

Klasifikasi Stunting Pada Balita Berdasarkan Status Gizi Menggunakan Pendekatan *Support Vector Machine* (SVM)

Adi Candra^{1*}, Moh. Erkamim², Muhammad Muharrom³, Edhi Prayitno⁴

Program Studi Sistem Komputer, Institut Teknologi dan Bisnis Bina Adinata¹
Program Studi Sistem Informasi Kota Cerdas, Universitas Tunas Pembangunan Surakarta²
Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana
Informatika³

Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri⁴
chandrakirana862@gmail.com^{1*}, erkamim@lecture.utp.ac.id², muhammad.muu@bsi.ac.id³,
edhi.epo@nusamandiri.ac.id⁴

*) Corresponding Author

(received: 15-08-24, revised: 11-09-24, accepted: 02-10-24)

Abstract

Stunting in infants is a serious nutritional issue that impacts the physical and cognitive development of children, particularly in developing countries like Indonesia. With the prevalence of stunting still high, early identification of at-risk infants is crucial to prevent long-term consequences. However, conventional methods for identifying stunting are often inaccurate and require significant resources. The objective of this study is to classify stunting in infants based on nutritional status using machine learning with the Support Vector Machine (SVM) algorithm. The choice of SVM is based on its advantages in handling complex multidimensional data and its capability to optimize the separation between data classes by maximizing the margin. This study also implements various data preprocessing techniques, such as feature standardization, categorical variable encoding, and duplicate data removal, to ensure optimal model performance. The results of the study reveal that the developed SVM model achieved an accuracy of 98.37%, indicating excellent performance in classifying infant nutritional status. These findings suggest that SVM has great potential to be applied in supporting medical decision-making and public health interventions, particularly in the context of monitoring and preventing stunting in infants.

Keywords: *Stunting Classification, Machine Learning, Nutritional Status, Support Vector Machine, SVM*

Abstrak

Stunting pada balita merupakan masalah gizi serius yang berdampak pada perkembangan fisik dan kognitif anak, terutama di negara berkembang seperti Indonesia. Dengan prevalensi stunting yang masih tinggi, identifikasi dini balita yang berisiko sangat penting untuk mencegah dampak jangka panjang. Namun, metode konvensional dalam mengidentifikasi stunting sering kali kurang akurat dan memerlukan banyak sumber daya. Tujuannya penelitian ini dilakukan yaitu untuk mengklasifikasikan stunting pada balita berdasarkan status gizi melalui pembelajaran mesin dengan algoritma *Support Vector Machine* (SVM). Pemilihan SVM didasarkan pada keunggulannya dalam mengolah data multidimensi yang rumit serta kapabilitasnya untuk mengoptimalkan pemisahan antar kelas data dengan memaksimalkan *margin*. Penelitian ini juga menerapkan berbagai teknik prapemrosesan data, seperti standarisasi fitur, pengkodean variabel kategorikal, dan penghapusan data duplikat, untuk memastikan performa optimal model. Hasil penelitian mengungkapkan bahwa model SVM yang dibangun memperoleh akurasi sebesar 98,37%, menandakan kinerja yang sangat baik dalam klasifikasi status gizi balita. Temuan ini mengindikasikan bahwa SVM memiliki potensi besar untuk diaplikasikan dalam mendukung pengambilan keputusan medis dan intervensi kesehatan masyarakat, terutama dalam konteks pemantauan dan pencegahan stunting pada balita.

Kata Kunci: *Klasifikasi Stunting, Pembelajaran Mesin, Status Gizi, Support Vector Machine, SVM*

I. Pendahuluan

Fenomena stunting pada anak di bawah lima tahun masih menjadi tantangan serius di skala global, khususnya di negara-negara berkembang, sebagai salah satu isu kritis dalam aspek nutrisi. Stunting ditandai dengan kelainan perkembangan fisik berupa tinggi badan yang tidak sebanding dengan usia, yang bisa berpengaruh buruk terhadap perkembangan kognitif dan kesehatan anak dalam jangka panjang [1]. Merujuk pada hasil Studi Status Gizi Indonesia (SSGI) yang dipublikasikan Kementerian Kesehatan di tahun 2022, tingkat kejadian stunting di kalangan anak balita Indonesia tercatat sebesar 21,6% [2]. Meskipun angka ini menunjukkan penurunan dari tahun-tahun sebelumnya, masih dibutuhkan intensifikasi strategi guna merealisasikan sasaran pemerintah dalam menekan prevalensi stunting menjadi 14% pada tahun 2024 [3]. Identifikasi dini balita yang berisiko stunting sangat penting untuk mencegah dampak lebih lanjut yang dapat mengganggu kualitas hidup anak di masa depan. Namun, proses klasifikasi dan identifikasi balita stunting berdasarkan status gizi sering kali dilakukan secara manual oleh tenaga kesehatan, yang memerlukan waktu dan sumber daya yang memadai. Disamping itu, metode konvensional seringkali kurang akurat dalam memprediksi stunting karena tidak mempertimbangkan faktor-faktor kompleks yang dapat mempengaruhi status gizi balita. Dalam menghadapi tantangan tersebut, penggunaan teknologi berbasis data dengan menggunakan pendekatan *machine learning* menjadi alternatif yang menjanjikan untuk meningkatkan akurasi dalam mengidentifikasi balita yang berisiko stunting

