JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58

ALGORITMA MACHINE LEARNING *RANDOM FOREST* DAN *DECISION TREE* DALAM KLASIFIKASI STATUS GIZI BALITA

Taufik Hidayat*[1)], Zalfa Ridha Febriyanti[2)], Sukisno[3)], Asep Hardiyanto Nugroho[4)], Robby Rizky[5)], Zaenal Hakim[6)]

1. Universitas Islam Syekh-Yusuf, Indonesia
2. Universitas Islam Syekh-Yusuf, Indonesia
3. Universitas Islam Syekh-Yusuf, Indonesia
4. Universitas Islam Syekh-Yusuf, Indonesia
5. Universitas Mathla'ul Anwar, Indonesia
6. Universitas Mathla'ul Anwar, Indonesia

Article Info
Kata Kunci: Gizi Balita, Machine Learning, Klasifikasi, Random Forest, Decision Tree, GridSearch

Keywords: *Toddler Nutrition, Machine Learning, Classification, Random Forest, Decision Tree, GridSearch*

Article history:
Received 25 February 2025
Revised 18 March 2025
Accepted 15 April 2025
Available 1 March 2026

DOI :
https://doi.org/10.29100/jipi.v11i1.7545

*Corresponding author.
Taufik Hidayat
E-mail address:
thidayat@unis.ac.id

ABSTRAK
Gizi pada balita merupakan sebuah isu strategis baik ditingkat nasional maupun daerah, terutama *stunting, wasting, underweight* dan *overweight* masih menjadi isu penting di Indonesia, termasuk di Kabupaten Tangerang. Periode usia 12-59 bulan merupakan masa kritis dalam pertumbuhan dan perkembangan anak, sehingga pemantauan dan intervensi gizi yang tepat sangat diperlukan. Tujuan penelitian ini mengembangkan model klasifikasi dengan machine learning pada gizi balita, serta memberikan kontribusi dalam identifikasi dini balita berisiko gizi buruk atau obesitas. Penelitian ini berfokus pada bagaimana menerapkan algoritma machine learning *Random Forest* dan *Decision Tree* untuk klasifikasi status gizi balita, serta bagaimana perbandingan kinerja kedua algoritma tersebut dalam deteksi dini masalah gizi. Dalam model yang telah dibuat, dengan algoritma *Random Forest* dan *Decision Tree* diperoleh hasil yang menunjukkan akurasi yang tinggi, masing-masing mencapai 92%. Dengan menggunakan *GridSearch* untuk menemukan nilai parameter terbaik, akurasi model meningkat menjadi 93% untuk *Random Forest* dan 94% untuk *Decision Tree*. Hasil ini menunjukkan potensi kedua algoritma dalam klasifikasi status gizi balita secara akurat.

ABSTRACT
Nutritional problems in toddlers, especially stunting, wasting, underweight and overweight are still important issues in Indonesia, including in Tangerang Regency. The age period of 12-59 months is a critical period in the growth and development of children, so that appropriate nutritional monitoring and intervention are needed. The purpose of this study is to develop an accurate toddler nutrition prediction model using both algorithms, and to contribute to the early identification of toddlers at risk of malnutrition or obesity. This study focuses on how to apply the Random Forest and Decision Tree algorithms to predict the nutritional status of toddlers, and how the performance of the two algorithms compares in early detection of nutritional problems. In the model that has been created, with the Random Forest and Decision Tree algorithms, the results obtained show high accuracy, each reaching 92%. By using GridSearch to find the best parameter values, the model accuracy increased to 93% for Random Forest and 94% for Decision Tree. These results show the potential of both algorithms in accurately predicting the nutritional status of toddlers.

.

I. PENDAHULUAN

Hasil Survei Status Gizi Indonesia (SSGI) Kementerian Kesehatan menunjukkan bahwa terdapat empat isu gizi yang dialami oleh anak-anak balita di Indonesia, termasuk *stunting, wasting, underweight*, dan *overweight*, angka stunting pada tahun 2022 mencapai 21,6%, melebihi level yang ditetapkan oleh WHO

sebesar 20%[1], meskipun demikian, angka tersebut mengalami penurunan sebesar 2,8 poin dibandingkan tahun sebelumnya, yang menunjukkan adanya perbaikan. Pada tahun 2019, jumlah balita di Indonesia mengalami penurunan sebesar 6,1 poin mencapai 27,7% pada saat itu. Selain itu, hasil survei SSGI 2022 menunjukkan bahwa angka *wasting* pada anak balita di Indonesia meningkat sebesar 0,6 poin dari 7,1% menjadi 7,7% pada tahun lalu[2]. Anak-anak usia 12 bulan - 59 bulan adalah kelompok usia yang paling rentan terhadap masalah gizi dan merupakan periode *golden age* atau periode emas pertumbuhan dan perkembangan otak anak, status gizi merupakan indikator keberhasilan pemberian gizi pada anak. Hal ini juga dapat dianggap sebagai ukuran keberhasilan dalam memenuhi kebutuhan gizi anak, yang tercermin dari berat badan dan tinggi badan mereka[3]. Kekurangan gizi mempengaruhi fungsi otot, melemahkan sistem kekebalan tubuh, mengganggu fungsi otak, dan dapat menyebabkan masalah perkembangan neurologis[4]. Ketidakpastian gizi secara signifikan mempengaruhi dan meningkatkan risiko kematian anak pola makan yang buruk pada balita juga dapat mengurangi kemampuan tubuh untuk melawan penyakit menular, sehingga seringkali mereka terserang penyakit[5].

