

Optimasi *Support Vector Machine* untuk Klasifikasi Status Gizi Balita

Optimization of Support Vector Machine for the Classification of Nutritional Status in Children

¹Emanuel Deftavalandra Rahmanto, ²Charitas Fibriani*

^{1,2}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah, Indonesia

^{1,2}Jl. Diponegoro No. 52-60, Kel. Salatiga, Kec. Sidorejo, Kota Salatiga, Prov. Jawa Tengah, Indonesia, 50711

*e-mail: charitas.fibriani@uksw.edu

(received: 25 February 2026, revised: 5 May 2026, accepted: 6 May 2026)

Abstrak

Masa balita merupakan periode kritis yang memerlukan pemantauan gizi presisi, namun klasifikasi otomatis sering terkendala oleh masalah ketidakseimbangan data (*imbalanced data*) yang menyebabkan kelas minoritas sulit dideteksi. Penelitian ini bertujuan mengoptimasi *Support Vector Machine* (SVM) dengan kernel *polynomial* untuk meningkatkan sensitivitas deteksi pada kelas kritis. Tanpa melibatkan fitur BMI guna menghindari redundansi, model yang diusulkan mencapai akurasi 98%. Poin kebaruan utama penelitian ini terletak pada capaian F1 Macro Score sebesar 0,86, yang mengonfirmasi bahwa model memiliki performa klasifikasi yang seimbang dan andal di seluruh kategori status gizi. Hasil ini membuktikan keunggulan model dalam mengenali kelas minoritas 'Gizi Buruk' dan 'Gizi Kurang' secara signifikan dibandingkan penelitian terdahulu, sehingga sangat layak direkomendasikan sebagai sistem pendukung keputusan deteksi dini stunting yang objektif.

Kata kunci: klasifikasi, pembelajaran mesin, optimasi, status gizi, balita

Abstract

Toddlerhood is a critical developmental period that requires precise nutritional monitoring. However, automated classification systems are often challenged by imbalanced data, which makes minority classes difficult to detect accurately. This study aims to optimize a Support Vector Machine (SVM) using a polynomial kernel to improve detection sensitivity for critical classes. By excluding BMI features to avoid redundancy, the proposed model achieved an accuracy of 98%. The main novelty of this research lies in its achievement of an F1 Macro Score of 0.86, confirming that the model provides balanced and reliable classification performance across all nutritional status categories. These results demonstrate the model's superiority in identifying the minority classes of Severe Malnutrition and Undernutrition more effectively than previous studies. Therefore, the model is highly recommended as an objective decision support system for the early detection of stunting.

Keywords: classification, machine learning, optimization, nutritional status, toddler

1 Pendahuluan

Masa balita merupakan periode kritis dalam kehidupan manusia, di mana pertumbuhan fisik, kognitif, dan emosional berlangsung sangat pesat. Masalah gizi pada anak usia dini tidak hanya berdampak pada perkembangan tubuh, tetapi juga pada kemampuan intelektual yang bersifat jangka panjang dan sulit diperbaiki di masa mendatang. Berdasarkan data dari Kementerian Kesehatan, prevalensi balita dengan masalah gizi di Indonesia masih tergolong tinggi, baik dalam bentuk kekurangan maupun kelebihan gizi [1]. Salah satu alternatif solusi yang berkembang untuk menangani persoalan ini adalah penggunaan pendekatan komputasi seperti machine learning, khususnya algoritma Support Vector Machine (SVM), untuk mengidentifikasi status gizi berdasarkan data antropometri dan demografi [2]. Berbeda dengan pendekatan tradisional yang seringkali bergantung

pada perhitungan BMI, model ini dilatih menggunakan variabel antropometri mentah untuk memberikan ruang bagi algoritma dalam mengekstraksi pola non-linear secara mandiri. Namun, proses pengukuran dan klasifikasi gizi ini masih sering dilakukan secara manual, sehingga rentan terjadi kesalahan klasifikasi, keterlambatan deteksi, dan rendahnya efisiensi pelayanan kesehatan di tingkat posyandu dan puskesmas.

Mengingat pentingnya akurasi dalam klasifikasi status gizi balita, maka muncul pertanyaan utama dalam penelitian ini: *Bagaimana membangun model klasifikasi yang efektif dan akurat untuk menentukan status gizi balita berdasarkan data antropometri menggunakan algoritma Support Vector Machine (SVM) dan metode perbaikan parameter yang sesuai?* Permasalahan ini muncul seiring dengan kebutuhan akan sistem cerdas yang mampu menangani volume data besar secara efisien dan dapat meningkatkan kecepatan serta ketepatan dalam proses pengambilan keputusan di bidang kesehatan masyarakat [3].

Penelitian ini bertujuan untuk membangun dan mengevaluasi model klasifikasi status gizi balita menggunakan algoritma Support Vector Machine (SVM) yang ditingkatkan melalui teknik tuning dan pengujian berbagai kernel. Pendekatan ini diharapkan dapat meningkatkan akurasi serta F1-score klasifikasi, khususnya pada kelas minoritas seperti gizi kurang dan gizi buruk. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi konfigurasi parameter terbaik agar model SVM dapat diimplementasikan sebagai solusi klasifikasi berbasis machine learning di bidang kesehatan [4].