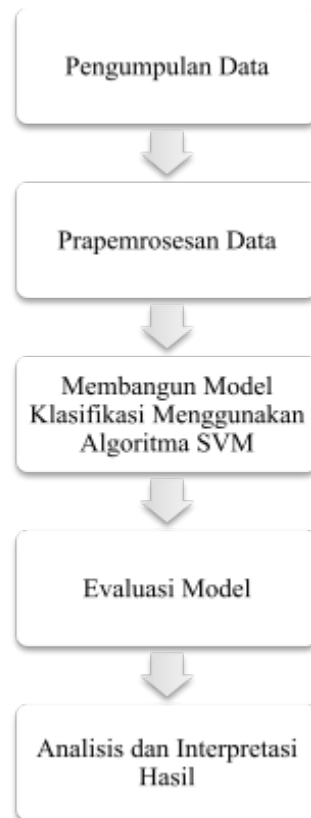
Penelitian terdahulu yang mengenai klasifikasi maupun prediksi stunting pada balita telah dilakukan oleh beberapa peneliti dengan berbagai metode *machine learning*. Terdapat penelitian mengenai prediksi anak stunting menggunakan algoritma *K-Nearest Neighbor* [4]. Pada penelitian ini, Hasil penelitian ini menunjukkan bahwa algoritma K-nearest neighbor menunjukkan kinerja yang baik dengan akurasi sebesar 97,16%. Penelitian selanjutnya, mengenai klasifikasi status gizi balita dengan menerapkan pendekatan *Random Forest* [5]. Evaluasi eksperimen mengindikasikan bahwa rasio pembagian dataset menjadi 90% untuk pelatihan dan 10% untuk validasi memberikan performa terbaik, menghasilkan tingkat akurasi sebesar 88,6%, nilai presisi 88,1%, skor *recall* 88,6%, serta *F1-Score* mencapai 88,2%. Terdapat juga penelitian yang menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan stunting pada balita [6]. Evaluasi yang dilakukan pada fase pengujian menggunakan metode *Naïve Bayes Classifier* menghasilkan performa sebagai berikut: tingkat presisi mencapai 85%, skor *recall* terukur pada 91%, sementara tingkat akurasi mencapai 80%. Berikutnya, penelitian tentang klasifikasi stunting pada balita dengan menerapkan pendekatan *Extreme Gradient Boosting* [7]. Seusai tahap pra-pemrosesan dan pemisahan dataset, dilanjutkan dengan pelatihan dan evaluasi model, studi ini menghasilkan kinerja dengan rincian sebagai berikut: tingkat akurasi mencapai 86%, nilai presisi terukur pada 89%, skor *recall* terhitung 95%, serta *F1-score* tercatat sebesar 92%.

Karakteristik dari data stunting ini yaitu data dengan dimensi tinggi, di mana fitur-fitur mungkin saling berkorelasi dan kompleks. Untuk itu, pada penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan data stunting berdasarkan status gizi. *Support Vector Machine* (SVM) merupakan pendekatan pembelajaran mesin yang dapat diimplementasikan dalam kasus klasifikasi, karena kemampuannya untuk mengatasi data dengan dimensi tinggi serta menghasilkan model yang kuat untuk klasifikasi biner dan multi kelas [8]. Inti dari SVM adalah menentukan *hyperplane* yang paling optimal untuk membedakan kelas-kelas dalam data dengan margin terbesar, yaitu jarak terluas yang memisahkan elemen data terdekat di setiap kelas [9]. SVM bekerja dengan sangat baik dalam ruang fitur yang tidak linear dengan memanfaatkan *kernel trick*, yang memungkinkan data diproyeksikan data ke ruang dimensi yang lebih tinggi, sehingga memungkinkan pemisahan linear data dengan lebih optimal [10]. Kelebihan utama SVM termasuk kemampuannya untuk menangani data yang kompleks dan non-linear, ketahanannya terhadap *overfitting* terutama dalam kasus-kasus di mana jumlah fitur lebih besar daripada jumlah sampel, dan efektivitasnya dalam bekerja dengan data berdimensi tinggi [11]. SVM juga relatif efisien dalam penggunaan memori karena hanya beberapa titik data, yang disebut support vectors, yang digunakan untuk menentukan *hyperplane* pemisah [12].

Penelitian ini bertujuan untuk mengembangkan model yang dapat mengklasifikasikan stunting pada balita berdasarkan status gizi menggunakan pendekatan *Support Vector Machine* (SVM). SVM digunakan dalam penelitian ini untuk melakukan klasifikasi dengan menemukan *hyperplane* yang optimal guna memisahkan kelas-kelas dalam data dengan margin maksimal. Pada penelitian ini algoritma SVM ditingkatkan melalui penerapan beberapa teknik *preprocessing* seperti standarisasi fitur, *encoding* variabel kategorikal, dan penghapusan duplikasi data. Kontribusi penelitian ini adalah pengembangan model klasifikasi stunting pada balita berdasarkan status gizi menggunakan metode SVM yang dioptimalkan melalui teknik *preprocessing* untuk meningkatkan akurasi dan efisiensi dalam identifikasi stunting. Hal ini memastikan bahwa model SVM yang dibangun dapat berfungsi secara optimal dengan data yang bersih dan terstruktur. Dengan adanya model ini, diharapkan proses identifikasi dan klasifikasi balita stunting dapat dilakukan dengan lebih efisien dan akurat, sehingga intervensi gizi dapat dilakukan lebih awal untuk mencegah dampak jangka panjang dari stunting.

II. Metodologi Penelitian

Penelitian yang dilakukan memerlukan panduan yang sistematis yang mengarahkan peneliti dalam merancang, mengimplementasikan, dan mengevaluasi penelitian. Untuk itu diperlukan metodologi penelitian yang memberikan kerangka kerja untuk setiap langkah dalam proses penelitian [13]. Langkah-langkah setiap tahapan penelitian dapat dilakukan diinterpretasikan pada Gambar 1.



Gambar 1. Prosedur Penelitian

Berdasarkan pada Gambar 1, maka secara rinci setiap tahapan yang dilaksanakan dijelaskan sebagai berikut.

Pengumpulan Data

Tahap pengumpulan data adalah langkah awal di mana data yang relevan untuk analisis dikumpulkan. Penelitian ini memanfaatkan dataset yang tersedia di platform Kaggle dengan nama dataaset "*Stunting Toddler Detection*" (yang dapat diakses di: <https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows/>) [14]. Dalam mengklasifikasikan status stunting, dataset ini mengadopsi metode *z-score* yang selaras dengan panduan *World Health Organization* (WHO). Fokus utamanya adalah identifikasi stunting untuk anak di bawah usia lima tahun. Dataset mencakup 121.000 entri, masing-masing memuat informasi mengenai usia, jenis kelamin, tinggi badan, serta status gizi balita. Klasifikasi status gizi dibagi menjadi empat kategori: Sangat Pendek, Pendek, Normal, dan Tinggi.