Dalam mengatasi permasalahan ini, maka diperlukan solusi dari permasalahan terkait dengan gizi balita berguna mempermudah dalam proses prediksi data yang akurat dengan pendekatan *machine learning*. Dalam pendekatan tersebut untuk mengolah data dalam jumlah yang besar dan dapat melakukan pencarian secara otomatis dengan memprediksi data status gizi balita, dimana selama ini masih menggunakan konvensional sebagai laporan. Klasifikasi merupakan salah satu metode yang terdapat pada *machine learning* atau pembelajaran mesin untuk mengelompokkan objek berdasarkan ciri-ciri yang dimiliki oleh objek klasifikasi[6], terdapat beberapa metode algoritma klasifikasi pada *machine learning*, diantaranya yaitu *K-Nearest Neighbors, Naïve Bayes, Support Vector Machine, Decision Tree, Logistic Regression* dan *Random Forest*[7]. Dalam penelitian ini mengusulkan algoritma *Random Forest* dan *Decision Tree*, menggunakan algoritma *Random Forest* yang meningkatkan akurasi dengan mengintegrasikan hasil dari beberapa pohon keputusan, pendekatan ini mengurangi tingkat kesalahan dibandingkan dengan model evaluasi tradisional dan menyediakan data evaluasi yang andal[8]. Pertama, makalah ini mengadopsi algoritma hutan acak untuk konstruksi model, yang dapat mengklasifikasikan dan mengevaluasi perkembangan gizi balita secara objektif menurut data dan standar yang relevan, mengurangi pengaruh faktor subjektif dan empiris, dan mengurangi kesalahan evaluasi. Kedua, dalam evaluasi lebih komprehensif dan ilmiah dan dapat menjadi perspektif multidimensi untuk mengambil perkembangan gizi balita. Ketiga, Orang tua dari balita dapat mengakses hasil setiap dimensi evaluasi secara independen dari hasil evaluasi, serta hasil komprehensif yang sesuai.

Untuk mengatasi klasifikasi status gizi balita, perlu dikembangkan sebuah algoritma untuk pengolahan data, klasifikasi dan evaluasi status gizi balita, algoritma ini digunakan untuk memfasilitasi penentuan langkah-langkah intervensi yang sesuai berdasarkan status gizi pada setiap balita. Beberapa penelitian sebelumnya telah membahas masalah yang serupa dengan yang penulis hadapi, sesuai dengan latar belakang masalah yang telah disampaikan. Penelitian-penelitian berikut menjadi referensi bagi penulis dalam melakukan penelitian ini.

Penelitian yang dilakukan oleh M Ramadani Akbar Ariyadi, Sri Lestanti, Sabitul Kiro, dengan judul Klasifikasi Balita Menggunakan *Random Forest Classifier* di Kabupaten Blitar, dengan 251 dataset balita dengan skenario terbaik yang ditemukan adalah menggunakan 80% data latih dan 20% data tes, dengan akurasi sebesar 90%, presisi sebesar 71.4%, dan recall sebesar 62.5%[9]. Perbedaan dengan penelitian ini yaitu tidak melakukan pra-pemrosesan data dengan tujuan untuk menangani *missing data*, tidak melakukan penanganan keseimbangan data dan dataset yang dipergunakan lebih sedikit dari penelitian ini.

Penelitian yang dilakukan oleh Indah Pratiwi Putri, Terttiaavini, Nur Arminarahmah, dengan judul Analisis Perbandingan Algoritma *Machine Learning* untuk Prediksi Stunting pada Anak, hasil evaluasi menunjukkan bahwa *Random Forest* memiliki akurasi 87.75% dan F1-score 0.922[10], menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Perbedaan dengan penelitian ini adalah pada penanganan data yang tidak seimbang.

Penelitian yang dilakukan Yunita Nur Aprilia, Dian Ahkam Sani, Nanda Martyan Anggadimas, dengan judul Klasifikasi Status Penderita Gizi Stunting Pada Balita Menggunakan Metode *Random Forest*, dengan hasil klasifikasi untuk mengukur serta mengevaluasi hasil kinerja dari metode *Random Forest* pada Klasifikasi Status Penderita Gizi Stunting Pada Balita dengan menggunakan dataset sebanyak 300 data sehingga menghasilkan rata-rata *accuracy* sebesar 81 %, hasil rata-rata *presicion* sebesar 76 %, hasil rata-rata recall sebesar 69 %, dan hasil rata-rata F1 score sebesar 72 %[11]. Yang membedakan dengan penelitian ini adalah penanganan pada data, penelitian ini menggunakan SMOTE dalam penanganan data yang tidak seimbang.

