

# Prediksi Prevalensi *Stunting* Balita dengan Pendekatan Algoritma Support Vector Machine dan Synthetic Minority Oversampling Technique (SMOTE)

## *Prediction of Stunting Prevalence in Toddlers Using Support Vector Machine Algorithm and Synthetic Minority Oversampling Technique (SMOTE)*

Taufik Hidayat<sup>1)</sup>, Irwan Sembiring<sup>2)</sup>, Hindriyanto Dwi Purnomo<sup>3)</sup>, Ade Iriani<sup>4)</sup>

<sup>1,2,3,4</sup> Universitas Kristen Satya Wacana  
<sup>1,2,3,4</sup> Jl. Diponegoro No.52-60, Makassar, 50711, Indonesia

thidayat@unis.ac.id<sup>1)</sup>, irwan@uksw.edu<sup>2)</sup>, hindriyanto.purnomo@uksw.edu<sup>3)</sup>, ade.iriiani@uksw.edu<sup>4)</sup>

Diterima: 11 Agustus 2024 || Direvisi: 15 Desember 2024 || Disetujui: 15 Maret 2025

**Abstrak** – *Stunting* pada balita merupakan sebuah status pada balita yang memiliki kekurangan gizi, data yang diusulkan sejumlah 6.879 data balita, hal ini tentu akan lebih kompleks bila tidak diketahui informasi terhadap data awal pada balita, dimana data tersebut sebelum diimplementasikan pada algoritma *machine learning* harus melalui *preprocessing* dan penyeimbangan data. Dalam prediksi prevalensi *stunting* pada balita ini menggunakan algoritma *machine learning* yaitu Support Vector Machine (SVM) dengan metode *supervised learning* dan *synthetic minority oversampling technique* (SMOTE) sebagai penyeimbang data serta *exploratory data analysis* (EDA) sebagai metode *preprocessing* terhadap *dataset* balita. Dari hasil implementasi penelitian ini diperoleh sebuah akurasi untuk prediksi sebesar 94%, terdiri dari *accuracy* 94%, *precision* 95%, *recall* 94%, dan *F1-score* 94%.

**Kata Kunci:** *stunting*, balita, support vector machine, data imbalanced, klasifikasi machine learning

**Abstract** – *Stunting in toddlers represents a condition where isa nutritional deficiency. This becomes more complex when there is insufficient information regarding stunting in toddlers available. Predicting the prevalence of stunting in toddlers involves studying a dataset of stunting prevalence among toddlers through a supervised learning model using Support Vector Machine (SVM) and synthetic minority oversampling technique (SMOTE). The use of SMOTE serves as a data balancing method, while exploratory data analysis (EDA) acts as the preprocessing method for the toddler dataset. From the research implementation on a dataset consisting of 6879 toddlers, an accuracy of 94% was obtained for predictions. This accuracy is comprised of 94% accuracy, 95% precision, 94% recall, and a 94% F1-score.*

**Keywords:** *stunting, toddlers, support vector machine, data imbalance, machine learning classification.*

## PENDAHULUAN

*Stunting* adalah kondisi gagal tumbuh pada balita akibat kekurangan asupan gizi dan infeksi yang berkepanjangan yang mengakibatkan tinggi badan yang lebih pendek dari standar usianya (Nadhiroh et al., 2022; Pohan et al., 2021). Survei Status Gizi Indonesia (SSGI) melaporkan bahwa prevalensi *stunting* pada tahun 2021 adalah 24,4% dan menurun menjadi 21,6% pada tahun 2022 (Syahrudin & Sari, 2023). *Stunting* memiliki pengaruh besar terhadap tumbuh kembangnya anak khususnya pada usia 0–59 bulan, dimana usia itu merupakan masa perkembangan dan pertumbuhan. *Data mining* adalah proses untuk

menemukan pola/informasi menarik pada data yang dipilih menggunakan teknik atau metode tertentu. Dalam metode penambangan data, teknik dan algoritma bervariasi secara signifikan. Metode atau algoritma yang dipilih bergantung pada tujuan yang ingin dicapai (Amanda & Negara, 2020). Menggunakan *Machine Learning* untuk melakukan prediksi. *Machine learning* cabang dari kecerdasan buatan (*Artificial Intelligence/AI*) yang berfokus pada pengembangan algoritma dan model yang memungkinkan komputer untuk belajar dari data tanpa perlu diprogram secara eksplisit. Dengan pendekatan ini, sistem dapat mengenali pola, membuat keputusan, dan melakukan prediksi berdasarkan data historis yang

tersedia, sehingga meningkatkan akurasi dan efisiensi dalam pengambilan prediksi.

Di dalamnya, mesin dibuat untuk belajar sendiri tanpa menggunakan pemrograman eksternal atau intervensi pengguna (Bharti et al., 2023).

*Machine learning* mampu menganalisis dan memproses data dengan banyak variabel untuk menemukan pola atau hubungan yang mungkin tidak tampak secara eksplisit, namun dapat membantu memprediksi kemungkinan terjadinya *stunting* pada balita dengan akurasi tinggi, hal tersebut berdasarkan data historis dan fitur-fitur terkait, penanganan volume data yang besar dan heterogen dengan cara yang efisien, serta dapat menangkap pola yang kompleks antara variabel-variabel yang berbeda dalam sebuah data balita.

