

Application of Artificial Neural Network (MLP) for Multivariate Analysis of Stunting Causes in Indonesia

Muhammad Diva Irnanda ^{1*}, Ananta Surya Pratama ^{2*}, Fawwaz Azhima Putra ^{3*}, Sugiyanto Sugiyanto ^{4**}

^{*}Program Studi Teknik Informatika, Universitas Dian Nuswantoro

^{**} Program Pendidikan Jarak Jauh Informatika, Universitas Dian Nuswantoro

111202315484@mhs.dinus.ac.id ¹, 111202315107@mhs.dinus.ac.id ², 111202314996@mhs.dinus.ac.id ³, sugiyanto@dsn.dinus.ac.id ⁴

Article Info

Article history:

Received 2025-07-13

Revised 2025-07-17

Accepted 2025-07-19

Keyword:

Food Security Index,
Multi-Layer Perceptron (MLP),
Stunting.

ABSTRACT

Stunting is a major public health challenge in Indonesia, primarily caused by prolonged malnutrition and recurrent infections during the First 1,000 Days of Life. This study utilizes the Multi-Layer Perceptron (MLP) neural network model to predict stunting, offering a new dimension in the analysis of complex data and identification of patterns influencing stunting. With its capabilities, the MLP model provides higher precision in detecting contributing factors to stunting. The evaluation results of the model show RMSE of 0.7231, MAE of 3.0313, and an R^2 value of 0.9463. The Food Security Index (IKP), feature X9, had the highest feature importance, followed by X5 (Lack of Clean Water) and X1 (NCPR). This study presents a novel approach to predicting stunting percentages and offers more objective insights to support evidence-based and effective health policies aimed at reducing stunting prevalence in Indonesia.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Stunting menjadi tantangan besar yang perlu diperhatikan dalam kesehatan masyarakat di Indonesia, dengan prevalensi mencapai 21,5% pada 2023 [1]. *Stunting* merupakan keadaan dimana pertumbuhan fisik balita terhambat akibat kurangnya nutrisi dalam jangka panjang, yang terlihat dari tinggi badan yang berada di bawah dua standar deviasi negatif dari median pertumbuhan yang ditetapkan oleh World Health Organization (WHO) [2]. Masalah ini terjadi akibat malnutrisi kronis dan infeksi yang terjadi berulang kali, khususnya pada waktu kritis 1.000 Hari Pertama Kehidupan, yang menyebabkan anak-anak memiliki tinggi badan lebih rendah dibandingkan rata-rata teman sebaya mereka [3]. Berdasarkan Kementerian Kesehatan Indonesia tahun 2023, bahwa meski ada penurunan prevalensi *stunting* dari 24,4% pada tahun 2021 menjadi 21,6% pada tahun 2022, angka tersebut masih jauh dari target nasional yang ditetapkan sebesar 14% pada tahun 2024. Dengan angka yang masih tinggi, upaya penanganan *stunting* di Indonesia membutuhkan perhatian dan tindakan yang lebih serius agar target tersebut dapat tercapai.

Selaras dengan *Sustainable Development Goals* (SDG) poin 2 (*Zero Hunger*) dan 3 (*Good Health and Well-being*)

yang fokus pada pengakhiran kelaparan dan kesehatan yang baik serta kesejahteraan bagi semua, penting untuk memastikan setiap insan, terutama anak-anak, memperoleh asupan gizi cukup dan akses terhadap pelayanan kesehatan yang memadai [4]. Untuk itu, diperlukan upaya strategis melalui analisis data multidimensi mengenai faktor-faktor penyebab *stunting*, seperti malnutrisi dan akses terbatas terhadap pelayanan kesehatan, sehingga kebijakan yang diambil dapat lebih tepat sasaran dan sesuai dengan kebutuhan daerah untuk mengurangi prevalensi *stunting* secara signifikan.

Sejumlah penelitian sebelumnya telah menerapkan *machine learning* untuk memprediksi *stunting*, seperti [5] yang membandingkan *Linear Regression*, *Support Vector Regression* (SVR), dan *Random Forest Regression* yang menemukan bahwa SVR yang unggul serta mengidentifikasi faktor penting seperti Berat Badan Rendah (BBR), indeks pembangunan manusia (IPM), sanitasi, dan indeks penduduk miskin. Di sisi lain, [6] mengeksplorasi menggunakan Naive Bayes, K-Nearest Neighbors, dan Random Forest di platform KNIME, dengan Random Forest unggul pada akurasi 87,75 % dan F1-score 0,922 sementara KNN lebih sensitif mendeteksi kasus positif. Serta [7] melakukan penelitian menggunakan Random Forest dengan 10-fold cross-

validation pada data Kecamatan Pitu, Ngawi yang mencapai akurasi rata-rata 97,87 % dan mengembangkan prototipe sistem *web & Android*.

Meskipun berbagai penelitian telah menunjukkan keberhasilan penggunaan *machine learning* dalam memprediksi *stunting*, masih terdapat beberapa kekurangan terkait fleksibilitas dan kemampuan dalam menangani data yang lebih kompleks serta dinamis. Selain itu, belum banyak studi yang mengaplikasikan MLP sebagai model regresi untuk prediksi *stunting*, terutama pada data multidimensi di sektor wilayah Indonesia.

Oleh karena itu, penelitian ini memiliki pendekatan berupa metode MLP, sebuah model jaringan syaraf tiruan *feedforward* yang terstruktur dari sejumlah neuron yang saling terhubung oleh bobot-bobot penghubung yang dianggap lebih unggul dibandingkan metode *machine learning* classic [8]. MLP terdiri dari tiga lapisan utama: lapisan *input*, lapisan tersembunyi (*hidden layer*), dan lapisan *output*. Lapisan *input* bertugas menerima sinyal masukan dari luar dan menyebarkannya ke seluruh neuron di lapisan tersembunyi. Selanjutnya, lapisan *output* akan mengolah dan menampilkan hasil akhir dari proses jaringan ini. Konsep dasar dari MLP, yang meniru sistem syaraf makhluk hidup, memberikan kemampuan kepada jaringan syaraf tiruan ini untuk belajar dan mengenali pola berdasarkan pengetahuan yang didapat melalui proses pelatihan [9]. Dengan kemampuan tersebut, model neural network berbasis Multilayer Perceptron dapat menjadi alat yang sangat berguna dalam menyelesaikan permasalahan kompleks, seperti identifikasi faktor-faktor penyebab *stunting*, sehingga dapat membantu merancang kebijakan yang lebih efektif dan tepat sasaran.