Penelitian yang dilakukan Nur Azizah Eka Budiartia, Abd. Azis Syama, Muh. Arsan Akbara, Muh. Nur Hidayat Al Amanah Muslimina, Asmaul Husnah Nasrullaha, dengan judul Klasifikasi Status Gizi Anak dengan *Decision Tree* Berdasarkan Data Stunting di Kabupaten Gowa, hasil penelitian menunjukkan bahwa model *Decision Tree* yang dibangun memiliki akurasi 90% dalam mengklasifikasikan status gizi anak[12]. Dengan penelitian ini memiliki perbedaan pada penerapan *hyperparameter tuning* dengan *GridSearch*.

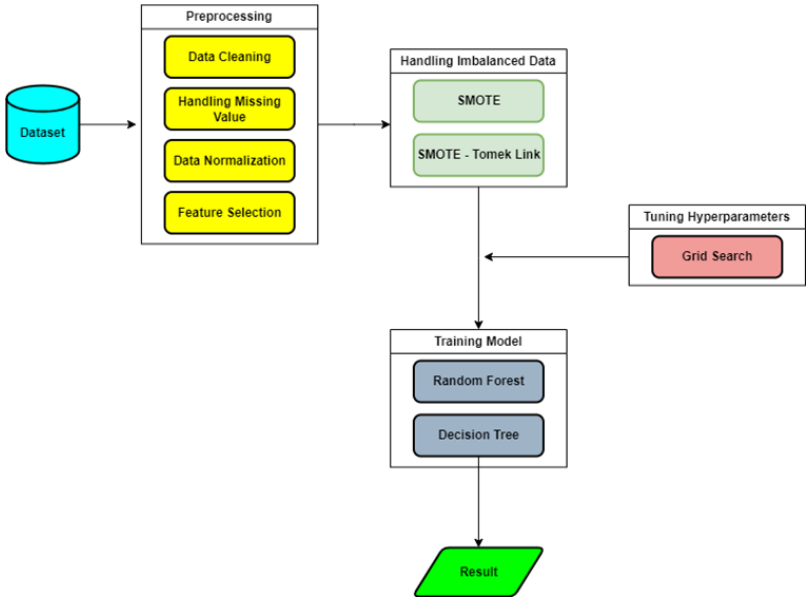
Penelitian yang dilakukan Mutammimul Ula, Ananda Faridhatul Ulva, Mauliza, Muhammad Abdullah Ali, Yumna Rilasmi Said, dengan judul Penerapan *Machine Learning* dalam Penentuan Klasifikasi Gizi Anak dengan

Decision Tree, hasil penelitian diperoleh nilai pengujian rata-rata accuracy berjumlah 81,25%, berdasarkan nilai variabel JK, BB, TB, Umur[13]. Perbedaan dari penelitian ini pada *handling unbalanced data* dan *hyperparameter tuning*.

Dari beberapa penelitian sebelumnya ada yang sama dengan penelitian ini menggunakan handling missing data, penanganan data tidak seimbang dan hyperparameter untuk meningkatkan performa model machine learning yang digunakan. Namun yang mencolok dari beberapa penelitian sebelumnya dengan penelitian ini adalah hasil akurasi, disamping beberapa penelitian sebelumnya juga ada yang tidak menggunakan penanganan data tidak seimbang. Penanganan data yang tidak seimbang sangat penting dalam klasifikasi data nutrisi balita karena memastikan bahwa kelas minoritas, seperti kekurangan gizi dan kekurangan gizi parah, terwakili secara memadai[14]. Penelitian sebelumnya yang menggunakan teknik *machine learning* untuk memprediksi stunting telah menunjukkan hasil yang menjanjikan, namun menghadapi beberapa keterbatasan, seperti kurangnya seleksi fitur yang relevan, yang mengakibatkan *overfitting* dan penurunan akurasi pada dataset besar[15]. *Gridsearch* telah terbukti secara signifikan meningkatkan akurasi model *Random Forest*[16]. *Hyperparameter Tuning* yang tepat dapat secara signifikan meningkatkan kinerja model[17]. Penelitian ini menggunakan pemrosesan data untuk penanganan *missing data*, menggunakan penanganan data tidak seimbang, dan peningkatan performa model dengan hyperparameter tuning yaitu *gridsearch*.

II. METODE PENELITIAN

Jenis penelitian ini menggunakan pendekatan kuantitatif, yaitu penelitian ilmiah yang menggunakan data kuantitatif untuk menguji hipotesis. Adapun tahapan – tahapan dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1.Skema Analisis Proses