Kontribusi penelitian mendukung peningkatan efisiensi program kesehatan dengan mengarahkan upaya pada kelompok yang paling rentan dengan cara memprediksi balita yang berisiko mengalami *stunting*. Dengan melakukan prediksi dan pencegahan *stunting* sejak awal, dapat mengurangi dampak jangka panjang bagi masyarakat. Juga dapat membantu mengurangi beban ekonomi negara, karena individu yang tumbuh dengan sehat dan berkembang dengan baik biasanya lebih produktif serta memiliki kualitas hidup yang lebih baik. *Machine learning* dapat memberikan wawasan tentang seberapa efektif intervensi dan membantu dalam merencanakan langkah-langkah berikutnya, serta menemukan keterkaitan yang rumit dalam data yang mungkin tidak terungkap melalui analisis tradisional.

Mustaqim dkk dengan judul *Kombinasi Synthetic Minority Oversampling Technique (SMOTE) dan Neural Network Backpropagation* untuk menangani data tidak seimbang pada prediksi pemakaian alat kontrasepsi *implant*, menghasilkan implementasi kombinasi SMOTE dan NN Backpropagation mampu memprediksi pada *imbalance class* dengan akurasi 96,1%. Teknik SMOTE mampu memprediksi dengan persentase 93% kelas minor pada *epoch* 10 dan 50, sedangkan pada *epoch* 100, sistem mampu memprediksi 94% serta 95% pada *epoch* ke 500 dan 1.000 (Mustaqim et al., 2019). Penelitian Muhajir dan Widiastuti dengan judul *Metode Random Forest untuk Klasifikasi Pelanggan Berdasarkan Pinjaman Bermasalah pada Usaha Mikro* menggunakan metode *random forest* untuk mengklasifikasikan calon pelanggan menggunakan metode *oversampling*. Berdasarkan penelitian, *random sampling* yang

diperbaiki dengan *oversampling* lebih baik daripada *random forest* yang menggunakan *undersampling* dan *random forest* yang menggunakan SMOTE. Hal ini terlihat dari matriks kebingungan dan skor AUC yang sebesar 66,69%, menggambarkan kinerja metode tersebut (Muhajir & Widiastuti, 2022). Laengstri dkk menerapkan lima pembelajaran mesin, termasuk *k-nearest neighbour* (K-NN), pohon keputusan, *random forest* (RF), jaringan syaraf tiruan (ANN) dan *support vector machine* (SVM), diterapkan untuk membangun model diskriminan mencapai hasil prediksi dengan akurasi eksternal, MCC dan AUC masing-masing sebesar 95,59, 0,87 dan 0,98 (Laengstri et al., 2019).

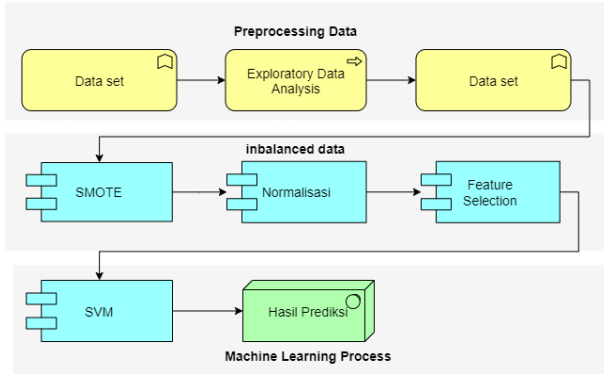
Tujuan dari penelitian saat ini yaitu untuk membuktikan dan menerapkan apakah pendekatan *supervised learning* ini dapat dipergunakan dan untuk melakukan sebuah prediksi dengan komputasi dari data yang telah diperoleh secara empiris, melakukan pengenalan pola dari sebuah data yang diperoleh, melakukan akurasi sebuah klasifikasi serta melakukan evaluasi terhadap model, dan memaksimalkan kinerja model dari data terbaik untuk memprediksi prevalensi *stunting* balita. Ruang lingkup penelitian ini hanya mencakup pada kontribusi sebuah algoritma *machine learning* dengan model *supervised* yang sering dipergunakan yaitu Support Vector Machine (SVM) dengan menerapkan algoritma Synthetic Minority Over-sampling Technique (SMOTE) untuk melakukan keseimbangan pada data.

Penelitian ini sangat penting karena memungkinkan pengembangan model klasifikasi yang secara efektif mengidentifikasi dan mengatasi gangguan pertumbuhan anak, yang akhirnya membantu pengurangan prevalensi *stunting*, untuk memprediksi balita berisiko, memungkinkan intervensi tepat waktu. Dengan model klasifikasi ini pihak terkait dapat melihat peta sebaran balita diseluruh wilayah zaskupannya (teritorial secara wilayah kerja). Karena yang diusulkan dalam penelitian ini tidak hanya kepada hasil akhir, namun juga dapat mengetahui hasil uji dari klasifikasi balita terhadap status gizi.

