G-Tech : Jurnal Teknologi Terapan

Vol. 8, No. 3, Juli 2024, hlmn. 1890-1900
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Prediksi Prevalensi Stunting di Indonesia dengan *Ordinary Least Square* (OLS)

Benny Putra[1]✉, Alva Hendi Muhammad[2]

[1, 2] Teknik Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 13-06-2024
Direvisi : 24-06-2024
Diterima : 28-06-2024

ABSTRAK

Stunting adalah masalah pertumbuhan pada anak-anak, menjadi perhatian serius di Indonesia dan global, dengan lebih dari 149 juta balita terdampak, termasuk 6,3 juta di Indonesia. Meski ada penurunan, mencapai target nasional 2024 masih menantang. Pemerintah telah mengeluarkan Peraturan Presiden untuk menangani stunting, fokus pada gizi keluarga dan kebersihan lingkungan. Penelitian ini bertujuan memprediksi prevalensi stunting, mengembangkan model prediksi yang lebih akurat, memberikan dasar kebijakan, dan berkontribusi pada literatur ilmiah mengenai stunting di Indonesia. Metode yang digunakan adalah membandingkan Algoritma Neural Network (NN), RBF Network, SVR kernel RBF, dan Ordinary Least Square (OLS). Evaluasi menunjukkan variasi kinerja signifikan Linear Regression menunjukkan nilai MAE sebesar 0.93 dan MSE sebesar 1.34, sementara SVR memiliki nilai MAE sebesar 0.91 dan MSE sebesar 1.30. Sebaliknya, OLS menampilkan kinerja terbaik dengan nilai MAE sebesar 0.020390, MSE sebesar 0.000816, RMSE sebesar 0.028561, R2 sebesar 0.923281, dan MAPE sebesar 0.044035. Hal ini menunjukkan bahwa OLS memberikan prediksi yang sangat akurat dengan kesalahan yang minimal dan korelasi yang tinggi, menjadikannya metode yang unggul dalam memprediksi prevalensi stunting.

Kata Kunci:

Stunting, Prediksi, Prevalensi, *Machine Learning, OLS*

ABSTRACT

Stunting is a growth issue in children, a serious concern in Indonesia and globally, with over 149 million affected children worldwide, including 6.3 million in Indonesia. Despite a decline, reaching the national target for 2024 remains challenging. The government has issued a Presidential Regulation to address stunting, focusing on family nutrition and environmental hygiene. This study aims to predict the prevalence of stunting, develop more accurate predictive models, provide a policy basis, and contribute to the scientific literature on stunting in Indonesia. The methods used involve comparing Neural Network (NN) algorithms, RBF Network, SVR kernel RBF, and Ordinary Least Square (OLS). The evaluation shows significant performance variations. Linear Regression shows an MAE value of 0.93 and MSE of 1.34, while SVR has an MAE of 0.91 and MSE of 1.30. Conversely, OLS demonstrates the best performance with an MAE of 0.020390, MSE of 0.000816, RMSE of 0.028561, R2 of 0.923281, and MAPE of 0.044035. This indicates that OLS provides very accurate predictions with minimal error and high correlation, making it a superior method for predicting stunting prevalence.

Keywords **:**

Stunting, Prediction, Prevalence, Machine Learning, OLS

Corresponding Author **:**
Benny Putra
Teknik Informatika, PJJ MTI, Universitas Amikom Yogyakarta, Indonesia
Jl. Ring Road Utara, Ngringin, Condongcatur, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yograkarta
Email: benny.putra@students.amikom.ac.id

This is an open access article under the CC BY license

PENDAHULUAN

Stunting merupakan permasalahan serius global, termasuk di Indonesia, yang ditandai dengan kerdil dan pertumbuhan terhambat pada balita akibat kekurangan gizi kronis. Statistik PBB Tahun 2020 mencatat bahwa lebih dari 149 juta balita di dunia mengalami stunting, dan di Indonesia, sekitar 6,3 juta anak usia dini terpengaruh. Menurut UNICEF, hal ini disebabkan oleh kekurangan gizi pada dua tahun pertama kehidupan anak, kekurangan nutrisi ibu selama kehamilan, dan sanitasi yang buruk. Meskipun prevalensi stunting di Indonesia menurun dari 24,4% pada 2021 menjadi 21,6% pada 2022, angka ini masih di atas target nasional sebesar 14% untuk tahun 2024. Stunting tidak hanya mempengaruhi tinggi badan, tetapi juga berhubungan dengan kemampuan belajar, keterbelakangan mental, dan risiko penyakit kronis (Kementerian Kesehatan, 2023).

Pemerintah Indonesia, melalui Perpres Nomor 72 Tahun 2021, menegaskan komitmennya untuk mempercepat penurunan stunting guna mencapai Sustainable Development Goals (SDGs). Upaya ini semakin penting di tengah dampak pandemi COVID-19 yang memperburuk masalah gizi anak. Strategi Nasional Percepatan Penurunan Stunting mencakup perluasan cakupan program, integrasi inisiatif, dan sinergi lintas sektor. Evaluasi kinerja program dilakukan melalui berbagai instrumen, termasuk Survei Status Gizi dan Evaluasi Kinerja 8 Aksi Konvergensi. Hasil Survei Status Gizi Indonesia 2021 menunjukkan penurunan prevalensi stunting dari 27,7% pada 2020 menjadi 24,4% pada 2021.