Data gizi balita dikumpulkan dari salah satu Puskesmas di daerah tersebut untuk penelitian ini. Sebelum teknik dataset digunakan, data menjalani *preprocessing*. Setelah itu, kumpulan data gizi balita diperiksa untuk mencari data mana saja yang akan digunakan dalam memproses data. Data tersebut diidentifikasi untuk menangani nilai yang kosong atau hilang, dan data tersebut kemudian diubah dalam bentuk skala yang sama dengan teknik normalisasi data serta dilakukan seleksi fitur untuk melihat kolerasi antar setiap data. Dataset ini memiliki 4976 baris dan lima kolom. Selanjutnya, metode *SMOTE* dan *SMOTE – Tomek Link* digunakan untuk mengatasi ketidakseimbangan data. Teknik Oversampling Minority Synthetic (SMOTE) untuk mengatasi ketidakseimbangan kelas dalam kumpulan data, memastikan bahwa model klasifikasi dilatih secara efektif pada kelas mayoritas dan minoritas[18]. Algoritma yang digunakan dalam penelitian ini menggunakan *Random Forest* dan *Decision Tree* digunakan untuk pelatihan dan pengujian. Penelitian ini menggunakan *GridSearch* untuk memilih kombinasi parameter yang optimal. Serta dilakukan *confusion matrix* pada hasil yang didapatkan.

A. Indikator Kategori

Pada penelitian menggunakan dua variabel, dapat dilihat pada tabel 1, menjelaskan terkait indikator kategori variabel terikat diantaranya variabel bebas (*independent variable*) dengan beberapa atribut yaitu

berat, tinggi, usia dan jenis kelamin, serta variabel terikat (*dependent variable*) yaitu sebagai label/ kelas target output pada penelitian ini yaitu atribut Berat Badan/Tinggi Badan (Status Gizi). Adapun kategori dalam status gizi balita diantaranya yaitu : gizi baik, gizi buruk, gizi kurang, gizi lebih, obesitas dan resiko gizi lebih, sedangkan Z-score adalah ukuran statistik yang digunakan untuk menilai status gizi balita berdasarkan indeks antropometri[19], dimana Z-score ini digunakan pula sebagai ambang batasan dalam menentukan status gizi balita, dimana dalam dataset sudah terdapat status gizi balita.

TABEL I
INDIKATOR KATEGORI

Indikator	Status Gizi	Z-Score
BB/TB	Gizi buruk	< -3.0 SD
	Gizi Kurang	-3.0 SD s/d <-2.0 SD
	Gizi Baik	-2.0 SD s/d 1.0 SD
	Resiko Gizi Lebih	1.0 SD s/d 2 SD
	Gizi Lebih	2.0 SD s/d 3.0 SD
	Obesitas	> 3.0 SD

B. *Preprocessing*

Preprocessing data adalah tugas penting dalam pembelajaran mesin yang secara signifikan meningkatkan hasil model[20]. Pengolahan data yang dilakukan dalam penelitian ini, dikarenakan untuk melakukan proses prediksi, dataset harus melalui tahap *preprocessing* untuk memperoleh format data yang sesuai.

1) Data Selection

Data selection sangat penting untuk meningkatkan akurasi model dan mengurangi kompleksitas komputasi. Dengan memilih fitur yang relevan, model dapat fokus pada aspek data yang paling informatif, yang mengarah ke prediksi dan wawasan yang lebih baik[21]. Pada tahap ini memilih atribut - atribut dari dataset gizi balita 12 -59 bulan sebagai indikator yang dapat digunakan dalam mengklasifikasi data untuk menghasilkan prediksi.

2) Data Cleaning

Pembersihan data, adalah untuk mengidentifikasi dan menghilangkan data yang tidak lengkap, duplikat, informasi usang, anomali, nilai yang hilang, dan kesalahan. Proses ini sangat penting karena secara signifikan mempengaruhi kualitas data, yang pada gilirannya mempengaruhi efektivitas model pembelajaran mesin[22].

3) Missing Value

Menangani nilai yang hilang dalam aliran data melibatkan penanganan masalah yang disebabkan oleh kesalahan peralatan, kesalahan jaringan, dan ketidakpastian data[23]. Pada tahap ini dilakukan jika data status gizi balita di Puskesmas Tigaraksa memiliki nilai yang kosong atau tidak terisi yang biasa disebut sebagai *missing value*. Namun setelah melakukan analisa data *checking missing value* tidak terdapat nilai yang kosong atau *null.* Maka dari itu, tidak dilakukan *handling missing value pada proses ini.*

4) Data Normalization

Data normalization merupakan tahapan untuk mengelompokkan dataset berdasarkan ukuran yang kemudian dinormalisasikan dalam rentang [0, 1] agar menjadi bentuk dengan skala yang sama[24].

$$(x\ norm) = \frac{(x-(x))}{((x)-((x)))} \quad (1)$$

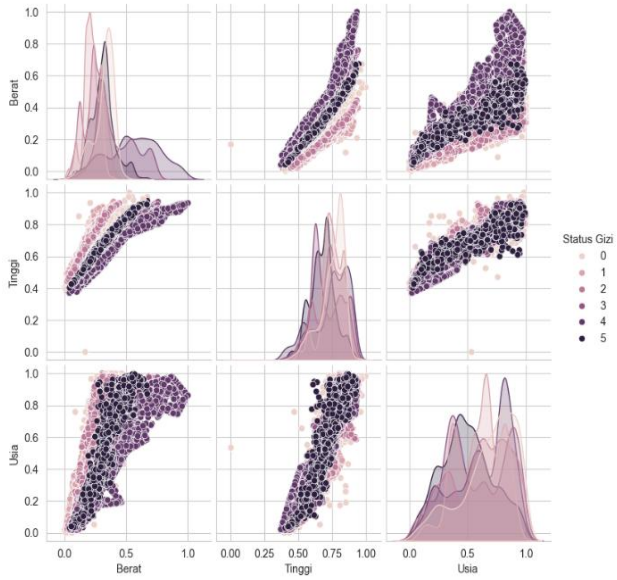
Menghitung data normalisasi dengan atribut dengan berat yaitu 9, nilai minimum yaitu 4,6 dan nilai maksimum berat yaitu 32.

