

# Machine Learning Klasifikasi Status Gizi Balita Menggunakan Algoritma Random Forest

Putri Handayani, Abd. Charis Fauzan, Harliana\*

Fakultas Ilmu Eksata, Program Studi Ilmu Komputer, Universitas Nahdlatul Ulama Blitar, Blitar, Indonesia

Email: <sup>1</sup>putrihandayani00000@gmail.com, <sup>2</sup>abdcharis@unublitar.ac.id, <sup>3</sup>harliana@unublitar.ac.id

Email Penulis Korespondensi: harliana@unublitar.ac.id

**Abstrak**—Tumbuh kembang balita di masa depan sangat dipengaruhi oleh permasalahan gizi pada usia 0-59 bulan. Untuk mencapai kesehatan yang optimal diperlukan status gizi yang tinggi. Perkembangan yang tidak tepat, energi yang tidak mencukupi untuk berolahraga, penurunan kekebalan tubuh, dan gangguan fungsi otak dalam jangka panjang semuanya dapat disebabkan oleh kekurangan gizi. Dalam hal ini, Pos Pusat Pelayanan Terpadu (Posyandu) bertugas mengawasi kesehatan gizi anak. Data antropometri, atau ukuran tubuh manusia, seperti tinggi dan berat badan, merupakan bagian dari prosedur pemantauan ini. Variabel lainnya meliputi pengukuran posisi dan keluhan yang telah disampaikan. Penelitian ini memiliki tujuan guna menggunakan algoritma *Random Forest* untuk mengklasifikasikan status gizi anak di Kecamatan Nglegok. Penelitian ini menggunakan matriks konfusi untuk mengevaluasi hasil hutan secara acak. Empat skenario, masing-masing dengan data pelatihan dan pengujian, dibuat dari data tersebut untuk melakukan pengujian. Hasil pengujian menunjukkan bahwa pembagian data pelatihan 90% dan 10% pengujian merupakan skenario optimal, dengan akurasi 88,6%, presisi 88,1%, recall 88,6%, dan F1-Score 88,2%.

**Kata Kunci:** Status Gizi; Random Forest; Klasifikasi; Confusion Matrix; Anthropometri.

**Abstract**—The future growth and development of toddlers is greatly influenced by nutritional problems at the age of 0-59 months. To achieve optimal health, high nutritional status is required. Improper development, insufficient energy for exercise, decreased immunity, and long-term impaired brain function can all be caused by malnutrition. In this case, the Integrated Service Center Post (Posyandu) is tasked with monitoring children's nutritional health. Anthropometric data, or human body measurements, such as height and weight, are part of this monitoring procedure. Other variables include position measurements and complaints that have been submitted. The aim of this research is to use the Random Forest algorithm to classify the nutritional status of children in Nglegok District. This study uses a confusion matrix to evaluate random forest yields. Four scenarios, each with training and test data, are created from the data to perform testing. The test results show that dividing 90% training data and 10% testing is the optimal scenario, with accuracy of 88.6%, precision of 88.1%, recall of 88.6%, and F1-Score of 88.2%.

**Keywords:** Nutritional Status; Random Forest; Classification; Confusion Matrix; Anthropometry

## 1. PENDAHULUAN

Gizi adalah istilah untuk makanan yang penting bagi perkembangan, pertumbuhan, dan kesehatan tubuh seseorang [1]. Memiliki status gizi yang tinggi diperlukan untuk mencapai kesehatan yang optimal. Gizi yang tidak memadai dapat mengakibatkan keterlambatan perkembangan, kekurangan energi untuk berolahraga, penurunan kekebalan tubuh, dan gangguan fungsi otak dalam jangka panjang, terutama pada balita antara usia 0 dan 59 bulan [2]. Sehingga, sangat penting untuk mengonsumsi makanan-makanan bernutrisi dan baik untuk tubuh mereka [3]. Teknik pengukuran tubuh manusia yang dikenal dengan nama "Antropometri" dapat digunakan untuk menilai kondisi gizi balita. Dengan mempertimbangkan usia serta jenis kelamin dari balita, pendekatan ini mengukur berat badan dan tinggi balita [4]. Pengukuran antropometri adalah gabungan dari banyak penanda yang banyak digunakan untuk mengevaluasi kesehatan gizi. Metrik yang digunakan dalam antropometri pada dasarnya adalah penilaian ukuran tubuh manusia. Diantaranya yaitu, indeks berat badan menurut tinggi badan, tinggi badan menurut umur, massa tubuh menurut umur, serta indeks berat badan menurut umur [5].

Berdasarkan temuan kajian kesehatan dasar Kemenkes RI pada tahun 2018, 17,7% anak usia di bawah lima tahun mengalami masalah gizi. Proporsi ini terdiri dari 3,9% dan 13,8% anak balita yang mengalami kekurangan gizi [6]. Untuk menangani permasalahan kesehatan dan status gizi balita, Di Kecamatan Nglegok dilakukan pengawasan status gizi balita melalui Pos Pusat Pelayanan Terpadu (Posyandu). Pengukuran tinggi badan, berat badan, posisi pengukuran, dan gejala yang dilaporkan merupakan bagian dari prosedur pengawasan ini. Informasi yang dikumpulkan kemudian dicatat secara manual di buku KMS untuk mengetahui status gizi balita dengan perbandingan menggunakan titik acuan yang berasal dari data pengukuran dan dilakukan proses pengecekan status gizi menggunakan aplikasi khusus yang digunakan untuk menganalisis status gizi. Di Kecamatan Nglegok mengalami peningkatan status gizi buruk pada balita. Deteksi dini terhadap kondisi gizi anak sangatlah penting, hal ini memicu penelitian algoritma *Machine Learning* [7]. Penelitian ini memiliki tujuan guna mengembangkan suatu model *Machine Learning* yang menggunakan pendekatan *Random Forest* guna mengklasifikasikan status gizi balita. Model ini diharapkan nantinya mampu memberikan status gizi secara cepat dan akurat untuk mendukung upaya mencapai target penurunan gizi buruk pada balita dengan memanfaatkan kecanggihan teknologi serta tingkat akurasi yang tinggi yang bisa diberikan oleh *Random Forest*.

