.

Tabel 1 Filtered Data

No	Filtered Data	
	Variabel	Keterangan
1	Nama	Nama balita yang ditimbang dan diukur
2	Jenis Kelamin (JK)	Jenis kelamin balita (L/P)
3	Tanggal Lahir (tgl lahir)	Tanggal lahir balita
4	Berat Badan (BB)	Berat badan balita (dalam kg)
5	Tinggi Badan (TB)	Tinggi badan (dalam cm)
6	Usia	Usia balita
7	Z-score TB/U	Status gizi: Sangat Pendek, Pendek, Normal, dan Tinggi

Hasil `feature selection` pada Tabel 1 ini kemudian disimpan dalam sebuah variabel

bernama `filtered_data`, sehingga hanya berisi data dengan kolom-kolom terpilih yang akan digunakan pada tahap selanjutnya. Nilai z-score TB/U dipertahankan sebagai acuan untuk menentukan kategori status gizi balita, yang kemudian dikodekan ke dalam format numerik agar dapat diproses oleh algoritma klasifikasi seperti KNN, Naive Bayes, dan SVM. Dengan demikian, `feature selection` membantu menyederhanakan dataset, meningkatkan efisiensi komputasi, dan memperkuat akurasi model dalam melakukan klasifikasi. Normalisasi juga diterapkan pada atribut numerik seperti berat badan, tinggi badan, dan usia menggunakan metode `Min-Max Scaling` agar setiap fitur berada dalam rentang nilai yang seragam. Setelah seluruh tahap `preprocessing` selesai, dataset dibagi menjadi dua bagian, yaitu data latih sebesar 80% dan data uji sebesar 20%, untuk memastikan model dapat menggeneralisasi dengan baik terhadap data yang belum pernah sebelumnya.

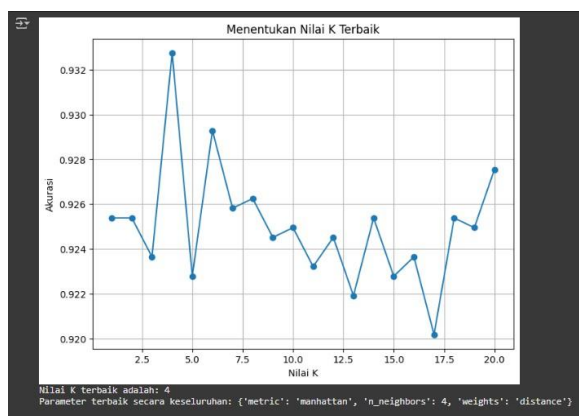
## 3. HASIL DAN PEMBAHASAN

Tuning hyperparameter adalah langkah penting dalam optimasi model machine learning yang bertujuan untuk meningkatkan performa algoritma dengan memilih nilai parameter yang tepat. Hyperparameter adalah parameter yang tidak dipelajari langsung dari data selama pelatihan, melainkan ditetapkan sebelum proses pelatihan dimulai. Oleh karena itu, pencarian nilai hyperparameter yang optimal sangat penting agar model dapat mencapai akurasi yang maksimal. Pada penelitian ini, tiga algoritma klasifikasi yang digunakan, yaitu K-Nearest Neighbors (KNN), Naive Bayes (NB), dan Support Vector Machine (SVM), memiliki beberapa hyperparameter yang dapat dioptimalkan. Untuk mencari kombinasi parameter yang terbaik, digunakan teknik `GridSearchCV` dari pustaka `scikit-learn`, yang secara otomatis melakukan pencarian secara grid (sistematis) pada kombinasi hyperparameter yang ditentukan. Dengan demikian, dapat ditemukan kombinasi terbaik yang memberikan kinerja optimal pada masing-masing model.