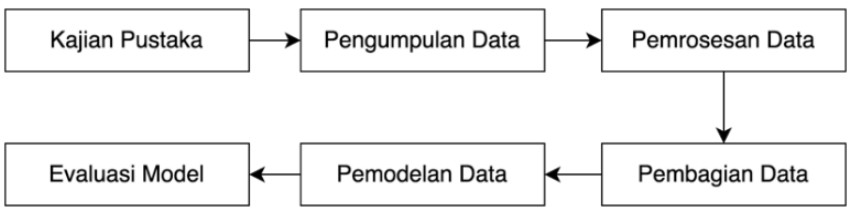
Penelitian sebelumnya oleh (Haris et al., 2022-2034) di Indonesia dan (Das et al., 2022) di Bangladesh menggunakan pendekatan machine learning dan analisis statistik untuk mengidentifikasi faktor-faktor yang berkontribusi terhadap stunting. Studi-studi ini menyoroti pentingnya meningkatkan Indeks Pembangunan Manusia, sanitasi, dan penanggulangan kemiskinan untuk mengatasi stunting.

Dalam penelitian ini, digunakan metode Algoritma *Neural Network*, *RBF Network, SVR kernel RBF*, dan *Ordinary Least Square* untuk memprediksi prevalensi stunting. Data Badan Pusat Statistik menjadi landasan penting untuk analisis. Penelitian ini bertujuan untuk memprediksi tingkat prevalensi stunting, mengidentifikasi faktor risiko signifikan, dan memberikan dasar bagi kebijakan dan intervensi kesehatan anak. Pendekatan OLS diharapkan dapat memberikan pemahaman holistik tentang dinamika kompleks stunting dan membantu dalam pengembangan kebijakan yang lebih efektif.

Namun, terdapat kelemahan dalam penggunaan algoritma prediktif, seperti ketergantungan pada kualitas dan kelengkapan data, risiko overfitting, dan pemilihan variabel yang dapat mempengaruhi hasil. Interpretasi hasil prediksi harus dilakukan dengan hati-hati untuk menghasilkan rekomendasi kebijakan yang tepat. Memahami kelemahan ini penting untuk memperbaiki algoritma dan memastikan penggunaan hasil analisis yang bijak untuk pengambilan keputusan kebijakan yang lebih efektif.

METODE PENELITIAN

Pada penelitan ini terdapat 6 (enam) tahapapan yang harus dilakukan yaitu Kajian Pustaka, Pengumpulan Data, Pemrosesan Data, Pembagian Data, Pemodelan Data dan Evaluasi Model seperti ditunjukan pada Gambar 1 berikut.



Gambar 1. Tahapan Penelitian

Kajian Pustaka

Stunting yaitu kondisi balita kerdil dan pendek, merupakan permasalahan serius dalam kesehatan anak di Indonesia. Penelitian terbaru oleh Haris et al. (2022) yang dipublikasi dalam Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK) menyoroti tingkat prevalensi stunting yang mencapai 26,8% di Provinsi Jawa Timur, melampaui standar minimal WHO sebesar 20%. Tujuan dari studi ini memberikan kontribusi signifikan dalam penanganan stunting di wilayah tersebut dengan menggunakan pendekatan *supervised machine learning*. Haris et al. (2022) tidak hanya membatasi diri pada satu metode, melainkan membandingkan beberapa diantaranya seperti linier regression, support vector regression, dan random forest regression. Hasil penelitian menunjukkan bahwa support vector regression memiliki tingkat galat yang lebih rendah, menjadikannya pilihan terbaik. Faktor-faktor penyebab stunting yang ditemukan, seperti Berat Badan Lahir Rendah (BBLR), Indeks Pembangunan Manusia, sanitasi, dan Indeks Penduduk Miskin, memberikan wawasan baru dengan mengungkapkan korelasi yang signifikan terhadap prevalensi stunting. Implikasi kebijakan yang muncul menunjukkan perlunya peningkatan pada Indeks Pembangunan Manusia, sanitasi, dan penanggulangan kemiskinan untuk menurunkan tingkat stunting di Provinsi Jawa Timur.

Seiringnya dengan tantangan gizi buruk di Indonesia, penelitian Das et al. (2022) di Bangladesh mengeksplorasi estimasi prevalensi gizi buruk pada tingkat distrik. Menggunakan data dari *Bangladesh Multiple Indicator Cluster Survey* 2019, penelitian ini memperhatikan disparitas spasial dan demografis dalam prevalensi stunting, *wasting*, dan *underweight*. Meskipun penelitian ini memiliki batasan, seperti tidak mempertimbangkan korelasi antar indikator gizi buruk, kontribusinya penting dalam pemahaman masalah gizi buruk anak di Bangladesh. Bab tinjauan pustaka ini memberikan latar belakang yang kuat, merinci tantangan kesehatan anak terkait stunting, dan memperkenalkan metode serta faktor yang menjadi fokus penelitian Das et al. (2022). Hasil penelitian ini juga memberikan wawasan yang berharga dalam konteks ketidaksetaraan dalam prevalensi gizi buruk di Bangladesh, mendukung upaya pemantauan dan perencanaan kebijakan kesehatan anak di tingkat distrik.

Sementara itu, Penelitian oleh Mambang et al. (2023) menganalisis prevalensi stunting pada anak balita di Kalimantan Selatan, Indonesia, menggunakan pendekatan pembelajaran mesin linier. Penelitian ini menggunakan berbagai algoritma, termasuk regresi linier, regresi ridge, dan regresi polinomial dengan model pipeline, untuk mengidentifikasi pola dan tren stunting. Hasilnya menunjukkan bahwa regresi polinomial dengan pipeline memberikan prediksi paling akurat, ditunjukkan oleh nilai MAPE dan RMSE yang rendah, serta nilai R-squared yang tinggi. Variabel minimum, maksimum, dan rata-rata terbukti signifikan dalam meningkatkan keakuratan prediksi. Temuan ini menekankan pentingnya variabel tersebut dalam pemodelan stunting dan memberikan kontribusi penting untuk penelitian lebih lanjut serta intervensi di bidang kesehatan dan perkembangan anak. Penelitian ini tidak hanya berkontribusi secara akademis tetapi juga memiliki implikasi praktis dalam menghadapi tantangan kesehatan anak.

