

# ANALISIS REGRESI LOGISTIK DENGAN SELEKSI FITUR DAN VALIDASI SILANG DALAM MENENTUKAN MODEL PREDIKSI STATUS GIZI ANAK

Dwi Wulandari Sari<sup>1</sup>, Kurnia Gusti Ayu<sup>2</sup>, Hariesa Budi Prabowo<sup>3</sup>

1,2,3) Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana, Indonesia

## Article Info

### Article history:

Received: 22 Juni 2025

Revised: 07 Juli 2025

Accepted: 16 Juli 2025

## ABSTRACT

### Abstrak

Dampak krisis ekonomi pascapandemi COVID-19 dan konflik Rusia-Ukraina telah memengaruhi stabilitas pangan di Indonesia, khususnya asupan gizi anak sekolah. Untuk mengatasi tantangan ini, penelitian ini mengembangkan model prediktif berbasis regresi logistik multivariat. Model ini bertujuan mengidentifikasi status gizi anak (normal, gizi buruk, atau obesitas) menggunakan data kuesioner dari orang tua anak usia 6–15 tahun sebanyak 308 data. Dalam metodologinya, sampel penelitian diperoleh melalui Stratified Random Sampling. Pendekatan ini memastikan representasi yang proporsional dari setiap jenjang pendidikan (SD, SMP, SMA) di wilayah studi. Kuesioner yang digunakan telah melalui uji validitas, termasuk Validitas Konvergen, untuk memastikan bahwa item-item seperti pola makan dan aktivitas fisik benar-benar berkorelasi positif dengan indikator kesehatan gizi yang relevan. Penting dicatat, data status gizi anak dikategorikan menggunakan Skala Nominal (normal, gizi buruk, atau obesitas), yang menjadi dasar untuk analisis model. Variabel yang dianalisis mencakup pola makan, aktivitas fisik, kebiasaan ngemil, serta konsumsi *fast food* dan minuman manis. Hasil analisis korelasi Pearson menunjukkan pola makan seimbang (kaya sayur, buah, dan protein) berkorelasi positif dengan status gizi normal. Sebaliknya, konsumsi *fast food* dan minuman manis berkorelasi negatif dengan kesehatan gizi. Kebiasaan sarapan dan aktivitas fisik juga terbukti memiliki pengaruh positif. Meskipun demikian, model prediksi yang dibangun menunjukkan akurasi terbatas. Spesifitasnya tinggi untuk kasus gizi normal (0,92), namun sensitivitasnya rendah (0,38) dalam mendeteksi gizi buruk/obesitas. Ini berarti model belum cukup andal untuk aplikasi medis kritis, tetapi dapat berfungsi sebagai alat skoring awal yang berguna. Temuan ini mengarah pada tiga rekomendasi praktis: (1) pengembangan sistem deteksi dini masalah gizi berbasis skor sederhana, (2) intervensi gizi terpersonalisasi (seperti mengurangi *fast food* dan meningkatkan aktivitas fisik), dan (3) penyempurnaan program Makan Bergizi Gratis agar lebih tepat sasaran. Meskipun masih dalam tahap *proof of concept* (TKT 3), penelitian ini menyediakan landasan ilmiah penting bagi kebijakan gizi berbasis data di layanan kesehatan primer, membantu upaya pencegahan malnutrisi anak sekolah agar lebih terarah dan adaptif di tengah dinamika sosial-ekonomi global.

**Kata Kunci:** Uji korelasi (pearson), regresi logistic, multivariant, heatmap, AUC- ROC, Odds

### Abstract

The impact of the economic crisis after the COVID-19 pandemic and the Russia-Ukraine conflict has affected food stability in Indonesia, especially the nutritional intake of school children. To address this challenge, this study developed a multivariate logistic regression-based predictive model. The model aims to identify children's nutritional status (normal, malnourished or obese) using questionnaire data from parents of children aged 6-15 years with 308 record data. In the methodology, the research sample was obtained through stratified random sampling. This approach ensured proportional representation of each education level (elementary, junior high, high school) in the study area. The questionnaire used underwent validity testing, including Convergent Validity, to ensure that items such as diet and physical activity were positively correlated with relevant nutritional health indicators. Importantly, child nutritional status data was categorized using a Nominal Scale (normal, malnourished, or obese), which formed the basis for the model analysis. The variables analyzed included diet, physical activity, snacking habits, and consumption of fast food and sugary drinks. The results of Pearson correlation analysis showed that a balanced diet (rich in vegetables, fruits, and protein) was positively correlated with normal

nutritional status. In contrast, consumption of fast food and sugary drinks was negatively correlated with nutritional health. Breakfast habits and physical activity were also shown to have a positive influence. Nonetheless, the prediction model built showed limited accuracy. The specificity was high for normal nutrition cases (0.92), but the sensitivity was low (0.38) in detecting malnutrition/obesity. This means that the model is not yet reliable enough for critical medical applications, but can serve as a useful initial scoring tool. These findings lead to three practical recommendations: (1) development of a simple score-based early detection system of nutrition problems, (2) personalized nutrition interventions (such as reducing fast food and increasing physical activity), and (3) refinement of the Free School Meal program to make it more targeted. Although still at the proof of concept stage (TKT 3), this study provides an important scientific foundation for data-driven nutrition policy in primary health care, helping school children's malnutrition prevention efforts to be more targeted and adaptive amid global socio-economic dynamics.