### 3.1. Tuning Hyperparameter Model K-Nearest Neighbors (KNN)

Proses tuning hyperparameter Neighbors (KNN) dengan menggunakan teknik `GridSearchCV`. Parameter utama yang diuji

adalah  $n\_neighbors$ , yaitu jumlah tetangga terdekat yang digunakan untuk klasifikasi, dengan nilai  $K$  yang diuji mulai dari 1 hingga 20. Selain itu, dilakukan pula pengujian pada parameter  $weights$  yang menentukan bobot perhitungan jarak, dengan dua opsi yaitu *uniform* (setiap tetangga memiliki bobot yang sama) dan *distance* (tetangga yang lebih dekat memiliki bobot lebih besar). Parameter lain yang dioptimalkan adalah *metric*, yang digunakan untuk mengukur jarak antar data, dengan opsi *Euclidean* dan *Manhattan*. Proses tuning dilakukan menggunakan *GridSearchCV* dengan *5-fold cross validation* dan fungsi evaluasi *scoring='accuracy'*, sehingga setiap kombinasi parameter diuji secara sistematis untuk memperoleh nilai akurasi terbaik. Hasil dari semua percobaan disimpan dalam variabel *results*, kemudian dihitung rata-rata akurasi dari setiap nilai  $K$ . Dari perhitungan ini dapat diperoleh nilai  $K$  terbaik (*best k*) dengan akurasi tertinggi, yang selanjutnya digunakan sebagai konfigurasi optimal dalam pelatihan model KNN. Dengan pendekatan ini, model KNN dapat dipastikan menggunakan kombinasi parameter yang paling optimal dalam mengklasifikasikan status gizi balita berdasarkan indikator TB/U, sehingga mampu meningkatkan performa model dibandingkan penggunaan parameter default.



Gambar 2 Visualisasi Akurasi Nilai K dan Parameter

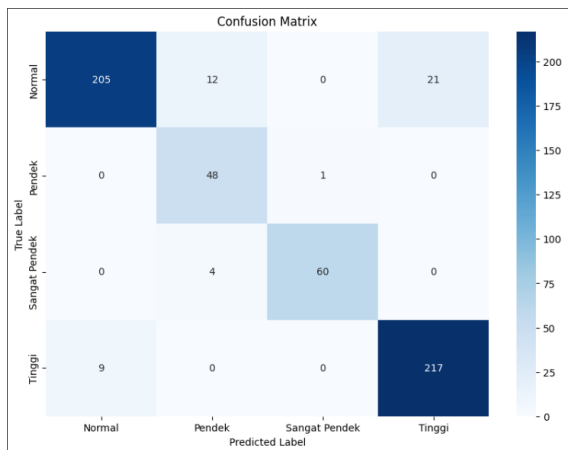
Hasil tuning hyperparameter pada Gambar 2 menunjukkan hubungan antara jumlah tetangga terdekat ( $K$ ) dengan akurasi model KNN. Terlihat bahwa nilai akurasi mengalami fluktuasi pada setiap perubahan nilai  $K$  dari 1 hingga 20. Puncak akurasi tertinggi diperoleh pada nilai  $K = 4$ , dengan akurasi mencapai sekitar 93,3%. Selain jumlah tetangga, parameter lain yang

berkontribusi terhadap hasil terbaik adalah penggunaan *metric = Manhattan* dan *weights = distance*. Kombinasi parameter ini memberikan hasil klasifikasi yang lebih optimal karena jarak *Manhattan* lebih sensitif dalam membedakan data yang memiliki distribusi tidak seragam, sedangkan bobot *distance* memberikan pengaruh lebih besar pada tetangga yang lebih dekat sehingga memperkuat akurasi prediksi. Dengan demikian, konfigurasi optimal model KNN dalam penelitian ini adalah  $K = 4$ , *metric = Manhattan*, dan *weights = distance*. Hasil ini membuktikan bahwa pemilihan hyperparameter yang tepat melalui *GridSearchCV* dapat meningkatkan kinerja model secara signifikan dalam mengklasifikasikan status gizi balita berdasarkan indikator TB/U.