Dalam studi yang dilakukan oleh Sathiya [8], penerapan metode *random forest* dalam klasifikasi penyakit Parkinson juga mencapai tingkat akurasi yang tinggi, mencapai 96%. Dalam studi yang dilakukan oleh Pahlevi [9] penerapan metode *random forest* dalam klasifikasi kelayakan kredit memberi tingkat akurasi kebenaran sebanyak 78,60% dan mempunyai nilai Area Under The Curve (AUC) sebanyak 0,0907. Studi yang dilakukan oleh Chauhan [10] yang membandingkan berbagai metode *machine learning* untuk mendeteksi dini penyakit pada tanaman menunjukkan bahwa

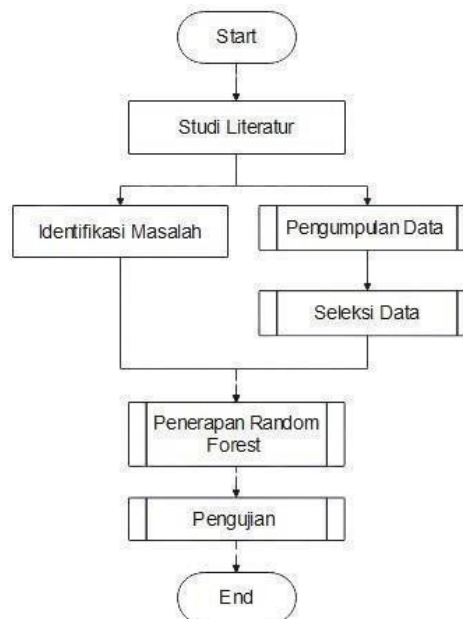
*random forest* mencapai tingkat akurasi yang superior dibandingkan dengan metode lain yang dievaluasi dalam penelitian tersebut, termasuk *naive bayes*, *decision tree*, *k-nearest neighbor* dan *support vector machine*. Studi yang dilakukan Akbar Ariyadi [7] skenario optimal dengan akurasi terbesar yaitu 90,1%, presisi sebesar 71,4% serta recall sebesar 62,5%, memakai 80% data untuk pelatihan lalu sebesar 20% untuk pengujian ketika pendekatan *random forest* digunakan untuk mengkategorikan balita yang mengalami *tunting*.

Berdasarkan permasalahan dan penjelasan yang diuraikan, maka *machine learning* dengan pendekatan *Random Forest* sangat tepat untuk mengkategorikan kondisi gizi balita karena tingkat akurasi metode yang tinggi. Untuk mengetahui persentase penurunan gizi, penelitian ini diharapkan dapat menghasilkan penilaian status gizi yang tepat di tahun mendatang khususnya di Kabupaten Nglegok sendiri. Selain mengkategorikan status gizi, penelitian ini juga fokus pada identifikasi variabel-variabel yang paling berpengaruh terhadap status gizi balita. Analisis ini memberikan wawasan yang lebih dalam yang dapat digunakan untuk intervensi yang lebih tepat sasaran.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Beberapa prosedur yang harus diikuti didalam suatu penelitian guna mewujudkan tujuan penelitian yang sudah ditentukan disebut dengan tahapan penelitian. Enam langkah diambil dalam pelaksanaan penelitian ini. Tinjauan literatur, identifikasi masalah, pengumpulan data, pengolahan data, dan representasi model hutan acak adalah langkah awal dalam proyek penelitian ini. selanjutnya hasil akan di evaluasi menggunakan *confusion matrix* [11].



Gambar 1. Diagram Alir Tahapan Penelitian

Gambar 1 menggambarkan alur tahapan penelitian yang akan dilaksanakan [12]. Didalam penelitian ini, langkah pertama yaitu melaksanakan studi literatur untuk mengumpulkan informasi yang berkaitan dengan topik penelitian serta beberapa referensi mengenai algoritma yang akan digunakan. Langkah selanjutnya adalah mengidentifikasi masalah untuk memahami topik penelitian dan menentukan faktor-faktor yang terkait. Kemudian mengumpulkan informasi yang diperlukan untuk penelitian ini. Penelitian ini akan memanfaatkan data Status Gizi yang dikumpulkan dari ahli gizi di Puskesmas Nglegok. Selanjutnya dilakukan seleksi data yang berpengaruh besar pada status gizi balita, menghilangkan nilai null, dan memilih atribut yang relevan dan melakukan pengolahan data yang sudah di seleksi dengan metode *random forest* yang dimana akan membagi data training dan data test untuk menghasilkan data yang bersifat prediktif. Nilai akurasi, presisi, recall, dan Skor F1 dihitung menggunakan matriks konfusi berdasarkan hasil prediksi, pengujian, dan data pelatihan.

### 2.2 Status Gizi

Kondisi gizi adalah suatu keadaan fisik individu yang bisa diketahui melalui asupan makanan yang dikonsumsi, dan hal ini bisa berdampak pada tingkat kognitif individu [13]. Untuk menafsirkan kondisi gizi balita, diperlukan ambang batas tertentu. Hasilnya, berat badan (BB) dan usia setiap balita diubah menjadi Z-score menggunakan rumus antropometri umum yang disediakan oleh WHO. Perhitungan Z-Score mengikuti persamaan 1 sebagai berikut [14]:

$$Z - Score = \frac{\text{Nilai Individual Subjek} - \text{Nilai Baku Rujukan}}{\text{Nilai Simpang Baku Rujukan}} \quad (1)$$

Untuk menilai status gizi anak, berbagai indeks dapat digunakan yang mencakup perbandingan berat badan terhadap umur (BB/U), tinggi badan terhadap umur (TB/U), dan tinggi badan terhadap berat badan (TB/BB). Masing-masing indeks ini memiliki ambang batas yang membantu dalam mengidentifikasi apakah status giziseorang anak tergolong buruk, kurang, normal, atau lebih. Pada tabel 1 dijelaskan detail mengenai ambang batas yang digunakan pada setiap indeks.