II. METODE

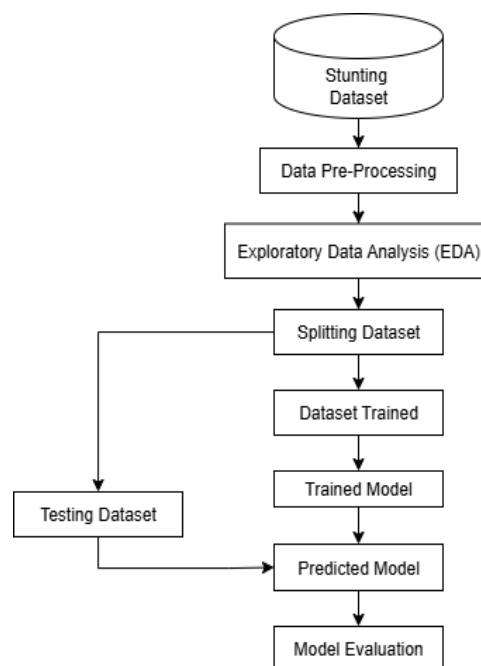
A. Tahapan Penelitian

Proses penelitian mencakup serangkaian langkah penelitian yang tertera pada Gambar 1.

B. Pengumpulan Data

Sumber data yang digunakan didapatkan dari Badan Pangan Nasional yang mencakup 514 kabupaten dan kota di seluruh Indonesia. Dataset yang digunakan mencakup 9 deskriptor utama, yaitu variabel-variabel yang berhubungan dengan kondisi sosial dan ekonomi daerah. Deskriptor-deskriptor ini akan menjadi input dalam model untuk membantu mengidentifikasi pola dan faktor-faktor yang memengaruhi persentase *stunting* di berbagai wilayah di Indonesia. Data yang digunakan mencakup berbagai aspek seperti konsumsi pangan, akses terhadap pelayanan dasar, dan tingkat pendidikan, yang semuanya relevan dalam memprediksi prevalensi *stunting* di suatu wilayah. Data ini diambil dari situs web resmi Badan Pangan Nasional di

<https://fsva.badanpangan.go.id/> dengan fokus pada data tahun 2024.



Gambar 1. Diagram Alir Penelitian

C. Variabel Riset

Variabel Dependen: Dalam penelitian ini, variabel yang ingin diprediksi ialah persentase *Stunting*. Wilayah yang memiliki tingkat *stunting* rendah mencerminkan proporsi anak kecil yang memiliki tinggi badan di bawah normalnya. *Stunting* mengacu pada kondisi gangguan pertumbuhan pada balita yang ditandai dengan tinggi badan yang lebih rendah dibandingkan dengan standar usia mereka. Dengan kata lain, *stunting* menggambarkan ketidaksesuaian pertumbuhan fisik anak yang disebabkan oleh kekurangan gizi kronis, infeksi berulang, atau faktor lingkungan lainnya yang menghambat perkembangan optimal. Secara umum, *stunting* tidak hanya berhubungan dengan masalah gizi pada masa bayi dan balita, tetapi juga mencerminkan faktor sosial ekonomi, akses terhadap perawatan kesehatan, serta kualitas lingkungan hidup yang mempengaruhi tumbuh kembang anak. Oleh karena itu, dalam penelitian ini, tingkat *stunting* menjadi variabel penting yang akan dianalisis untuk mendapatkan pemahaman lebih dalam mengenai faktor-faktor yang mempengaruhinya serta intervensi yang dapat dilakukan untuk menurunkannya.

Variabel Independen: Deskriptor-deskriptor yang digunakan untuk memengaruhi variabel dependen, dalam hal ini *stunting*. Berikut adalah penjelasan mengenai setiap variabel independen yang diterapkan dalam studi ini pada Tabel 1.

TABEL 1
VARIABEL INDEPENDEN

Kode	Variabel	Definisi
X1	NCPR (Normative Consumption to Net Production Ratio)	Rasio konsumsi normatif per kapita terhadap produksi bersih komoditas padi, jagung, ubi kayu, ubi jalar, dan sagu, serta stok beras pemerintah daerah.
X2	Kemiskinan (%)	Persentase rumah tangga kemiskinan dibawah rata – rata.
X3	Pengeluaran Pangan (%)	Persentase rumah tangga dengan proporsi pengeluaran untuk pangan lebih dari 65% terhadap total pengeluaran.
X4	Tanpa Listrik (%)	Persentase rumah tangga tanpa akses listrik.
X5	Tanpa Air Bersih (%)	Persentase rumah tangga tanpa akses ke air bersih.
X6	Lama Sekolah Perempuan (tahun)	Rata-rata lama sekolah perempuan di atas 15 tahun.
X7	Rasio Tenaga Kesehatan	Rasio jumlah penduduk per tenaga kesehatan terhadap tingkat kepadatan penduduk
X8	Angka Harapan Hidup (tahun)	Angka harapan hidup pada saat lahir.
X9	IKP (Indeks Ketahanan Pangan)	Ukuran tingkat ketahanan pangan daerah.