KNN Classification Report (Tuned):				
	precision	recall	f1-score	support
0	0.99	1.00	1.00	204
1	1.00	1.00	1.00	125
2	1.00	1.00	1.00	103
3	0.73	0.69	0.71	32
4	0.67	0.15	0.25	13
5	0.91	1.00	0.95	100
accuracy			0.96	577
macro avg			0.88	577
weighted avg			0.96	577
Accuracy : 0.9636				
Precision : 0.9585				
Recall : 0.9636				
Specificity : 0.9929				
F1 Score : 0.9570				
ROC AUC : 0.9925				

Gambar 3 Classification Report KNN

Selanjutnya pada Gambar 3, *Classification Report* terhadap model *K-Nearest Neighbors (KNN)* dengan parameter terbaik berdasarkan *GridSearchCV*, yang menghasilkan *Accuracy 91,85%*, *Precision 0,9250*, *Recall 0,9185*, *Specificity 0,9703*, *F1 score 0,9191*, serta *ROC AUC 0,9766*. Hasil ini menunjukkan bahwa KNN memiliki kinerja yang cukup baik dalam klasifikasi status gizi balita berdasarkan indikator TB/U, dengan generalisasi yang stabil pada seluruh kelas meskipun terdapat sedikit variasi nilai *precision* dan *recall*, khususnya pada kelas dengan jumlah data yang lebih kecil.



Gambar 4 Confusion Matrix KNN

Gambar 4 menunjukkan confusion matrix dari hasil klasifikasi model KNN setelah dilakukan tuning hyperparameter. Berdasarkan confusion matrix tersebut, dapat dilihat bahwa kelas Normal berhasil diprediksi dengan baik sebanyak 205 data, meskipun masih terdapat kesalahan prediksi ke kelas Tinggi sebanyak 21 data dan ke kelas Pendek sebanyak 12 data. Pada kelas Pendek, sebanyak 48 data berhasil diprediksi dengan benar, namun terdapat 1 data yang salah diprediksi sebagai Sangat Pendek. Kelas Sangat Pendek juga memiliki performa yang cukup baik dengan 60 data terklasifikasi dengan benar, meskipun ada 4 data yang salah diprediksi sebagai kelas Pendek. Sedangkan pada kelas Tinggi, sebanyak 217 data berhasil diprediksi dengan benar, namun masih terdapat 9 data yang salah diprediksi sebagai kelas Normal. Secara keseluruhan, confusion matrix ini memperlihatkan bahwa model KNN mampu mengklasifikasikan mayoritas data dengan benar, khususnya pada kelas Normal dan Tinggi yang memiliki jumlah sampel lebih besar. Kesalahan klasifikasi relatif kecil dan cenderung terjadi pada kelas yang memiliki jumlah sampel lebih sedikit seperti Pendek dan Sangat Pendek, yang menunjukkan adanya ketidakseimbangan data. Meskipun demikian, performa model KNN tetap menunjukkan hasil yang cukup baik dalam mengklasifikasikan status gizi balita berdasarkan indikator TB/U.

### 3.2. STuning Hyperparameter Model Naive Bayes (NB)

Proses tuning hyperparameter untuk algoritma Naive Bayes dengan menggunakan metode GridSearchCV. Parameter utama yang diuji adalah `var_smoothing`, yaitu nilai kecil

yang ditambahkan ke varians setiap fitur agar model lebih stabil dalam menghadapi data dengan distribusi yang sangat bervariasi atau mengandung nilai nol. Parameter ini diuji dengan beberapa skala, mulai dari  $1e-12$  hingga  $1e-6$ . Proses tuning dilakukan menggunakan validasi silang (cross-validation) sebanyak 5 lipatan dengan fungsi evaluasi `scoring='accuracy'`, sehingga setiap nilai parameter dapat dibandingkan berdasarkan akurasi yang diperoleh. Dari hasil pengujian, nilai `var_smoothing = 1e-12` memberikan performa terbaik dengan akurasi tertinggi. Dengan pemilihan parameter ini, model Naive Bayes menjadi lebih optimal dalam menangani variasi data dan mampu meningkatkan akurasi klasifikasi status gizi balita berdasarkan indikator TB/U.

```

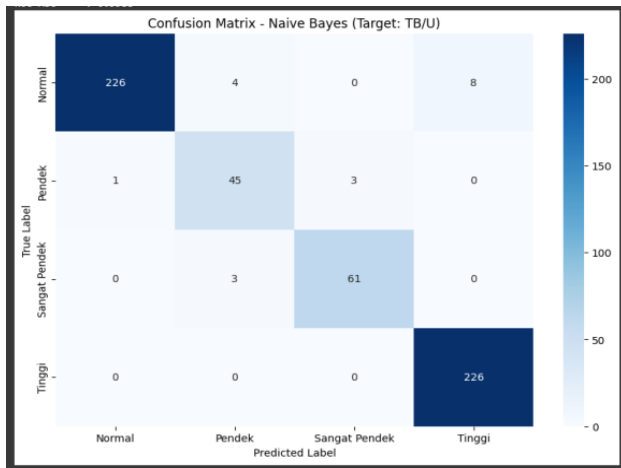
Naive Bayes Classification Report (Tuned):
precision  recall  f1-score  support
0         1.00    1.00    1.00     204
1         1.00    1.00    1.00     125
2         1.00    1.00    1.00     103
3         0.94    1.00    0.97      32
4         1.00    0.85    0.92      13
5         1.00    1.00    1.00     100

accuracy          1.00     577
macro avg         0.99    0.97    0.98     577
weighted avg      1.00    1.00    1.00     577

Accuracy          : 0.9965
Precision         : 0.9967
Recall            : 0.9965
Specificity       : 0.9994
F1 Score          : 0.9964
ROC AUC          : 0.9999
    
```