**Tabel 1.** Kategori status gizi balita

| Indeks                                    | Status Gizi   | Ambang Batas                |
|---|---------------|-----------------------------|
| BBerat Badan Terhadap Umur (BB/U)         | Gizi Buruk    | <-3,0 SD                    |
|   | Gizi Kurang   | <-2,0 SD Sampai >= -3,0 SD  |
|   | Gizi Normal   | >= -2 SD sampai 2,0 SD      |
|   | Gizi Lebih    | >2,0 SD                     |
| TiTinggi Badan terhadap Umur (TB/U)       | Sangat Pendek | <-3,0 SD                    |
|   | Pendek        | <-2,0 SD Sampai >= -3,0 SD  |
|   | Normal        | >= -2,0 Sampai 2,0 SD       |
|   | Tinggi        | > 2,0 SD                    |
| Tinggi Badan terhadap Berat Badan (TB/BB) |               | <-3,0 SD                    |
|   | Gizi Buruk    | <-2,0 SD Sampai >= - 3,0 SD |
|   | Gizi Kurang   | >= -2 SD sampai 2,0 SD      |
|   | Gizi Normal   | > 2,0 SD                    |
|   | Gizi Lebih    |                             |

Status gizi anak dikategorikan menurut indeks masing-masing, sebagaimana dijelaskan secara rinci pada tabel di atas. Penggunaan bagan ini sangat penting untuk intervensi gizi yang optimal dan melacak pertumbuhan dan perkembangan anak.

Tabel 1 menunjukkan bahwa klasifikasi Status Gizi Balita dibedakan menjadi empat, yaitu :

a. Gizi Buruk

Gizi buruk adalah kondisi di mana seseorang tidak mendapatkan asupan nutrisi yang cukup untuk memenuhi kebutuhan tubuhnya. Hal ini bisa disebabkan oleh kekurangan kalori, protein, vitamin, dan mineral yang diperlukan untuk pertumbuhan dan fungsi tubuh yang normal.

b. Gizi Kurang

Gizi kurang adalah kondisi di mana seseorang tidak mendapatkan asupan nutrisi yang cukup untuk memenuhi kebutuhan tubuhnya secara optimal, tetapi kondisinya tidak seberat gizi buruk. Ini bisa terjadi ketika asupan makanan sehari-hari tidak mencukupi jumlah atau kualitas nutrisi yang dibutuhkan tubuh, termasuk karbohidrat, protein, lemak, vitamin, dan mineral.

c. Gizi Normal

Gizi normal adalah kondisi di mana seseorang mendapatkan asupan nutrisi yang cukup dan seimbang untuk memenuhi kebutuhan tubuhnya, sehingga dapat menjalankan fungsi-fungsi tubuh dengan baik. Kondisi gizi normal mendukung pertumbuhan yang optimal, perkembangan fisik dan mental, serta menjaga kesehatan dan kesejahteraan secara keseluruhan.

d. Gizi Lebih

Gizi lebih adalah kondisi di mana seseorang mengonsumsi lebih banyak kalori, lemak, gula, atau nutrisi tertentu dari yang dibutuhkan tubuh, yang dapat menyebabkan kelebihan berat badan atau obesitas dan masalah kesehatan lainnya. Kondisi ini biasanya disebabkan oleh pola makan yang tidak seimbang, konsumsi makanan tinggi kalori dan rendah nutrisi, serta kurangnya aktivitas fisik.

## 2.3 Balita

Anak-anak yang berusia antara 1 sampai 3 tahun disebut balita. Ini adalah tahap perkembangan pertama setelah masa bayi dan sebelum usia prasekolah. Balita tumbuh dan berkembang dengan cepat pada masa ini di sejumlah bidang, termasuk fisik, kognitif, sosial, dan emosional. Mereka mulai berjalan, mengucapkan kata-kata pertama, dan memperoleh kemampuan sosial dasar.

Pada Usia ini, balita memerlukan pertimbangan khusus dalam hal kesehatan, pola makan, dan stimulasi di sekitar agar perkembangannya maksimal. Interaksinya dengan orang tua, pengasuh lain, dan lingkungannya berdampak besar terhadap perkembangannya di usia ini. Selain itu, penting untuk melakukan pemeriksaan kesehatan secara rutin, termasuk vaksinasi dan kunjungan medis, untuk menjamin pertumbuhan dan perkembangan yang sehat.

## 2.4 Machine Learning

Pembelajaran mesin (ML) sebuah penerapan kecerdasan buatan (AI), adalah untuk menciptakan sistem pembelajaran

mandiri yang tidak memerlukan pemrograman terus-menerus. Sebelum menghasilkan hasil, pembelajaran mesin (ML) memerlukan data, atau data pelatihan. Sederhananya, pembelajaran mesin adalah proses mengajar komputer untuk mencapai tujuan tertentu atau menyelesaikan tugas tertentu berdasarkan data pelatihan atau pengalaman sebelumnya .

Ada beberapa algoritma pembelajaran mesin yang tersedia saat ini yang dapat dibuat dan digunakan untuk berbagai tugas. Sebuah penelitian dilakukan untuk mengevaluasi kinerja beberapa algoritma pembelajaran mesin, seperti *Naïve Bayes* (NB), *Radial Basis Function* (RBF), dan *Support Vector Machine* (SVM), dan *Random Forest* [15].