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan di bidang kesehatan, khususnya untuk klasifikasi status gizi balita secara otomatis menggunakan algoritma Support Vector Machine (SVM). Dengan pemanfaatan model ini, proses identifikasi gizi yang sebelumnya dilakukan secara manual dapat diotomatisasi dan ditingkatkan akurasinya. Sistem ini berpotensi digunakan oleh tenaga medis, kader posyandu, hingga pengembang aplikasi kesehatan anak di masa depan [5].

Pendekatan ini sejalan dengan penelitian sebelumnya yang juga menekankan pentingnya optimasi parameter untuk meningkatkan kinerja SVM dalam kasus klasifikasi yang kompleks. Sebagai contoh, penelitian oleh Fibriani dan Kristiyani (2025) menerapkan algoritma Support Vector Machine (SVM) yang dioptimasi menggunakan Grid Search CV untuk melakukan klasifikasi kondisi mental siswa berdasarkan berbagai fitur. Penelitian tersebut, yang juga menangani ketidakseimbangan data menggunakan SMOTE [6], menunjukkan bahwa proses optimasi parameter merupakan langkah krusial dalam melatih model. Meskipun demikian, model yang mereka hasilkan masih menunjukkan tantangan dalam mengenali kelas minoritas (*Presence*), yang ditandai dengan nilai *recall* yang rendah. Hal ini memperkuat justifikasi penelitian ini untuk fokus pada *improved SVM*, di mana eksplorasi kernel dan *tuning* parameter yang cermat sangat penting untuk mencapai performa yang tidak hanya akurat tetapi juga seimbang di semua kelas, termasuk kelas minoritas seperti "gizi buruk" dan "gizi kurang".

2 Tinjauan Literatur

Penelitian oleh Nadroh memanfaatkan dataset balita dari posyandu dengan fitur antropometri dan demografis, yang kemudian diproses melalui skema encoding, normalisasi, dan pembagian data. Model SVM dengan kernel polynomial dioptimasi melalui Grid Search, dan hasil terbaik dibandingkan menggunakan metrik akurasi serta F1-score untuk mengukur ketepatan prediksi pada kelas minoritas. Konfigurasi optimal dari parameter kernel menunjukkan bahwa model SVM yang ditingkatkan mampu menghasilkan prediksi yang lebih stabil dan akurat dibandingkan baseline KNN atau SVM default, sehingga dapat menjadi solusi dalam otomatisasi klasifikasi status gizi balita di layanan kesehatan tingkat pertama [7].

Support Vector Machine (SVM) adalah algoritma klasifikasi *supervised learning* yang bekerja dengan prinsip fundamental untuk menemukan sebuah *hyperplane* atau batas keputusan yang paling optimal antar kelas data, seringkali dengan memetakannya ke dalam ruang berdimensi tinggi untuk menangani pemisahan yang kompleks. Keunggulan utama SVM terletak pada kemampuannya untuk memaksimalkan margin pemisah, yaitu jarak antara *hyperplane* dengan titik data terdekat (disebut *support vector*) dari setiap kelas, yang menghasilkan model dengan kemampuan generalisasi yang baik dan lebih tahan *overfitting*. Selain itu, kapabilitas unggul SVM lainnya adalah kemampuannya menangani data non-linear secara efisien melalui mekanisme "kernel trick"; dengan

bantuan fungsi kernel, algoritma ini dapat menemukan batas keputusan yang melengkung dan kompleks seolah-olah data tersebut berada di dimensi yang lebih tinggi, tanpa perlu melakukan transformasi data secara eksplisit [8].

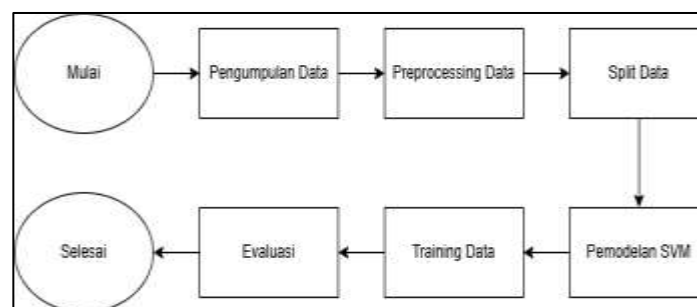
Studi oleh Gaffar menunjukkan bahwa penggunaan kernel polynomial pada SVM dapat meningkatkan performa klasifikasi status gizi balita secara signifikan, terutama ketika dikombinasikan dengan validasi silang berbasis K-Fold yang menghasilkan rata-rata akurasi mencapai 96,98% [9]. Hal ini menunjukkan potensi tinggi dari penggunaan kernel yang sesuai pada kasus klasifikasi dengan fitur kompleks. Penelitian lain oleh Banurea menambahkan bahwa peran teknik hyperparameter tuning seperti Grid Search juga memberikan kontribusi penting dalam menaikkan akurasi dari SVM, sebagaimana dibuktikan melalui peningkatan skor F1 makro dalam klasifikasi stunting pada data berskala besar [10].