**Keywords:** Uji korelasi (pearson), regresi logistic, multivariant, heatmap, AUC- ROC, Odds

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



**Corresponding Author:**

E-mail : [dwi.wulandari@mercubuana.ac.id](mailto:dwi.wulandari@mercubuana.ac.id)

## 1. PENDAHULUAN

Asupan gizi anak merupakan faktor penting yang menentukan status kesehatan serta perkembangan fisik dan kognitif mereka [1]. Kurangnya gizi yang cukup dapat menyebabkan berbagai masalah kesehatan, termasuk gangguan pertumbuhan, penurunan daya tahan tubuh, dan risiko penyakit kronis di masa mendatang [2]. Faktor-faktor seperti pola makan, kebiasaan konsumsi, serta akses terhadap makanan bergizi menjadi penentu utama dalam menciptakan status gizi yang baik bagi anak-anak [3]. Oleh karena itu, pemahaman mengenai pola asupan gizi anak menjadi semakin penting dalam upaya mencegah permasalahan kesehatan yang berkaitan dengan kekurangan atau kelebihan gizi [4].

Anak usia 7-12 tahun membutuhkan nutrisi seimbang untuk mendukung pertumbuhan, perkembangan otak, dan aktivitas fisik. Kebutuhan kalori harian anak-anak Indonesia berdasarkan Kementerian Kesehatan berkisar antara 1.800-2.050 kalori, tergantung pada usia, jenis kelamin, dan tingkat aktivitas.[5]. Pola makan seimbang harus memenuhi kebutuhan anak agar kelak tidak menjadi permasalahan Kesehatan. Komposisi Makanan Seimbang diantaranya Karbohidrat: Sumber energi utama. Protein: Penting untuk pertumbuhan otot dan jaringan. Lemak: Pilih lemak sehat seperti alpukat,

kacang-kacangan, biji-bijian, dan minyak zaitun. Batasi lemak jenuh dan hindari lemak trans. Serat: Bantu pencernaan dan cegah sembelit. Vitamin dan Mineral: Pastikan anak mendapatkan cukup vitamin A, C, D, kalsium, zat besi, dan zinc dari makanan alami [6]. Porsi makan harian juga perlu diperhatikan seperti Buah dan Sayuran: 4-5 porsi per hari (1 porsi = 1 buah ukuran sedang atau 1 mangkuk kecil sayuran). Protein: 2-3 porsi per hari (1 porsi = 1 potong daging ukuran sedang, 1 butir telur, atau 1/2 mangkuk kacang-kacangan). Karbohidrat: 4-6 porsi per hari (1 porsi = 1 potong roti atau 1/2 mangkuk nasi). Susu dan Produk Olahan Susu: 2-3 porsi per hari (1 porsi = 1 gelas susu atau 1 potong keju). Hidrasi yang Cukup Anak perlu minum 6-8 gelas air per hari. Hindari minuman manis seperti soda atau jus kemasan yang tinggi gula. Sarapan penting untuk memberikan energi dan meningkatkan konsentrasi di sekolah. Contoh sarapan sehat: oatmeal dengan buah, roti gandum dengan telur, atau susu dengan sereal gandum. Berikan camilan sehat seperti buah segar, yogurt, kacang-kacangan, atau sayuran rebus. Hindari camilan tinggi gula dan lemak. Dorong anak untuk aktif secara fisik setidaknya 60 menit per hari melalui olahraga, bermain di luar, atau kegiatan fisik lainnya [7].

Salah satu metode yang dapat digunakan untuk menganalisis pola konsumsi dan memprediksi dampaknya terhadap kesehatan anak adalah regresi logistik. Apalagi kondisi ekonomi dunia saat ini semakin memburuk dimulai sejak covid-19 menyerang seluruh dunia tanpa terkecuali Indonesia. Semakin, sulit orang mendapatkan pekerjaan sehingga perekonomian rumah tangga ikut terganggu sehingga pemenuhan makan bergizi belum tentu tercukupi. Program pemerintah yang menggalakkan tentang Makan Bergizi Gratis memberikan peluang untuk memperbaiki gizi anak-anak usia sekolah, agar mereka dapat terpenuhi asupannya. Namun, pelaksanaannya masih mendapatkan pro dan kontra terkait target yang tepat menerima program ini.

Regresi logistik merupakan metode analisis statistik yang digunakan untuk memprediksi probabilitas suatu kejadian berdasarkan variabel independen yang ada [8]. Dalam konteks penelitian ini, regresi logistik digunakan untuk memahami hubungan antara pola asupan gizi anak dengan berbagai faktor yang memengaruhi status kesehatannya. Metode regresi logistik dalam penelitian ini diterapkan untuk menganalisis hubungan antara variabel-variabel seperti data demografi anak, pola makan harian anak, asupan nutrisi anak, aktivitas fisik dan pola hidup, dan riwayat status kesehatan anak. Dengan menggunakan data primer yang diperoleh melalui kuesioner,

yang disebar ke orangtua yang memiliki anak usia sekolah 7-19 tahun, analisis dilakukan untuk memprediksi kemungkinan seorang anak mengalami masalah kesehatan berdasarkan pola asupan yang diamati. Penelitian yang dilakukan oleh Rahmadeni et al. (2023) menunjukkan bahwa teknik ini mampu mengungkap pola distribusi data gizi dengan lebih akurat dibandingkan metode konvensional seperti analisis deskriptif sederhana [9].