Pembelajaran mesin (*Machine Learning* ) dapat dibagi menjadi dua konsep utama dalam pembelajaran. Yang pertama adalah *Supervised Learning*, yaitu teknik pembelajaran mesin yang membuat suatu fungsi berdasarkan data yang ada. Dalam pengertian ini dapat dikatakan bahwa teknik ini menggunakan data yang tersedia secara rinci dan terklasifikasi dengan baik, yang akan digunakan untuk membuat model data pada saat proses uji dilakukan dengan menggunakan data baru beserta data yang telah dikumpulkan sebelumnya.

Yang kedua adalah *Unsupervised Learning*, yaitu teknik pembelajaran mesin yang bertujuan untuk merepresentasikan pola masukan yang berasal dari data pelatihan. Perbedaan antara Pembelajaran Tanpa Supervisi dan Pembelajaran yang Diawasi adalah bahwa Pembelajaran Tanpa Supervisi tidak menyertakan klasifikasi data masukan apa pun. Unsupervised learning merupakan teknik penting dalam machine learning karena didasarkan pada sistem kerja yang sama dengan pekerjaan manusia, artinya dalam proses pembelajaran tidak tersedia role model, informasi, atau contoh yang dapat dijadikan model untuk melaksanakan pembelajaran. proses uji coba untuk menyelesaikan suatu masalah dengan data baru [16].

## 2.5 Metode Random Forest

Teknik pembelajaran mesin yang disebut *Random Forest* menggabungkan data dari banyak pohon keputusan untuk mendapatkan hasil yang lebih akurat. Dibandingkan dengan pendekatan Classification and Regression Tree (CART), algoritma ini memiliki beberapa keunggulan. Ini merupakan perpanjangan dari metode CART. Karena Random Forest membuat beberapa pohon keputusan berdasarkan data dan atribut yang dipilih secara acak, kadang-kadang disebut sebagai "hutan" karena banyaknya pohon keputusan. Selanjutnya, semua pohon keputusan yang dibuat digunakan untuk menguji data di *Random Forest*, dan hasilnya ditentukan dengan melihat sebagian besar respons yang diberikan oleh pohon *random forest* [16]. Proses pembentukan pohon keputusan di Random Forest meliputi penentuan atribut akar yang didasarkan pada nilai gain yang dihitung dari entropi. Proses tersebut diulangi pada setiap atribut sampai seluruh instance mempunyai kelas yang sama [7].

Adapun alur dan rumus untuk menentukan pohon keputusan adalah sebagai berikut:  
Rumus entrophy. Adapun rumus entrophy tertera pada persamaan 2.

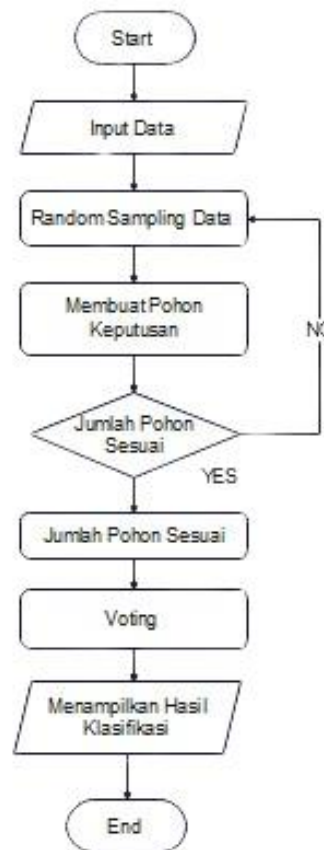
$$Entropy (S) = \sum_{i=1}^n - P_i * \log_2 P_i \quad (2)$$

Persamaan 2 menunjukkan rumus entrophy untuk menentukan pohon keputusan. Dimana Himpunan kasus (S) merupakan kumpulan data atau fenomena yang diamati dalam suatu studi. Fitur (A) dalam konteks ini adalah karakteristik atau variabel yang diamati pada setiap kasus dalam S. Jumlah partisi (n) merujuk pada pembagian atau segmentasi dari himpunan kasus S. Proporsi (Pi) adalah persentase atau bagian dari setiap partisi Si terhadap total himpunan S, menggambarkan distribusi atau komposisi dari S dalam setiap n partisi.

Menentukan Gain menggunakan persamaan berikut:

$$Gain (S, A) = Entrophy (S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entrophy (S_i) \quad (3)$$

Persamaan 3 menunjukkan rumus entrophy untuk menentukan pohon keputusan. Dimana Himpunan kasus (S) merujuk pada total entitas atau data yang diamati dalam konteks analisis. Atribut (A) merupakan karakteristik atau variabel yang diukur atau diamati pada setiap kasus dalam S. Jumlah partisi atribut A (n) adalah pembagian atau segmen yang dibuat berdasarkan nilai-nilai atribut yang berbeda dalam A. |Si| mengindikasikan jumlah kasus yang termasuk dalam partisi ke-idari atribut A, sementara |S| mewakili jumlah total kasus dalam himpunan S secara keseluruhan.



**Gambar 2.** Alur Proses Klasifikasi Random Forest

Alur proses klasifikasi *random forest* ditunjukkan pada gambar 2. Dimana langkah pertama dimulai dari input data, Random sampling data, membuat pohon keputusan. Setelah membuat pohon keputusan dan jumlah pohon keputusan belum sesuai maka akan terjadi proses perulangan random sampling data sampai menemukan pohon keputusan yang diinginkan. Jika sudah menemukan jumlah pohon yang sesuai maka dilanjutkan dengan voting pohon yang terbaik dan diakhiri dengan menampilkan hasil klasifikasi model *random forest*.

## 2.6 Pengujian

Pengujian Evaluasi kinerja model klasifikasi dapat memakai *confusion matrix* [17]. *Confusion matrix* ialah satu diantara teknik yang dipakai guna menilai seberapa baik model klasifikasi bekerja yang mengukur seberapa efektif model memprediksi kelas data. Empat komponen utama *confusion matrix* ialah *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Komponen-komponen ini memungkinkan peneliti untuk menghitung metrik penilaian kinerja seperti presisi, akurasi, f1-Score, dan perolehan [7].