Terakhir, pendekatan klasifikasi gizi berbasis metode Naïve Bayes sebagaimana digunakan oleh Harliana et al. telah menunjukkan akurasi sebesar 86% dengan pengolahan spasial berbasis GIS. Namun, metode tersebut belum mengintegrasikan teknik optimasi parameter lanjutan maupun strategi penanganan distribusi data yang tidak seimbang [11]. Oleh karena itu, penelitian ini hadir dengan fokus pada bagian yang belum ditangani secara mendalam, yaitu eksplorasi sistematis terhadap tuning kernel SVM dan dampaknya terhadap kestabilan prediksi di kelas-kelas minoritas. Dengan menggabungkan pendekatan SVM dan optimasi parameter, diharapkan kontribusi penelitian ini mampu memperkuat strategi klasifikasi status gizi yang lebih adaptif dan presisi dalam konteks implementasi dunia nyata.

Berdasarkan tinjauan dari beberapa literatur tersebut, belum ada yang secara komprehensif mengintegrasikan *tuning* parameter tingkat lanjut untuk menangani ketidakseimbangan kelas ekstrem. Oleh karena itu, kebaruan (*novelty*) dari penelitian ini terletak pada pemanfaatan algoritma SVM dengan kernel polynomial yang dioptimasi secara spesifik untuk meningkatkan sensitivitas (F1 Macro Score) pada pendeteksian kelas minoritas yang kritis, yaitu 'Gizi Buruk' dan 'Gizi Kurang'.

3 Metode Penelitian

Metode penelitian yang digunakan dalam studi ini dirancang secara sistematis untuk memastikan proses klasifikasi status gizi balita berjalan secara objektif dan terukur. Alur kerja penelitian secara menyeluruh, mulai dari tahap awal pengumpulan data hingga tahap evaluasi akhir model, disajikan pada Gambar 1.



Gambar 1 Tahapan penelitian

Berdasarkan Gambar 1, penelitian ini berfokus pada pengembangan dan evaluasi model klasifikasi status gizi balita dengan algoritma Support Vector Machine (SVM) yang dioptimasi. Data yang digunakan dalam penelitian ini merupakan dataset sekunder yang bersumber dari platform repositori data publik, <https://www.kaggle.com/datasets/alfiyahabdillah/balita>. Dataset asli yang diunduh memiliki fitur-fitur yang tidak sepenuhnya selaras dengan parameter antropometri standar yang dibutuhkan, sehingga dilakukan penyesuaian nilai menggunakan teknik pemetaan kategori (*categorical mapping*) berdasarkan rentang nilai standar kesehatan. Sebagai contoh, rentang umur 0-1 tahun dipetakan menjadi kategori 'U1', dan berat badan 0-3 kg dipetakan menjadi 'BB1'. Proses standarisasi ini menghasilkan dataset final yang lebih representatif untuk penelitian. Proses ini menghasilkan dataset final (DATA_BALITA.xlsx) yang lebih representatif untuk penelitian, yang terdiri dari total 1115 baris data. Dataset ini mencakup lima kolom utama yang digunakan untuk pemodelan. Empat kolom berfungsi sebagai variabel fitur (independen), yaitu Umur balita dalam

bulan (numerik), Jenis Kelamin (kategorikal), Berat Badan dalam kg (numerik), dan Tinggi Badan dalam cm (numerik). Satu kolom sisanya adalah variabel target (dependen), yaitu Status Gizi (kategorikal), yang diklasifikasikan ke dalam empat label: "Gizi Buruk", "Gizi Kurang", "Gizi Normal", dan "Gizi Lebih". Meskipun variabel BMI dihapus sebagai fitur masukan model guna mencegah redundansi, penentuan label target 'Status Gizi' tetap didasarkan pada standar antropometri medis yang berlaku. Labeling ini merujuk pada pemetaan rentang nilai berat badan dan tinggi badan sesuai dengan standar pertumbuhan anak dari Kementerian Kesehatan, sehingga validitas klasifikasi medisnya tetap terjaga meskipun model SVM bekerja menggunakan fitur antropometri mentah.

Tahap preprocessing data merupakan langkah krusial untuk menyiapkan dataset mentah sebelum proses pemodelan. Proses ini diawali dengan pengecekan integritas data, yang meliputi pemeriksaan *missing values* (data hilang) dan data duplikat untuk memastikan kualitas dan keunikan setiap baris data. Setelah dipastikan data lengkap dan unik, dilakukan pemilihan fitur dengan menghapus kolom-kolom yang tidak relevan atau bersifat redundan, seperti NO, NAMA BALITA, TINGGI(M), dan TINGGI M2. Pada tahap pemilihan fitur, variabel BMI (Indeks Massa Tubuh) serta turunannya seperti TINGGI(M) dan TINGGI M2 dihapus dari dataset. Hal ini dilakukan karena variabel tersebut memiliki korelasi linear yang sangat tinggi dengan variabel berat badan dan tinggi badan, sehingga penghapusan BMI bertujuan untuk mencegah bias model akibat redundansi fitur. Selanjutnya, dilakukan encoding kategorikal, di mana fitur JENIS KELAMIN dan STATUS GIZI diubah menjadi format numerik menggunakan LabelEncoder. Terakhir, untuk menyeragamkan rentang nilai dan mencegah bias, semua fitur numerik dinormalisasi menggunakan StandardScaler. Keseluruhan tahapan preprocessing ini penting untuk meminimalkan bias dan menjaga kestabilan model selama pelatihan.[12].