Penerapan regresi logistik dalam studi kesehatan telah banyak digunakan dalam berbagai penelitian sebelumnya. Misalnya, penelitian oleh Kwandy Chanda et al. (2024) yang memprediksi digital penyakit jantung coroner, variabel yang memiliki hubungan kuat dengan penyakit jantung adalah `General_Health`, `Checkup`, `Exercise`, `Diabetes`, dan `Smoking_History`. memiliki pengaruh signifikan [10]. Menurut studi yang dilakukan oleh Sylvia Sujana et al. (2024), penerapan regresi logistik dalam memprediksi kanker payudara. Metode ini tepat untuk memprediksi data yang digunakan karena kemampuannya dalam menangani variabel dependen yang bersifat kategorikal serta memberikan output dalam bentuk probabilitas. [11]. Selain itu, penelitian lain oleh Aditya Pratama et al. (2023) menunjukkan bahwa model regresi logistik dapat digunakan dalam sistem cerdas untuk memprediksi risiko diabetes. Dalam penelitian ini, telah dibuat suatu model prediksi dengan menggunakan `logistic regression` di Python IDE (Jupyter Notebook) untuk tujuan deteksi dini, dengan fokus pada kemampuan untuk memprediksi apakah seseorang mungkin mengalami penyakit diabetes [12].

Penelitian regresi logistik juga digunakan dalam beberapa penelitian lain tidak hanya dibidang Kesehatan, seperti penelitian oleh Nurmalitasari et al (2022) bertujuan untuk memprediksi keberhasilan mahasiswa Universitas Duta Bangsa Surakarta (UDB) berdasarkan performanya [13]. Selain itu, penelitian K. Kutha Ardana1(2023) menunjukkan perbandingan performa tiga algoritma pembelajaran mesin, yaitu KNN (K-Nearest Neighbors), naive Bayes, dan regresi logistik binomial, dalam mengklasifikasi status ekonomi negara sebagai maju atau berkembang [14]

Berdasarkan latar belakang tersebut, penelitian ini berupaya untuk mengeksplorasi dan memvisualisasikan hubungan antara data demografi, pola makan harian anak, asupan nutrisi anak, aktivitas fisik dan pola hidup, dan riwayat status kesehatan menggunakan metode regresi logistik. Rumusan masalah yang menjadi fokus utama dalam penelitian ini mencakup pertanyaan bagaimana pola konsumsi gizi anak (dalam

konteks krisis ekonomi) memengaruhi status kesehatan mereka? faktor apa saja yang paling signifikan berkontribusi terhadap masalah gizi (gizi buruk/obesitas) berdasarkan analisis regresi logistik? bagaimana visualisasi data (heatmap, OR plot, ROC curve) dapat mengidentifikasi pola asupan berisiko dan mendukung pengambilan keputusan?

Beberapa penelitian sebelumnya telah mengimplementasikan algoritma Regresi Logistik untuk memprediksi berbagai penyakit kronis, seperti diabetes, kanker payudara, penyakit kardiovaskular, dan COVID-19. Namun, masih terdapat beberapa gap (kesenjangan penelitian) yang belum terjawab, terutama dalam konteks prediksi kesehatan anak berdasarkan asupan gizi. Penelitian oleh [15] dan [19] menggunakan Regresi Logistik untuk memprediksi risiko diabetes pada orang dewasa dengan variabel seperti tekanan darah, indeks massa tubuh (IMT), dan riwayat keluarga. Hasilnya menunjukkan bahwa normalisasi data dan pemilihan fitur (seperti umur dan tekanan darah) meningkatkan akurasi prediksi. Namun, penelitian ini tidak membahas asupan gizi sebagai faktor dominan, padahal nutrisi memainkan peran penting dalam perkembangan penyakit metabolik seperti diabetes. Studi [16] tentang kanker payudara dan [18] tentang COVID-19 membandingkan Regresi Logistik dengan metode lain seperti Random Forest, menunjukkan bahwa performa Regresi Logistik cukup baik meskipun tidak selalu menjadi yang terbaik. Namun, penelitian ini hanya berfokus pada gejala klinis atau citra medis, bukan pada faktor nutrisi atau pola makan. Penelitian [17] membandingkan Regresi Logistik dengan model *machine learning* lain (seperti *Neural Network* dan *Gradient Boosting*) untuk memprediksi penyakit kardiovaskular, diabetes, dan hipertensi. Hasilnya menunjukkan bahwa Regresi Logistik tetap kompetitif dalam kasus dengan prediktor klinis sederhana. Sedangkan Penelitian [22] menggunakan pendekatan deskriptif korelasional dengan desain potong lintang (cross-sectional), dengan variabel independen berupa Tingkat Kecukupan Energi (TKE) dan Tingkat Kecukupan Protein (TKP), serta variabel dependen berupa status gizi anak. Hasil penelitian menunjukkan adanya hubungan yang signifikan antara status gizi dengan tingkat kecukupan energi ( $p = 0.000$ ) dan protein ( $p = 0.002$ ). Temuan ini mengindikasikan bahwa konsumsi makanan di tingkat rumah tangga sangat memengaruhi status gizi anak. Jika konsumsi makanan keluarga baik—yakni beragam, bergizi, dan seimbang—maka status gizi anak juga cenderung baik. Penelitian [23] pada tahun 2024 menunjukkan bahwa dari 100 responden, 52% memiliki aktivitas fisik

rendah, 38% memiliki pola makan berlebihan, dan 31% mengalami obesitas. Hasil uji chi-square menunjukkan adanya hubungan yang signifikan antara aktivitas fisik ( $p = 0,002$ ) dan pola makan ( $p = 0,000$ ) dengan kejadian obesitas pada anak. Temuan ini mengindikasikan bahwa pola makan yang tidak seimbang dan kurangnya aktivitas fisik berkontribusi terhadap obesitas anak.