Berdasarkan *confusion matrix*, sejumlah ukuran dapat digunakan untuk menilai performa model, termasuk f1-Score, akurasi, presisi, dan perolehan [18]. Adapun rumus *confusion matrix* dapat dilihat melalui persamaan berikut:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1 - Score} = 2x \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

- Presisi mengacu pada seberapa dekat hasil pengukuran satu sama lain atau seberapa konsisten hasil pengukuran
- Akurasi mengacu pada seberapa dekat hasil pengukuran dengan nilai sebenarnya atau nilai referensi
- Recall digunakan untuk menilai seberapa baik suatu model bisa mendeteksi sebagian besar sampel positif yang sebenarnya [7].
- F1 Score adalah metrik kinerja yang digunakan untuk mengevaluasi model klasifikasi, terutama ketika data memiliki distribusi kelas yang tidak seimbang.

### 3. HASIL DAN PEMBAHASAN

Pada penelitian ini, kondisi gizi balita di Posyandu Kecamatan Nlegok Kabupaten Blitar dikategorikan menggunakan algoritma *Random Forest*. Tahapan algoritma random forest dalam menyelesaikan masalah meliputi pengumpulan data, pengolahan data, dan representasi model hutan acak, dan melakukan evaluasi hasil dengan menggunakan *confusion matrix*. Beberapa faktor yang berkaitan terhadap status gizi, seperti jenis kelamin, berat badan, tinggi badan, usia, ZS menurut TB/U, BB/U, dan TB/U, dan label status gizi, digunakan dalam penelitian ini sebagai karakteristik kategorisasi. Berdasarkan karakteristik tersebut, algoritma *Random Forest* digunakan untuk menilai data dan mengkategorikan status gizi. Penelitian ini bertujuan guna mengidentifikasi variabel-variabel yang mempengaruhi status gizi balita dan membuat model klasifikasi yang dapat mengkategorikan status gizi secara akurat dengan menggunakan informasi yang ada. Adapun hasil yang diperoleh dari algoritma random forest adalah 88,6%, presisi 88,1%, recall 88,6%, dan F1-Score 88,2%.

#### 3.1 Pengumpulan Data

Tabel 2 menunjukkan Data Status Gizi yang akan dipergunakan didalam penelitian. Data status gizi balita tersebut didapat dari ahli gizi di Puskesmas Nlegok. Data yang dikumpulkan berjumlah 2.542 informasi status gizi balita. NO, Nama Balita, Jenis Kelamin, Usia, Berat Badan, Tinggi Badan, Nilai ZS (tinggi badan/umur, berat badan/tinggi badan, berat badan/umur) dan label status gizi (tinggi badan/umur, berat badan/tinggi badan, berat badan/umur) adalah atribut yang terdapat dalam kumpulan data.

Tabel 2. Data gizi balita

| NO   | NAMA                           | Jenis Kelamin | Tgl.Lahir |     |      | Umur                       | BB    | TB   | NILAI Z SCORE |       |       | STATUS GIZI   |                    |            |
|------|--------------------------------|---------------|-----------|-----|------|----------------------------|-------|------|---------------|-------|-------|---------------|--------------------|------------|
|      |                                |               | Tgl       | Bln | Thn  |                            |       |      | TB/U          | BB/U  | BB/TB | TB/U          | BB/U               | BB/TB      |
| 1    | Aldebara                       | L             | 30        | 6   | 2021 | 0 Tahun - 6 Bulan - 12Hari | 7     | 65   | -1,52         | -1,31 | -0,46 | Normal        | Berat Badan Normal | Gizi Baik  |
| 2    | Rafaeza                        | L             | 31        | 7   | 2019 | 2 Tahun - 5 Bulan - 11Hari | 13,9  | 97   | 1,64          | 0,45  | -0,57 | Normal        | Berat Badan Normal | Gizi Baik  |
| 3    | M Alik                         | L             | 30        | 11  | 2017 | 4 Tahun - 1 Bulan - 10Hari | 14,9  | 101  | -0,73         | -0,83 | -0,59 | Normal        | Berat Badan Normal | Gizi Baik  |
| 4    | Izma Lubna                     | L             | 31        | 5   | 2018 | 3 Tahun - 7 Bulan - 12Hari | 13,5  | 83   | -4,4          | -1,16 | 2,2   | Sangat Pendek | Berat Badan Normal | Gizi Lebih |
| 2537 | ELGHAFI JALALUDIN R            | L             | 25        | 11  | 2019 | 2 Tahun - 1 Bulan - 18Hari | 13,5  | 94   | 1,75          | 0,69  | -0,29 | Normal        | Berat Badan Normal | Gizi Baik  |
| 2538 | DIKE VIRGINNIS SHOLIHAH        | P             | 20        | 3   | 2019 | 2 Tahun - 9 Bulan - 24Hari | 13,7  | 98,5 | 1,35          | 0,15  | -0,87 | Normal        | Berat Badan Normal | Gizi Baik  |
| 2539 | JAZDIEN BARRATA                | L             | 11        | 4   | 2019 | 2 Tahun - 8 Bulan - 30Hari | 14,5  | 99   | 1,39          | 0,4   | -0,48 | Normal        | Berat Badan Normal | Gizi Baik  |
| 2540 | UAURELLIA GYSKA NAYSILLA PUTRI | P             | 9         | 6   | 2018 | 3 Tahun - 7 Bulan - 4Hari  | 16,95 | 104  | 1,03          | 0,8   | 0,32  | Normal        | Berat Badan Normal | Gizi Baik  |
| 2541 | NABILA RAMADHANI P             | P             | 27        | 5   | 2018 | 3 Tahun - 7 Bulan - 21Hari | 19,3  | 100  | -0,03         | 1,66  | 2,49  | Normal        | Risiko Lebih       | Gizi Lebih |
| 2542 | MALWI MUSTOFA                  | L             | 18        | 10  | 2018 | 3 Tahun - 2 Bulan - 24Hari | 13    | 97,5 | -0,1          | -1,07 | -1,53 | Normal        | Berat Badan Normal | Gizi Baik  |