Penelitian ini dirancang dalam bentuk supervised learning yang memanfaatkan pembagian data sebesar 80% untuk pelatihan dan 20% untuk pengujian, dengan pendekatan stratified sampling guna menjaga proporsi label kelas minoritas seperti "gizi buruk" atau "gizi kurang". Evaluasi hasil dilakukan dengan confusion matrix dan klasifikasi multikelas berbasis F1 macro untuk memperhatikan seimbangannya prediksi di semua kelas. Pendekatan ini juga memperhatikan variasi dalam distribusi kelas, sehingga metode ini diklaim lebih adaptif untuk diterapkan pada data yang tidak seimbang[13].

Pemodelan dalam penelitian ini berpusat pada algoritma Support Vector Machine (SVM), yang diimplementasikan menggunakan *library* scikit-learn di lingkungan Python. SVM dipilih karena kemampuannya yang unggul dalam menemukan *hyperplane* optimal untuk memisahkan kelas data, termasuk pada kasus data minoritas. Mengingat data antropometri balita cenderung memiliki pola non-linear, maka diadopsi kernel polynomial yang mampu menangani hubungan data yang kompleks secara efektif. Untuk mencapai performa model yang maksimal, dilakukan proses optimasi parameter (*hyperparameter tuning*) menggunakan teknik Grid Search Cross-Validation. Teknik ini secara sistematis menguji berbagai kombinasi parameter meliputi C (regularisasi), degree (derajat kernel), gamma (koefisien kernel), dan coef0 (koefisien independen) untuk menemukan konfigurasi terbaik. Proses evaluasi dalam *Grid Search* ini menggunakan metrik F1 Macro Score dengan validasi silang 5-fold, untuk memastikan model yang dihasilkan memiliki performa yang seimbang di semua kelas, termasuk kelas minoritas. Seluruh eksperimen komputasi dijalankan pada platform Google Colab untuk menjamin efisiensi dan kemudahan reproduktifitas penelitian[14]

Dasar matematis dari algoritma Support Vector Machine (SVM) dalam membangun *hyperplane* optimal dengan menyertakan variabel toleransi kesalahan (*slack variable*) dapat direpresentasikan melalui Persamaan (1) berikut:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i \quad (1)$$

Keterangan dari variabel-variabel yang terdapat pada Persamaan (1) adalah sebagai berikut:

- w (Vektor Bobot): Ini adalah vektor yang menentukan kemiringan atau orientasi dari garis pemisah (*hyperplane*). Intinya, w mengarahkan ke mana garis pemisah harus menghadap.
- b (Bias): Ini adalah sebuah nilai skalar (angka biasa) yang berfungsi sebagai intersep. Ia menggeser garis pemisah ke atas atau ke bawah tanpa mengubah kemiringannya, memastikan posisinya optimal.
- ξ_i (dibaca "ksi-i" / *Slack Variable*): Ini adalah variabel toleransi kesalahan untuk setiap titik data ke- i .

- Jika $\xi_i = 0$, data terklasifikasi dengan benar dan di luar margin.
 - Jika $0 < \xi_i \leq 1$, data terklasifikasi benar tetapi berada di dalam margin (terlalu dekat dengan garis).
 - Jika $\xi_i > 1$, data tersebut salah diklasifikasikan (berada di sisi yang salah).
- x_i (Vektor Fitur): Ini adalah representasi numerik dari data Anda. Untuk setiap balita ke- i , x_i adalah serangkaian angka yang mewakili fitur-fiturnya (misal: [umur, jenis kelamin, berat, tinggi]).
- y_i (Label Kelas): Ini adalah label status gizi untuk setiap balita ke- i . Dalam SVM biner, nilainya adalah +1 atau -1.

Proses pelatihan model (*training*) merupakan tahap inti dalam penelitian ini, di mana data latih (80% dari total dataset) digunakan untuk "mengajari" model SVM. Pada tahap ini, fungsi `.fit()` dari *library* scikit-learn dieksekusi pada model SVM yang telah dikonfigurasi dengan parameter kernel polynomial terbaik. Selama proses `.fit()`, algoritma secara matematis mempelajari pola dan hubungan non-linear yang kompleks antara variabel-variabel predictor meliputi umur balita (bulan), jenis kelamin, berat badan (kg), dan tinggi badan (cm) dengan kelas target status gizi, yaitu "GIZI BURUK", "GIZI KURANG", "GIZI NORMAL", dan "GIZI LEBIH". Tujuan dari proses ini adalah agar algoritma mampu membangun sebuah *hyperplane* pemisah yang optimal di dalam ruang fitur berdimensi tinggi, sehingga model dapat mengklasifikasikan data baru yang belum pernah dilihat sebelumnya secara akurat.[15].

Analisis dan evaluasi model dilakukan untuk mengukur kinerja dan efektivitas algoritma SVM setelah melalui proses *tuning* parameter. Evaluasi ini berpusat pada penggunaan Confusion Matrix (Matriks Kebingungan), yang merupakan sebuah matriks tabulasi untuk membandingkan hasil prediksi model (y_{pred}) dengan label aktualnya (y_{test})[16].

Confusion Matrix mengkuantifikasi performa model ke dalam empat kuadran utama:

- True Positive (TP): Data positif yang diprediksi dengan benar.
- True Negative (TN): Data negatif yang diprediksi dengan benar.
- False Positive (FP): Data negatif yang salah diprediksi sebagai positif (Kesalahan Tipe I).
- False Negative (FN): Data positif yang salah diprediksi sebagai negatif (Kesalahan Tipe II).