D. Pengolahan dan Analisis Data

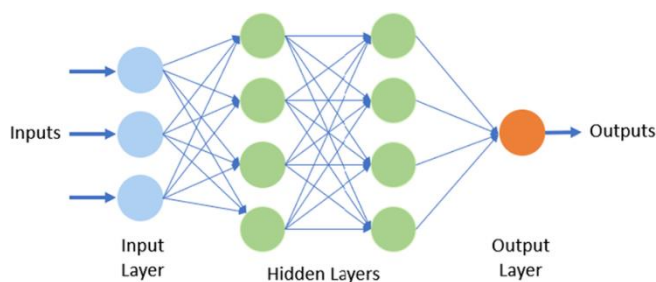
1) *Data Pre-Processing*: Langkah pertama dalam membangun model *machine learning* adalah data *pre-processing*, sebagai tahap krusial untuk memastikan data yang digunakan dalam model sudah siap untuk diproses. Proses ini melibatkan beberapa tahapan, salah satunya adalah pembersihan data, yang berfokus pada penanganan *outliers* atau nilai-nilai ekstrim yang dapat memengaruhi akurasi dan performa model. *Outlier* seringkali disebabkan oleh kesalahan sensor, ketidakpastian model, perubahan lingkungan, kehilangan data, atau serangan siber [10]. Oleh karena itu, langkah ini sangat penting untuk menjaga integritas data agar model dapat belajar secara efektif. Selain itu, tahap *pre-processing* juga mencakup normalisasi dan standarisasi data, yang bertujuan untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang seragam. Normalisasi umumnya dilakukan untuk mengubah nilai fitur dalam rentang tertentu, seperti [0, 1], sementara standarisasi mengubah data agar memiliki distribusi dengan rata-rata 0 dan deviasi standar 1. Tanpa proses ini, fitur dengan skala yang lebih besar bisa mendominasi proses pembelajaran, sementara fitur dengan rentang nilai yang lebih kecil tidak akan diperhatikan secara maksimal. Dengan melakukan normalisasi atau standarisasi, model dapat mempelajari data

secara adil, sehingga menghasilkan performa yang lebih optimal dan akurat.

2) *Exploratory Data Analysis (EDA)*: Tahapan dalam analisis data yang bertujuan untuk menggali dan memahami struktur dataset secara mendalam. Proses ini mencakup analisis statistik dan visualisasi data untuk menggali berbagai pola, hubungan, serta distribusi data yang mungkin tersembunyi. EDA memungkinkan analisis mendalam terhadap data sebelum model *machine learning* dibangun. Salah satu tujuan utama dari EDA adalah untuk mengidentifikasi distribusi yang tidak merata pada data. Pada dataset genomik skala besar, variabel prediktor seringkali menunjukkan distribusi yang tidak normal, misalnya sangat miring (*skewed*), yang dapat menurunkan kinerja model *machine learning*. Oleh karena itu, transformasi data seperti logaritmik atau Box-Cox sering diperlukan untuk menormalkan distribusi tersebut. Selain itu, EDA juga digunakan untuk mengeksplorasi hubungan antar fitur dalam dataset. Proses ini membantu dalam memahami bagaimana masing-masing fitur (variabel) saling berinteraksi satu sama lain. Dalam EDA, kita dapat melihat apakah ada fitur yang sangat terkait atau saling mempengaruhi, yang mungkin dapat membantu model dalam memprediksi variabel target dengan lebih akurat. Selain hubungan antar fitur, EDA juga bertujuan untuk mengidentifikasi korelasi yang tinggi antara fitur dan variabel target [11][12]. Korelasi yang tinggi antara fitur dan target menandakan bahwa fitur tersebut sangat relevan untuk digunakan dalam model, sedangkan fitur dengan korelasi rendah mungkin bisa diabaikan [13]. Selain itu, EDA juga membantu untuk mengidentifikasi redundansi data, yaitu fitur-fitur yang memiliki informasi yang hampir sama atau sangat mirip, sehingga dapat dihapus untuk mengurangi kompleksitas model. Secara keseluruhan, EDA berfungsi untuk memberikan wawasan yang lebih jelas mengenai data yang dimiliki, membantu dalam pemilihan fitur yang relevan, dan mengurangi redundansi data. Hal ini pada akhirnya akan meningkatkan kualitas model yang dibangun dan mengoptimalkan hasil analisis.

3) *Splitting Data*: Prosedur pembagian data dilakukan di mana 80% dari keseluruhan data dimanfaatkan untuk fase pelatihan dan 20% yang tersisa digunakan untuk menguji model. Pembagian ini penting untuk memastikan model dapat dilatih dengan data yang cukup dan diuji pada data yang tidak terlihat sebelumnya, sehingga dapat mengevaluasi kemampuannya untuk generalisasi. Rasio ini dipilih karena menawarkan komposisi yang optimal antara volume data yang memadai untuk melatih model dan volume data yang cukup untuk mengevaluasi performanya. Hal ini sesuai dengan studi yang menunjukkan bahwa pemisahan data yang benar, seperti 80% untuk pelatihan dan 20% untuk pengujian, sangat mempengaruhi kualitas dan keandalan model yang dikembangkan [14]

4) *Model Prediksi*: Dalam penelitian ini, algoritma MLP Regressor digunakan untuk memprediksi persentase *stunting*, di mana regresi adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen yang bersifat kontinu, seperti halnya prediksi nilai persentase *stunting* dalam model ini. MLP merupakan model jaringan syaraf tiruan dengan arsitektur *feedforward* yang terdiri dari kumpulan neuron yang saling terhubung melalui bobot-bobot yang menghubungkan mereka [8]. Algoritma ini berbasis Neural Network yang terdiri dari *input layer*, *hidden layer* dan *output layer*. Lapisan *input* berperan untuk menerima sinyal masukan dari luar dan kemudian mendistribusikannya ke seluruh neuron di lapisan tersembunyi. Sementara itu, lapisan *output* akan menerima sinyal hasil dari lapisan tersembunyi dan menghasilkan *output* dari jaringan secara keseluruhan. JST MLP memiliki kemampuan untuk mengidentifikasi dan memahami pola berdasarkan pengetahuan yang diperoleh melalui proses pelatihan [15][16]. Model ini dibangun menggunakan *library* Python scikit-learn untuk pengembangan jaringan syaraf tiruan. Pada penelitian ini, arsitektur yang digunakan terdiri dari dua lapisan tersembunyi (*hidden layers*), dengan jumlah neuron yang berbeda di masing-masing lapisan. Lapisan pertama memiliki 13 neuron, sedangkan lapisan kedua memiliki 27 neuron. Fungsi aktivasi yang digunakan pada setiap lapisan tersembunyi adalah ReLU (Rectified Linear Unit), yang memungkinkan model untuk menangkap hubungan non-linier dalam data. Selain itu, model ini menggunakan L-BFGS sebagai algoritma optimasi untuk meminimalkan fungsi kerugian, dengan kecepatan pembelajaran yang adaptif untuk meningkatkan konvergensi selama pelatihan. Model ini dilatih dengan parameter *max_iter*=1000, yang berarti model akan melakukan hingga 1000 iterasi selama proses pelatihan, dengan tujuan untuk memastikan konvergensi yang optimal. *Random_state*=42 digunakan untuk memastikan bahwa eksperimen ini dapat direproduksi dengan hasil yang konsisten.