## METODOLOGI

Dalam penelitian ini untuk melakukan suatu penerapan metode *machine learning* dengan model *supervised* diperlukan sebuah tahapan terdiri dari pemilihan data, eksplorasi data *analysis*, keseimbangan data, penerapan pada *supervised machine learning*

serta sampai tahapan pengujian prediksi. Dapat dilihat pada gambar 1.



**Gambar 1** Arsitektur dari Sistem yang Dibangun

Dari gambar 1 dari *dataset* dilakukan sebuah pra-pemrosesan tujuan untuk menangani data hilang, data keliru atau data kurang lengkap. Dalam pra-pemrosesan ini kami mengusulkan pembersihan data, setelah data diketahui sebaran datanya dengan melihat hasil pra-pemrosesan (*Exploratory Data Analysis*) maka selanjutnya diketahui apakah data tersebut seimbang atau tidak seimbang, jika dinyatakan tidak seimbang dilakukan penyeimbangan data, setelah itu dilakukan *feature selection* untuk mengetahui hubungan antar atribut, variabel, untuk mempermudah proses klasifikasi sebelum masuk pada klasifikasi dengan SVM. Dimana hasil akan diketahui setelah dilakukan uji model.

**EDA (Exploratory Data Analysis)**

Melalui EDA, peneliti dapat menghasilkan visualisasi data yang informatif, seperti *scatterplot*, *line chart*, dan *heatmap*; yang memungkinkan pemahaman yang lebih baik tentang hubungan antar variabel dan tren waktu (Austin et al., 2022; Doshi et al., 2021; Mayasari et al., 2023). Analisis Eksploratif adalah teknik penting dalam setiap profesi karena meramalkan masa depan dan pola tersembunyi. Analisis data dianggap sebagai suatu penyelidikan data (EDA) yang mendeteksi kesalahan, menemukan data yang tepat, memeriksa asumsi, dan menentukan korelasi di antara teknologi yang hemat biaya dalam waktu yang baru-baru ini dan memainkan peran penting dalam perawatan kesehatan yang meliputi temuan penelitian baru, situasi darurat (Indrakumari et al., 2020).

Adapun dalam penelitian ini kami menerapkan metode EDA untuk melakukan pengenalan pola dan melakukan pemecahan masalah untuk beberapa kasus dari sebuah *dataset*. Tahapan dalam *exploratory data analysis* yang kami terapkan meliputi:

1. Melakukan *cleaning* data dari anomali pada data, dan imputasi *missing value*.
2. Melakukan transformasi label dari *string* ke *integer* contohnya gizi baik = 0, gizi buruk = 1, gizi kurang = 2, gizi lebih = 3, obesitas = 4, dan risiko gizi lebih = 5.
3. Melakukan deskriptif data secara statistik dengan memperhatikan bagaimana data menyebar.
4. Langkah selanjutnya adalah normalisasi data.  
Langkah di atas dilakukan sebelum dengan tujuan supaya data diketahui polanya dan data lebih baik untuk dilakukan langkah selanjutnya.

**SMOTE (Synthetic Minority Oversampling Technique)**

Metode *Synthetic Minority Oversampling Technique* (SMOTE) sebagai solusi untuk menangani *imbalance class* (Mustaqim et al., 2019). Untuk mengatasi data tidak seimbang, dengan teknik *oversampling* (Austin et al., 2022; Doshi et al., 2021; Malau et al., 2023; Siambaton & Husein, 2022; Wibowo, 2022). Setelah dilakukan menyeimbangkan data, maka data dapat diterapkan pada *machine learning*.

**Supervised Machine Learning**

*Machine learning* menggunakan komputer untuk menyimulasikan pembelajaran manusia dan memungkinkan mereka mengidentifikasi serta memperoleh pengetahuan dari dunia nyata dan meningkatkan kinerja pada beberapa tugas berdasarkan pengetahuan baru ini (Nofriani, 2019). *Supervised machine learning* mempermudah proses *learning* prediksi atau klasifikasi yang efektif pada data, dikarenakan data telah memiliki label sebelumnya (Barros et al., 2022; Dritsas & Trigka, 2023; Tiwari, 2022; Uddin et al., 2019; Wittek et al., 2023). Dengan data yang telah memiliki label sebelumnya, akan lebih akurat pada saat dilakukan proses mulai dari *pre-processing*, keseimbangan data, *feature selected*, hubungan antar variabel, dan klasifikasi serta prediksi.

**Data Source**

Tabel 1 di bawah ini adalah kelas dan atribut dari sebuah *dataset* gizi balita dari usia 0–59 bulan dari Puskesmas Teluknaga Kabupaten Tangerang.