$$(x\ norm) = \frac{(x-(x))}{((x)-((x)))}\ \ x\ norm\ (9) = \frac{(9-4{,}6)}{(-4{,}6)} = 0{,}16$$

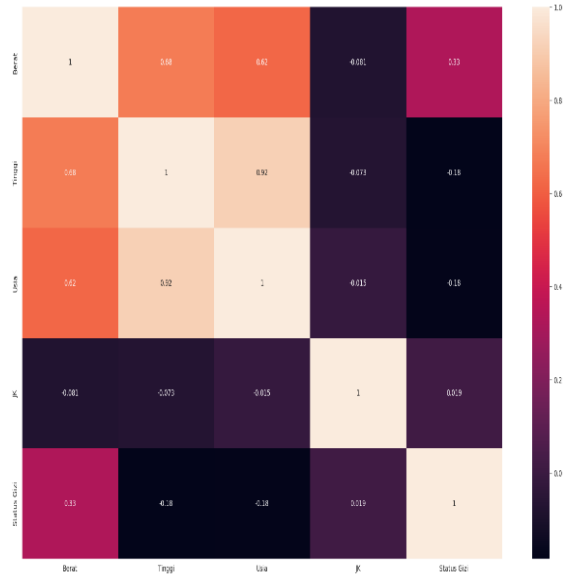
5) Feature Selection

Pemilihan fitur meningkatkan akurasi dan efisiensi model *machine learning* dengan berfokus pada fitur yang paling informatif, yang mengarah pada kinerja prediktif yang lebih baik dan pengurangan kelengkapan[25]. Pada tahapan kolerasi variabel data ini, proses dilakukan dengan menampilkan akurasi model analitik keselurahan antar atribut atau variabel dengan menggunakan *heatmap* sebagai visualisasi data dapat dilihat pada gambar 2 dan gambar 3.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58



Gambar 2. *Pair Grid*



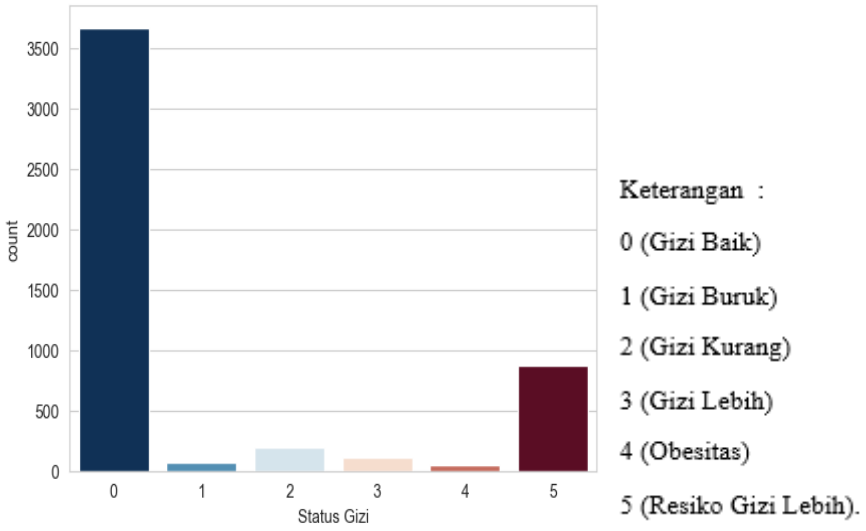
Gambar 3. Kolerasi Data Original

C. *Imbalanced Data*

Proses analisa data dengan menyeimbangkan kelas yang tidak seimbang dengan menggunakan dua metode yaitu SMOTE dan SMOTE – Tomek Link.