Prapemrosesan Data

Prapemrosesan data merupakan langkah di mana data yang diperoleh diolah dan disiapkan untuk analisis selanjutnya [15]. Ini mencakup penghapusan duplikasi, penanganan nilai yang hilang, dan transformasi data yang diperlukan, seperti encoding variabel kategorikal dan standarisasi fitur numerik. Misalnya, jenis kelamin balita mungkin perlu diubah menjadi format numerik, dan fitur seperti umur atau tinggi badan perlu distandarisasi agar model SVM dapat bekerja lebih efektif. Setelah data dibersihkan dan ditransformasi, langkah penting lainnya adalah melakukan pemisahan dataset menjadi dua bagian: subset untuk pelatihan model (*training data*) dan subset untuk evaluasi kinerja model (*testing data*). Pembagian data yang digunakan yaitu 80:20. Dengan jumlah data pelatihan yang lebih besar, model dapat menangkap variasi dalam data lebih baik, yang meningkatkan kemampuan generalisasi model ketika dihadapkan pada data baru [16]. Proses pembagian

data ini dilakukan untuk memastikan representasi yang akurat dari setiap kategori status gizi dalam kedua subset data, sehingga menghindari bias dalam proses pelatihan dan evaluasi model.

Membangun Model Klasifikasi Menggunakan Algoritma SVM

Pada tahap ini, model klasifikasi dibangun menggunakan algoritma *Support Vector Machine* (SVM). Prinsip kerja SVM terletak pada pencarian *hyperplane* yang mampu memisahkan kelas-kelas data secara optimal dengan memaksimalkan jarak (*margin*) ke titik-titik data terdekat [17]. Pendekatan ini bertujuan untuk menekan risiko *overfitting* dan meningkatkan kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya dengan lebih akurat [18]. Selain itu, SVM dapat menangani *outlier* dengan baik dan cocok untuk berbagai jenis data, dengan menggunakan *kernel trick* untuk mempermudah pemisahan antar kelas dalam ruang dimensi yang lebih tinggi tanpa menghitung transformasi secara eksplisit. [19]. SVM bekerja dengan baik dalam kasus di mana data memiliki dimensi tinggi, yaitu ketika jumlah fitur lebih besar dari jumlah sampel, serta ketika batas antara kelas-kelas tidak dapat dipisahkan secara linear [20]. Melalui penggunaan kernel, SVM dapat memetakan data ke dimensi yang lebih tinggi, memungkinkan pemisahan yang lebih efektif pada data yang kompleks. SVM memformulasikan masalah klasifikasi sebagai masalah optimasi. Masalah optimisasi yang harus diselesaikan SVM adalah meminimalkan jarak antara *hyperplane* melalui persamaan (1).

$$\frac{1}{2} \|w\|^2 \quad (1)$$

dimana w menunjukkan vektor bobot yang mendefinisikan orientasi *hyperplane*, b merujuk pada bias atau *offset hyperplane* dari asal.

Dengan syarat bahwa semua titik data harus dipisahkan oleh *hyperplane* dengan margin yang benar, maka digunakan persamaan (2).

$$y_i(w \cdot x_i - b) \geq 1 \quad (2)$$

dimana x_i merujuk pada titik data ke- i , dan y_i merepresentasikan label kelas (+1 atau -1) dari titik data ke- i .

Setelah *hyperplane* ditemukan, SVM mengklasifikasikan titik data baru berdasarkan posisinya relatif terhadap *hyperplane*. Titik data akan diprediksi sebagai kelas +1 jika terletak pada satu sisi *hyperplane* serta sebagai kelas -1 jika terdapat pada sisi lainnya.

Evaluasi Model

Pasca-konstruksi model SVM, evaluasi performa dilakukan dengan memanfaatkan set data uji (*testing data*) yang independen dari data pelatihan. Evaluasi model berfungsi untuk menilai kinerja dan keandalan model yang telah dibangun [21]. Untuk evaluasi, digunakan *confusion matrix*, yang merupakan tabel yang menampilkan jumlah prediksi benar maupun salah yang dihasilkan oleh model untuk masing-masing kelas. Berdasarkan *confusion matrix* diperoleh parameter-parameter pengujian seperti *accuracy*, *precision*, *recall*, *F1-score*, dan ROC-AUC yang digunakan untuk menilai seberapa baik model dapat memprediksi kelas yang benar. Akurasi menghitung persentase prediksi yang benar dari total prediksi yang dihasilkan oleh model [22]. Meskipun akurasi menyajikan pandangan umum tentang performa model, metrik ini perlu perhatian khusus apabila data tidak seimbang, di mana satu kelas lebih dominan daripada kelas lainnya. Presisi adalah metrik yang menunjukkan proporsi prediksi positif yang akurat dari keseluruhan prediksi positif yang dihasilkan oleh model, dan sangat penting dalam situasi di mana kesalahan positif palsu (*false positives*) harus diminimalkan [22]. *Recall*, atau sensitivitas, mengukur kemampuan model untuk mendeteksi semua *instance* positif yang sebenarnya, hal ini perlu perhatian karena kegagalan mendeteksi positif memiliki konsekuensi serius [22]. *F1-score* merupakan rata-rata harmonis dari *precision* dan *recall*, yang menciptakan keseimbangan di antara keduanya dan sangat berguna saat kita perlu mempertimbangkan kedua metrik tersebut secara setara, terutama dalam situasi di mana terdapat kompromi antara *precision* dan *recall* [22]. Terakhir, ROC-AUC (*Receiver Operating Characteristic - Area Under Curve*) mengukur sejauh mana model mampu membedakan antara kelas positif dan negatif [23]. AUC memberikan ukuran area di bawah kurva ROC, di mana nilai yang lebih tinggi memperlihatkan kinerja model yang lebih optimal dalam memisahkan kelas-kelas yang berbeda [23]. Metrik ini sangat bermanfaat dalam situasi di mana kita perlu memahami keseimbangan antara *true positive rate* maupun *false positive rate*.

Analisis dan Interpretasi Hasil

Langkah akhir dalam proses penelitian ini melibatkan analisis dan interpretasi hasil yang diperoleh dari evaluasi model. Pada tahap ini, peneliti secara mendalam meninjau parameter uji yang digunakan seperti akurasi, *precision*, *recall*, *F1-score*, dan ROC-AUC untuk mendapatkan pemahaman yang komprehensif tentang kinerja

model. Selain itu, melalui analisis ini, peneliti dapat mengidentifikasi area-area spesifik di mana model mungkin memerlukan perbaikan atau penyempurnaan lebih lanjut. Hasil dari analisis ini juga memberikan wawasan berharga yang dapat digunakan untuk melakukan penyesuaian lebih lanjut pada model. Lebih jauh lagi, wawasan yang diperoleh dari analisis ini juga dapat menginformasikan keputusan berbasis data dalam konteks yang lebih luas, seperti pengembangan strategi intervensi atau pengambilan keputusan kebijakan yang didasarkan pada hasil prediksi model. Dengan demikian, analisis dan interpretasi hasil tidak hanya membantu meningkatkan kualitas model, tetapi juga memastikan bahwa model tersebut memberikan kontribusi nyata terhadap tujuan penelitian dan aplikasi praktisnya.