**Tabel 1** *Dataset* Kelas dan Atribut

Variabel	Keterangan
Status	0 = gizi baik
kelas	1 = gizi buruk
	2 = gizi kurang
	3 = gizi lebih

Variabel	Keterangan
	4 = obesitas
	5 = risiko gizi lebih
Usia saat diukur	Kondisi usia balita pada saat dilakukan pengukuran
BB lahir	Berat badan balita pada saat dilahirkan
TB lahir	Tinggi badan balita pada saat dilahirkan
Berat	Berat badan balita pada saat dilakukan pengukuran
Tinggi	Tinggi badan balita pada saat dilakukan pengukuran
LiLa	Lingkar tangan atas
LiPa	Lingkar kepala atas
JK	Jenis kelamin balita
BB/TB	Target dari prevalensi <i>stunting</i> balita

*Data source* ini memudahkan dalam pengenalan pola dan analisa data, dikarenakan setiap variabel atau atribut memiliki fungsi dan nilai, untuk nantinya dimasukan dalam *machine learning*.

### Data Pre-Processing

Tahap ini kami melakukannya dengan metode *exploratory data analysis* dengan menggunakan *dataset* seperti Tabel 2.

**Tabel 2** Transformasi Nilai dari Kelas yang Dipilih

Kelas yang dipilih	Transformasi ke integer
gizi baik	0
gizi buruk	1
gizi kurang	2
gizi lebih	3
obesitas	4
risiko gizi lebih	5

Tujuan dari melakukan transformasi ini yaitu untuk memudahkan langkah *pre-processing*. *Pre-processing* ini mengurangi ketidakseimbangan data, *missing value*, *outlier*, meningkatnya jumlah data yang dihasilkan, dan perbaikan data (Booeshaghi et al., 2023; Erol et al., 2022; Kong et al., 2023; Torres-Martos et al., 2023). Dengan dilakukannya *preprocessing* data pada nilai sebuah kelas, maka akan dapat dilakukan proses selanjutnya dengan data yang lebih baik.

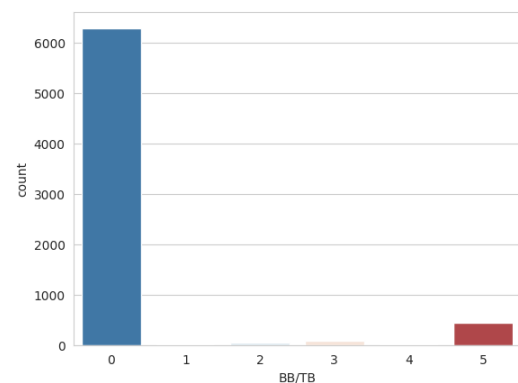
### HASIL DAN PEMBAHASAN

Penelitian ini menuangkan beberapa hasil meliputi beberapa proses dari prediksi prevalensi *stunting* balita, diantaranya *preprocessing* data untuk mengenal pola data dengan metode *exploratory data analysis*, melakukan cek *missing value* dan *handling missing value*, melakukan deskriptif data secara statistik dengan memperhatikan sebaran data, nilai minimum, maksimum, nilai rata-rata, simpangan baku, serta

rentang kuartil tertentu. Melakukan validasi data, normalisasi data *handling imbalanced data* dengan metode SMOTE, memeriksa sebaran data setiap variabel, memvisualisasikan hubungan antar variabel/atribut, melakukan *feature selection*, pembagian data uji dan data latih dengan menggunakan teknik *dataset splitting*, selanjutnya melakukan prediksi dengan SVM diawali dengan melakukan klasifikasi *support vector classifier*, dengan menampilkan akurasi, presisi, *recall* dan *f1-score*. Data yang diolah terdiri dari 6.879 data balita.

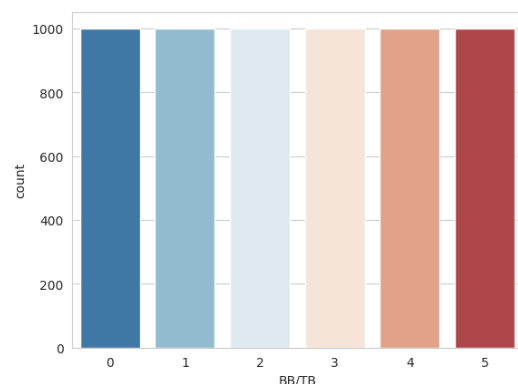
### Data tidak seimbang dan data seimbang

Data awal yang dimiliki tidak seimbang seperti pada Gambar 2



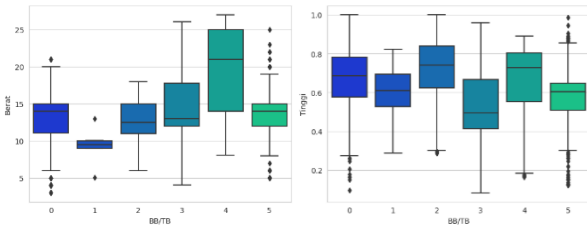
**Gambar 2** Data Tidak Seimbang

Setelah dilakukan penanganan data tidak seimbang dengan SMOTE, maka hasilnya dapat dilihat pada Gambar 3.