Gambar 5 Classification Report Naive Bayes

Selanjutnya Pada Gambar 5 Classification Report terhadap model Naive Bayes (NB) dengan parameter terbaik berdasarkan GridSearchCV yaitu `var_smoothing = 1e-12`, yang menghasilkan Accuracy 99,65%, Precision 0,9967, Recall 0,9965, dan F1 score 0,9964, dengan hampir semua kelas memiliki nilai Precision, Recall, dan F1-score sebesar 1,00, kecuali pada salah satu kelas yang memperoleh nilai F1-score 0,92 akibat penurunan Recall sebesar 0,85. Selain itu, nilai ROC AUC mencapai 0,9999, yang menunjukkan kinerja model sangat baik meskipun masih terdapat sedikit kelemahan pada kelas tertentu dengan jumlah sampel yang lebih kecil.



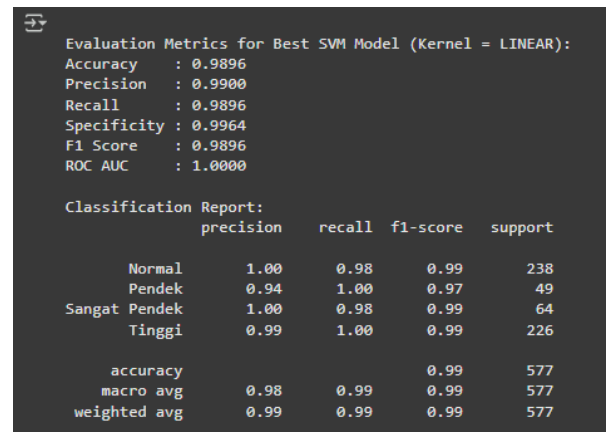
Gambar 6 Confusion Matrix Naive Bayes

Gambar 6 menunjukkan confusion matrix dari hasil klasifikasi model Naive Bayes. Berdasarkan confusion matrix tersebut, dapat dilihat bahwa kelas Sangat Pendek memiliki performa klasifikasi yang cukup baik dengan 61 data berhasil diprediksi dengan benar, meskipun masih terdapat 3 data yang salah diprediksi sebagai kelas Pendek. Pada kelas Pendek, jumlah kesalahan klasifikasi relatif kecil jika dibandingkan dengan kelas lainnya. Sementara itu, kelas Tinggi menunjukkan performa yang sangat baik dengan 226 data berhasil diprediksi dengan benar tanpa adanya kesalahan prediksi ke kelas lain. Secara keseluruhan, confusion matrix ini memperlihatkan bahwa model Naive Bayes mampu mengklasifikasikan sebagian besar data dengan benar. Kesalahan klasifikasi terjadi dalam jumlah kecil dan terutama muncul pada kelas Pendek dan Sangat Pendek yang memiliki karakteristik serupa, sehingga meningkatkan potensi terjadinya misklasifikasi.