Dari komponen-komponen matriks ini, dihitunglah beberapa metrik evaluasi kuantitatif untuk mengukur kinerja model secara komprehensif, sesuai dengan yang diimplementasikan dalam penelitian ini:

1. Akurasi (Accuracy) = Metrik ini mengukur proporsi total prediksi yang benar dari keseluruhan data uji, yang dihitung menggunakan Persamaan (2).

$$Akurasi = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

2. Presisi (Precision) = Mengukur tingkat ketepatan dari prediksi yang diklasifikasikan sebagai positif (seberapa banyak yang *sebenarnya* positif dari yang *diprediksi* positif) melalui Persamaan (3).

$$Presisi = TP / (TP + FP) \quad (3)$$

3. Perolehan (Recall) = Mengukur kemampuan model untuk menemukan kembali (mendeteksi) semua data yang *sebenarnya* positif sesuai dengan Persamaan (4).

$$Perolehan = TP / (TP + FN) \quad (4)$$

4. F1-Score = Merupakan rata-rata harmonik dari presisi dan perolehan yang diseimbangkan melalui Persamaan (5).

$$F1 - Score = 2 * (Presisi * Perolehan) / (Presisi + Perolehan) \quad (5)$$

Secara khusus, penelitian ini berfokus pada F1 Macro Score. Metrik ini menghitung F1-Score untuk setiap kelas (misal: "Gizi Buruk", "Gizi Kurang", dll.) secara terpisah, lalu mengambil nilai rata-ratanya tanpa memandang jumlah data di tiap kelas. Pendekatan ini sangat penting karena dataset yang digunakan tidak seimbang, dan F1 Macro Score mampu memberikan gambaran yang lebih jujur tentang kemampuan model dalam mengklasifikasi kelas minoritas secara efektif.

4 Hasil dan Pembahasan

Dataset berhasil di-load.

	NO	NAMA BALITA	JENIS KELAMIN	UMUR	BERAT(KG)	TINGGI (CM)	TINGGI(M)	\
0	1	M. ADRIEL	L	49	17.4	102.1	1.021	
1	2	ALISA SALSA	P	44	11.5	91.6	0.916	
2	3	AULIA R	P	45	12.5	96.7	0.967	
3	4	M. AIDIL	L	48	14.9	94.7	0.947	
4	5	RATU AQILAH HSB	P	37	12.7	94.5	0.945	

	TINGGI M2	BMI	STATUS GIZI
0	1.042	16.70	Gizi Baik
1	0.839	13.71	Gizi Baik
2	0.935	13.37	Gizi Baik
3	0.897	16.61	Gizi Baik
4	0.893	14.22	Gizi Baik

Gambar 2 Contoh data

Dari Gambar 2, terlihat lima contoh dataset awal yang terdiri dari beberapa kolom seperti NO, NAMA BALITA, JENIS KELAMIN, UMUR, BERAT(KG), TINGGI (CM), TINGGI(M), TINGGI M2, BMI, dan STATUS GIZI. Identifikasi awal ini mengonfirmasi keberadaan kolom-kolom yang tidak relevan secara langsung untuk proses pemodelan (misalnya NO, NAMA BALITA), serta kolom-kolom turunan (TINGGI(M), TINGGI M2, BMI) yang berpotensi menimbulkan redundansi atau multi-kolinearitas. Oleh karena itu, kolom-kolom ini akan ditangani secara spesifik pada tahapan *preprocessing* data untuk memastikan dataset yang optimal untuk pelatihan model

```
# === 3. Preprocessing ===
df = df.drop(columns=["NO", "NAMA BALITA", "TINGGI(M)", "TINGGI M2", "BMI"])

le_gender = LabelEncoder()
df["JENIS KELAMIN"] = le_gender.fit_transform(df["JENIS KELAMIN"])

le_status = LabelEncoder()
df["STATUS GIZI"] = le_status.fit_transform(df["STATUS GIZI"])
```

Gambar 3 Pembersihan kolom

Sebagaimana ditunjukkan pada Gambar 3, tahap awal pengolahan hasil difokuskan pada pembersihan data melalui penghapusan kolom yang tidak relevan dan berpotensi menimbulkan redundansi. Kolom administratif seperti 'NO' dan 'NAMA BALITA' dihapus karena tidak memiliki nilai informatif bagi model, sementara fitur turunan seperti 'TINGGI(M)', 'TINGGI M2', dan 'BMI' juga dieliminasi dari daftar fitur utama. Khusus untuk variabel BMI, penghapusan dilakukan secara sengaja untuk mencegah masalah multi-kolinearitas dalam model SVM, karena informasi indeks tersebut sudah direpresentasikan secara intrinsik melalui variabel berat badan dan tinggi badan. Langkah ini memastikan bahwa model murni mengandalkan fitur antropometri dasar untuk mengekstraksi pola secara mandiri, sehingga meningkatkan objektivitas dan efisiensi dalam proses klasifikasi status gizi.

```
# 4.5. Pembagian Data (Split Data)
# 80% data latih, 20% data uji, dengan stratify=y
X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42, stratify=y
)
print(f"Data telah dibagi: {len(X_train)} data latih dan {len(X_test)} data uji.")
print("Preprocessing selesai.\n")
```