Berdasarkan tinjauan di atas, terdapat beberapa kesenjangan penelitian yang menjadi dasar penelitian ini yaitu Fokus pada Orang Dewasa – Mayoritas penelitian masih banyak menganalisis penyakit pada orang dewasa, sementara prediksi kesehatan anak (terkait gizi dan tumbuh kembang) belum banyak dieksplorasi, Kurangnya Penekanan pada Asupan Gizi – Variabel yang digunakan umumnya bersifat klinis (tekanan darah, IMT, riwayat keluarga), sedangkan faktor nutrisi dan pola makan anak belum menjadi fokus utama dan Metode Regresi Logistik untuk Data Anak – Belum ada penelitian yang menguji secara khusus efektivitas Regresi Logistik dalam memprediksi status kesehatan anak berdasarkan data gizi, yang mungkin memiliki karakteristik berbeda dari data orang dewasa.

Oleh karena itu, penelitian ini akan menganalisis data asupan gizi anak untuk memprediksi status kesehatan mereka menggunakan Regresi Logistik. Dengan mempertimbangkan variabel seperti jenis makanan, frekuensi makan, berat badan, tinggi badan, dan riwayat kesehatan, penelitian ini diharapkan dapat memberikan model prediktif yang akurat untuk deteksi dini masalah gizi anak, seperti stunting, obesitas, atau defisiensi nutrisi. Hasilnya dapat menjadi dasar rekomendasi bagi orang tua dan tenaga kesehatan dalam pencegahan masalah gizi anak sejak dini di masa krisis ekonomi pasca covid-19. Meskipun Regresi Logistik telah terbukti efektif dalam prediksi berbagai penyakit, penerapannya pada analisis gizi dan kesehatan anak masih terbatas. Penelitian ini akan mengisi gap tersebut dengan membangun model prediksi berbasis asupan gizi anak, sehingga memberikan kontribusi baru dalam bidang *health analytics* dan kesehatan masyarakat.

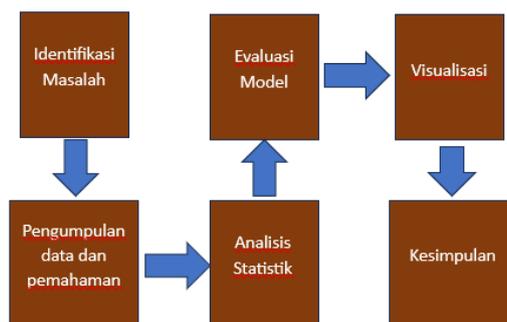
Hasil penelitian ini tidak hanya menjawab tantangan gizi anak di era krisis ekonomi, tetapi juga menyediakan kerangka berbasis data untuk 1) *Prevensi*: Identifikasi risiko gizi sejak dini melalui model prediktif. 2) *Intervensi*: Strategi terarah sesuai profil risiko (contoh: subsidi pangan untuk keluarga berpendapatan rendah). 3) *Evaluasi Kebijakan*: Dasar ilmiah untuk memastikan program Makan Bergizi Gratis tepat sasaran. Temuan ini

diharapkan dapat menjadi solusi implementatif untuk mengurangi double burden of malnutrition di Indonesia.

## 2. METODE PENELITIAN

Penelitian ini menerapkan Pendekatan Kuantitatif-Eksplanatif, fokus: Mengukur hubungan sebab-akibat antara variabel independen (asupan gizi, aktivitas fisik, dll.) dan variabel dependen (status gizi: normal/gizi buruk/obesitas). Sampel penelitian diperoleh melalui Stratified Random Sampling. Pendekatan ini memastikan representasi yang proporsional dari setiap jenjang pendidikan (SD, SMP, SMA) di wilayah studi. Kuesioner yang digunakan telah melalui uji validitas, termasuk Validitas Konvergen, untuk memastikan bahwa item-item seperti pola makan dan aktivitas fisik benar-benar berkorelasi positif dengan indikator kesehatan gizi yang relevan. Penting dicatat, data status gizi anak dikategorikan menggunakan Skala Nominal (normal, gizi buruk, atau obesitas), yang menjadi dasar untuk analisis model. Model yang digunakan adalah Regresi Logistik Multivariat / Multinomial: Untuk memprediksi probabilitas kategori status gizi berdasarkan kombinasi variabel dan Uji Korelasi Pearson: Mengidentifikasi kekuatan dan arah hubungan antar variabel. Pendekatan Prediktif-Berbasis Data, fokus: Membangun model untuk mengklasifikasikan risiko gizi anak berdasarkan pola data historis. Alat yang digunakan: Seleksi Fitur (Backward Elimination): Memilih variabel paling berpengaruh (misal: frekuensi fast food, asupan sayur). Validasi Silang (10-Fold Cross Validation): Memastikan model tidak overfitting. Output: Skoring Risiko Gizi: Skala sederhana untuk deteksi dini di puskesmas/sekolah. Pendekatan Visualisasi Data, fokus: Mentransformasi hasil statistik menjadi insight visual yang mudah dipahami. Alat yang digunakan: Heatmap Korelasi: Menunjukkan pola hubungan (misal: korelasi negatif antara konsumsi buah dan obesitas). Odds Ratio Plot: Menyoroti prediktor dominan (contoh: OR minuman manis = 2.5). ROC Curve: Mengevaluasi akurasi model (AUC > 0.7 dianggap baik). Penelitian ini diharapkan dapat memberikan pemahaman terkait dampak krisis ekonomi pada gizi anak, prediksi risiko melalui model statistik, mengkomunikasikan temuan melalui visualisasi dan mengarahkan kebijakan berbasis bukti.