Penelitian oleh Haris et al. (2023) menyoroti masalah serius stunting pada balita di Provinsi Jawa Timur, Indonesia, di mana prevalensinya mencapai 26,8%, melebihi standar WHO. Studi ini menggunakan algoritma machine learning Random Forest untuk memprediksi prevalensi stunting. Keunggulan Random Forest adalah kemampuannya mempelajari pola dari data berlabel, menjadikannya metode potensial untuk peramalan data kesehatan. Metodologi penelitian mencakup pengumpulan, persiapan, dan analisis data, pelatihan, serta pengujian model. Setelah proses pemilihan fitur, ditemukan bahwa dari 20 faktor kandidat, hanya 12 yang memiliki korelasi signifikan dengan prevalensi stunting. Meskipun Random Forest digunakan, hasil penelitian menunjukkan bahwa Multi Linear Regression (MLR) memiliki kinerja yang lebih baik dengan nilai error MAE 0,93 dan MSE 1,34, dibandingkan Random Forest yang memiliki MAE 1,02 dan MSE 1,64. Hal ini kemungkinan disebabkan oleh jumlah fitur yang terbatas (12 fitur) dan dataset yang kecil (38 data), di mana model machine learning cenderung bekerja lebih baik dengan dataset yang lebih besar. Penelitian ini menunjukkan bahwa metode statistik seperti MLR masih relevan dan dapat bersaing dengan model machine learning dalam kondisi dataset terbatas. Implikasi dari penelitian ini adalah perlunya penambahan data dari provinsi lain untuk meningkatkan akurasi model Random Forest dalam penelitian lanjutan.Penelitian oleh Bitew et al. (2021) mengevaluasi

algoritma pembelajaran mesin untuk memprediksi kekurangan gizi pada anak-anak di Ethiopia. Mereka menemukan bahwa penggunaan algoritma XGBTree lebih efektif dalam memprediksi kondisi gizi anak-anak, dengan faktor-faktor seperti akses air, ketahanan pangan, dan kebijakan kesuburan memiliki peran yang penting. Namun, studi tersebut tidak secara rinci membahas kelemahan metodologi yang digunakan.

Penelitian oleh Darnila et al. (2022) bertujuan untuk memprediksi stunting di Aceh Timur menggunakan pendekatan pembelajaran mesin. Mereka menemukan bahwa model Random Forest secara efektif dapat memprediksi status stunting di wilayah tersebut, dan berhasil mengidentifikasi zona-zona risiko tinggi di bagian timur Aceh. Namun, diskusi dalam penelitian ini terbatas pada algoritma *Random Forest* saja, tanpa membandingkan dengan metode statistik tradisional untuk prediksi, serta kurangnya perbandingan dengan model-model lain seperti NN, RBF Network, dan SVR kernel RBF.

Secara keseluruhan, temuan dari berbagai penelitian ini memberikan pemahaman mendalam tentang faktor-faktor yang berkaitan dengan prevalensi stunting dan memberikan dasar untuk pengembangan kebijakan dan intervensi kesehatan anak yang lebih efektif di berbagai konteks regional. Diperlukan kolaborasi lintas sektor dan pengumpulan data yang lebih luas untuk meningkatkan akurasi model dan memahami dinamika stunting secara menyeluruh.

Pengumpulan Data

Pada penelitian ini menggunakan data sekunder yang diperoleh dari publikasi Badan Pusat Statistik di Indonesia yang berupa data Statistik Kesejahteraan Rakyat Tahun 2013-2023 serta Data dari Kementerian Kesehatan Republik Indonesia tentang Pemantauan Status Gizi Tahun 2013-2017, Laporan Nasional Riskesdas Tahun 2018 dan Studi Status Gizi Balita Indonesia Tahun 2019-2022. Dari seluruh data tersebut dilakukan kompilasi menjadi beberapa variabel yaitu, Data Prevalensi Stunting (target), Wilayah Provinsi (wilayah), Tahun (tahun), Persentase Penduduk Tinggal Di Rumah (x1), Persentase Penduduk Kemampuan Membaca dan Menulis (x2), Persentase Penduduk Kuliah (x3), Persentase Penduduk Mendapat Imunisasi Lengkap (x4), Persentase Balita ASI Eksklusif (x5) dan Persentase Rumah Tangga Akses Air Minum Layak (x6).

Pra-Pemrosesan Data

Tahap pemrosesan data terlebih dahulu sebelum dilakukan pemodelan. Tahapan *data cleaning* dilakukan untuk mengidentifikasi data yang tidak valid atau memiliki nilai kosong. Langkah ini mencakup teknik pembersihan data yang berguna untuk menangani nilai kosong atau data yang tidak lengkap. Hasil dari pembersihan data ini membantu mempersiapkan data untuk analisis lebih lanjut. Selanjutnya, dilakukan *feature selection* untuk menentukan fitur atau atribut yang paling berpengaruh dalam pemodelan data. Proses ini melibatkan pengukuran setiap fitur untuk mengidentifikasi fitur yang dominan yang selanjutnya digunakan dalam pemodelan.