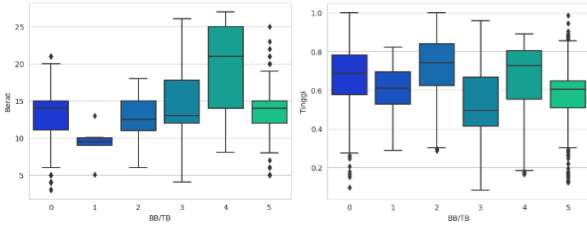


**Gambar 3** Data Seimbang

Kemudian dilakukan pemeriksaan sebaran data pada setiap atribut/variabel, tujuannya apakah terdapat perbedaan sebaran nilai yang signifikan, dapat dilihat dari 2 kelas seperti pada gambar di bawah ini;

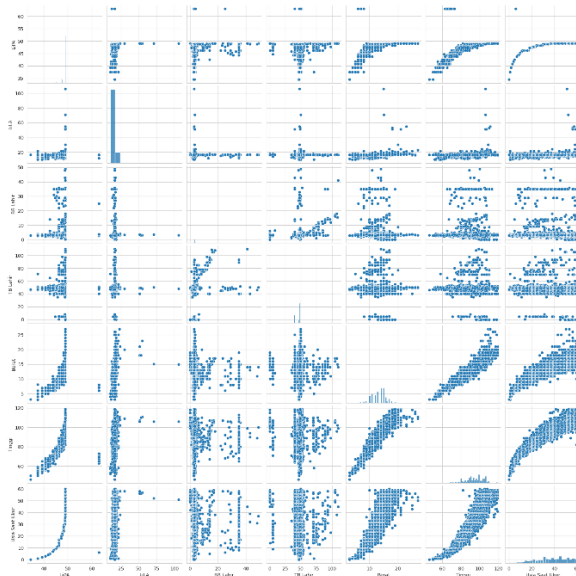


Gambar 4 Nilai Sebaran Data Awal/Original



Gambar 5 Nilai Sebaran Data Setelah Data Seimbang

Selanjutnya menampilkan bagaimana hubungan antar atribut atau variabel dari enam atribut.



Gambar 6 Hubungan Antar Atribut

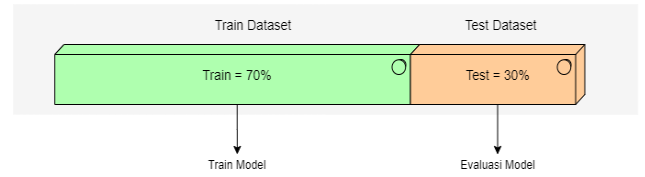
Untuk mengetahui tingkat korelasi antar variabel dalam sebuah data maka dilakukan sebuah *feature selection* seperti Gambar 7.



Gambar 7 Korelasi Antar Variabel Data Seimbang

Pembagian *dataset* dengan model *splitting* ini merupakan upaya pembagian antara data latih (*train*) dan data uji (*test*), dengan komposisi 70% sebagai data latih, dan 30% sebagai data uji.

DATA SET



Gambar 8 Teknik *Splitting* dari Sebuah Data

Confusion Matrik

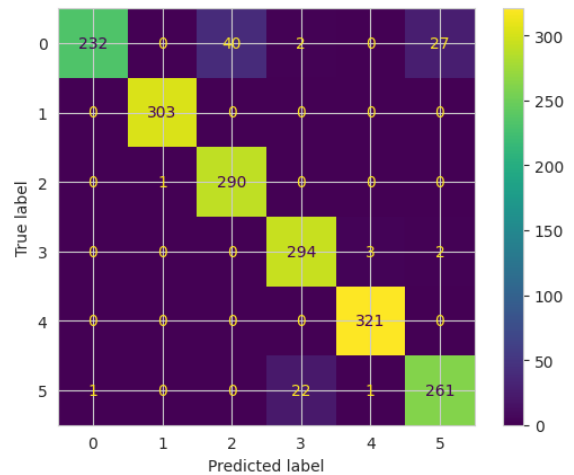
Dari data yang telah dibagi dan selanjutnya dilakukan perhitungan matrik evaluasi kinerja model seperti:

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn} \tag{1}$$

$$Precision = \frac{TP}{FP+TP} \tag{2}$$

$$Recall = \frac{TP}{FN+TP} \tag{3}$$

$$F\ measure = \frac{2 \times precision \times recall}{(precision+recall)} \tag{4}$$



Gambar 9 Confusion Matrik

Gambar 9 memvisualisasikan dari kelas (*predicated label*) dengan kebenaran atau akurasi, dimana hal tersebut sebagai hasil dari evaluasi kinerja model.

Evaluasi Hasil

Melakukan perhitungan untuk mengetahui nilai dari klasifikasi model.