### 3.3. Tuning Hyperparameter Model Support Vector Machine (SVM)

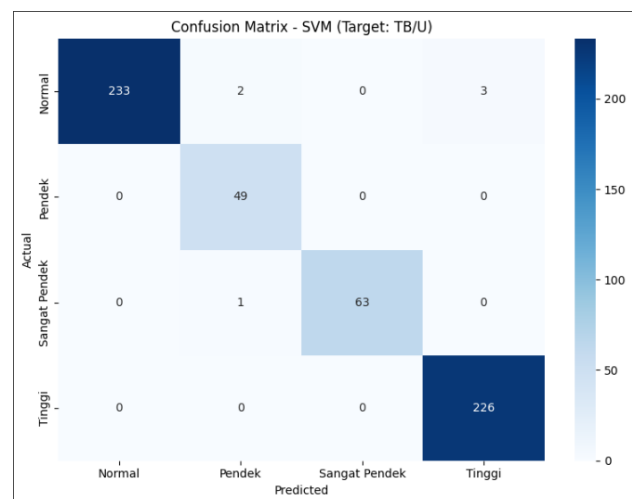
Proses tuning hyperparameter untuk algoritma Support Vector Machine (SVM) dengan menggunakan GridSearchCV. Parameter utama yang diuji meliputi C yang mengontrol margin pemisah, kernel yang menentukan fungsi pemetaan data (linear, polynomial, dan sigmoid), serta gamma yang mengatur pengaruh data pada ruang fitur. Untuk kernel polynomial ditambahkan pula parameter degree untuk mengatur kompleksitas polinomial. Proses tuning dilakukan dengan validasi silang 5-fold dan fungsi evaluasi scoring='accuracy'. Dari hasil pengujian, kombinasi kernel tertentu dengan nilai C dan gamma yang sesuai

menghasilkan performa terbaik dan digunakan sebagai konfigurasi optimal pada model SVM dalam klasifikasi status gizi balita berdasarkan indikator TB/U.



Gambar 7 Classification Report SVM

Pada Gambar 3.6 Classification Report terhadap model Support Vector Machine (SVM) dengan parameter terbaik berdasarkan GridSearchCV yaitu kernel linear, yang menghasilkan Accuracy 98,96%, Precision 0,9900, Recall 0,9896, dan F1-score 0,9896, dengan hampir semua kelas memiliki nilai Precision, Recall, dan F1-score yang sangat tinggi, kecuali pada kelas Pendek yang memperoleh nilai F1-score 0,97 akibat precision yang sedikit lebih rendah sebesar 0,94. Selain itu, nilai Specificity 0,9964 dan ROC AUC 1,0000 menunjukkan bahwa model SVM memiliki kinerja paling optimal dalam mengklasifikasikan status gizi balita berdasarkan indikator TB/U, dengan performa yang konsisten pada hampir seluruh kelas.



Gambar 8 Confusion Matrix SVM

Gambar 8 menunjukkan confusion matrix dari hasil klasifikasi model SVM setelah dilakukan tuning hyperparameter. Berdasarkan confusion matrix tersebut, dapat dilihat bahwa kelas Normal berhasil diprediksi dengan baik sebanyak 233 data, meskipun masih terdapat kesalahan prediksi ke kelas Pendek sebanyak 2 data dan ke kelas Tinggi sebanyak 3 data. Pada kelas Pendek, sebanyak 49 data berhasil diprediksi dengan benar, tanpa terdapat kesalahan prediksi ke kelas lain. Kelas Sangat Pendek juga memiliki performa yang cukup baik dengan 63 data terklasifikasi dengan benar, meskipun ada 1 data yang salah diprediksi sebagai kelas Pendek. Sedangkan pada kelas Tinggi, sebanyak 226 data berhasil diprediksi dengan benar, dan tidak terdapat kesalahan prediksi ke kelas lainnya. Secara keseluruhan, confusion matrix ini memperlihatkan bahwa model SVM mampu mengklasifikasikan mayoritas data dengan benar, khususnya pada kelas Normal dan Tinggi yang memiliki jumlah sampel lebih besar. Kesalahan klasifikasi relatif kecil dan cenderung terjadi pada kelas yang memiliki jumlah sampel lebih sedikit seperti Sangat Pendek, yang menunjukkan adanya ketidakseimbangan data. Meskipun demikian, performa model SVM tetap menunjukkan hasil yang cukup baik dalam mengklasifikasikan status gizi balita berdasarkan indikator TB/U.