Gambar 2. Arsitektur MLP dengan 2 *hidden layer*

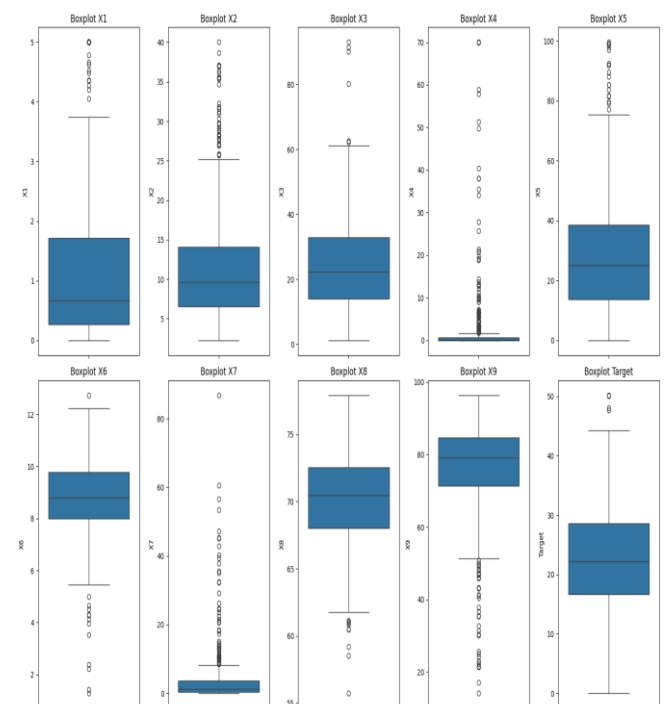
5) *Model Evaluation*: Hasil prediksi dalam penelitian ini akan dievaluasi menggunakan beberapa matriks evaluasi yang umum dalam analisis regresi, yaitu RMSE (*Root Mean Squared Error*), MAE (Mean Absolute Error), dan R^2 (*R Square*). RMSE mengukur rata-rata kesalahan prediksi

model dengan memberikan bobot lebih besar pada kesalahan yang lebih besar, sehingga lebih sensitif terhadap prediksi yang jauh dari nilai sebenarnya dan memberikan penalti lebih tinggi untuk kesalahan besar. MAE menghitung rata-rata perbedaan mutlak antara estimasi dan nilai yang sesungguhnya, tanpa memberikan bobot tambahan pada kesalahan besar, yang menjadikannya metrik yang lebih sederhana untuk menilai akurasi model secara keseluruhan. Sedangkan, R^2 mengukur sejauh mana model dapat menjelaskan variasi dalam data, dengan nilai mendekati 1 menunjukkan model yang mampu menjelaskan sebagian besar variasi data dengan baik. Nilai R^2 yang rendah menunjukkan bahwa model kurang mampu menjelaskan variasi data, yang mengindikasikan adanya faktor-faktor lain yang belum diperhitungkan dalam model yang dapat meningkatkan akurasi prediksi.

III. HASIL DAN PEMBAHASAN

A. Hasil EDA

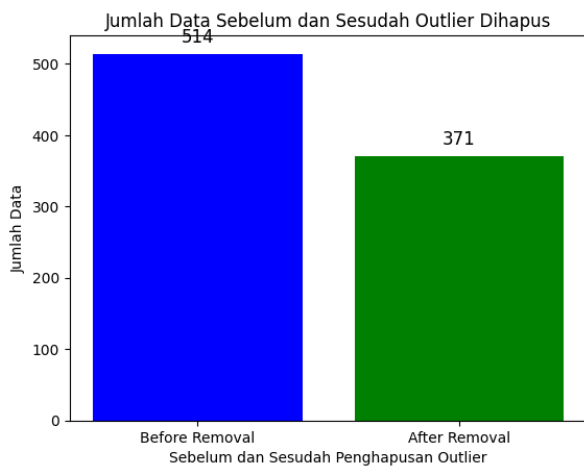
Subbagian ini melakukan analisis statistik deskriptif untuk memahami kondisi sosial wilayah yang tertera pada Gambar 3.



Gambar 3. Diagram hasil EDA

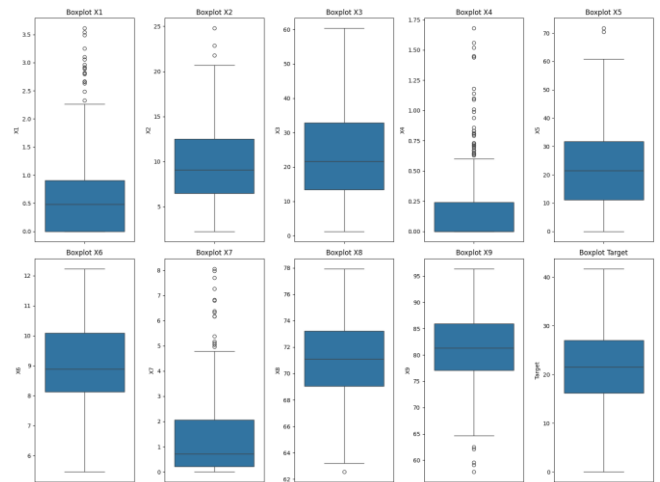
Berdasarkan visualisasi boxplot yang ditampilkan menunjukkan pada distribusi data dari berbagai variabel sosial-ekonomi dan kesehatan terdapat *outlier* yang dapat mempengaruhi hasil model yang dibangun. Pada variabel kemiskinan (X2), pengeluaran pangan (X3), dan akses dasar

(X4, X5), mencerminkan wilayah dengan masalah ekonomi dan infrastruktur yang memerlukan perhatian khusus. Angka harapan hidup (X8) dan Indeks Ketahanan Pangan (X9) menunjukkan wilayah dengan kondisi yang lebih baik. Disisi lain dengan tingkat *outlier* rendah, seperti NCPR (X1), rasio tenaga kesehatan (X7), dan lama sekolah perempuan (X6) menunjukkan masalah seperti ketidakseimbangan sumber daya, kekurangan tenaga kesehatan, dan rendahnya akses pendidikan, yang membutuhkan intervensi kebijakan yang tepat. Berikut diagram perbandingan jumlah data sebelum dan sesudah dihapus pada Gambar 4.