**Tabel 3** Tabel Perhitungan *Accuracy, Precision, Recall, dan F1-Score*

Kelas	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
0	1.00	0.77	0.87	301
1	1.00	1.00	1.00	303
2	0.88	1.00	0.93	291
3	0.92	0.98	0.95	299
4	0.99	1.00	0.99	321
5	0.90-	0.92	0.91	285
accuracy			0.94	1800
macro avg	0.95	0.94	0.94	1800
weighted avg	0.95	0.94	0.94	1800

Dari hasil performa diperoleh akurasi 94%, dimana hasil ini telah diperoleh dari pengolahan *dataset* balita dengan kelas atribut yang digunakan sebanyak 6 atribut terdiri dari BB lahir, TB lahir, berat, tinggi, LiLa, LiPa. Hasil ini kemudian didistribusikan pada 5 kelas yaitu gizi baik, gizi buruk, gizi kurang, obesitas, dan risiko gizi lebih. Akurasi sebesar 94% dalam konteks praktis menunjukkan bahwa model yang dikembangkan mampu mengklasifikasikan status gizi balita dengan tingkat ketepatan yang sangat tinggi berdasarkan enam atribut yang dianalisis. Artinya, dari seluruh prediksi yang dilakukan, sebanyak 94% balita berhasil dipetakan ke dalam kategori status gizi yang sesuai, yaitu gizi baik, gizi buruk, gizi kurang, obesitas, atau risiko gizi lebih. Capaian ini menunjukkan bahwa model memiliki potensi untuk digunakan sebagai alat bantu skrining awal yang andal di lapangan, khususnya dalam konteks percepatan deteksi risiko *stunting*. Dengan memanfaatkan model ini, tenaga kesehatan dapat secara lebih efisien dan cepat mengidentifikasi balita yang berisiko mengalami *stunting* atau masalah gizi lainnya, sehingga intervensi yang bersifat preventif maupun kuratif dapat segera dilakukan. Selain mempercepat proses deteksi, penggunaan model ini juga dapat membantu alokasi sumber daya yang lebih tepat sasaran, terutama di daerah dengan keterbatasan tenaga ahli atau fasilitas kesehatan. Namun demikian, meskipun tingkat akurasi yang dicapai tinggi, tetap diperlukan konfirmasi hasil prediksi melalui pemeriksaan fisik atau klinis oleh tenaga kesehatan untuk meminimalisir risiko kesalahan klasifikasi, terutama pada kasus gizi buruk dan obesitas yang memiliki implikasi serius terhadap kesehatan balita. Nilai 94% ini adalah nilai dari akurasi yang diperoleh dari kinerja model dalam pengolahan data, dimana sebelumnya melalui beberapa tahapan seperti pra-pemrosesan data (EDA), *handling missing data* sampai

implementasi model SVM sehingga nilai akurasi itu diperoleh. Bahwa SVM ini dapat di implementasikan dalam klasifikasi *stunting* balita dengan 6 atribut dan 5 kelas. Dalam konteks praktis diusulkan untuk memfasilitasi sebuah aplikasi untuk interaksi antara *user* dengan pengembangan data.

Hasil penelitian ini menunjukkan bahwa kombinasi algoritma Support Vector Machine (SVM) dengan teknik Synthetic Minority Oversampling Technique (SMOTE) menghasilkan performa yang kompetitif dalam prediksi status gizi balita, dengan akurasi 94%, precision 95%, recall 94%, dan F1-score 94%. Jika dibandingkan dengan studi sebelumnya oleh Khusna, yang menggunakan Random Forest untuk klasifikasi kasus *stunting* pada balita, diperoleh akurasi sebesar 77,3% sebelum optimasi dan meningkat menjadi 95,78% setelah dilakukan *hyperparameter tuning* menggunakan Grid Search (Khusna et al., 2024). Sementara itu, penelitian oleh Subadi dan Kusri menunjukkan bahwa algoritma K-Nearest Neighbor (K-NN) mencapai akurasi sebesar 79,43% dalam mendiagnosis *stunting* berdasarkan gejala medis. Perbandingan ini mengindikasikan bahwa pendekatan SVM+SMOTE dalam penelitian ini mampu memberikan hasil yang kompetitif dibandingkan dengan metode lain, serta menunjukkan efektivitas dalam menangani ketidakseimbangan data dan meningkatkan ketepatan klasifikasi dalam konteks prediksi prevalensi *stunting* (Subadi, 2020). Penelitian yang dilakukan Khansa dan Gunawan, mengusulkan untuk memprediksi *stunting* pada balita menggunakan algoritma KNN dan Naïve Bayes, Algoritma KNN menunjukkan performa yang sangat baik, dengan peningkatan skor F1 dari 67,20% menjadi 95,62%, serta akurasi mencapai 95,67%. Algoritma Naive Bayes juga mengalami perbaikan, meningkatkan Skor F1 dari 71,22% menjadi 95,62%, dengan akurasi sebesar 94%. Penelitian ini memberikan kontribusi pada pengembangan metode klasifikasi *stunting* yang lebih efektif. Dengan memanfaatkan pengukuran langsung terhadap tinggi dan berat badan, penelitian ini mendeteksi *stunting*, penelitian ini menggunakan algoritma SMOTE untuk menangani data ketidakseimbangan. Ketidakseimbangan dapat dikelola dengan menggunakan teknik *undersampling* dan *oversampling* (Khansa & Gunawan, 2024). Penelitian yang dilakukan Juwariyem menggunakan komparasi antara algoritma Bagging dan algoritma Random Forest (RF), Berdasarkan hasil pengujian menggunakan metode Bagging dan algoritma Random



Forest pada penelitian ini diperoleh hasil yaitu kelas *accuracy* sebesar 91,98%, penelitian ini menggunakan *dataset* dengan total 10.001 *record* data, 7 atribut dan 1 kelas atribut (Juwariyem et al., 2024).