Gambar 4 Encoding

Pada Gambar 4, kolom kategorikal seperti “JENIS KELAMIN” dan “STATUS GIZI” diubah menjadi bentuk numerik menggunakan teknik label encoding. Data numerik kemudian dinormalisasi dengan StandardScaler untuk menyamakan skala antar fitur. Visualisasi distribusi data menunjukkan bahwa kelas status gizi tidak seimbang; kelas “GIZI NORMAL” dan “GIZI LEBIH” mendominasi, sedangkan “GIZI BURUK” dan “GIZI KURANG” hanya mencakup sebagian kecil data. Oleh karena itu, digunakan teknik pembagian data secara stratified split untuk menjaga proporsi antar kelas tetap seimbang di data latih dan uji.

```
# 4.2. Encoding Kategorikal
# Mengubah 'JENIS KELAMIN' menjadi angka
le_gender = LabelEncoder()
df["JENIS KELAMIN"] = le_gender.fit_transform(df["JENIS KELAMIN"])

# Mengubah 'STATUS GIZI' (target) menjadi angka
le_status = LabelEncoder()
df["STATUS GIZI"] = le_status.fit_transform(df["STATUS GIZI"])
# Menyimpan nama kelas untuk laporan akhir
target_names = le_status.classes_
print("Fitur kategorikal telah di-encode.")
```

Gambar 5 Pembagian data

Pada Gambar 5, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Model awal menggunakan algoritma Support Vector Machine (SVM) dengan kernel polynomial karena data tidak linear. Pada tahap ini, model dilatih tanpa tuning, hanya dengan parameter default. Hasil pelatihan menunjukkan bahwa model sudah cukup mampu mengenali pola, namun performanya masih bisa ditingkatkan terutama pada kelas minoritas. Oleh karena itu, dilanjutkan dengan optimasi parameter untuk memaksimalkan performa klasifikasi di seluruh kelas.

```
# 5.1. Definisikan Parameter Grid
param_grid_poly = {
    'C': [0.1, 1, 10],          # Parameter regularisasi
    'degree': [2, 3, 4],       # Pangkat kernel polynomial
    'gamma': ['scale', 'auto'], # Koefisien kernel
    'coef0': [0.0, 0.5, 1.0]  # Koefisien independen
}

# 5.2. Inisialisasi Model SVM
svm_poly = SVC(kernel='poly', probability=True)

# 5.3. Inisialisasi Grid Search
# Menggunakan cv=5 (5-fold cross-validation)
# Menggunakan scoring='f1_macro' untuk fokus pada kelas minoritas
grid_search_poly = GridSearchCV(
    svm_poly,
    param_grid=param_grid_poly,
    cv=5,
    scoring='f1_macro',
    n_jobs=-1 # Menggunakan semua core CPU agar lebih cepat
)

# 5.4. Latih Grid Search
grid_search_poly.fit(X_train, y_train)

# 5.5. Dapatkan Model Terbaik
best_poly_svm = grid_search_poly.best_estimator_
print("Optimasi Grid Search selesai.")
print(f"Parameter terbaik ditemukan: {grid_search_poly.best_params_}\n")
```

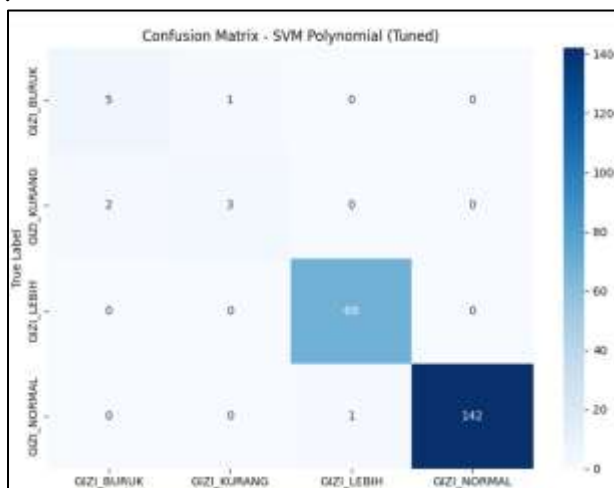
Gambar 6. Pemodelan optimasi SVM

Pada Gambar 6, proses optimasi menggunakan teknik *Grid Search Cross Validation* dilakukan untuk mencari kombinasi parameter terbaik bagi model SVM kernel polynomial tanpa melibatkan variabel BMI menggunakan persamaan (1). Hasil pengujian menunjukkan bahwa kombinasi parameter paling optimal adalah nilai $C = 10$, $degree = 4$, $gamma = 'auto'$, dan $coef0 = 1.0$. Parameter ini dipilih karena menghasilkan nilai *F1 Macro* tertinggi pada proses validasi silang, yang menandakan kemampuan generalisasi model yang baik pada seluruh kelas status gizi.

	precision	recall	f1-score	support
GIZI_BURUK	0.714	0.833	0.769	6.000
GIZI_KURANG	0.750	0.600	0.667	5.000
GIZI_LEBIH	0.986	1.000	0.993	69.000
GIZI_NORMAL	1.000	0.993	0.996	143.000
accuracy	0.982	0.982	0.982	0.982
macro avg	0.862	0.857	0.856	223.000
weighted avg	0.982	0.982	0.982	223.000