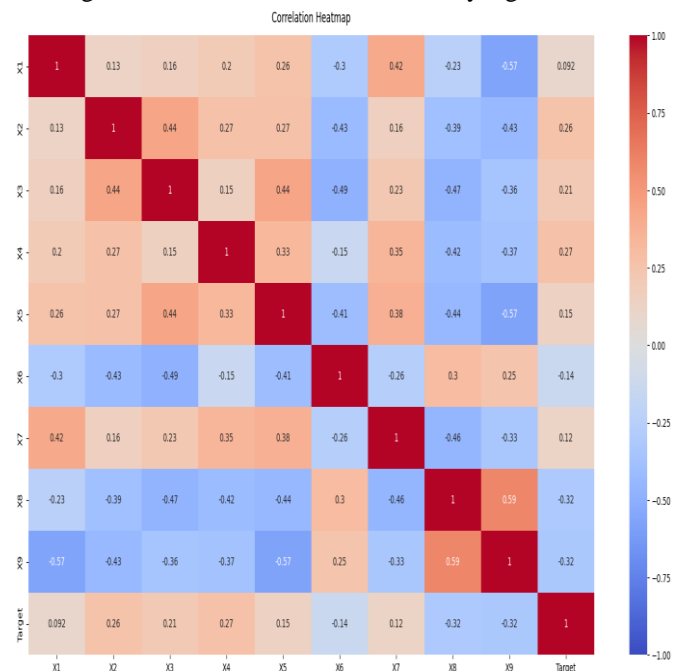
Gambar 4. Perbandingan jumlah data sebelum dan sesudah *outliers* dihapus

Dengan *outlier* yang sangat signifikan, dilakukan penanganan *outlier* untuk mendapatkan model yang cukup optimal sehingga representasi keadaan yang sebenarnya menjadi lebih akurat. Penanganan dilakukan berupa penghapusan *outlier* yang dimana jumlah data mencapai 514 data. Setelah dilakukan penanganan yang dianggap sebagai *outlier*, jumlah data berkurang menjadi 371 data. Hal ini menunjukkan bahwa dengan menghapus *outlier*, data yang tersisa lebih representatif dan dapat memberikan hasil analisis yang lebih akurat. Penghapusan *outlier* dilakukan untuk mengurangi dampak data yang ekstrim yang bisa mengganggu analisis statistik dan model yang lebih stabil.



Gambar 5. Diagram data setelah *outliers* dihapus

Gambar 5 menunjukkan perbedaan yang signifikan setelah penanganan *outlier* pada distribusi data. Jumlah *outlier* pada sebagian besar variabel mengalami pengurangan seperti pada X1, X2, X4, X5, dan X7. Hal ini meningkatkan konsistensi data, membuat distribusi lebih terpusat pada rentang interkuartil (IQR) dan mengurangi pengaruh nilai *outlier*. Variabel seperti X3 dan X9 menunjukkan distribusi yang lebih stabil. Penghapusan *outlier* juga membuat data lebih representatif, memungkinkan analisis yang lebih akurat tanpa gangguan dari data yang tidak wajar, terutama pada Target (*Stunting*). Secara keseluruhan, penanganan *outlier* meningkatkan kualitas data untuk analisis yang lebih valid.



Gambar 6. Visualisasi *heatmap*

Berdasarkan visualisasi heatmap pada Gambar 6, korelasi antara variabel-variabel independen dengan variabel target (*stunting*), diperoleh beberapa temuan penting sebagai berikut.

1) *Variabel dengan Korelasi Kuat terhadap Stunting*: Indeks Ketahanan Pangan (X9) menunjukkan korelasi negatif sebesar -0,32 dengan persentase stunting, yang menandakan adanya hubungan negatif yang cukup kuat antara ketahanan pangan dan persentase stunting. Hal ini berarti bahwa wilayah dengan ketahanan pangan yang lebih baik (IKP tinggi) cenderung memiliki persentase stunting yang lebih rendah, kemungkinan besar karena ketahanan pangan yang baik mendukung ketersediaan gizi yang memadai bagi anak-anak. Sementara itu, Angka Harapan Hidup (X8) juga menunjukkan korelasi negatif sebesar -0,32, yang berarti bahwa semakin tinggi angka harapan hidup suatu wilayah, semakin rendah persentase stunting. Hal ini mencerminkan bahwa peningkatan kualitas hidup, yang berkaitan dengan akses yang lebih baik terhadap layanan kesehatan, gizi, dan perawatan, berkontribusi pada penurunan angka stunting.

2) *Variabel dengan Korelasi Sedang terhadap Stunting*: Tingkat Kemiskinan (X2) memiliki korelasi positif sebesar 0,26 dengan angka stunting, yang menunjukkan semakin besar tingkat kemiskinan, semakin tinggi juga persentase stunting. Kemiskinan berperan sebagai faktor yang memengaruhi persentase stunting akibat keterbatasan akses terhadap makanan bergizi, layanan kesehatan, serta kebutuhan dasar lainnya. Sementara itu, Akses terhadap Air Bersih (X5) menunjukkan korelasi positif sebesar 0,15, yang berarti wilayah dengan proporsi penduduk tanpa akses air bersih cenderung memiliki tingkat stunting yang lebih tinggi. Meskipun akses terhadap air bersih penting, faktor-faktor lain seperti sanitasi yang buruk dan tingginya angka penyakit infeksi juga berperan dalam memengaruhi tingkat stunting secara keseluruhan.