## KESIMPULAN

Dapat disimpulkan dari hasil prediksi prevalensi *stunting* balita memiliki akurasi yang baik sekali, dari penerapan model *supervised learning*. Bagi peneliti selanjutnya dapat dikembangkan dengan menggunakan model *semi supervised learning* untuk mempersiapkan dataset yang akan dipergunakan dengan teori *handling missing data* dengan metode imputer K-NN atau K-Means kedua metode ini untuk melakukan *clustering* data sebelum data diimplementasikan ke model *machine learning*, serta dapat diterapkan dengan *hyperparameter tuning grid search, random search, cross validation*, untuk mendapatkan hasil yang akurasi yang baik (Hidayat et al., 2024). Selanjutnya program implementasi model dapat dikembangkan ke sebuah desktop atau berbasis *website*, untuk memudahkan *user* dalam pekerjaan terkait data *stunting*.

## UCAPAN TERIMA KASIH

Terima kasih kepada Dinas Kesehatan Kabupaten Tangerang yang telah memberikan izin penelitian ini, serta pembimbing dari dosen Universitas Kristen Satya Wacana (UKSW) Salatiga dalam memberikan masukan dan arahan dalam penelitian ini.

## DAFTAR PUSTAKA

- Amanda, R., & Negara, E. S. (2020). Analysis and Implementation Machine Learning for YouTube Data Classification by Comparing the Performance of Classification Algorithms. *Jurnal Online Informatika*, 5(1), 61–72. <https://doi.org/10.15575/join.v5i1.505>
- Austin, R. R., Mathiason, M. A., & Monsen, K. A. (2022). Using data visualization to detect patterns in whole-person health data. *Research in Nursing and Health*, 45(4), 466–476. <https://doi.org/10.1002/nur.22248>
- Barros, M. T., Siljak, H., Mullen, P., Papadias, C., Hyttinen, J., & Marchetti, N. (2022). Objective Supervised Machine Learning-Based Classification and Inference of Biological Neuronal Networks. *Molecules*, 27(19), 1–23. <https://doi.org/10.3390/molecules27196256>
- Bharti, Gill, N. S., & Gulia, P. (2023). Exploring machine learning techniques for fake profile detection in online social networks. *International Journal of Electrical and Computer Engineering*, 13(3), 2962–2971. <https://doi.org/10.11591/ijece.v13i3.pp2962-2971>
- Booeshaghi, A. S., Sullivan, D. K., & Pachter, L. (2023). Universal preprocessing of single-cell genomics data. *BioRxiv*, 2023.09.14.543267. <https://www.biorxiv.org/content/10.1101/2023.09.14.543267v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.09.14.543267v1.abstract>
- Doshi, N., Gundam, S., & Chaudhury, B. (2021). *Strategizing University Rank Improvement using Interpretable Machine Learning and Data Visualization*. <http://arxiv.org/abs/2110.09050>
- Dritsas, E., & Trigka, M. (2023). Supervised Machine Learning Models to Identify Early-Stage Symptoms of SARS-CoV-2. *Sensors*, 23(1). <https://doi.org/10.3390/s23010040>
- Erol, G., Uzbaş, B., Yücelbaş, C., & Yücelbaş, Ş. (2022). Analyzing the effect of data preprocessing techniques using machine learning algorithms on the diagnosis of COVID-19. *Concurrency and Computation: Practice and Experience*, 34(28), 1–16. <https://doi.org/10.1002/cpe.7393>
- Hidayat, T., Manongga, D., Hendry, Nataliani, Y., Wijono, S., Prasetyo, S. Y., Maria, E., Raharja, U., & Sembiring, I. (2024). Performance Prediction Using Cross Validation (GridSearchCV) for Stunting Prevalence. *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, 1–6. <https://doi.org/https://doi.org/10.1109/AIMS61812.2024.10512657>
- Indrakumari, R., Poongodi, T., & Jena, S. R. (2020). Heart Disease Prediction using Exploratory Data Analysis. *Procedia Computer Science*, 173(2019), 130–139. <https://doi.org/10.1016/j.procs.2020.06.017>
- Juwariyem, J., Sriyanto, S., Lestari, S., & Chairani, C. (2024). Prediction of Stunting in Toddlers Using Bagging and Random Forest Algorithms. *Sinkron*, 8(2), 947–955. <https://doi.org/https://doi.org/10.33395/sinkron.v8i2.13448>
- Khansa, G. A. F., & Gunawan, P. H. (2024). Predicting Stunting in Toddlers Using KNN and Naïve Bayes Methods. *2024 International Conference on Data Science and Its Applications (ICoDSA)*, 17–21. <https://doi.org/10.1109/ICoDSA62899.2024.10651676>
- Khusna, N. F., Rahmah, A., & Nur, R. K. (2024). *Implementasi Random Forest dalam Klasifikasi Kasus Stunting pada Balita dengan Hyperparameter Tuning Grid Search*. *2024(Senada)*, 791–801.
- Kong, X., Ravikumar, V., Mulpuru, S. K., Roukoz, H., & Tolkacheva, E. G. (2023). A Data-Driven Preprocessing Framework for Atrial Fibrillation Intracardiac Electrocardiogram Analysis. *Entropy*, 25(2), 1–15. <https://doi.org/10.3390/e25020332>
- Laengsri, V., Shoombuatong, W., Adirojananon, W., Nantasenamart, C., Prachayasittikul, V., & Nuchnoi, P. (2019). ThalPred: A web-based prediction tool for discriminating thalassemia trait and iron deficiency