Gambar 7 Evaluasi

Gambar 7 menyajikan metrik evaluasi akhir dari model SVM yang telah dioptimasi, yang diuji pada 223 data uji tanpa menyertakan variabel turunan (BMI). Evaluasi model ini menggunakan Persamaan (1), (2), (3) dan (4). Berdasarkan hasil evaluasi model SVM, model ini berhasil mencapai tingkat akurasi keseluruhan yang sangat tinggi, yaitu 98%. Selain metrik akurasi, performa model pada dataset yang tidak seimbang dievaluasi secara ketat menggunakan *F1 Macro Score*, yang mencapai nilai 0,86. Angka ini mencerminkan kinerja yang solid dan realistis dari algoritma SVM dalam mengekstrak pola kompleks langsung dari fitur antropometri mentah. Secara spesifik, model menunjukkan keandalan yang mumpuni dalam menangani kelas minoritas kritis, dibuktikan dengan nilai *recall* sebesar 0,83 (mampu mendeteksi 83% kasus aktual) untuk label 'Gizi Buruk' dan *precision* 0,75 untuk 'Gizi Kurang'.



Gambar 8 Confusion matrix

Analisis lebih mendalam terhadap performa model divisualisasikan melalui *Confusion Matrix* pada Gambar 8. Secara keseluruhan, model menunjukkan akurasi yang luar biasa dalam mengklasifikasikan kelas mayoritas, yaitu 'Gizi Normal' (142 prediksi benar dari total 143 data) dan 'Gizi Lebih' (69 prediksi benar dari 69 data, atau akurasi 100% pada kelas ini). Namun, tantangan terbesar terletak pada klasifikasi kelas minoritas, di mana distribusi data sangat terbatas. Pada kelas 'Gizi Buruk', dari total 6 data aktual, model berhasil memprediksi 5 data dengan benar, sementara 1 data salah diklasifikasikan sebagai 'Gizi Kurang'. Pada kelas 'Gizi Kurang' (total 5 data aktual), model memprediksi 3 data dengan tepat, namun keliru mengenali 2 data sebagai 'Gizi Buruk'. Kesalahan klasifikasi (*misclassification*) yang terjadi secara eksklusif hanya terjadi antara kelas 'Gizi Buruk' dan 'Gizi Kurang'. Hal ini sangat wajar dari sudut pandang medis dan komputasi, mengingat batas (*threshold*) antropometri antara kedua kategori tersebut sangat tipis dan rentan terhadap *noise* data mentah. Meskipun terdapat sedikit penurunan *recall* pada kelas minoritas setelah penghapusan variabel indeks (BMI), model ini tetap dinilai sangat efektif karena tidak ada kasus malnutrisi ('Gizi Buruk' atau 'Gizi Kurang') yang diklasifikasikan secara fatal sebagai kondisi 'Gizi Normal' atau sebaliknya. Keandalan ini membuktikan bahwa algoritma SVM mampu mempelajari batasan keputusan (*decision boundary*) yang aman dan relevan untuk aplikasi klinis.

Model yang dikembangkan dalam penelitian ini menunjukkan peningkatan performa signifikan dibandingkan studi sebelumnya. Huswaton Hasanah dkk. melaporkan akurasi 81% setelah tuning Grid Search[14], sedangkan penelitian ini mencapai 98%. Keunikan penelitian ini tidak hanya pada peningkatan akurasi, tetapi juga pada evaluasi menyeluruh terhadap semua kelas, termasuk minoritas, dan penggunaan dataset terbuka (Kaggle) yang dapat direplikasi oleh peneliti lain. Selain itu, penggunaan SVM dengan kernel polynomial memberikan keunggulan dalam mengatasi data non-linear yang umum pada data klasifikasi status gizi balita.

Pencapaian akurasi 98% dan F1 Macro Score sebesar 0,86 mengukuhkan keunggulan implementasi model yang diusulkan. Sebagai perbandingan, penelitian oleh Fibriani dan Kristiyani (2025) yang juga menggunakan optimasi SVM untuk menangani ketidakseimbangan data, melaporkan kesulitan yang sangat besar dalam mengenali kelas minoritas (*Presence*), dengan nilai *recall* yang anjlok di angka 0,26 dan F1-score 0,35. Hal ini membuktikan bahwa pendekatan optimasi parameter pada SVM dengan kernel polynomial dalam penelitian ini terbukti jauh lebih adaptif, stabil, dan superior dalam menjaga sensitivitas prediksi kelas minoritas dibandingkan pendekatan pada studi kasus tersebut.

5 Kesimpulan

Penelitian ini berhasil mengembangkan dan mengevaluasi model klasifikasi status gizi balita berbasis algoritma Support Vector Machine (SVM) kernel polynomial yang dioptimasi dengan konfigurasi parameter $C=10$, $degree=4$, $gamma='auto'$, dan $coef0=1.0$. Dengan secara tegas menghindari penggunaan fitur konversi redundan seperti Indeks Massa Tubuh (BMI), model ini murni mengandalkan fitur antropometri dasar untuk mencapai tingkat akurasi yang sangat baik, yaitu sebesar 98%, disertai dengan nilai F1 Macro Score 0,86. Keseimbangan performa ini membuktikan kemampuan algoritma dalam mengatasi tantangan distribusi data yang tidak seimbang, khususnya dalam mendeteksi kelas minoritas yang krusial seperti gizi buruk dan gizi kurang secara efektif. Berdasarkan komparasi dengan pendekatan terdahulu, model pembelajaran mesin ini terbukti lebih superior dan sangat direkomendasikan untuk diimplementasikan sebagai sistem pendukung keputusan cerdas guna membantu tenaga kesehatan di posyandu dan puskesmas dalam mempercepat serta meningkatkan akurasi deteksi dini stunting secara digital.