Pembagian Data

Data dibagi menjadi dua bagian yaitu data latih dan data uji. Data latih digunakan untuk melatih model dalam *machine learning*, sementara data uji digunakan untuk menguji hasil dari pembelajaran model tersebut. Pembagian data pada penelitian ini menggunakan rasio 80% untuk data latih dan 20% untuk data uji. Pembagian data ini bertujuan untuk meningkatkan kualitas pembelajaran mesin. Semakin banyak data yang digunakan untuk melatih model, diharapkan model dapat belajar dari variasi yang lebih besar dari data yang ada. Hasilnya, diharapkan model dapat memberikan prediksi atau hasil yang lebih akurat saat diuji dengan data yang belum pernah dilihat sebelumnya.

Pemodelan Data

Teknik pemodelan yang dilakukan pada penelitian ini adalah Algoritma *Neural Network* (NN), *RBF Network*, SVR kernel RBF, dan *Ordinary Least Square* (OLS).

1. *Neural Network* (NN)
   Neural Network (NN) adalah sebuah model komputasi yang terinspirasi oleh cara kerja jaringan saraf biologis. NN terdiri dari sejumlah unit pemrosesan yang disebut neuron, yang saling terhubung dengan bobot. NN dapat belajar dari data dan menyesuaikan bobotnya secara otomatis dengan menggunakan algoritma pembelajaran tertentu. NN dapat digunakan untuk berbagai tujuan, seperti klasifikasi, regresi, pengenalan pola, dan optimasi. Persamaannya seperti pada (1) (Hagan, Demuth, Beale, & De Jesús, 2014).
   Dimana y_i adalah keluaran neuron ke-i, f adalah fungsi aktivasi, wij adalah bobot dari neuron ke-j ke neuron ke-i, x_j adalah masukan dari neuron ke-j, dan bi adalah bias dari neuron ke-i.

$$y_i = f(\sum_{j=1}^{n} w_{ij}x_j + b_i) \quad (1)$$

2. RBF Network
   RBF Network adalah sebuah jenis NN yang menggunakan fungsi basis radial sebagai fungsi aktivasi. Fungsi basis radial adalah fungsi yang nilainya bergantung pada jarak antara input dan pusat tertentu. RBF Network biasanya memiliki tiga lapisan: lapisan input, lapisan tersembunyi dengan fungsi aktivasi RBF, dan lapisan output dengan fungsi linear. RBF Network dapat digunakan untuk menyelesaikan masalah pendekatan fungsi, prediksi deret waktu, klasifikasi, dan kontrol sistem [28]. Rumus untuk menghitung keluaran dari setiap lapisan tersembunyi adalah sebagai berikut (2):

$$h_i = \phi(||x - c_i||) \quad (2)$$

   Di mana hi adalah keluaran neuron ke-i, ϕ adalah fungsi basis radial, x adalah vektor masukan, dan ci adalah pusat dari neuron ke-i. Rumus untuk menghitung keluaran dari setiap neuron di lapisan keluaran adalah sebagai berikut (3):

$$y_i = \sum_{j=1}^{n} w_{ij}h_j + b_i \quad (3)$$

   Di mana yi adalah keluaran neuron ke-i, wij adalah bobot dari neuron ke-j ke neuron ke-i, hj adalah keluaran dari neuron ke-j di lapisan tersembunyi, dan bi adalah bias dari neuron ke-i.

3. SVR kernel RBF
   SVR kernel RBF adalah sebuah metode regresi yang menggunakan *Support Vector Machine* (SVM) dengan kernel RBF. SVM adalah sebuah metode klasifikasi yang mencari *hyperplane* terbaik yang memisahkan dua kelas data. SVR adalah sebuah variasi dari SVM yang digunakan untuk regresi, yaitu mencari fungsi yang paling sesuai dengan data. Kernel RBF adalah sebuah fungsi yang menghitung kemiripan antara dua vektor input dengan menggunakan jarak Euclidean. SVR kernel RBF dapat digunakan untuk menyelesaikan masalah regresi non-linear dengan akurasi yang tinggi (Chang & Lin, 2011). Rumus untuk menghitung fungsi prediksi adalah sebagai berikut (4):

$$f(x) = \sum_{j=1}^{n} (a_i - a_i^*)K(x_i, x) + b \quad (4)$$

   di mana *f(x)* adalah nilai prediksi untuk vektor masukan x, α_i dan α_i* adalah variabel Lagrange, $K(x_i,x)$ adalah fungsi kernel RBF, dan b adalah bias. Fungsi kernel RBF didefinisikan sebagai berikut (5):

$$K(x_i, x) = exp\,(-\gamma||\,x_i - x\,||^2) \quad (5)$$

   di mana γ adalah parameter kernel yang harus dipilih.

4. *Ordinary Least Square* (OLS)
   Ordinary Least Square (OLS) adalah sebuah metode regresi linier yang mencari parameter yang meminimalkan jumlah kuadrat kesalahan antara nilai observasi dan nilai prediksi. OLS adalah metode yang sederhana, mudah diimplementasikan, dan dapat memberikan estimasi yang tidak bias dan efisien jika asumsi-asumsi regresi terpenuhi. OLS dapat digunakan untuk menyelesaikan masalah regresi linier sederhana maupun berganda (Gujarati & Porter, 2009). Rumus untuk menghitung parameter adalah sebagai berikut (6):

$$\beta = (X^T X)^{-1} X^T y \quad (6)$$

   di mana β adalah vektor parameter, X adalah matriks masukan, dan y adalah vektor keluaran. Rumus untuk menghitung fungsi prediksi adalah sebagai berikut (7):

$$y\hat{} = X\beta \quad (7)$$

   di mana y^ adalah vektor prediksi.