#### 3.2 Pre-Processing Data

Langkah awal dan paling penting dalam menyiapkan data untuk pemodelan klasifikasi adalah pre-processing. Beberapa tahapan persiapan yang dilakukan dalam penelitian ini meliputi [18]:

##### a. Seleksi data.

Sejumlah faktor antara lain jenis kelamin, umur, BB, TB, Nilai ZS (tinggi badan/umur, berat badan/tinggi badan, berat badan/umur) dan label status gizi (tinggi badan/umur, berat badan/tinggi badan, berat badan/umur), mempunyai pengaruh yang signifikan terhadap pengolahan klasifikasi kondisi gizi.

##### b. Pembersihan data

Pada tahap ini memastikan Menghapus kolom yang tidak relevan terhadap klasifikasi dan mengatasi data yang kosong maupun data yang duplikat sehingga tidak mempengaruhi proses klasifikasi.

##### c. Transformation Data

Pada tahap ini melakukan proses mengubah data non-numerik menjadi format numerik agar dapat digunakan dalam analisis kemudian membagi dataset ke variabel X dan Y.

#### 3.3 Implementasi Random Forest

Algoritma Random Forest diimplementasikan menggunakan Google Colab dan bahasa pemrograman python serta modul *Scikit-learn*. Seratus pohon dihasilkan untuk digunakan dalam teknik ini. Untuk menemukan cabang yang optimal, entropi dan gain harus dihitung selama proses pembangunan pohon. Seratus pohon digunakan dalam klasifikasi Hutan Acak ini, dengan delapan cabang per pohon. Pembagian optimal ditentukan dengan mengambil

Akar kuadrat dari total fitur. Kualitas pembelahan pohon diukur menggunakan entropi. Setidaknya ada dua sampel per leaf node [19]. Kelas mayoritas menentukan hasil dalam langkah prediksi, yang ditentukan melalui "voting" menggunakan temuan dari semua pohon. Penting diketahui bahwasanya instrumen ini tidak menyajikan data dari tiap pohon keputusan.

#### 3.4 Pengujian

Evaluasi *confusion matrix* dapat digunakan untuk pengujian model *Random Forest*. *Confusion Matrix* menawarkan informasi penting tentang perolehan model, akurasi, presisi, dan skor f1 [19]. Empat skenario pemisahan data dengan persentase pelatihan dan pengujian yang bervariasi digunakan dalam pengujian hutan acak untuk menilai kinerja model dalam kondisi yang beragam. Empat skenario yang digunakan adalah 60% data training dan 40% data testing, 70% data training dan 30% data testing, 80% data training dan 20% data testing, dan skenario terakhir adalah 90% data training dan 10% data testing [19].

Hasil pengujian model *Random Forest* untuk empat skenario tersebut menghasilkan empat *Confusion Matrix* yang berbeda. Masing-masing *Confusion Matrix* menampilkan akurasi, presisi, perolehan, dan f1-score dari pengujian model yang ditampilkan oleh *confusion matrix*.

a. Skenario 60% data training dan 40% data testing.

```
Precision: 0.8400654161863984
Recall: 0.8550295857988166
F1 Score: 0.8435491262528148
Accuracy: 0.8550295857988166
```

**Gambar 2.** Hasil Skenario 60% data training dan 40% data testing

Gambar 2 menunjukkan bahwa evaluasi terhadap model yang dibangun menggunakan skenario pemisahan data training sebesar 60% dan data testing sebesar 40%, diperoleh hasil yang cukup baik. Model yang dikembangkan mampu mencapai tingkat akurasi sebesar 85.5%, yang menunjukkan seberapa baik model dalam melakukan prediksi yang benar secara keseluruhan. Selain itu, nilai precision sebesar 84.0% mengindikasikan bahwa dari semua prediksi positif yang dibuat oleh model, sebanyak 84.0% di antaranya adalah benar-benar positif. Recall atau tingkat keberhasilan dalam mengidentifikasi keseluruhan kelas positif mencapai 85.5%, sementara nilai F1-score sebesar 84.4% menunjukkan keseimbangan antara precision dan recall. Hasil ini menegaskan bahwa model yang dikembangkan memiliki kinerja yang baik dalam melakukan klasifikasi data.

b. Skenario 70% data training dan 30% data testing.

```
Precision: 0.8445642180859
Recall: 0.8580814717477004
F1 Score: 0.8480111406951836
Accuracy: 0.8580814717477004
```

**Gambar 3.** Hasil Skenario 70% data training dan 30% data testing

Gambar 3 menunjukkan skenario di mana data training sebesar 70% dan data testing sebesar 30%, hasil evaluasi terhadap model menunjukkan performa yang sangat baik. Model ini berhasil mencapai tingkat akurasi sebesar 85.8%, yang mengindikasikan tingkat kebenaran prediksi secara keseluruhan. Precision, yang mengukur proporsi prediksi positif yang benar dari keseluruhan prediksi positif, mencapai 84.5%. Recall, atau tingkat keberhasilan dalam mengidentifikasi keseluruhan kelas positif, juga mencapai 85.8%. Selain itu, nilai F1-Score sebesar 84.8% menunjukkan keseimbangan yang baik antara precision dan recall. Hasil ini menunjukkan bahwa model memiliki kemampuan yang kuat dalam melakukan klasifikasi pada data, serta memberikan keyakinan bahwa model dapat digunakan secara efektif untuk tujuan klasifikasi di masa mendatang.

c. Skenario 3 dengan 80% data training dan 20% data testing.

```
Precision: 0.8710159294623107
Recall: 0.8777120315581854
F1 Score: 0.8733529594413392
Accuracy: 0.8777120315581854
```