1) *Imbalanced Data* dengan *SMOTE*

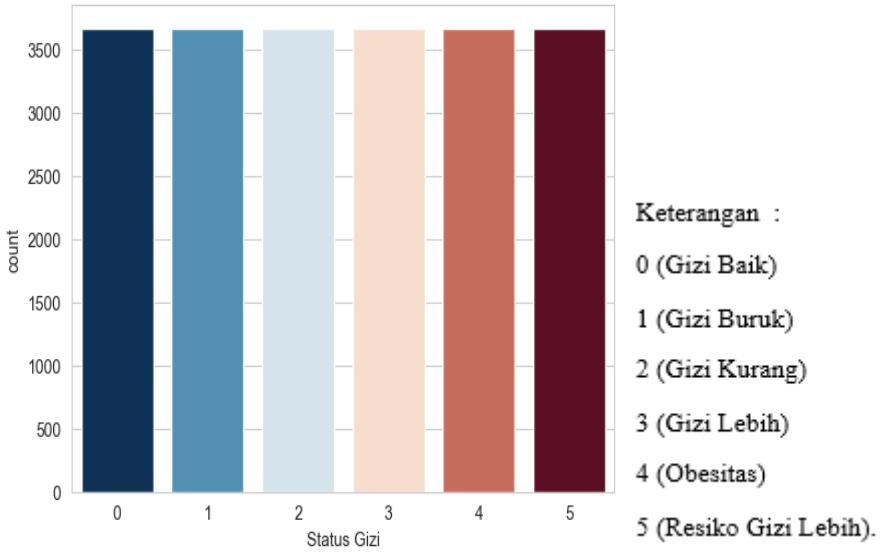
Dalam menggunakan teknik ini data akan diseimbangkan dengan kelas mayoritas dengan cara menambahkan jumlah data yang lebih sedikit hingga jumlah datanya sama ke jumlah data yang lebih banyak.



Gambar 4. *Before SMOTE*

Pada gambar 4 menampilkan hasil yang sangat tidak seimbang kelas antar atribut target. Perbedaan sangat signifikan antar data dengan jumlah data paling banyak yaitu status gizi baik dengan status gizi lainnya. Maka dari itu diperlukan *imbalanced data* untuk menyeimbangkan data yang tidak seimbang pada data status gizi balita.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58

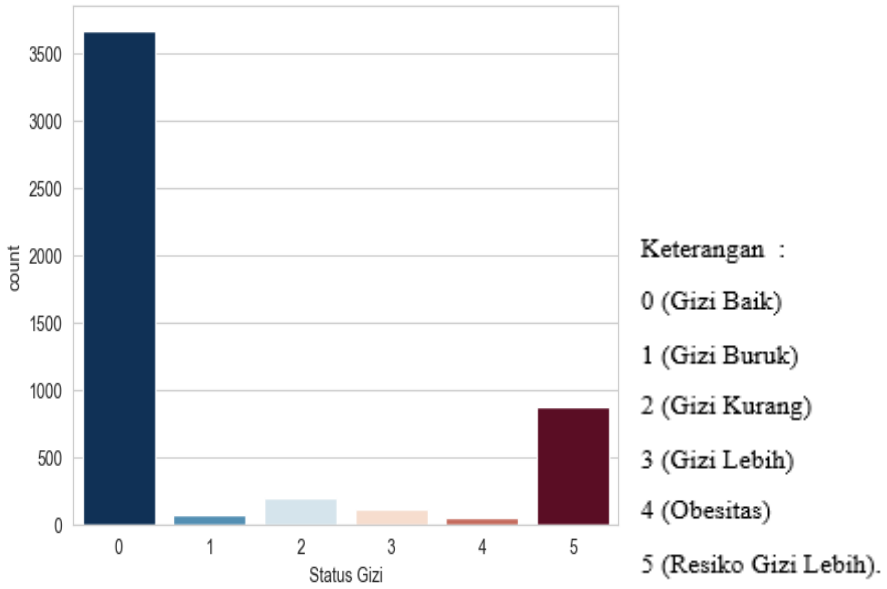


Gambar 5. *After SMOTE*

Pada gambar 5, dapat dilihat hasil setelah dilakukan penanganan data tidak seimbang dengan *SMOTE* terdapat perbedaan yaitu pada sebelum dan sesudah data diseimbangkan, Jumlah sampel per kelas setelah *SMOTE: Counter*({1: 3671, 0: 3671, 2: 3671, 5: 3671, 3: 3671, 4: 3671}).

2) *Imbalanced Data* dengan *SMOTE – Tomek Link*

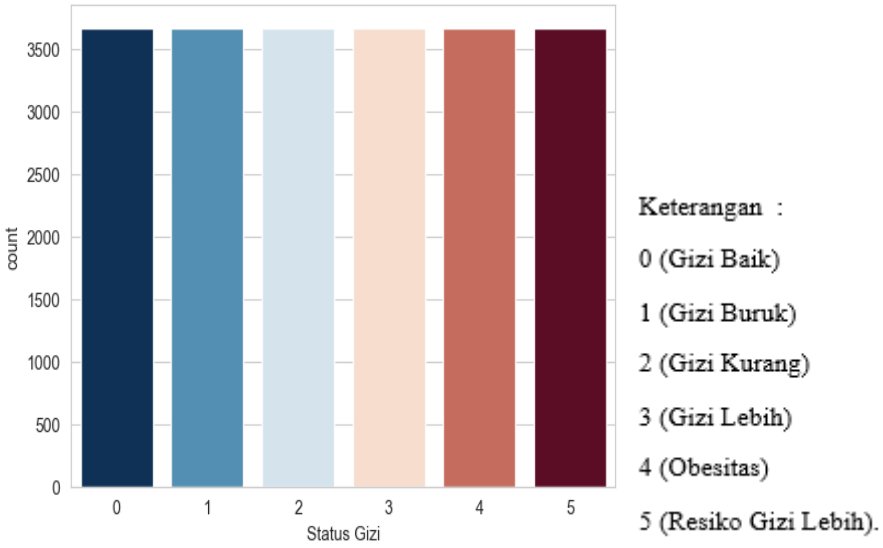
Dalam tahap ini menyimbangkan data yang tidak seimbang dengan *SMOTE – Tomek Link*. Teknik ini menggabungkan dua metode, yaitu *oversampling* dan *undersampling*, untuk membersihkan data dan meningkatkan akurasi klasifikasi, khususnya untuk kelompok data yang lebih kecil.