Evaluasi Model

Melakukan analisis terhadap hasil data mining dari proses pemodelan yang dilakukan sebelumnya. Evaluasi model dilakukan untuk memastikan model tersebut sesuai dengan tujuan yang telah ditetapkan di awal penelitian. Indikator kinerja yang digunakan dalam penelitian ini meliputi MAE (*Mean Absolute Error*), MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*), dan R2.

HASIL DAN PEMBAHASAN

Dataset

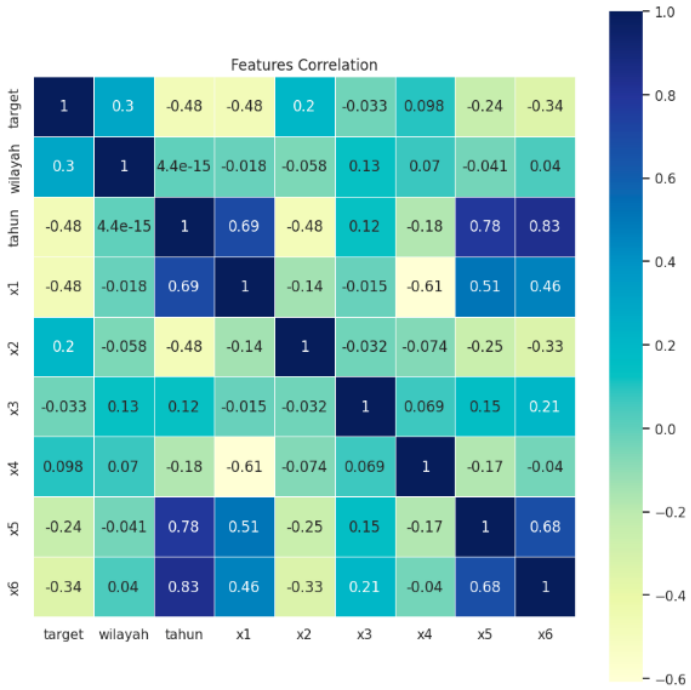
Setelah data dikumpulkan, selanjutnya data diintegrasikan, sehingga didapatkan data sebanyak 340 record yang merupakan data 34 provinsi di Indonesia berdasarkan faktor penyebab stunting pada balita. Adapun tampilan dataset tersebut dapat dilihat pada Tabel 1.

Tabel 1. Dataset

	target	wilayah	tahun	x1	x2	x3	x4	x5	x6
0	41.5	1	2013	51.11	95.66	4.05	54.40	40.89	27.80
1	42.5	2	2013	54.34	97.43	2.49	60.99	31.20	39.52
2	39.2	3	2013	55.04	96.61	3.27	63.01	48.08	31.88
3	36.8	4	2013	58.62	97.04	2.65	65.17	44.21	37.43
…	…	…	…	…	…	…	…	…	…
336	26.1	31	2022	96.75	3.29	5.17	53.39	79.54	92.10
337	26.1	32	2022	97.78	3.21	4.23	51.22	73.38	88.10
338	30.0	33	2022	96.79	7.75	3.45	53.85	72.11	81.57
339	34.6	34	2022	97.81	1.64	1.98	49.12	79.29	65.39

Dalam penelitian ini, data yang digunakan adalah data prevalensi stunting di Indonesia dari tahun 2013 sampai 2023. Namun, terdapat beberapa data yang kosong atau tidak tersedia, sehingga perlu dilakukan pengisian data dengan cara tertentu. Data prevalensi stunting tahun 2014 kosong, sehingga diisi dengan cara mengambil rata-rata dari data prevalensi tahun 2013 dan 2015. Data provinsi Kalimantan Utara pada tahun 2013 juga kosong, karena provinsi ini belum ada pada saat itu. Oleh karena itu, data provinsi Kalimantan Utara pada tahun 2013 diisi dengan data provinsi Kalimantan Timur, karena provinsi ini merupakan provinsi asal pemekaran. Dengan demikian, data yang digunakan dalam penelitian ini menjadi lebih lengkap dan dapat dianalisis lebih lanjut.

Pada tahap ini dilakukan uji coba pembuatan model untuk mengetahui hubungan antar parameter x1-x6 atau tanpa x1-x6. Hasil dari uji coba tersebut dapat dilihat pada gambar 3. Dari hasil tersebut dapat dilihat bahwa penggunaan x1-x6 dapat meningkatkan performa model. Hubungan korelasi antar variabel dapat dilihat di gambar dibawah ini :



Gambar 2. Korelasi Fitur Berkelanjutan

Dalam tahap pra-pemrosesan data, normalisasi merupakan langkah penting untuk memastikan bahwa fitur-fitur dalam dataset memiliki skala yang seragam. Proses normalisasi fitur-fitur pada dataset menggunakan metode *Standar Scaler* yang dapat dilihat pada Tabel 2 dan *MinMaxScaler* pada Tabel 3.