### 3.3. Evaluasi Kinerja Ketiga Model

Tabel 2 Evaluasi Kinerja Ketiga Model

Model	Accuray	Precision	Recall	F1 - Score	ROC AUC
KNN	91.85%	0.9250	0.9185	0.9191	0.9766
NB	99.65%	0.9967	0.9965	0.9964	0.9999
SVM	98.96%	0.9900	0.9896	0.9896	1.0000

Berdasarkan hasil pengujian pada Tabel 2, dapat dilihat bahwa algoritma Naive Bayes (NB) memberikan performa terbaik dibandingkan dengan K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM). Naive Bayes memperoleh akurasi tertinggi sebesar 99,65% dengan nilai F1-Score 99,64% dan ROC AUC 0,9999, menunjukkan kemampuan klasifikasi yang sangat baik pada seluruh kelas data. Algoritma SVM dengan kernel linear juga menunjukkan kinerja yang sangat baik dengan akurasi 98,96% dan nilai ROC AUC sempurna

(1,0000), menandakan kemampuan pemisahan kelas yang sangat kuat. Sementara itu, KNN menghasilkan akurasi 91,85% dengan performa yang stabil pada kelas mayoritas namun mengalami penurunan pada kelas minoritas. Hasil ini mengindikasikan bahwa Naive Bayes paling sesuai untuk klasifikasi status gizi balita dalam dataset ini karena mampu mengatasi variasi distribusi data dengan baik.

## 4. PENUTUP

### 4.1. Kesimpulan

Berdasarkan hasil pengujian terhadap tiga algoritma klasifikasi, yaitu K-Nearest Neighbors (KNN), Naive Bayes (NB), dan Support Vector Machine (SVM), dapat disimpulkan bahwa algoritma Naive Bayes memberikan hasil paling optimal dengan akurasi 99,65% dan nilai ROC AUC 0,9999. Sementara itu, algoritma SVM menunjukkan performa yang juga sangat baik dengan akurasi 98,96%, sedangkan KNN menghasilkan akurasi yang relatif lebih rendah. Secara keseluruhan, hasil ini menunjukkan bahwa Naive Bayes merupakan algoritma paling optimal dalam klasifikasi status gizi balita untuk dataset penelitian ini, dengan SVM sebagai alternatif yang juga sangat kompetitif. Di sisi lain, algoritma KNN masih dapat digunakan pada kondisi tertentu dengan dataset berukuran sedang, meskipun tingkat akurasinya lebih rendah dibandingkan dua algoritma lainnya. Dengan demikian, Naive Bayes direkomendasikan sebagai model terbaik untuk klasifikasi status gizi balita pada dataset ini.

## 5. DAFTAR PUSTAKA

- Dwinanto, R.W., Sandi A, A.S., Ardianto, R., 2024. Klasifikasi Berisiko Stunting pada Balita: Perbandingan K-Nearest Neighbor, Naive Bayes, Support Vector Machine. *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi* 8, 264–273. <https://doi.org/10.46880/jmika.Vol8No2.p264-273>
- El-Sofany, H.F., 2024. Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques. *IEEE Access* 12, 106146–106160. <https://doi.org/10.1109/ACCESS.2024.3437181>