Referensi

- [1] E. N. Candra, I. Cholissodin, dan R. C. Wihandika, "Klasifikasi Status Gizi Balita menggunakan Metode Optimasi Random Forest dengan Algoritme Genetika (Studi Kasus: Puskesmas Cakru)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, Vol. 6, No. 9, hlm. 4467–4475, 2022. [Online].
- [2] H. Hananti dan K. Sari, "Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita Studi Kasus pada Puskesmas Salisingan," *Seminar Nasional Official Statistics*, Vol. 2021, No. 1, hlm. 1036-1043, 2021.
- [3] S. Larissa, Z. Lewoema, dan P. T. Prasetyaningrum, "Implementasi Data Mining pada Klasifikasi Status Gizi Bayi dengan Metode Decision Tree CHAID (Studi Kasus: Puskesmas Godean 1 Yogyakarta)," *Journal of Information Technology Ampere*, Vol. 5, No. 1, 2024.
- [4] A. W. Septyanto dan H. L. Hariyanto, "Perbandingan Teknik Klasifikasi Catatan Medis untuk Indeks Antropometri Status Gizi Balita," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, Vol. 6, No. 1, hlm. 229–235, Jan 2024.
- [5] S. Kenia, P. Loka, dan A. Marsal, "Comparison Algorithm of K-Nearest Neighbor and Naïve Bayes Classifier for Classifying Nutritional Status in Toddlers," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, Vol. 3, No. 1, hlm. 8–14, 2023.
- [6] C. Fibriani dan D. N. Kristiyani, "Model Klasifikasi Mental Siswa menggunakan Algoritma Support Vector Machine," *Progresif: Jurnal Ilmiah Komputer*, Vol. 21, No. 2, hlm. 472, Agu 2025.
- [7] A. Nadroh, D. N. Triwibowo, dan R. B. B. Sumantri, "Klasifikasi Status Gizi Balita menggunakan Algoritma Support Vector Machine dengan Optimasi Grid Search Cross-Validation," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, Vol. 8, No. 2, hlm. 250–257, Okt 2024.
- [8] A. Jalil, A. Homaidi, dan Z. Fatah, "Implementasi Algoritma Support Vector Machine untuk

<http://sistemasi.ftik.unisi.ac.id>

- Klasifikasi Status *Stunting* pada Balita,” *G-Tech: Jurnal Teknologi Terapan*, Vol. 8, No. 3, hlm. 2070–2079, Jul 2024.
- [9] A. W. M. Gaffar, A. M. Halis, P. Purnawansyah, dan S. R. Jabir, “Penerapan *Algoritma Support Vector Machine* untuk Klasifikasi *Stunting* pada Balita di Kabupaten Enrekang,” *Jurnal Minfo Polgan*, Vol. 13, No. 1, hlm. 286–292, Apr 2024.
- [10] M. Banurea, D. Betaria Hutagaol, dan O. Sihombing, “Klasifikasi Penyakit *Stunting* dengan menggunakan *Algoritma Support Vector Machine* dan *Random Forest*,” *Jurnal TEKINKOM*, Vol. 6, No. 2, hlm. 2023, 2023.
- [11] R. D.R. Yusron dan I. Machfud, “Klasifikasi dan Monitoring Status Gizi Balita melalui Penerapan Metode *Naïve Bayes Classification* berbasis *GIS*,” *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, Vol. 4, No. 02, hlm. 161–168, 2022.
- [12] L. Pebrianti *et al.*, “Perbandingan Metode *Algoritma Supervised Naive Bayes* dan *Support Vector Machine (SVM)* dalam Klasifikasi Penderita *Stunting* di Kabupaten Deli Serdang,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 9, No. 3, hlm. 4327-4333, 2025.
- [13] M. F. Muchtar, R. Laila, D. Shinta, dan H. M. Y. Pusadan, “Perbandingan *Algoritma Naïve Bayes* dan *Support Vektor Machine* untuk Klasifikasi Status *Stunting* pada Balita,” *The Indonesian Journal of Computer Science*, Vol. 13, No. 4, Jul 2024.
- [14] E. Ramon *et al.*, “Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba menggunakan *Algoritma Support Vector Machine (SVM)*,” *Jurnal Sistem Informasi dan Informatika (Simika)*, Vol. 5, hlm. 2622–6901, 2022.
- [15] K. H. Hanif dan N. R. Muntiari, “Penerapan *Algoritma Decision Tree, SVM, Naive Bayes* dalam Deteksi *Stunting* pada Balita,” *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*, Vol. 8, No. 1, hlm. 105-109, 2024.
- [16] M. Amirudin dan A. D. Wowor, “Analisis Perbandingan Klasifikasi Balita Beresiko *Stunting* menggunakan Metode *Support Vector Machine* dan *Decision Tree*,” *Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media*, hlm. 1-11, 2023.