1. *Standard Scaler*

Tabel 2. Tabel Hasil Normalisasi *Standard Scaler*

	target	wilayah	tahun	x1	x2	x3	x4	x5	x6
0	1.7860	-1.6818	-1.5667	-2.4953	0.3526	1.1131	0.0634	-1.3653	-1.3482
1	1.9296	-1.5799	-1.5667	-2.2865	0.4189	-0.4313	0.4529	-2.0010	-0.8735
…	…	…	…	…	…	…	…	…	…
338	0.1355	1.5799	1.5667	0.4585	-2.9399	0.5191	0.0309	0.6830	0.8297
339	0.7957	1.6818	1.5667	0.5245	-3.1687	-0.9362	-0.2487	1.1541	0.1743

2. *MinMax Scaler*

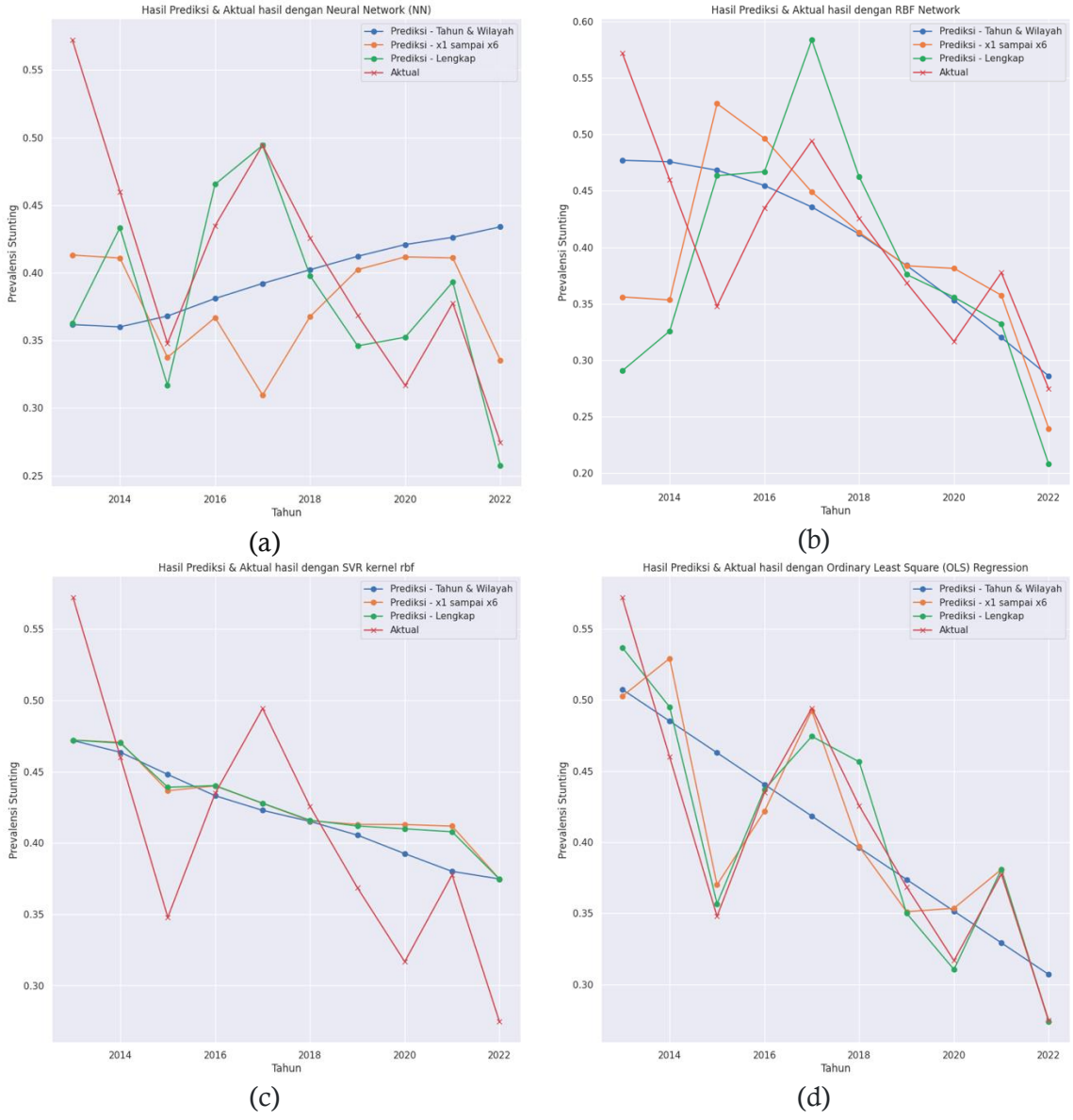
Tabel 3. Tabel Hasil Normalisasi *MinMaxScaler*

	target	wilayah	tahun	x1	x2	x3	x4	x5	x6
0	0.767	0.000	0.0	0.000	0.958	0.535	0.509	0.270	0.147
1	0.789	0.030	0.0	0.067	0.976	0.266	0.597	0.096	0.286
…	…	…	…	…	…	…	…	…	…
338	0.503	0.969	1.0	0.953	0.069	0.432	0.502	0.831	0.783
339	0.609	1.000	1.0	0.974	0.007	0.178	0.438	0.960	0.592

Berdasarkan perhitungan Accuracy MSE with standard scaler yaitu 0.609 dan Accuracy MSE with MinMax scaler : 0.016, dapat disimpulkan bahwa MinMax Scaler memiliki kinerja yang lebih baik daripada model yang menggunakan Standar Scaller.

Pemodelan

Pada penelitian ini metode yang digunakan adalah regresi dengan menguji beberapa *Algoritma Neural Network* (NN), *RBF Network*, SVR kernel rbf, dan *Ordinary Least Square* (OLS) dengan Data Jawa Timur. Pengujian dilakukan dengan tiga skenario, yaitu Skenario 1 : variabel yang digunakan hanya tahun dan wilayah. Skenario 2 : variabel yang digunakan hanya x1, x2, x3, x4, x5 dan x6. Sementara Skenario 3 : menggunakan semua variabel yaitu tahun, wilayah, x1, x2, x3, x4, x5 hingga x6. Hasil dari pengujian ketiga skenario tersebut dapat dilihat pada Gambar 3.