III. Hasil dan Pembahasan

Untuk membangun model klasifikasi stunting pada balita berdasarkan status gizi dengan mengimplementasikan algoritma SVM (*Support Vector Machine*) dibangun menggunakan editor teks Google Collab melalui bahasa pemrograman Python. Pembuatan model diawali dengan menyiapkan dataset, dimana dataset yang digunakan mencakup 121.000 entri. Pada dataset tersebut masing-masing memuat informasi mengenai usia, jenis kelamin, tinggi badan, serta status gizi balita. Klasifikasi status gizi dibagi menjadi empat kategori: Sangat Pendek, Pendek, Normal, dan Tinggi. Setelah itu dilakukan prapemrosesan untuk memastikan bahwa data telah siap digunakan dalam melakukan klasifikasi. Pertama, data diimpor dari file CSV dan diperiksa untuk informasi umum, termasuk jumlah baris dan kolom, serta adanya nilai-nilai yang hilang atau duplikasi. Setelah itu, duplikasi data dihapus untuk menghindari bias dalam pelatihan model. Selanjutnya, fitur kategorikal seperti Jenis Kelamin dan Status Gizi diubah menjadi format numerik menggunakan “*LabelEncoder*”, yang memungkinkan model SVM untuk memprosesnya dengan lebih baik. Setelah encoding, fitur numerik seperti Umur (bulan) dan Tinggi Badan (cm) diolah melalui standarisasi menggunakan “*StandardScaler*”, yang bertujuan untuk mengatur fitur-fitur ini dalam skala yang sama, agar tidak terdapat fitur yang mendominasi yang lain selama pelatihan model. *Output* dari operasi yang dilakukan di lingkungan Python menggunakan Google Collab mengenai informasi penting tentang *Data Frame* yang digunakan dalam analisis data pada penelitian ini yang tersaji pada Gambar 2.

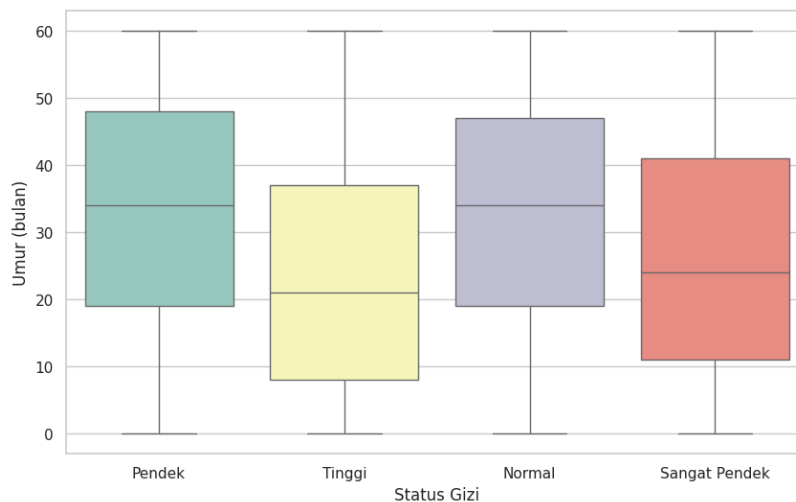
```
      Umur (bulan)  Jenis Kelamin  Tinggi Badan (cm)  Status Gizi
0                0      laki-laki    44.591973         Pendek
1                0      laki-laki    56.705203         Tinggi
2                0      laki-laki    46.863358         Normal
3                0      laki-laki    47.508026         Normal
4                0      laki-laki    42.743494         Sangat Pendek
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 120999 entries, 0 to 120998
Data columns (total 4 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Umur (bulan)          120999 non-null  int64
1   Jenis Kelamin         120999 non-null  object
2   Tinggi Badan (cm)    120999 non-null  float64
3   Status Gizi           120999 non-null  object
dtypes: float64(1), int64(1), object(2)
memory usage: 3.7+ MB
(120999, 4)
Umur (bulan)          0
Jenis Kelamin         0
Tinggi Badan (cm)    0
Status Gizi           0
dtype: int64
81574
(39425, 4)
```

Gambar 2. Informasi *Data Frame* Dari Datasets Yang Digunakan

Pasca tahap pra-pemrosesan, dataset dipartisi menjadi dua bagian: set pelatihan dan set pengujian, dengan rasio pembagian 80:20, di mana mayoritas data (80%) dialokasikan untuk proses pembelajaran model, sedangkan sisanya (20%) diperuntukkan bagi evaluasi kinerja model. Pembagian ini dilakukan dengan menggunakan metode “*train_test_split*” dari “*scikit-learn*”, untuk memastikan bahwa proporsi kelas dalam data pelatihan dan pengujian tetap konsisten dengan proporsi aslinya dalam dataset.

Proses selanjutnya setelah prapemrosesan adalah melakukan visualisasi data untuk memahami distribusi fitur-fitur utama dalam dataset dan bagaimana mereka terkait dengan status gizi balita. Visualisasi pertama yaitu

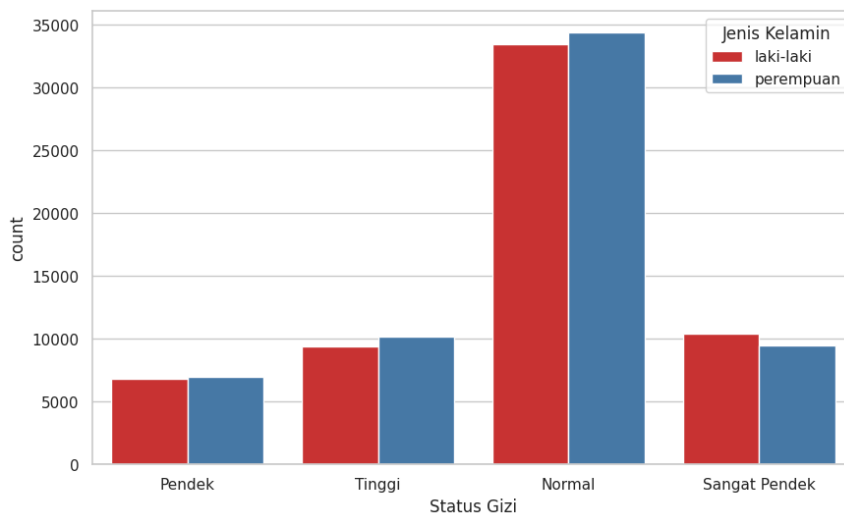
fitur umur berdasarkan status gizi. Visualisasi ini bertujuan untuk melihat bagaimana distribusi umur balita terkait dengan kategori status gizinya. Untuk memberikan gambaran mengenai distribusi data umur berdasarkan status gizi ditampilkan menggunakan *boxplot* pada Gambar 3.