3) *Variabel dengan Korelasi Rendah terhadap Stunting*: Rata-rata Lama Sekolah Perempuan (X6) memiliki korelasi negatif sebesar -0,14 terhadap angka stunting, yang menunjukkan pengaruh yang lemah. Meskipun pendidikan perempuan sering dikaitkan dengan peningkatan pengetahuan gizi dan perawatan anak, dalam konteks ini pengaruhnya terhadap stunting tampak terbatas, kemungkinan karena faktor ekonomi dan akses layanan kesehatan yang lebih dominan. Sementara itu, Pengeluaran Pangan (X3) dan Persentase Rumah Tangga Tanpa Akses Listrik (X4) menunjukkan korelasi yang sangat lemah terhadap tingkat stunting. Hal ini mengindikasikan bahwa meskipun pengeluaran pangan dan akses listrik memengaruhi kualitas hidup secara umum, keduanya tidak memiliki hubungan yang cukup kuat untuk dijadikan indikator utama terhadap persentase stunting.

B. Hasil Evaluasi Model

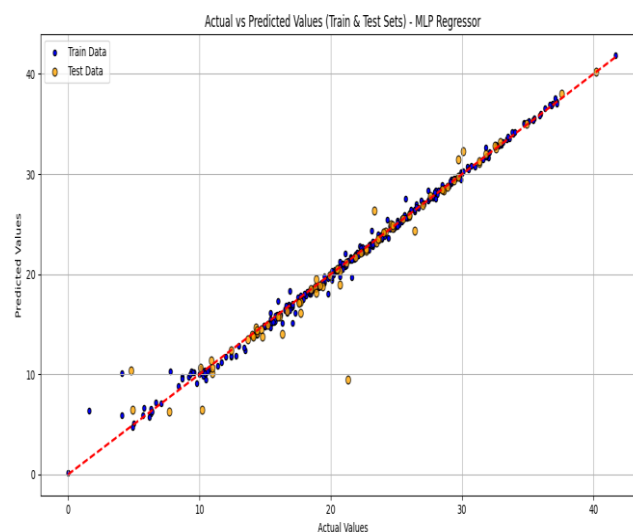
Bagian ini membahas pengembangan model Regresi untuk mengidentifikasi persentase *stunting* dengan mengevaluasi algoritma MLP Regresor.

TABEL 2

HASIL EVALUASI MODEL MLP REGRESSOR

Metrik Performa	Performa Model
RMSE	0.7231
MAE	3.0313
R2	0.9463

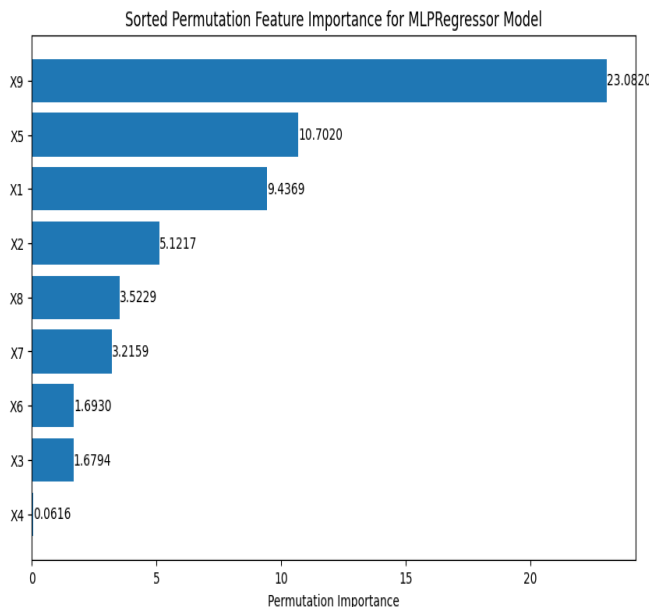
Berdasarkan hasil evaluasi model pada Tabel II yang menggunakan metrik performa yang umum, yaitu RMSE, MAE, dan R², dapat disimpulkan bahwa model ini memberikan kinerja maksimal. Nilai RMSE (0,7231) dan MAE yang bernilai 3,0313 yang mengindikasikan bahwa model mampu meminimalkan perbedaan antara nilai prediksi dan nilai aktual serta memberikan prediksi yang akurat dengan selisih yang terbatas dari nilai aktual. Selain itu, nilai R² sebesar 0,9463 menunjukkan bahwa model sangat efektif dalam menggambarkan hubungan antara variabel *input* dan *output*, dengan hanya sedikit variasi yang tidak dapat dijelaskan. Secara umum, hasil evaluasi mengindikasikan bahwa model yang digunakan memiliki kemampuan prediksi yang sangat baik, dengan tingkat akurasi tinggi, dan dapat diandalkan untuk analisis lebih lanjut.



Gambar 7. Perbandingan nilai prediksi dan nilai aktual model

Gambar 7. menampilkan grafik yang membandingkan antara nilai prediksi dan nilai aktual model diperlihatkan dengan jelas. Dari grafik ini, dapat dilihat bahwa sebagian besar titik prediksi berada sangat dekat dengan garis ideal, menunjukkan bahwa model mampu memberikan hasil prediksi yang akurat dan sesuai dengan nilai aktual yang

diobservasi. Namun demikian, terdapat beberapa titik yang agak terpisah dari garis ideal, terutama pada rentang nilai aktual yang lebih rendah. Hal ini menunjukkan adanya ketidaktepatan yang lebih besar pada prediksi untuk nilai-nilai kecil, meskipun untuk nilai yang lebih tinggi model menunjukkan performa yang sangat baik dan akurat. Secara keseluruhan, hasil yang ditunjukkan oleh grafik ini mengindikasikan bahwa model ini secara umum memberikan hasil prediksi yang baik, dengan sedikit kesalahan pada nilai rendah.