Gambar 3. Grafik Perbandingan Hasil Prediksi dan Aktual

Evaluasi

Berdasarkan grafik hasil prediksi yang ditunjukan pada Gambar 3, terlihat bahwa algoritma OLS pada skenario 3 (Prediksi - Lengkap) yang memiliki grafik paling balik dalam memprediksi data prevalensi stunting daerah dengan data uji wilayah Provinsi Jawa Timur tahun 2013 hingga 2022.

Dapat dilihat juga bahwa performa model meningkat jika menggunakan variabel lengkap, sehingga dapat disimpulkan variabel-variabel yang diuji diatas dapat meningkatkan kemampuan model dalam memprediksi prevalensi stunting. Hasil evaluasi skenario 3 terhadap setiap algoritma dapat dilihat pada Tabel 4.

Tabel 4. Hasil Evaluasi Algoritma dengan Skenario 3

No.	Algoritma	MAE	MSE	RMSE	R2	MAPE
1	Neural Network (NN)	0.089266	0.012422	0.111453	-0.168264	0.227117
2	RBF Network	0.035051	0.002403	0.049025	0.773960	0.073242
3	SVR kernel RBF	0.061366	0.004850	0.069644	0.543838	0.150302
4	Ordinary Least Square (OLS)	0.020390	0.000816	0.028561	0.923281	0.044035

Dari Tabel 4 dapat disimpulkan bahwa performa beberapa model prediksi dapat diukur menggunakan empat matrik evaluasi utama: MAE, MSE, RMSE, dan MAPE. Dari tabel tersebut, dapat disimpulkan bahwa algoritma OLS memiliki performa yang paling baik dalam memprediksi data, karena memiliki nilai MAE, MSE, RMSE, dan MAPE yang paling rendah, serta nilai R2 yang paling tinggi. Algoritma lain seperti SVR kernel RBF, RBF Network, dan Neural Network memiliki performa yang cukup baik namun dibawah performa OLS.

Tabel 5. Perbandingan Algoritma dengan Penelitian Sebelumnya

No.	Algoritma	MAE	MSE	RMSE	R2	MAPE
1	Linear Regression [3]	0.93	1.34	-	-	-
2	SVR [3]	0.91	1.30	-	-	-
3	OLS	0.020390	0.000816	0.028561	0.923281	0.044035

Tabel 5 menunjukan bahwa algoritma OLS dapat dikatakan lebih unggul dibandingkan algoritma *Linear Regression* dan SVR dari penelitian sebelumnya.

Setelah mendapatkan algoritma terbaik, kami kembali melakukan pengujian algoritma OLS ini dengan data 33 provinsi lainnya dengan 100% data untuk pengujian prediksi dan hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil Evaluasi Algoritma OLS untuk semua Provinsi diurutkan dari performa terbaik

No.	Provinsi	MAE	MSE	RMSE	R2	MAPE
1	Kalimantan Barat	0.001350	0.000004	0.002006	0.999249	0.002353
2	DI Yogyakarta	0.002657	0.000011	0.003342	0.997544	0.009571
	…	…	…	…	…	…
33	Jambi	0.031154	0.002272	0.047665	0.871460	0.084662
34	Nusa Tenggara Barat	0.039760	0.002359	0.048571	0.755765	0.060490

Dari hasil Tabel 6 menunjukan bahwa OLS mampu melakukan prediksi nilai prevalensi stunting di Indonesia, bahkan performanya sangat baik jika menggunakan data Provinsi Kalimantan Barat dengan nilai MAE = 0.001350, MSE=0.000004, RMSE=0.002006, R2=0.999249 dan MAPE=0.002353. Oleh karena itu, algoritma OLS dapat direkomendasikan sebagai algoritma regresi yang paling sesuai untuk memprediksi prevalensi stunting Indonesia.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan penelitian yang dilakukan beberapa algoritma machine learning untuk memprediksi prevalensi stunting menunjukkan variasi dalam kinerja masing-masing model. *Neural Network* (NN) menunjukkan kinerja yang cukup baik dengan nilai *Mean Absolute Error* (MAE) yang cukup tinggi dan R2 negatif. Sementara itu, RBF Network menampilkan performa lebih baik dengan tingkat kesalahan yang lebih rendah dan korelasi yang cukup kuat. SVR kernel RBF menunjukkan kinerja baik, namun model *Ordinary Least Square* (OLS) memperlihatkan kualitas prediksi yang sangat baik dengan tingkat kesalahan yang minim dan korelasi tinggi. Dari hasil evaluasi ini dapat disimpulkan bahwa penggunaan algoritma *Ordinary Least Square* (OLS)

memberikan hasil prediksi prevalensi stunting yang paling akurat, dengan tingkat kesalahan yang rendah dan nilai R2 yang tinggi serta lebih unggul dibandingkan algoritma *Linear Regression* dan SVR dari penelitian sebelumnya. Variabel-variabel yang mendukung peningkatan performa antara lain Wilayah Provinsi, Tahun, Persentase Penduduk Tinggal Di Rumah, Persentase Penduduk Kemampuan Membaca dan Menulis, Persentase Penduduk Kuliah, Persentase Penduduk Mendapat Imunisasi Lengkap, Persentase Balita ASI Eksklusif dan Persentase Rumah Tangga Akses Air Minum Layak.