**Gambar 4.** Hasil Skenario 80% data training dan 20% data testing

Gambar 4 menunjukkan skenario di mana data training sebesar 80% dan data testing sebesar 20%, untuk pengujian dalam klasifikasi status gizi balita, model yang dikembangkan menunjukkan performa yang sangat baik. Tingkat akurasi model mencapai 87.8%, mencerminkan kemampuan model dalam membuat prediksi yang tepat secara keseluruhan. Precision, yang mengukur ketepatan prediksi positif dari semua prediksi positif yang dibuat, mencapai 87.1%, menunjukkan bahwa sebagian besar prediksi positif yang dibuat oleh model adalah benar. Recall, yang mengukur seberapa baik model dalam mengidentifikasi keseluruhan kelas positif, juga tinggi yaitu 87.8%. Selain itu, nilai F1-Score sebesar 87.3% menunjukkan keseimbangan yang baik antara precision dan recall. Hasil ini menegaskan bahwa model memiliki kinerja yang sangat baik dalam mengklasifikasikan data, baik pada data yang digunakan untuk melatih model maupun pada data yang diuji yang belum pernah dilihat sebelumnya. Dengan demikian, model ini dapat diandalkan untuk aplikasi praktis dalam berbagai bidang seperti pengenalan pola, prediksi, klasifikasi dan pengambilan keputusan

berbasis data.

d. Skenario 4 dengan 90% data training dan 10% data testing.

```
Precision: 0.8812709619678711
Recall: 0.8858267716535433
F1 Score: 0.8820016816079809
Accuracy: 0.8858267716535433
```

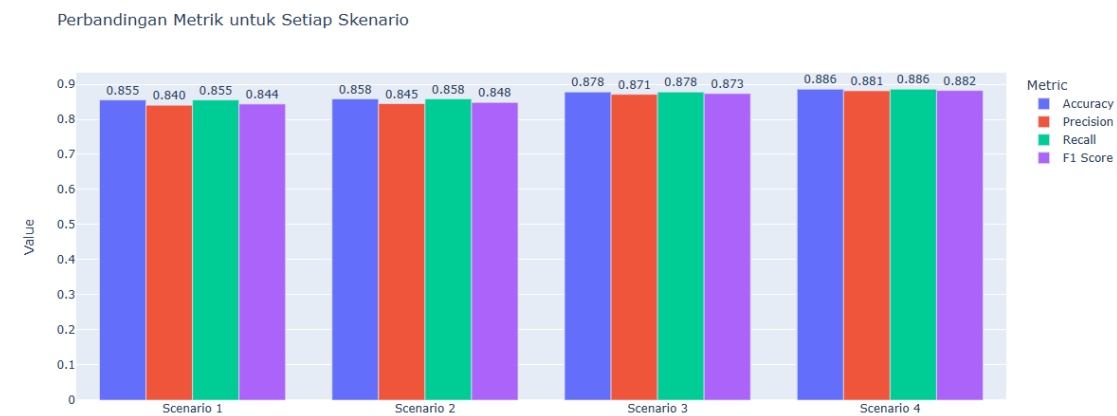
Gambar 5. Hasil Skenario 90% data training dan 10% data testing

Gambar 5 menunjukkan skenario di mana 90% data digunakan untuk melatih model dan 10% sisanya untuk pengujian, model yang dikembangkan berhasil memberikan hasil evaluasi yang sangat baik dalam mengklasifikasikan status gizi balita. Akurasi model mencapai 88.6%, menunjukkan tingkat kebenaran prediksi secara keseluruhan dalam mengidentifikasi status gizi balita. Precision sebesar 88.1% menunjukkan proporsi prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model, sementara recall mencapai 88.6%, mengindikasikan keberhasilan model dalam mengidentifikasi sebagian besar kasus status gizi yang sebenarnya. Selain itu, F1- Score sebesar 88.2% menunjukkan keseimbangan yang baik antara precision dan recall, yang penting untuk mengukur performa model dalam mengklasifikasikan status gizi balita dengan akurat. Hasil ini memberikan keyakinan bahwa model yang dikembangkan dapat menjadi alat yang efektif dalam mendukung upaya pencegahan dan intervensi gizi balita, serta membantu pengambilan keputusan yang tepat dalam konteks kesehatan masyarakat.

Hasil pengujian menunjukkan seberapa baik kinerja model dalam kasus pemisahan data. Model Random Forest memiliki kinerja terbaik ketika terdapat 90% data pelatihan dan 10% data pengujian [7]. Hal ini menunjukkan bahwa skenario ini memberikan hasil terbaik dalam mengkategorikan status gizi balita. Meskipun tidak sebaik skenario dengan 90% data pelatihan dan 10% data pengujian, skenario lain masih memiliki kinerja yang baik. Secara keseluruhan, hasil pengujian menunjukkan bahwa model *Random Forest* dapat digunakan secara efektif untuk mengklasifikasikan status gizi balita dan kinerja model mungkin dipengaruhi oleh kondisi pemisahan data yang sesuai.

### 3.5 Perbandingan Nilai Confusion Matrix

Setelah melakukan pengujian model *Random Forest* menggunakan *confusion matrix* dengan membandingkan empat skenario pemisahan data latih dan data uji. Gambar 4 menunjukkan perbedaan hasil evaluasi model *Random Forest* dalam mengkategorikan status gizi balita. Metrik yang digunakan termasuk akurasi, precision, recall, dan F1-score.



Gambar 6. Perbandingan confusion matrix

Gambar 6 menunjukkan bahwa perbandingan empat skenario yang membagi data pelatihan dan data pengujian untuk mengevaluasi model *Random Forest* menggunakan *confusion matrix* [20]. Meskipun nilai-nilai dalam keempat skenario tersebut berbeda-beda, namun tidak banyak berubah. Namun skenario dengan distribusi data latih 90% dan distribusi data pengujian 10% sehingga memperoleh nilai paling besar, dengan akurasi 88,6%, presisi 88,1%, recall 88,6%, dan F1-Score 88,2%.