Gambar 8. *Feature importance* pada tiap fitur

Ditemukannya peran kuat Indeks Ketahanan Pangan (IKP) terhadap *stunting* menunjukkan bahwa program intervensi berbasis ketahanan pangan lokal, dapat memberikan dampak yang lebih besar dalam mengurangi persentase *stunting*. Penjelasan mengenai *Feature importance* pada Gambar 8 yang menggambarkan kontribusi masing-masing variabel dalam model prediksi *stunting* sebagai berikut:

1) *Variabel dengan Feature importance Kuat*: Variabel X9 (Indeks Ketahanan Pangan - IKP) memiliki kontribusi fitur yang paling tinggi sebesar 23.0820, yang menunjukkan bahwa variabel ini sangat berpengaruh terhadap kinerja model. Ketahanan pangan yang lebih baik berhubungan dengan tingkat *stunting* yang lebih rendah, sehingga ketahanan pangan menjadi aspek yang sangat signifikan terhadap *stunting* yang secara langsung mempengaruhi kesehatan dan status gizi, khususnya pada anak-anak yang rentan terhadap gangguan pertumbuhan. Kebijakan yang mendukung ketahanan pangan lokal, seperti Program Pemberdayaan Petani atau Distribusi Pangan Bergizi, dapat membantu meningkatkan akses masyarakat terhadap pangan yang bergizi, yang berkontribusi pada pencegahan *stunting*. Dalam literatur, ketahanan pangan

rumah tangga telah terbukti mengurangi prevalensi malnutrisi dan *stunting* pada anak [17].

2) *Variabel dengan Feature importance Sedang*: Fitur X5 (Tanpa Air Bersih) dan X1 (NCPR) menunjukkan kontribusi fitur sedang, masing-masing sebesar 10.7020 dan 9.4369. Keterbatasan akses terhadap air bersih meningkatkan risiko penyakit yang berhubungan dengan sanitasi, yang dapat berkontribusi pada angka *stunting* yang lebih tinggi. Penurunan akses terhadap air bersih meningkatkan beban penyakit terkait sanitasi, seperti diare, yang dapat mengganggu status gizi anak dan meningkatkan risiko *stunting*. Hal ini sejalan dengan kebijakan kesehatan seperti Program Sanitasi Total Berbasis Masyarakat (STBM) yang bertujuan untuk meningkatkan akses masyarakat terhadap air bersih dan sanitasi. Selain itu, NCPR (National Consumer Price Ratio) yang tidak seimbang dapat mengindikasikan ketidakstabilan dalam pasokan pangan, yang dapat mempengaruhi pola makan dan status gizi, serta berdampak pada tingkat *stunting* [18].

3) *Variabel dengan Feature importance Rendah*: X2 (Kemiskinan), X8 (Angka Harapan Hidup), dan X7 (Rasio Tenaga Kesehatan) memiliki kontribusi rendah terhadap model, meskipun kemiskinan berhubungan dengan akses terbatas terhadap gizi dan kesehatan. Dalam konteks kebijakan, Program Keluarga Harapan (PKH) yang diberikan kepada keluarga miskin dapat membantu meningkatkan akses mereka terhadap layanan kesehatan dan gizi yang baik, yang berpotensi menurunkan angka *stunting*. Namun, kontribusinya terhadap model prediksi *stunting* lebih kecil dibandingkan dengan variabel yang berkaitan langsung dengan ketahanan pangan dan sanitasi. Selain itu, meskipun Angka Harapan Hidup dan Rasio Tenaga Kesehatan memberikan informasi penting tentang kualitas hidup, pengaruhnya terhadap *stunting* lebih kecil dibandingkan variabel yang berkaitan langsung dengan pangan dan sanitasi, sesuai dengan temuan dalam Studi Gizi dan Kesehatan Global yang menunjukkan bahwa faktor-faktor yang berhubungan langsung dengan gizi memiliki dampak yang lebih besar terhadap *stunting* daripada faktor sosial ekonomi lainnya [19].

X6 (Lama Sekolah Perempuan), X3 (Pengeluaran Pangan), dan X4 (Tanpa Listrik) memiliki importance yang sangat rendah. Meskipun pendidikan perempuan (X6) dapat mempengaruhi status gizi, pengaruhnya terhadap *stunting* tidak sebesar faktor lain yang lebih langsung berhubungan dengan akses pangan dan sanitasi. Pengeluaran Pangan (X3) juga memberikan kontribusi kecil terhadap prediksi *stunting* dalam model ini. Hal ini menunjukkan bahwa meskipun pendidikan dan pengeluaran rumah tangga terkait pangan memiliki peran, faktor-faktor langsung seperti ketahanan pangan dan sanitasi memiliki kontribusi yang lebih besar. Kebijakan yang meningkatkan Akses Energi untuk Rumah Tangga (X4) dapat berkontribusi pada peningkatan kualitas hidup, namun pengaruhnya terhadap *stunting* lebih kecil

dibandingkan kebijakan yang mendukung ketahanan pangan dan akses sanitasi.

IV. KESIMPULAN

Berdasarkan analisis dari penelitian yang telah kami lakukan, dapat disimpulkan bahwa hal-hal berikut ini:

1) *Model Multi-Layer Perceptron (MLP)*: MLP memberikan performa yang memuaskan dalam memprediksi tingkat risiko stunting yang menunjukkan evaluasi matrik RMSE, MAE, dan R^2 masing-masing sebesar 0.7231, 3.0313 dan 0.9463 pada data testing yang mengindikasikan bahwa model berhasil memprediksi tepat sebagian besar data yang diuji.

2) *Faktor Utama dan Pendukung*: Faktor utama pada fitur X9 sebagai IKP (Indeks Ketahanan Pangan) yang menunjukkan bahwa ketahanan pangan memiliki peran krusial dalam identifikasi presentase stunting. Dan diikuti oleh X5 (Tanpa Air Bersih) dan NCPR (Normative Consumption to Net Production Ratio) sebagai X1.