Gambar 3. Distribusi Data Umur Terhadap Status Gizi

Visualisasi yang ditampilkan pada Gambar 3, menunjukkan bahwa distribusi umur balita cenderung bervariasi di setiap kategori status gizi. Visualisasi ini berguna untuk mengidentifikasi pola atau tren yang mungkin ada antara umur dan status gizi balita, yang dapat dimanfaatkan dalam mempelajari pola-pola dari data yang tersedia.

Visualisasi data selanjutnya yaitu fitur jenis kelamin berdasarkan status gizi. Visualisasi ini membantu dalam memahami apakah ada perbedaan distribusi status gizi berdasarkan jenis kelamin. Dengan memplot jumlah balita laki-laki dan perempuan dalam setiap kategori status gizi, maka dapat melihat apakah jenis kelamin mempengaruhi status gizi atau apakah distribusi status gizi serupa di antara kedua jenis kelamin. Untuk memberikan deskripsi yang jelas mengenai distribusi data jenis kelamin berdasarkan status gizi, maka dibuat menggunakan *count plot* pada Gambar 4.

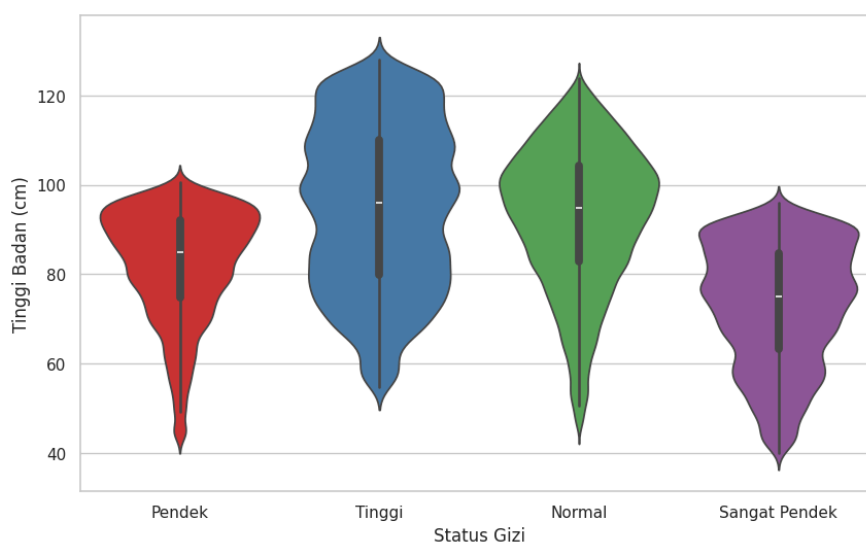


Gambar 4. Distribusi Data Jenis Kelamin Terhadap Status Gizi

Visualisasi yang ditampilkan pada Gambar 3 menjelaskan mengenai distribusi jumlah balita berdasarkan jenis kelamin dalam setiap kategori status gizi. Pada sumbu *horizontal* (sumbu x), terdapat kategori status gizi yang dikelompokkan ke dalam beberapa kelas, sementara sumbu vertikal (sumbu y) menunjukkan jumlah balita dalam setiap kategori tersebut. Warna atau pola dalam setiap batang mewakili jenis kelamin balita, dimana, warna merah menunjukkan jenis kelamin laki-laki serta warna biru merupakan jenis kelamin perempuan.

Analisis data menunjukkan bahwa persebaran status gizi tidak menunjukkan perbedaan signifikan antara balita laki-laki dan perempuan, meskipun terdapat kecenderungan sedikit lebih tinggi pada balita perempuan untuk memiliki status gizi yang normal. Visualisasi ini berguna untuk mengidentifikasi apakah terdapat perbedaan atau kecenderungan tertentu dalam status gizi berdasarkan jenis kelamin, yang dapat menjadi dasar bagi analisis lebih lanjut atau intervensi yang lebih terarah.

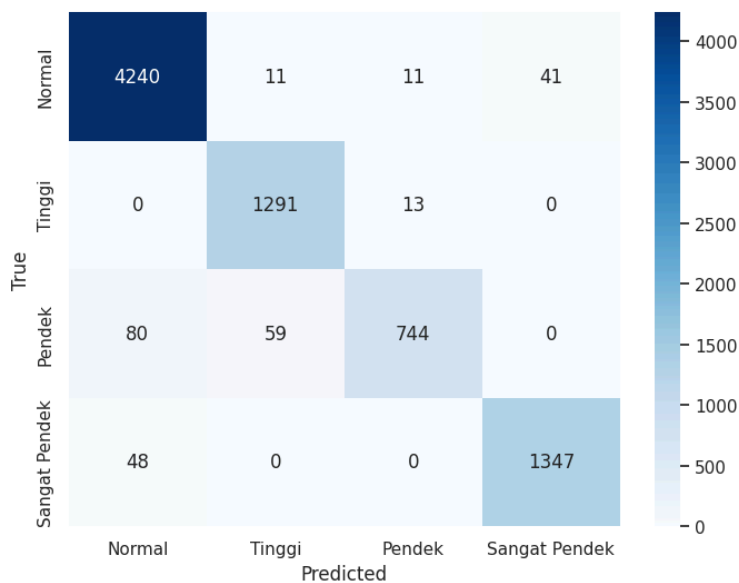
Untuk visualisasi persebaran data berikutnya yaitu fitur tinggi badan berdasarkan status gizi. Visualisasi tinggi badan berdasarkan status gizi bertujuan untuk menunjukkan apakah ada hubungan antara tinggi badan dan status gizi. Hal ini dapat mengungkap pola penting dalam data yang mungkin tidak terlihat hanya dengan melihat statistik deskriptif. Untuk memberikan gambaran mengenai distribusi data tinggi badan berdasarkan status gizi ditampilkan menggunakan *violin plot* pada Gambar 5.