## 4. KESIMPULAN

Bisa diambil kesimpulan bahwasanya algoritma Random Forest mempunyai akurasi yang baik didalam mengidentifikasi status gizi anak berdasarkan analisis dan pembahasan yang telah dilakukan. Metode *Random Forest* diuji menggunakan data status gizi balita, dan hasilnya memperlihatkan bahwasanya dengan memakai 90% data training serta 10% data testing dapat memberikan hasil skenario terbaik dengan akurasi 88,6%, presisi 88,1%, recall 88,6%, dan F1-Score 88,2%. Untuk meningkatkan universalitas dan keakuratan model, lebih banyak data dari berbagai demografi dan wilayah harus



dikumpulkan, sesuai dengan saran penelitian. Model ini dapat lebih memahami variasi status gizi balita yang lebih luas dengan menyertakan data yang lebih beragam. Menggunakan teknik alternatif adalah cara lain untuk meningkatkan efektivitas.

## REFERENCES

- [1] M. K-means, O. Purwaningrum, Y. Y. Putra, and A. A. Arifiyanti, "Penentuan Kelompok Status Gizi Balita dengan Menggunakan," vol. 15, no. 2, pp. 129–136, 2021.
- [2] E. N. Candra, I. Cholissodin, and R. C. Wihandika, "Klasifikasi Status Gizi Balita menggunakan Metode Optimasi Random Forest dengan Algoritme Genetika (Studi Kasus: Puskesmas Cakru)," vol. 6, no. 5, pp. 2188–2197, 2022.
- [3] H. Saleh, M. Faisal, and R. I. Musa, "Klasifikasi Status Gizi Balita Menggunakan Metode K-Nearest Neighbor," *Simtek J. Sist. Inf. dan Tek. Komput.*, vol. 4, no. 2, pp. 120–126, 2019, doi: 10.51876/simtek.v4i2.60.
- [4] H. S. T. B. Hafizan and A. N. S. T. B. Putri, "Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi," *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 1, no. 2, pp. 68–72, 2020.
- [5] A. Y. Atmanegara, I. A. E. Zaeni, D. Lestari, and Y. S. A. Gumilang, "Alat Pengukur Status Gizi Balita Berdasarkan Berat dan Panjang Badan Menggunakan Indeks Antropometri Dengan Metode Logika Fuzzy," *JASIEK (Jurnal Apl. Sains, Informasi, Elektron. dan Komputer)*, vol. 4, no. 1, pp. 9–22, 2022, doi: 10.26905/jasiek.v4i1.8352.
- [6] Kemenkes RI, *Injeksi 2018*. 2020.
- [7] M. R. Akbar Ariyadi, S. Lestanti, and S. Kirom, "Klasifikasi Balita Stunting Menggunakan Random Forest Classifier Di Kabupaten Blitar," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3846–3851, 2024, doi: 10.36040/jati.v7i6.7822.
- [8] T. Sathiya, R. Reenadevi, and B. Sathiyabhama, "Random Forest Classifier based detection of Parkinson's disease," vol. 25, no. 5, pp. 2980–2987, 2021.
- [9] O.- Pahlevi, A.- Amrin, and Y.- Handrianto, "Implementasi Algoritma Klasifikasi Random Forest Untuk Penilaian Kelayakan Kredit," *J. Infortech*, vol. 5, no. 1, pp. 71–76, 2023, doi: 10.31294/infortech.v5i1.15829.
- [10] D. Chauhan, R. Walia, C. Singh, M. Deivakani, M. Kumbhkar, and A. Professor, "Detection of Maize Disease Using Random Forest Classification Algorithm," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 9, pp. 715–720, 2021.
- [11] E. R. Pratama and J. B. B. Darmawan, "Klasifikasi Status Gizi Balita Menggunakan Jaringan Syaraf," *Ris. dan Teknol. Terap.*, pp. 1–10, 2021.
- [12] P. R. Togatorop, M. Sianturi, D. Simamora, and D. Silaen, "Optimizing Random Forest using Genetic Algorithm for Heart Disease Classification," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 13, no. 1, p. 60, 2022, doi: 10.24843/lkjiti.2022.v13.i01.p06.
- [13] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 8–14, 2023, doi: 10.57152/malcom.v3i1.474.
- [14] R. Setiawan and A. Triayudi, "Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web," *J. Media Inform. Budidarma*, vol. 6, no. 2, p. 777, 2022, doi: 10.30865/mib.v6i2.3566.
- [15] C. Chazar and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 12, no. 1, pp. 67–80, 2020, doi: 10.37424/informasi.v12i1.48.
- [16] S. Devella, Y. Yohannes, and F. N. Rahmawati, "Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 310–320, 2020, doi: 10.35957/jatisi.v7i2.289.
- [17] T. Purwa, "Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017)," *J. Mat. Stat. dan Komputasi*, vol. 16, no. 1, p. 58, 2019, doi: 10.20956/jmsk.v16i1.6494.
- [18] E. Prasetyo and K. Nugroho, "Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data," *Techno.Com*, vol. 23, no. 1, pp. 1–10, 2024, doi: 10.62411/tc.v23i1.9779.
- [19] N. A. Pramudhyta and M. S. Rohman, "Perbandingan Optimasi Metode Grid Search dan Random Search dalam Algoritma XGBoost untuk Klasifikasi Stunting," *J. Media Inform. Budidarma*, vol. 8, no. 1, pp. 19–29, 2024, doi: 10.30865/mib.v8i1.6965.
- [20] A. N. Rachmi, "Implementasi Metode Random Forest Dan Xgboost Pada Klasifikasi Customer Churn," pp. 1–101, 2020.