Gambar 6. Atribut target sebelum *SMOTE – Tomek Link*

Pada gambar 6. yaitu target sebelum di *SMOTE-Tomek Link* menampilkan hasil yang sangat tidak seimbang kelas antar atribut target. Dengan menggunakan teknik *SMOTE-Tomek Link* dengan gabungan teknik *oversampling* dan *undersampling. Oversampling* dengan *SMOTE* menambahkan data baru ke kelompok kecil, namun hal ini dapat menyebabkan kelompok data saling tumpang tindih. Oleh karena itu, *Tomek-Link* dengan *undersampling* digunakan untuk menghapus data yang tumpang tindih ini dengan cara mengidentifikasi pasangan titik data dari kelompok yang berbeda yang berdekatan satu sama lain.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58



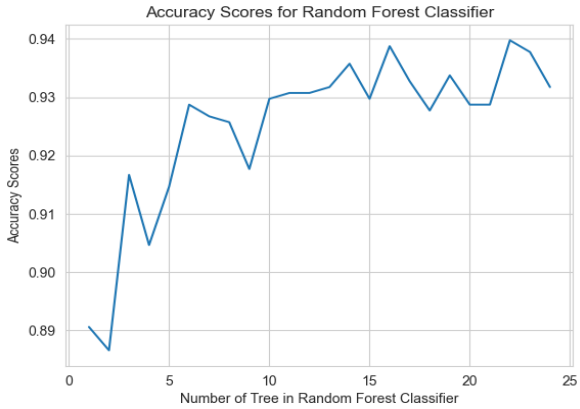
Gambar 7. Atribut target setelah *SMOTE – Tomek Link*

Pada gambar 7. Hasil setelah dilakukan penanganan ketidak seimbangan data dengan teknik *SMOTE – Tomek Link* terdapat perbedaan yaitu pada sebelum dan sesudah data diseimbangkan, *After Counter* ({0: 3600, 1: 3600, 3: 3600, 4: 3600, 2: 3600, 5: 3600,}).

III. HASIL DAN PEMBAHASAN

Penerapan dilakukan secara bertahap. Pertama akan dilakukan dengan menggunakan algoritma *Random Forest* kemudian menggunakan *Decision Tree*. Berikut contoh kinerja dari masing masing algoritma sebagai berikut.

A. Kinerja Model *Random Forest*

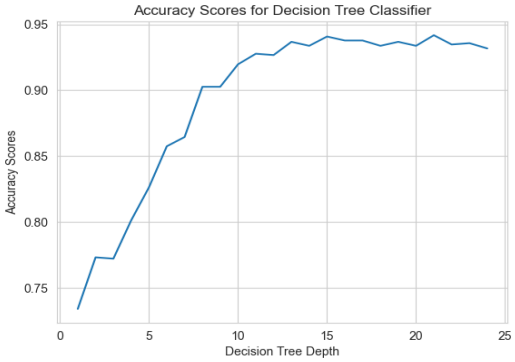


Gambar 8. Model *Random Forest*

Pada gambar 8. Menampilkan grafik dengan mengidentifikasi jumlah pohon terbaik dalam mengaplikasikan model *Random Forest*. Semakin banyak pohon yang ditambahkan, akurasi model cenderung meningkat pesat karena setiap pohon menganalisis bagian data dan atribut yang berbeda, sehingga mengurangi *overfitting* dan meningkatkan kemampuan model dalam memprediksi data baru. Berdasarkan grafik tersebut, jumlah pohon ideal berkisar pada kelipatan 22 karena akurasi sudah optimal dan stabil pada rentang tersebut dengan nilai akurasi 0,92.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58