Saran

Untuk penelitian mendatang, disarankan untuk mempertimbangkan penggunaan Time Series Analysis dengan memanfaatkan *Model Radial Basis Function Neural Network* (RBFNN). Penggunaan RBFNN dalam konteks ini diharapkan dapat memberikan hasil yang lebih presisi dalam memodelkan pola perubahan faktor penyebab stunting pada balita seiring berjalannya waktu. Hal ini dapat menjadi kontribusi penting untuk mendukung kebijakan pemerintah dalam menangani masalah stunting di berbagai wilayah di Indonesia secara lebih efektif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Amikom Yogyakarta yang telah memberi dukungan terhadap penelitian ini, dan terima kasih kepada Pak Alva Hendi Muhammad selaku Dosen Pembimbing.

REFERENSI

Badan Pusat Statistik. (2019). Statistik Kesejahteraan Rakyat 2019. Jakarta, Indonesia: Badan Pusat Statistik. ISSN: 0215-4641. Katalog BPS: 4101002. No. Publikasi: 04210.1913

Badan Pusat Statistik. (2020). Statistik Kesejahteraan Rakyat 2020. Jakarta, Indonesia: Badan Pusat Statistik. ISSN: 0215-4641. Katalog BPS: 4101002. No. Publikasi: 04210.1913

Badan Pusat Statistik. (2021). Statistik Kesejahteraan Rakyat 2021. Jakarta, Indonesia: Badan Pusat Statistik. ISSN: 0215-4641. Katalog BPS: 4101002. No. Publikasi: 04210.2116

Badan Pusat Statistik. (2022). Statistik Kesejahteraan Rakyat 2022. Jakarta, Indonesia: Badan Pusat Statistik. ISSN: 0215-4641. Katalog BPS: 4101002. No. Publikasi: 04210.2216

Badan Pusat Statistik. (2023). Laporan Indeks Khusus Penanganan Stunting Kabupaten/Kota 2020-2021, Nomor Katalog: 4207003, Nomor Publikasi: 04200.2306

Badan Pusat Statistik. (2023). Statistik Kesejahteraan Rakyat 2023. Jakarta, Indonesia: Badan Pusat Statistik. ISSN: 0215-4641. Katalog BPS: 4101002. No. Publikasi: 04210.2316

Bitew, F., & others. (2021). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutrition.

Cholid, F., Trishnanti, D., & Azies, H. Al. (2019). Pemetaan Faktor-Faktor yang Mempengaruhi Stunting pada Balita dengan Geographically Weighted Regression (GWR) Mapping of The Factors Affecting Stunting on Toddlers with Geographically Weighted Regression (GWR). Dalam SEMNAkes 2019 "Improving the Quality of Health Tharough Advances in Research of Health Sciences," (hal. 156–165). https://doi.org/10.17605/OSF.IO/9MZU7

Damodar Gujarati, (2019). Basic Econometrics, 4th Edition. Singapore: McGrawth Hill.

Darnila, E., & others. (2022). Supervised models to predict the Stunting in East Aceh. International Journal of Engineering, Science and Information Technology.

Das, S., Bafour, B., & Richardson, A. (2022). Prevalence of child undernutrition measures and their spatio-demographic inequalities in Bangladesh: an application of multilevel Bayesian modelling. BMC Public Health, 22(1008). https://doi.org/10.1186/s12889-022-13170-4.

Galasso, E., & Wagstaff, A. (2019). The Aggregate Income Losses from Childhood Stunting and The Returns to A Nutrition Intervention Aimed at Reducing Stunting. Economics and Human Biology, 34(August), 225–238. https://doi.org/10.1016/j.ehb.2019.01.010

Ginantra, N. L. W. S. R., et al. (2021). Data Mining dan Penerapan Algoritma. Medan: Yayasan Kita Menulis.

Haris, M. S., Anshori, M., & Khudori, A. N. (2023). Prediction of stunting prevalence in East Java Province with Random Forest algorithm. Jurnal Teknik Informatika (JUTIF), volume, nomor halaman. DOI: 10.20884/1.jutif.2023.4.1.614.

Haris, M. S., Khudori, A. N., & Kusuma, W. T. (2022). Perbandingan Metode Supervised Machine Learning untuk Prediksi Prevalensi Stunting di Provinsi Jawa Timur. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), volume, nomor halaman. doi: 10.25126/jtiik.202296744.

Indarwati, T., Irawati, T., & Rimawati, E. (2019). Penggunaan Metode Linear Regression untuk Prediksi Penjualan Smartphone. Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN), https://doi.org/10.30646/tikomsin.v6i2.369.

Kementerian Kesehatan. (2021). Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi, dan Kabupaten/Kota Tahun 2021. Jakarta, Indonesia.

Kementerian Kesehatan. (2021). Prediksi Angka Stunting Tahun 2020. Jakarta, Indonesia.

Kementerian Kesehatan. (2021). Studi Status Gizi Balita Indonesia (SSGBI) Tahun 2019 dan Studi Determinan Status Gizi (SDSG) Pada Masa Pendemi Covid-19 Tahun 2020. Jakarta, Indonesia.

Kementerian Kesehatan. (2023). Hasil Survei Status Gizi Indonesia (SSGI) 2022. Jakarta, Indonesia.

Mambang, F. D., Marleny, & Zulfadhilah, M. (2023). Prediction of linear model on stunting prevalence with machine learning approach. Bulletin of Electrical Engineering and Informatics, volume, nomor halaman. DOI: 10.11591/eei.v12i1.4028.

Paudpedia. (2023). 149 Juta Anak di Dunia Alami Stunting Sebanyak 6,3 Juta di Indonesia, Wapres Minta Keluarga Prioritaskan Kebutuhan Gizi.

Rokom. (2023). Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%.

UNICEF Indonesia. Ringkasan Kajian Gizi Ibu dan Anak. Diakses dari www.unicef.org pada bulan April 2024.

World Health Organization (WHO). Diakses dari www.who.int pada bulan April 2024.