Metode penelitian yang akan dilakukan oleh peneliti dituangkan dalam bentuk diagram berikut ini :



Gambar 1. Diagram Alir Penelitian

Penjelasan dari tahapan pada diagram alir penelitian :

1) Identifikasi masalah

Mengidentifikasi masalah dengan cara melihat permasalahan yang terjadi di masyarakat khususnya asupan gizi anak usia sekolah setelah pasca covid-19 dan terutama pasca pecahnya perang Russia – Ukraina serta menguji ketepatan sasaran program MBG

2) Pengumpulan data

Pengumpulan data dalam penelitian ini dilakukan dengan menyebarkan kuesioner secara langsung ke orangtua di sekitar DKI Jakarta yang memiliki anak usia sekolah 6-15 tahun. Kuisisioner berisi data demografi anak, pola makan harian anak, asupan nutrisi anak, aktivitas fisik dan pola hidup, dan riwayat status kesehatan anak. Jumlah data yang didapatkan sebanyak 308 data / record. Pemahaman data dilakukan dengan mengidentifikasi setiap variable yang ada, apakah ada variable yang terisi tidak seimbang?

3) Analisis Statistik

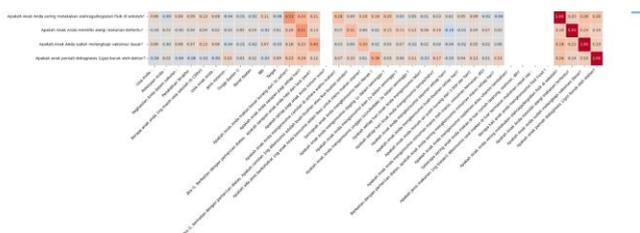
Dilakukan dengan melakukan uji pearson terlebih dahulu, untuk menguji antar variabel. Lalu seleksi fitur dengan Backforward Elimination untuk mencari fitur yang paling relevan. Kemudian, Regresi Logistik untuk mencari sesuai target output yaitu Normal, Gizi Buruk atau Obesitas.

4) Evaluasi Model

Di tahapan ini, akan dilakukan perhitungan Akurasi, sensitivitas, spesifisitas, dan AUC-ROC dimana kinerja model dengan ROC Curve untuk Evaluasi akurasi klasifikasi ( $AUC > 0.8 = \text{model baik}$ )







Gambar 4. Grafik HeatMap Uji Korelasi

Setelah melakukan analisis korelasi Pearson, berikut beberapa temuan penting: Korelasi Terkuat Positif terdapat pada Konsumsi sayuran harian dengan konsumsi buah-buahan harian ( $r = 0.42$ ), Konsumsi ikan 2x seminggu dengan konsumsi unggas 3x seminggu ( $r = 0.38$ ) dan Berat badan dengan tinggi badan ( $r = 0.72$ ). Sedangkan Korelasi Terkuat Negatif terdapat pada Konsumsi fast food dengan konsumsi sayuran harian ( $r = -0.31$ ) dan Konsumsi minuman manis dengan konsumsi air putih ( $r = -0.29$ ). Untuk Korelasi Signifikan Lainnya membuktikan bahwa Anak yang sarapan pagi cenderung memiliki berat badan lebih ideal ( $r = 0.25$ ), Konsumsi nasi merah berhubungan dengan konsumsi karbohidrat selain nasi ( $r = 0.27$ ) dan Aktivitas fisik di sekolah berkorelasi positif dengan konsumsi protein (daging/ikan/unggas) ( $r = 0.23$ ). Analisa pada Fitur yang Tidak Berkorelasi Signifikan diantaranya Jumlah anak usia sekolah dalam keluarga, Jenis kelamin anak, Status vaksinasi dasar dan Diagnosis gizi buruk sebelumnya.

### 3) Seleksi Fitur: Backward-Forward Elimination

Langkah ini akan mencari fitur yang paling relevan dalam prediksi target. Variabel target akan didefinisikan berdasarkan kolom "Apakah anak pernah didiagnosis obesitas/gizi buruk oleh dokter?". Adapun fitur yang dipilih dari 35 fitur diantaranya adalah fitur :

- Berapa kali anak anda mengkonsumsi Fast Food ?
- Apakah Anak Anda sering melakukan olahraga/kegiatan fisik di sekolah?
- Apakah Anak Anda mengkonsumsi minuman manis (teh manis, minuman kemasan, dll)?
- Berkaitan dengan pertanyaan diatas, apakah Anak Anda sering mengkonsumsi minuman manis setiap hari?
- Apakah Anak Anda mengkonsumsi suplemen atau vitamin tambahan?

Penerapan evaluasi model dengan menggunakan python sebagai berikut :

```

# Pilih fitur prediktor
features = [
    "Berapa kali anak anda mengkonsumsi Fast Food ?",
    "Apakah Anak Anda sering melakukan olahraga/kegiatan fisik di sekolah?",
    "Apakah Anak Anda mengonsumsi minuman manis (teh manis, minuman kemasan, dll)?",
    "Berkaitan dengan pertanyaan diatas, apakah Anak Anda sering mengonsumsi minuman manis setiap hari?",
    "Apakah Anak Anda mengonsumsi suplemen atau vitamin tambahan?"
]
X = sm.add_constant(data[features])
y = data["Target_encoded"]

# Regresi Logistik Multinomial
model = sm.Logit(y, X)
result = model.fit()

# Hitung Odds Ratio
odds_ratios = np.exp(result.params)
conf = np.exp(result.conf_int())
conf.columns = ["Lower CI", "Upper CI"]
p_values = result.pvalues

# Gabungkan hasil
odds_result = pd.concat([odds_ratios, conf, p_values], axis=1)
odds_result.columns = ["Odds Ratio", "Lower CI", "Upper CI", "p-value"]
print(odds_result)