- Fawwaz, T., Ihsan, F., Ramadhan, I., Akbar, D.R., Ismanto, E., 2025. Perbandingan Model Machine Learning Untuk Klasifikasi Deteksi Penyakit Jantung 6, 199–205. <https://doi.org/10.37859/coscitech.v6i2.9811>
- Fitrylia Azis, M., Kaesmetan, Y.R., Uyelindo Kupang, S., Perintis Kemerdekaan, J.I., Putih, K., Oebobo, K., Kupang, K., Tenggara Tim, N., 2024. Sistematis : Jurnal Ilmiah Sistem Informasi Penerapan K-NN (K-Nearest Neighbors) Pada Sistem Pakar Diagnosa Gejala Stunting Pada Balita Menggunakan Naïve Bayes Classifier. Oktober 1. <https://doi.org/10.69533>
- Hardiani, T., Putri, R.N., 2024. Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita. *Digital Transformation Technology* 4, 621–627. <https://doi.org/10.47709/digitech.v4i1.4481>
- Homepage, J., Kenia, S., Loka, P., Marsal, A., 2023. MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier Untuk Klasifikasi Status Gizi Pada Balita 3, 8–14.
- Laksono, A.D., Wulandari, R.D., Amaliah, N., Wisnuwardani, R.W., 2022. Stunting among children under two years in Indonesia: Does maternal education matter? *PLoS One* 17. <https://doi.org/10.1371/journal.pone.0271509>
- Lestari, E., Siregar, A., Hidayat, A.K., Yusuf, A.A., 2024. Stunting and its association with education and cognitive outcomes in adulthood: A longitudinal study in Indonesia. *PLoS One* 19. <https://doi.org/10.1371/journal.pone.0295380>
- Muh. Faried Muchtar, Rahma Laila, Dwi Shinta, H. M. Yazdi Pusadan, 2024. Perbandingan Algoritma Naïve bayes Dan Support Vektor Machine Untuk Klasifikasi Status Stunting Pada Balita. *The Indonesian Journal of Computer Science* 13. <https://doi.org/10.33022/ijcs.v13i4.4055>
- Permatasari, T.A.E., Chadirin, Y., 2022. Assessment of undernutrition using the composite index of anthropometric failure (CIAF) and its determinants: A cross-sectional study in the rural area of the Bogor District in Indonesia. *BMC Nutr* 8. <https://doi.org/10.1186/s40795-022-00627-3>
- Pramulia Fitri, R., Saputra, R., Kaswa Putri, M., Kesehatan Payung Negeri Pekanbaru, I., Institut Kesehatan Payung Negeri Pekanbaru, K., Masyarakat Institut Kesehatan Payung Negeri Pekanbaru, K., 2024. Pemberdayaan Masyarakat Pencegahan Stunting pada Balita Melalui Penyuluhan di Desa Bukit Kratai. *Jurnal Medika: Medika* 3.
- Rizal, M., Damanik, M., Muliadi, T., Ahmad, A., 2024. Insights into Stunting Policy Implementation: A Qualitative Analysis in Aceh Province, in: *IOP Conference Series: Earth and Environmental Science*. Institute of Physics. <https://doi.org/10.1088/1755-1315/1359/1/012141>
- Trisnawati, E., Veronica, S., Isnaini, M., Wulandari, E., 2022. Badan Ibu Hamil Dengan Kejadian Stunting Pada Anak Balita Usia 12-59 Bulan Di Wilayah Kerja Puskesmas Gunung 12-59 Months At the Work Area of Public Health Center. *Jurnal Maternitas Aisyah* 3, 229–239.
- Wahyu Dwinanto, R., Setia Sandi, A.A., Ardianto, R., 2024. METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi KLASIFIKASI BERISIKO STUNTING PADA BALITA: PERBANDINGAN K-NEAREST NEIGHBOR, NAÏVE BAYES, SUPPORT VECTOR MACHINE 8. <https://doi.org/10.46880/jmika.Vol8No2.p264-273>
- Wahyuni, S.D., Kusumodestoni, R.H., 2024. Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting. *Bulletin of Information Technology (BIT)* 5, 56–64. <https://doi.org/10.47065/bit.v5i2.1247>
- Widyawati, F., Suhito, H.P., Yassin, W., Agus Santoso, H., 2025. CLASSIFICATION OF TODDLER NUTRITIONAL STATUS USING SUPPORT VECTOR MACHINE

AND RANDOM FOREST TECHNIQUES  
WITH OPTIMAL FEATURE  
SELECTION. Jurnal Teknik Informatika  
(Jutif) 5, 1893–1904.  
<https://doi.org/10.52436/1.jutif.2024.5.6.4>  
162

Wulandari, A.R., Avianto, D., 2024. Sistem  
Pakar Diagnosa Kelainan Stunting Balita  
Menggunakan Metode KNN Berbasis Web.  
Jurnal Indonesia : Manajemen Informatika  
dan Komunikasi 5, 1064–1072.  
<https://doi.org/10.35870/jimik.v5i1.587>