Gambar 5. Distribusi Data Tinggi Badan Terhadap Status Gizi

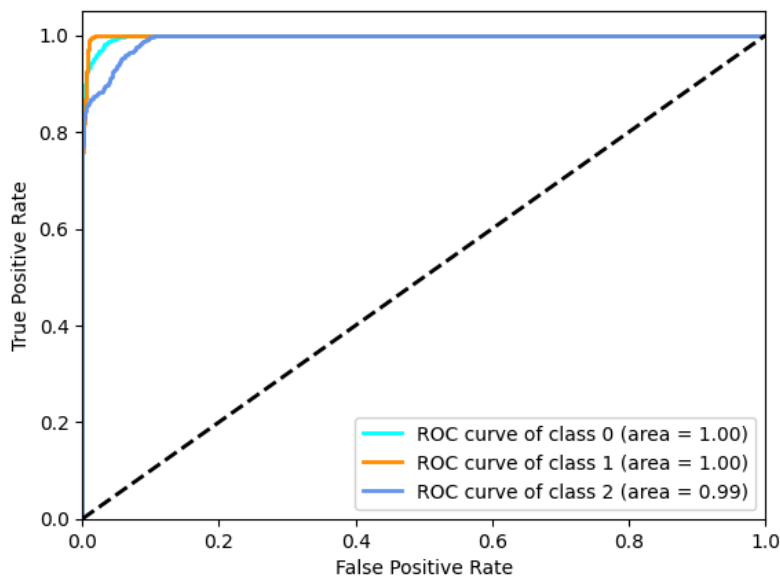
Visualisasi yang ditampilkan pada Gambar 5 memperlihatkan distribusi tinggi badan balita dalam berbagai kategori status gizi. Pada sumbu horizontal (sumbu x), terdapat kategori status gizi yang dikelompokkan ke dalam beberapa kelas, sementara sumbu vertikal (sumbu y) menunjukkan tinggi badan balita dalam satuan *centimeter* (cm). Analisis data menunjukkan bahwa balita yang dikategorikan memiliki status gizi tinggi menampilkan variasi tinggi badan yang lebih luas dan nilai tengah yang lebih tinggi dibandingkan kelompok lainnya. Di sisi lain, balita yang tergolong pendek dan sangat pendek memiliki tinggi badan yang secara nyata lebih rendah, dengan beberapa kasus ekstrem yang sangat pendek (*outlier*). Pola distribusi tinggi badan dalam kaitannya dengan status gizi ini menyediakan petunjuk penting bagi model dalam proses klasifikasi status gizi balita.

Model SVM kemudian dilatih pada data latih menggunakan “SVC” dari “*scikit-learn*”. “SVC” adalah singkatan dari *Support Vector Classification*, yang merupakan salah satu kelas di pustaka “*scikit-learn*” yang digunakan untuk melakukan klasifikasi menggunakan pendekatan *Support Vector Machine* (SVM). Pustaka “SVC” bertujuan untuk menemukan hyperplane yang membedakan data dari berbagai kelas dengan margin seluas mungkin. Dalam kasus ini, model dilatih untuk memprediksi status gizi balita berdasarkan umur, jenis kelamin, dan tinggi badan. Model dilatih untuk mengenali pola-pola dalam data yang memungkinkannya memisahkan kelas-kelas yang berbeda. Setelah model dilatih, model digunakan untuk memprediksi data uji. Prediksi yang dihasilkan dibandingkan dengan label asli untuk menilai kinerja model melalui berbagai metrik evaluasi. *Confusion matrix* digunakan untuk memvisualisasikan performa klasifikasi model dalam bentuk matriks, yang menunjukkan jumlah prediksi yang tepat maupun yang salah pada masing-masing kelasnya. Ini memberikan gambaran yang lebih detail tentang dimana model melakukan kesalahan. *Confusion matrix* dari model klasifikasi stunting berdasarkan status gizi tervisualisasi pada Gambar 5.



Gambar 6. *Confusion matrix* Model SVM

Confusion matrix yang ditampilkan pada Gambar 6 memvisualisasikan bagaimana prediksi model dibandingkan dengan nilai sebenarnya dalam bentuk matriks, yang membantu dalam mengidentifikasi kelas-kelas yang salah diklasifikasikan. Evaluasi selanjutnya yaitu menggunakan *ROC Curve (Receiver Operating Characteristic Curve)* untuk model klasifikasi multi-kelas. *ROC Curve* adalah grafik yang menampilkan hubungannya antara *True Positive Rate (TPR)* serta *False Positive Rate (FPR)* pada beberapa penilaian *threshold* klasifikasi. *ROC Curve* berguna dalam mengevaluasi kemampuan model klasifikasi dalam membedakan antara kelas-kelas yang berbeda. Tampilan hasil evaluasi *ROC Curve* ditunjukkan pada Gambar 7.



Gambar 7. *ROC Curve* Model SVM

Gambar 6 menunjukkan *ROC Curve*, dimana kurva ini membantu dalam menilai kualitas prediksi model dengan menunjukkan seberapa baik model dapat memisahkan antara kelas-kelas yang berbeda. Semakin mendekati kurva *ROC* ke sudut kiri atas grafik, semakin baik model tersebut, karena menunjukkan tingkat *True Positive* yang tinggi dengan *False Positive* yang rendah. *Area Under Curve (AUC)* yang tertera pada grafik adalah ukuran kuantitatif dari kualitas ini, di mana nilai 1.0 menunjukkan pemisahan sempurna oleh model, dan nilai 0.5 menunjukkan kinerja model yang tidak lebih baik dari tebakan acak. *ROC Curve* memperlihatkan bahwa model klasifikasi memiliki kinerja yang sangat baik dalam memprediksi setiap kelas. *AUC* yang mendekati 1

untuk setiap kelas mengindikasikan bahwa model mampu membedakan antara kelas-kelas dengan sangat baik. Secara keseluruhan, model ini menunjukkan kemampuan prediksi yang tinggi dengan sedikit kesalahan dalam klasifikasi.

Selanjutnya, dari matrik yang ada model dievaluasi menggunakan *precision*, *recall*, *F1-score*, dan akurasi untuk mendapatkan gambaran yang komprehensif tentang kinerja model klasifikasi. Hasil dari evaluasi ini disusun kedalam Tabel 1.