```

Dari penerapan diatas didapatkan hasil sebagai berikut :

Tabel 2. Fitur Pilihan Back Forward

| Pertanyaan  | Nilai Odds | Keterangan                         |
|---|------------|------------------------------------|
| Berapa kali anak anda mengkonsumsi fast food?   | 4,99       | Mampu meningkatkan resiko obesitas |
| Apakah anak anda sering melakukan / kegiatan fisik di sekolah ?                                     | -0,0023    | Mampu menurunkan resiko obesitas   |
| Apakah anda anda mengkonsumsi minuman manis (teh manis, minuman kemasan, dll)?                      | 4,58       | Mampu meningkatkan resiko obesitas |
| Berkaitan dengan pertanyaan diatas, apakah anak anda sering mengkonsumsi minuman manis setiap hari? | 3,55       | Mampu meningkatkan resiko obesitas |
| Apakah anak anda mengkonsumsi suplemen atau vitamin tambahan ?                                      | 0,0047     | Mampu menurunkan resiko obesitas   |

#### 4) Model Prediksi : Regresi Logistik Multinomial

Karena target terdiri dari tiga kelas, kita akan menggunakan regresi logistik multinomial untuk memprediksi status gizi anak (Normal, Gizi Buruk, Obesitas).

Hasil yang didapatkan setelah olah data diatas menggunakan python bahwa Faktor yang Signifikan Mempengaruhi Status Gizi diantaranya Konsumsi sayuran setiap hari mengurangi kemungkinan gizi buruk (OR=0.70) dan meningkatkan kemungkinan obesitas (OR=1.30), Konsumsi buah-buahan setiap hari mengurangi risiko obesitas (OR=0.75) dan meningkatkan kemungkinan status gizi normal (OR=1.48), Konsumsi minuman manis secara signifikan meningkatkan risiko obesitas (OR=1.60 untuk status normal vs OR=0.71 untuk obesitas) dan Vaksinasi dasar lengkap berhubungan dengan status gizi normal (OR=1.69). Pola makan yang berpengaruh diantaranya Sarapan pagi teratur berhubungan dengan status gizi normal (OR=1.29), Konsumsi fast food meningkatkan risiko obesitas (OR=1.40) dan Aktivitas fisik di sekolah

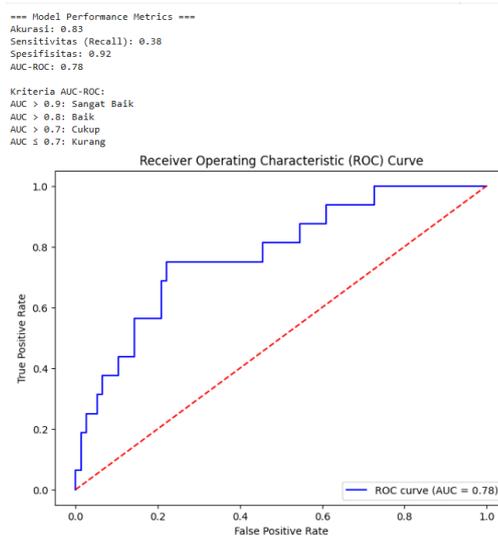
mengurangi risiko obesitas (OR=0.74). Implikasi terhadap kesehatan menyimpulkan bahwa Pola makan seimbang dengan sayur dan buah serta mengurangi minuman manis dan fast food penting untuk mencegah masalah gizi dan Aktivitas fisik dan vaksinasi tampaknya juga berperan dalam status gizi anak. Model ini memiliki pseudo R-squared sebesar 0.3524, yang menunjukkan bahwa sekitar 35% variasi dalam status gizi dapat dijelaskan oleh variabel-variabel yang dimasukkan dalam model. Dari data yang diolah menunjukkan bahwa anak-anak banyak yang berada di status Normal.

## 5) Evaluasi Model dan Visualisasi

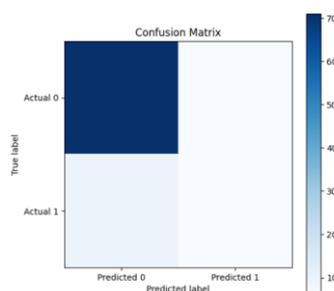
Tahapan berikutnya yaitu evaluasi model dimana akan menghitung :

- Akurasi: Proporsi prediksi benar.
- Sensitivitas (Recall): Kemampuan model mendeteksi kelas tertentu.
- Spesifisitas: Kemampuan model mendeteksi bukan kelas tertentu.
- AUC-ROC: Area di bawah kurva ROC (AUC > 0.8 menunjukkan model baik)

Dengan menggunakan sintaks python diuraikan sebagai berikut :