Berdasarkan temuan ini, disarankan untuk mengembangkan model regresi tingkat *stunting* lebih lanjut dengan menambahkan fitur baru atau algoritma alternatif guna meningkatkan evaluasi matriks. Salah satu pengembangan yang dapat dilakukan adalah penambahan variabel spasial, seperti curah hujan dan jarak ke fasilitas kesehatan, yang dapat diuji untuk memperkaya informasi yang memengaruhi tingkat *stunting*. Selain itu, optimasi fitur kunci dalam prediksi *stunting*, khususnya yang memiliki *Feature importance* tinggi, perlu dilakukan untuk menghasilkan informasi yang lebih relevan dan akurat.

Selain pengembangan teknis, kolaborasi yang lebih kuat antara pemerintah, lembaga swadaya masyarakat, dan komunitas lokal sangat diperlukan. Sinergi ini akan memperkuat implementasi program pembangunan yang berfokus pada keberlanjutan dan partisipasi aktif masyarakat, sehingga dapat menanggulangi *stunting* secara lebih efektif. Dengan pendekatan yang lebih inklusif dan berbasis data, diharapkan upaya pengurangan prevalensi *stunting* di Indonesia dapat dilakukan dengan lebih tepat sasaran dan berkelanjutan.

DAFTAR PUSTAKA

- [1] M. Soffiudin, W. Putri, R. Amanda, D. Pratiwi, dan F. Makrufardi, "Community Based Stunting Prevention: Learning from Blue Collar Workers' Children in Indonesia," 2025.
- [2] I. C. A. Soleha dan R. Riya, "Faktor – Faktor yang berhubungan dengan Kejadian Stunting pada Balita: Literature Review," *Jurnal Akademika Baiturrahim Jambi*, vol. 13, no. 1, hlm. 158–167, Mar 2024, doi: 10.36565/jab.v13i1.821.
- [3] B. Astria Paramashanti, H. Hadi, dan I. Made Alit Gunawan, "Pemberian ASI eksklusif tidak berhubungan dengan stunting pada anak usia 6-23 bulan di Indonesia Exclusive breastfeeding practice was not related with stunting in young children 6-23 months in Indonesia," 2016.
- [4] Ishartono dan Raharjo, "Sustainable Development Goals (SDGs) dan Pengentasan Kemiskinan," 2016, [Daring]. Tersedia pada: <http://www.bappenas.go.id/id/berita-dan-siaran>
- [5] M. S. Haris, A. N. Khudori, dan W. T. Kusuma, "Perbandingan Metode Supervised Machine Learning Untuk Prediksi Prevalensi Stunting Di Provinsi Jawa Timur," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 2022, doi: 10.25126/jtiik.202296744.
- [6] I. P. Putri, Terttiaavini, dan N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, hlm. 257–265, Jan 2024, doi: 10.57152/malcom.v4i1.1078.
- [7] A. Y. Perdana, R. Latuconsina, dan A. Dinimaharawati, "Prediksi Stunting pada Balita dengan Algoritma Random Forest," 2021.
- [8] D. Sinaga, "Jaringan Saraf Tiruan Infeksi Mata Dengan Menggunakan Metode Beraksitektur Multi Layer Perceptron," 2020.
- [9] T. Bikkur, "Multi-layered deep learning perceptron approach for health risk prediction," *J Big Data*, vol. 7, no. 1, Des 2020, doi: 10.1186/s40537-020-00316-7.
- [10] H. Fang, M. A. Haile, dan Y. Wang, "Robust Extended Kalman Filtering for Systems with Measurement Outliers," Mar 2019, [Daring]. Tersedia pada: <http://arxiv.org/abs/1904.00335>
- [11] S. Dhumad, "The Imperative of Exploratory Data Analysis in Machine Learning," *Scholars Journal of Engineering and Technology*, vol. 13, no. 01, hlm. 30–44, Jan 2025, doi: 10.36347/sjet.2025.v13i01.005.
- [12] A. Gnanavelu, C. Venkataramu, dan R. Chintakunta, "Cardiovascular Disease Prediction Using Machine Learning Metrics," *Journal of Young Pharmacists*, vol. 17, no. 1, hlm. 226–233, Jan 2025, doi: 10.5530/jyp.20251231.
- [13] Z. Noroozi, A. Orooji, dan L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," *Sci Rep*, vol. 13, no. 1, Des 2023, doi: 10.1038/s41598-023-49962-w.
- [14] M. K. Uçar, M. Nour, H. Sindi, dan K. Polat, "The Effect of Training and Testing Process on Machine Learning in Biomedical Datasets," *Math Probl Eng*, vol. 2020, 2020, doi: 10.1155/2020/2836236.
- [15] P. Lingga, "Implementasi Diagnosa Penyakit Panleukopenia Pada Kucing Dengan Menggunakan Jaringan Saraf Tiruan Multi Layer Perceptron (Studi Kasus:Clinic Sasmita Pet Shop)," 2020.
- [16] M. Yanto, F. Hadi, dan S. Arlis, "Optimization of Machine Learning Classification Analysis of Malnutrition Cases in Children," *Jurnal RESTI*, vol. 7, no. 6, hlm. 1378–1386, Des 2023, doi: 10.29207/resti.v7i6.5278.
- [17] Najwa Zahira Shofa dkk., "Food Security on The Incidence of Stunting in Agricultural Areas," *Nursing and Health Sciences Journal (NHSJ)*, vol. 4, no. 4, hlm. 451–458, Des 2024, doi: 10.53713/nhsj.v4i4.432.
- [18] S. N. Munthe, R. M. Silalahi, K. C. Pertiwi, V. Y. Permasari, dan H. Andriani, "Kajian Literatur: Pengaruh Konsumsi Air Bersih Terhadap Stunting Pada Anak yang Berasal dari Keluarga Berpendapatan Rendah," *MAHESA : Malahayati Health Student Journal*, vol. 4, no. 2, hlm. 566–580, Feb 2024, doi: 10.33024/mahesa.v4i2.13301.
- [19] N. M. Sahara, D. Arianto, dan M. Santoso, "Dampak Program Bantuan Sosial sebagai Intervensi Gizi Sensitif terhadap Status Stunting pada Balita," *Jurnal Ekonomi Kesehatan Indonesia*, vol. 9, no. 1, Jun 2024, doi: 10.7454/eki.v9i1.1034.