Tabel 1. Hasil Evaluasi Model

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Accuracy</i>
Normal	0.9873	0.9931	0.9902	
Sangat Pendek	0.9745	0.9909	0.9827	
Pendek	0.9755	0.9363	0.9555	0.9837
Tinggi	0.9861	0.977	0.9815	

Pada Tabel 1 menunjukkan hasil evaluasi model klasifikasi menggunakan SVM memiliki kinerja yang sangat baik dalam mengklasifikasikan status gizi anak ke dalam empat kategori: Normal, Tinggi, Pendek, dan Sangat Pendek. Model ini mencapai akurasi keseluruhan sebesar 98.37%, yang berarti hampir semua prediksi yang dihasilkan oleh model ini benar. Namun, untuk memahami kinerja model secara lebih mendalam, perlu dilihat metrik lain seperti presisi, *recall*, serta *F1-score* pada masing-masing kelasnya. Kelas "Normal" dan "Tinggi" menunjukkan nilai *precision*, *recall*, dan *F1-score* yang sangat tinggi. Ini menunjukkan bahwa model sangat andal dalam mengidentifikasi kasus-kasus dari kedua kelas ini. Artinya, ketika model memprediksi bahwa seorang anak memiliki status gizi "Normal" atau "Tinggi", prediksi ini hampir selalu benar. Hal ini sangat penting mengingat bahwa anak-anak dengan status gizi normal dan tinggi mungkin merupakan mayoritas dalam dataset, yang dapat mempengaruhi kinerja model secara keseluruhan. Namun, ada sedikit penurunan kinerja yang terlihat pada kelas "Pendek", di mana *recall* mencapai 93.63%. Meskipun *precision* untuk kelas ini masih tinggi (97.55%), *recall* yang lebih rendah menunjukkan bahwa model kadang-kadang gagal mengidentifikasi anak-anak yang sebenarnya memiliki status gizi "Pendek". Ini berarti bahwa ada sejumlah anak dengan status gizi "Pendek" yang diprediksi sebagai kelas lain, walaupun tingkat kesalahannya tergolong kecil. Kelas "Sangat Pendek" menunjukkan kinerja yang sangat baik dengan *precision* dan *recall* yang hampir sempurna. Ini sangat penting karena anak-anak dengan status gizi "Sangat Pendek" berada pada risiko yang lebih tinggi dan memerlukan intervensi yang cepat.

Hasil akurasi yang diperoleh dari model yang dikembangkan menggunakan algoritma SVM dan proses *preprocessing* yang telah dilakukan kemudian dibandingkan dengan hasil akurasi dari penelitian terdahulu mengenai klasifikasi atau prediksi stunting terhadap status gizi. Tabel 2 menunjukkan hasil akurasi dari penelitian terdahulu dengan penelitian yang dilakukan.

Tabel 1. Perbandingan Model *Machine Learning* Penelitian Terdahulu

Metode	<i>Accuracy</i>
<i>K-Nearest Neighbor</i> [4]	97,16%
<i>Random Forest</i> [5]	88,60%
<i>Naive Bayes Classifier</i> [6]	85,00%
<i>Extreme Gradient Boosting</i> [7]	86,00%
<i>Support Vector Machine</i> (Penelitian Yang Dilakukan)	98.37%

Berdasarkan Tabel 1 yang membandingkan berbagai metode *machine learning* dari penelitian sebelumnya, menunjukkan bahwa metode Support Vector Machine (SVM) yang digunakan dalam penelitian ini menunjukkan kinerja paling tinggi dengan tingkat akurasi sebesar 98,37%. Hasil ini lebih unggul dibandingkan dengan algoritma *K-Nearest Neighbor* sebesar 97,16% [4]; pendekatan *Random Forest* sebesar 88,6% [5]; metode *Naive Bayes Classifier* sebesar 88,6% [6]; dan pendekatan *Extreme Gradient Boosting* dengan tingkat akurasi 86% [7].

Performa yang dihasilkan oleh metode Support Vector Machine (SVM) yang dikembangkan, mengindikasikan bahwa model dapat diandalkan dalam mengklasifikasikan stunting berdasarkan status gizinya. Faktor keberhasilan ini dikarenakan SVM memiliki kelebihan dalam menangani data dengan dimensi yang tinggi, bahkan ketika jumlah fitur melebihi jumlah sampel, karena pendekatan ini berusaha untuk mendapatkan *hyperplane* optimal yang membedakan kelas melalui *margin* yang terbesar. Selain itu, algoritma SVM didukung dengan teknik prapemrosesan seperti standarisasi fitur, *encoding* variabel kategorikal, serta penghapusan data duplikat. Langkah-langkah ini memastikan bahwa model SVM yang dibangun dapat beroperasi secara optimal dengan data yang bersih dan terorganisir. Meskipun model SVM ini sangat efektif secara keseluruhan, perhatian khusus perlu diberikan untuk meningkatkan *recall* pada kelas "Pendek". Salah satu faktor utama bisa jadi adalah ketidakseimbangan data, di mana jumlah sampel untuk kelas "Pendek" lebih sedikit dibandingkan dengan kelas lainnya.

IV. Kesimpulan

Penelitian ini telah berhasil menunjukkan bahwa model *Support Vector Machine* (SVM) dapat digunakan secara efektif untuk mengklasifikasikan status gizi balita ke dalam beberapa kategori, termasuk Normal, Tinggi, Pendek, dan Sangat Pendek. Dengan akurasi keseluruhan sebesar 98,37%, model tersebut memperlihatkan performa yang sangat baik dalam mengidentifikasi status gizi balita. Algoritma SVM mampu mencapai akurasi yang tinggi dalam klasifikasi stunting pada balita karena kemampuannya dalam menentukan *hyperplane* terbaik yang membagi data dari bermacam-macam kelas dengan *margin* terbesar. SVM bekerja dengan memaksimalkan *margin* antara kelas-kelas, yang membuatnya sangat efektif dalam mengatasi data yang memiliki dimensi tinggi atau kompleksitas tinggi, seperti dalam kasus klasifikasi status gizi balita. Selain itu, penelitian ini menerapkan beberapa teknik prapemrosesan seperti standarisasi fitur, pengkodean variabel kategorikal, serta penghapusan data duplikat. Langkah-langkah ini memastikan bahwa model SVM yang dikembangkan dapat bekerja secara optimal dengan data yang bersih dan terorganisir. Hasil penelitian ini menunjukkan potensi besar penggunaan SVM dalam aplikasi nyata untuk mendukung pengambilan keputusan medis atau intervensi kesehatan masyarakat, khususnya dalam konteks pemantauan dan pencegahan stunting pada balita. Namun, ada beberapa aspek yang bisa dipertimbangkan untuk penelitian di masa mendatang. Pada penelitian ini terdapat kelas yang memiliki *precision* dan *recall* tidak setinggi kelas lainnya. Untuk itu dapat mencoba melakukan penyesuaian *hyperparameter* model, menggunakan teknik *oversampling* untuk mengatasi ketidakseimbangan kelas. Selain itu, penelitian berikutnya juga dapat menerapkan teknik *ensemble* atau mengkombinasikan dengan teknik *boosting* guna meningkatkan performa model.