Gambar 5. ROC dan AUC



Gambar 6. Cofusion Matrix

Kinerja model setelah diuji menunjukkan Akurasi (Accuracy) = 0.83 artinya Model memprediksi 83% kasus dengan benar. Nilai ini cukup baik, tetapi perlu dilihat bersama metrik lain karena bisa menyesatkan jika ada ketidakseimbangan kelas. Sensitivitas (Recall) = 0.38 artinya Model hanya mampu mendeteksi 38% kasus positif (misalnya: pasien sakit, transaksi fraud, dll.). Rendahnya sensitivitas menunjukkan bahwa model sering gagal mendeteksi kasus positif (banyak False Negatives). Spesifisitas (Specificity) = 0.92 artinya Model sangat baik dalam mengidentifikasi 92% kasus negatif (misalnya: pasien sehat, transaksi normal). Tingginya spesifisitas berarti model jarang salah memprediksi negatif sebagai positif (False Positives rendah). AUC-ROC = 0.78 artinya termasuk kategori "Cukup" berdasarkan kriteria AUC-ROC. Kurva ROC menunjukkan bahwa model masih memiliki kemampuan diskriminasi, tetapi tidak optimal. Kelebihan Model ini Spesifisitas tinggi (0.92) - Baik untuk menghindari alarm palsu (False Positives). Akurasi relatif baik (0.83) - Secara umum, prediksi cukup akurat. Sedangkan Kelemahan Model ini Sensitivitas sangat rendah (0.38) - Model gagal mendeteksi banyak kasus positif, yang bisa berbahaya dalam aplikasi seperti diagnosa medis atau deteksi fraud. AUC-ROC (0.78) tidak mencapai standar "Baik" (AUC > 0.8) sehingga kemampuan klasifikasi masih terbatas.

#### **4. SIMPULAN**

Dari Uji Pearson dapat ditarik kesimpulan sebagai berikut bahwa : Intervensi gizi sebaiknya fokus pada peningkatan konsumsi makanan sehat secara komprehensif (bukan hanya satu jenis makanan) dan mengurangi konsumsi fast food serta minuman manis. Program promosi kesehatan bisa menekankan pentingnya sarapan pagi dan aktivitas fisik. Pola Makan Seimbang : Terdapat hubungan positif antara berbagai komponen makanan sehat (sayur, buah, protein hewani), menunjukkan bahwa anak yang mengonsumsi satu jenis makanan sehat cenderung mengonsumsi jenis makanan sehat lainnya. Kebiasaan Sarapan: Anak yang rutin sarapan pagi menunjukkan pola gizi yang lebih baik secara keseluruhan. Dampak Fast Food : Konsumsi fast food berhubungan negatif dengan pola makan sehat, terutama konsumsi sayuran. Faktor Fisik : Tinggi dan berat badan menunjukkan korelasi kuat seperti yang diharapkan, namun tidak banyak berkorelasi dengan faktor pola makan. Keterbatasan: Banyak variabel gaya hidup tidak

menunjukkan korelasi kuat, mungkin karena kompleksitas faktor yang mempengaruhi status gizi anak. Untuk Seleksi Fitur – digunakan untuk melihat nilai Odds yang dapat meningkatkan dan menurunkan obesitas, didapat 3 hal yang dapat meningkatkan obesitas diantaranya fast food, minuman manis dan frekuensi minuman mas, sedangkan 2 fitur yang diuji ternyata dapat menurunkan kemungkinan obesitas yaitu kegiatan fisik dan suplemen/vitamin yang dikonsumsi. Model prediksi yaitu Regresi Logistik Multivarian – didapatkan model ini cukup baik dalam mengidentifikasi kasus negatif (spesifisitas tinggi) tetapi buruk dalam mendeteksi kasus positif (recall sangat rendah). Namun tidak direkomendasikan untuk aplikasi kritis (seperti medis atau keamanan) di mana False Negatives berbahaya dan perlu perbaikan, terutama dalam menangani ketidakseimbangan data dan pemilihan fitur/model yang lebih tepat. Solusi yang ditawarkan : Atasi Class Imbalance yaitu gunakan teknik resampling (oversampling minoritas/undersampling mayoritas). Coba weighted loss function (misal: `class_weight='balanced'` di LogisticRegression). Untuk Optimasi Model coba algoritma lain seperti Random Forest, XGBoost, atau SVM yang lebih robust terhadap imbalance. Lakukan tuning hyperparameter untuk meningkatkan recall. Lakukan evaluasi dengan Metrik Lain, gunakan F1-Score (harmonik dari precision & recall) karena sensitivitas rendah. Lalu periksa Precision-Recall Curve (lebih relevan jika kelas positif langka).

## REFERENCES

- [1] Rahayu, S., dan Widodo, T, “Pengaruh Asupan Gizi terhadap Perkembangan Kognitif dan Fisik Anak Usia Dini,” : Jurnal Gizi dan Kesehatan Indonesia, Vol. 15, No. 2, 2023
- [2] Meo, L, Dinatha, N.M, Maku. K.R.M, Meo. G, Rabu. L.A, “Permasalahan Gizi Pada Anak Usia Dini”, Jurnal Edukasi Citra Olah Raga, Vol.4 No.3, Oktober 2024.
- [3] Leviana, S. dan Agustina, Y., “Analisis Pola Makan Dengan Status Gizi Pada Siswa-Siswi Kelas V Di SDN Jatiwaringin XII Kota Bekasi”, Malahayati Nursing Journal, Vol.6 No.4, Maret 2024
- [4] Manaf, S.A.R., Amelia. R., Indahwati, “Faktor-faktor yang memengaruhi Permasalahan Stunting di Jawa Barat menggunakan Regresi Logistik Biner”, Jurnal Ilmiah Teori dan Aplikasi Statistika, Vol.15 No.2, 2022 [4]
- [5] Tabel Angka Gizi, <https://p2ptm.kemkes.go.id/infographic/tabel-angka-kecukupan-gizi-rata-rata-yang-dianjurkan-per-orang-per-hari-untuk-anak-umur-7-12-tahun>
- [6] Komposisi Makanan Seimbang, [https://www.medicalogy.com/blog/pentingnya-nutrisi-seimbang-untuk-menjaga-kesehatan-jangka-panjang/?srsltid=AfmBOopa79lmHdCeZEWf8AfNrWRQwDhVJAmWpPbGbSZGnBA\\_49EzKMDL](https://www.medicalogy.com/blog/pentingnya-nutrisi-seimbang-untuk-menjaga-kesehatan-jangka-panjang/?srsltid=AfmBOopa79lmHdCeZEWf8AfNrWRQwDhVJAmWpPbGbSZGnBA_49EzKMDL)
- [7] Porsi Makanan Sehat, <https://raisingchildren.net.au/school-age/nutrition-fitness/daily-food-guides/school-age-food-groups#:~:text=School%2Dage%20children%20need%20to,%2Doxidants%2C%20fibre%20and%20water>
- [8] Fahmuddin, M., Aidid, M.K., Taslim, M.J., “Implementasi Analisa Regresi Logistik dengan Metode Machine Learning Untuk Mengklasifikasi Berita di Indonesia”, Journal of Statistics and Its Application on Teaching and Research, Vol. 5 No. 03 (2023)