- anemia. *BMC Medical Informatics and Decision Making*, 19(1), 1–14. <https://doi.org/10.1186/s12911-019-0929-2>
- Malau, C. G. M., Sholihah, B., & Salim, A. (2023). Visualisasi Data Pembelian Barang dan Jasa Pada PT. Transcoal Pacific Menggunakan Exploratory Data Analysis. *Intelmatiks*, 3(1), 7–12. <https://doi.org/10.25105/itm.v3i1.16302>
- Mayasari, R., Nugraha, B., Juwita, A. R., & Heryana, N. (2023). Analisis Produktifitas Padi di Pulau Sumatera menggunakan Exploratory Data Analysis (EDA). *Jurnal Elektronik Sistem Informasi Unsika*, 1(1), 17–24.
- Muhajir, M., & Widiastuti, J. (2022). Random Forest Method Approach to Customer Classification Based on Non-Performing Loan in Micro Business. *Jurnal Online Informatika*, 7(2), 177–183. <https://doi.org/10.15575/join.v7i2.842>
- Mustaqim, M., Warsito, B., & Surarso, B. (2019). Kombinasi Synthetic Minority Oversampling Technique (SMOTE) dan Neural Network Backpropagation untuk menangani data tidak seimbang pada prediksi pemakaian alat kontrasepsi implan. *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, 5(2), 128. <https://doi.org/10.26594/register.v5i2.1705>
- Nadhiroh, S. R., Riyanto, E. D., & Salsabil, I. S. (2022). Potensi Balita Risiko Stunting dan Hubungannya dengan Keluarga Pra-Sejahtera di Jawa Timur : Analisis Data PK-21. *Media Gizi Indonesia (National Nutrition Journal) Special Issue: International Conference on Stunting 2022*, 17(1), 112–119. <https://doi.org/10.20473/mgi.v17i1SP.112-119>
- Nofriani, N. (2019). Comparations of Supervised Machine Learning Techniques in Predicting the Classification of the Household's Welfare Status. *Journal Pekommas*, 4(1), 43. <https://doi.org/10.30818/jpkm.2019.2040105>
- Pohan, H., Zarlis, M., Irawan, E., Okprana, H., & Pranayama, Y. (2021). Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia. *JUKI : Jurnal Komputer Dan Informatika*, 3(2), 97–104. <https://doi.org/10.53842/juki.v3i2.69>
- Siambaton, M. Z., & Husein, A. M. (2022). Menganalisis Data Kesehatan Global : Pendekatan Analisis Data Eksplorasi Visual. *Data Sciences Indonesia (DSI)*, 1(2), 41–49. <https://doi.org/10.47709/dsi.v1i2.1315>
- Subadi, A. (2020). *Diagnosa Stunting Berdasarkan Gejala Medis*. 501–510.
- Syahrudin, A. N., & Sari, N. P. (2023). *Water Sanitation and Hygiene ( WASH ) and feeding patterns : Linkages with stunting among children aged 6-23 months Water Sanitation and Hygiene ( WASH ) dan pola pemberian makan : Hubungannya dengan stunting pada anak usia 6-23 bulan Abstrak*. 8(3), 466–477.
- Tiwari, S. (2022). Supervised Machine Learning: A Brief Introduction. *Proceedings of the International Conference on Virtual Learning*, 17(5), 219–230. <https://doi.org/10.58503/icvl-v17y202218>
- Torres-Martos, Á., Bustos-Aibar, M., Ramírez-Mena, A., Cámara-Sánchez, S., Anguita-Ruiz, A., Alcalá, R., Aguilera, C. M., & Alcalá-Fdez, J. (2023). Omics Data Preprocessing for Machine Learning: A Case Study in Childhood Obesity. *Genes*, 14(2). <https://doi.org/10.3390/genes14020248>
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 1–16. <https://doi.org/10.1186/s12911-019-1004-8>
- Wibowo, A. (2022). Analisa Dan Visualisasi Data Penjualan Menggunakan Exploratory Data Analysis Pada PT. Telkominfra. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(3), 2292–2304. <https://doi.org/10.35957/jatisi.v9i3.2737>
- Wittek, N., Wittek, K., Keibel, C., & Güntürkün, O. (2023). Supervised machine learning aided behavior classification in pigeons. *Behavior Research Methods*, 55(4), 1624–1640. <https://doi.org/10.3758/s13428-022-01881-w>