Daftar Pustaka

- [1] M. R. Nugroho, R. N. Sasongko, and M. Kristiawan, "Faktor-Faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 5, no. 2, pp. 2269–2276, 2021, doi: 10.31004/obsesi.v5i2.1169.
- [2] M. E. Setiyawati, L. P. Ardhiyanti, E. N. Hamid, N. A. T. Muliarta, and Y. J. Raihanah, "Studi Literatur: Keadaan dan Penanganan Stunting di Indonesia," *IKRAITH-HUMANIORA*, vol. 8, no. 2, pp. 179–186, 2022.
- [3] H. Rahman, M. Rahmah, and N. Saribulan, "Upaya Penanganan Stunting di Indonesia: Analisis Bibliometrik dan Analisis Konten," *J. Ilmu Pemerintah. Suara Khatulistiwa*, vol. VIII, no. 01, pp. 44–59, 2023.
- [4] A. F. Amida, S. E. Permana, D. Pratama, K. Anam, and A. R. Rinaldi, "Prediction of Stunted Toddlers Using K-Nearest Neighbor Algorithm in Kamarang Lebak Village," *Instal J. Komput.*, vol. 15, no. 02, pp. 345–355, 2023.
- [5] P. Handayani, A. C. Fauzan, and H. Harliana, "Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 6, pp. 3064–3072, 2024, doi: 10.30865/klik.v4i6.1909.
- [6] T. Hardiani and R. N. Putri, "Implementasi Metode Naïve Bayes Classifier Untuk Klasifikasi Stunting Pada Balita," *Digit. Transform. Technol.*, vol. 4, no. 1, pp. 621–627, 2024.
- [7] M. Fikri, "Klasifikasi Status Stunting Pada Anak Bawah Lima Tahun Menggunakan Extreme Gradient Boosting," *Merkurius J. Ris. Sist. Inf. dan Tek. Inform.*, vol. 2, no. 4, pp. 173–184, 2024.
- [8] I. M. D. P. Asana and N. P. D. T. Yanti, "Sistem Klasifikasi Pengajuan Kredit Dengan Metode Support Vector Machine (SVM)," *J. Sist. Cerdas*, vol. 06, no. 02, pp. 123–133, 2023.
- [9] R. Sistem, K. Mahasiswa, and T. Waktu, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no.

- 2, pp. 386–392, 2021.
- [10] U. Amelia *et al.*, “Implementasi Algoritma Support Vector Machine (SVM) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh,” *Sci. Student J. Information, Technol. Sci.*, vol. III, no. 2, pp. 254–259, 2022.
- [11] A. W. Mucholladin, F. A. Bachtiar, and M. T. Furqon, “Klasifikasi Penyakit Diabetes menggunakan Metode Support Vector Machine,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 2, pp. 622–633, 2021.
- [12] M. D. A. Rosyid and S. Subektiningsih, “Klasifikasi Tingkat Risiko Kesehatan Ibu Hamil Menggunakan Algoritma Support Vectore Machine,” *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 2798–2807, 2023.
- [13] R. I. Borman, R. Napianto, N. Nugroho, D. Pasha, Y. Rahmanto, and Y. E. P. Yudoutomo, “Implementation of PCA and KNN Algorithms in the Classification of Indonesian Medicinal Plants,” in *International Conference on Computer Science, Information Technology and Electrical Engineering (ICOMITEE)*, 2021, pp. 46–50.
- [14] R. P. Pradana, “Stunting Toddler Detection,” *Kaggle*, 2024. <https://www.kaggle.com/datasets/rendiputra/stunting-balita-detection-121k-rows/>
- [15] R. I. Borman and M. Wati, “Penerapan Data Maining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes,” *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 25–34, 2020.
- [16] I. O. Muraina, “Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts,” in *International Mardin Artuklu Scientific Researches Conference*, 2022, pp. 496–505.
- [17] C. M. Sitorus, A. Rizal, and M. Jajuli, “Prediksi Risiko Perjalanan Transportasi Online Dari Data Telematik Menggunakan Algoritma Support Vector Machine,” *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 254–265, 2020.
- [18] N. G. Ramadhan and A. Khoirunnisa, “Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine,” *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1580–1584, 2021, doi: 10.30865/mib.v5i4.3347.
- [19] M. Singla and K. K. Shukla, “Robust statistics-based support vector machine and its variants: a survey,” *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11173–11194, 2020, doi: 10.1007/s00521-019-04627-6.
- [20] S. S. Arifin, A. M. Siregar, A. R. Juwita, and T. Al Mudzakir, “Klasifikasi Penyakit Kanker Serviks Menggunakan Algoritma Support Vector Machine (SVM),” in *Conference on Innovation and Application of Science and Technology (CIASTECH)*, 2021, pp. 521–528.
- [21] Y. Fernando, R. Napianto, and R. I. Borman, “Implementasi Algoritma Dempster-Shafer Theory Pada Sistem Pakar Diagnosa Penyakit Psikologis Gangguan Kontrol Impuls,” *Insearch Inf. Syst. Res. J.*, vol. 2, no. 2, pp. 46–54, 2022.
- [22] G. Naidu, T. Zuva, and E. M. Sibanda, “A Review of Evaluation Metrics in Machine Learning Algorithms,” in *Artificial Intelligence Application in Networks and Systems*, 2023, pp. 15–25.
- [23] Y. Liu, Y. Li, and D. Xie, “Implications of imbalanced datasets for empirical ROC-AUC estimation in binary classification tasks,” *J. Stat. Comput. Simul.*, vol. 94, no. 1, pp. 183–203, Jan. 2024, doi: 10.1080/00949655.2023.2238235.