- [9] Rahmadeni dan Puspita, S., “Aplikasi Regresi Logistik Ordinal dalam Pemodelan Status Gizi Balita (Studi Kasus : Puskesmas Limapuluh di Kota Pekanbaru)”, *Jurnal Sains Matematika dan Statistika*, Vol.9, No.2, Juli 2023
- [10] Chandra, K. dan Prasetyo, J.S., “Prediksi Penyakit Jantung Koroner Menggunakan Metode K-NN dan Regresi Logistik berdasarkan Kerangka Kerja CRISP-DM”, *Prosiding SENAM 2024 : Sistem Informasi & Informatika Vo.4*, September 2024
- [11] Sujana, S., Juwita, A.R., Rahmat, Faisal, S., “Penerapan Metode Regresi Logistik Untuk Memprediksi Peristiwa Biner Pasien Pasca Operasi Kanker payudara”, *Journal of Information System Research (JOSH)*, Vol.5, No.4, Juli 2024
- [12] Pratama, A., Nurcahyo, A.C, Figria, L., “Penerapan Machine Learning dengan Algoritma Logistik Regresi untuk Memprediksi Diabetes”, *Seminal Nasional CORISINDO*, Agustus 2023
- [13] N. Nurmalitasari and E. Purwanto, “Prediksi Performa Mahasiswa Menggunakan Model Regresi Logistik,” *J. Deriv. J. Mat. dan Pendidik. Mat.*, vol. 9, no. 2, pp. 145–152, 2022, doi: 10.31316/jderivat.v9i2.2639
- [14] N. K. K. Ardana et al., “Perbandingan Metode KNN, Naive Bayes, dan Regresi Logistik Binomial dalam Pengklasifikasian Status Ekonomi Negara,” *Jambura J. Math.*, vol. 5, no. 2, pp. 404–418, 2023, doi: 10.34312/jjom.v5i2.21103.
- [15] Cahyani Q, Finandi M, Rianti J et al., “Prediksi Risiko Penyakit Diabetes menggunakan Algoritma Regresi Logistik Diabetes Risk Prediction using Logistic Regression Algorithm ” *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, (2022), 2828-9099, 1(2)
- [16] A. D. Achmad, “Klasifikasi Breast Cancer Menggunakan Metode Logistic Regression,” *Jtriste*, vol. 9, no. 1, pp. 143-148, 2022, [Online]. Available: <https://jurnal.kharisma.ac.id/jtriste/article/download/384/201>
- [17] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., Wong, T. Y., & Cheng, C. (2020). Logistic Regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*, 122, 56–69. <https://doi.org/10.1016/j.jclinepi.2020.03.002>
- [18] Firmansyah, I., Samudra, J.T, Pardede, D., Situmorang, Z., “Komparasi Random Forest dan Logistic Regression dalam Klasifikasi Penderita Covid-19 Berdasarkan Gejalanya”, *Journal of Science and Social Research*, V (3): 595 – 601, Oct 2022.
- [19] Gunawan, S., Astuti, R., Priharto, W., Hamonangan, R., “Prediksi Diabetes Mellitus Tipe 2 Dengan Algoritma Logistic Regression Untuk Pendeteksi Dini”, *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, Vol. 13 No. 1, Januari 2025.
- [20] Ayu, K.G, Sari, D.W., “Developing KSAM (Kenyang Bersama) Application Using Extreme Programming Methodology”, *International Journal of Computer Techniques — Volume 8 Issue 2*, April 2021
- [21] Kekalih, A., Chandra, D.N., Mirtha, L.T., “Dietary intakes, nutritional and biochemical status of 6 months to 12-year-old children before the COVID-19 pandemic era: the South East Asian Nutrition Survey II Indonesia (SEANUTS II) study in Java and Sumatera Islands, Indonesia”, *Journal Public Health Nutrition*, Volume 28, Issue1, 2025.
- [22] Maya, S., Kusharto, C. M., Nurdin, N.M., “Kualitas Konsumsi Pangan berdasarkan *Healthy Eating Index* dihubungkan dengan Status Gizi Anak Sekolah Dasar Kabupaten Kerinci”, *Jurnal Gizi Kerja dan Produktivitas*, Vol.4 No.1, 2023.
- [23] Agita, N.M.A.D., Karyus, A., Astuti, D.W., “Hubungan Antara Aktivitas Fisik dan Pola Makan Dengan Kejadian Obesitas pada Anak Sekolah Dasar SD Negeri 1 Setia Bakti Lampung Tengah Tahun 2024”, *Journal of Education Teknologi Information, Social Science and Health*, Vol4, No.1, 2025.