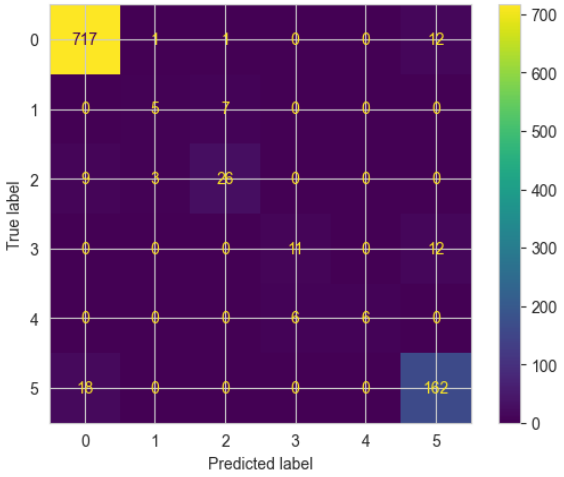
B. Kinerja Model *Decision Tree*



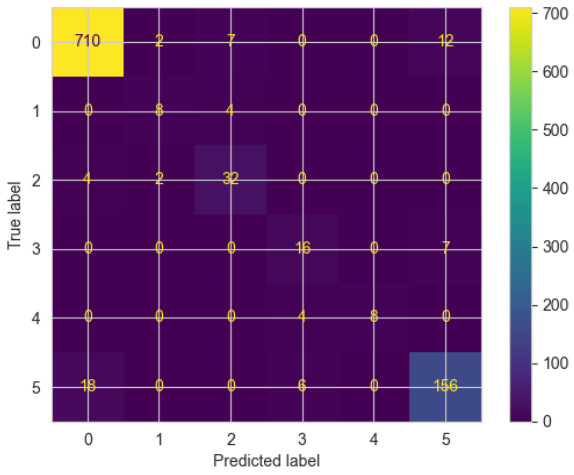
Gambar 9. Model *Decision Tree*

Pada gambar 9 menampilkan grafik dengan mengidentifikasi jumlah pohon terbaik dalam model *Decision Tree*. Berdasarkan grafik tersebut, jumlah pohon ideal berkisar antara 10-15 karena akurasi sudah optimal dan stabil pada rentang tersebut dengan nilai akurasi 0,92 pada lipatan ke-15.

C. Parameter *GridSearch*



Gambar 10. *GridSearch Random Forest*



Gambar 11. *GridSearch Desicison Tree*

Pada tahap ini dilakukan untuk menemukan kombinasi parameter terbaik dengan *GridSearch* akan menjalankan perulangan yang sama. Setelah memiliki kombinasi terbaik, lalu dilakukan *running* pada semua data latih untuk membangun satu model baru menggunakan pengaturan kombinasi parameter terbaik. Tahap ini dilakukan dengan pemodelan yang sama yaitu algoritma *Random Forest* dan *Decision Tree*. Namun perbedaannya terletak dalam menentukan parameter ini.

D. *Result Predict*

Berdasarkan hasil perbandingan dari hasil optimal antara sebelum dan sesudah menggunakan *Gridsearch*, hasilnya dapat dillihat pada tabel 2 menunjukan bahwa hasil *Random Forest* sebelum menggunakan *Gridsearch* menghasilkan nilai *accuracy* 92%, *precision* 78%, *recall*(*sensitivity*) 65%, *F-1 Score* (*harmonic mean* antara *precision* dan *recall*) 70%. dan *Decision Tree* sebelum menggunakan *Gridsearch* menghasilkan nilai *accuracy* 92%, *precision* 78%, *recall*(*sensitivity*) 70%, *F-1 Score* (*harmonic mean* antara *precision* dan *recall*) 74%. Dari kedua algoritma memiliki hasil *accuracy* dan *precision* yang sama, namun berbeda pada nilai hasil recall dan F-1 Score dikarenakan pada algoritma *random forest* lebih sedikit menemukan sampel positifnya daripada algoritma *decision tree* sehingga nilai *recall* dan *F-1 Score* dari decision tree lebih tinggi.

Untuk mengetahui hasil setelah menggunakan *Gridsearch* dapat dilihat pada tabel 3, ada peningkatan nilai hasil yang diperoleh masing-masing algoritma, untuk *random forest* menjadi 93% untuk nilai *accuracy*, 82% untuk nilai *precision*, 66% untuk nilai , dan 72% untuk nilai *F-1 score*. Sedangkan pada algoritma *decision tree* nilai *accuracy* sebesar 94%, nilai *precision* sebesar 82%, nilai *recall* 78%, dan *F-1 score* sebesar 79%. Penelitian ini mengusulkan menggunakan *Gridsearch cross validation*, namun untuk menghindari overfitting pada penelitian ini belum dibahas detail, dan pada penelitian ini belum menerapkan post-pruning untuk mengurangi kompleksitas struktur pohon

sambil mempertahankan akurasi klasifikasi, dan belum menguji dengan data independen. Karen overfitting itu terjadi bilamana model terlalu kompleks memiliki banyak parameter atau terlalu dalam pada pohon keputusan, jumlah data pelatihan terlalu sedikit dibandingkan dengan kompleksitas model, dan jumlah fitur terlalu banyak.

TABEL II
RESULT PREDICT WITHOUT GRIDSEARCH

Model	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F-1 Score (%)*
Random Forest	92%	78%	65%	70%
Decision Tree	92%	78%	70%	74%

TABEL III
RESULT PREDICT WITH GRIDSEARCH

Model	*Accuracy (%)*	*Precision (%)*	*Recall (%)*	*F-1 Score (%)*
Random Forest	93%	82%	66%	72%
Decision Tree	94%	82%	78%	79%

Precision (Ketepatan) mengukur seberapa banyak dari prediksi positif yang benar-benar benar. Nilai precision tinggi berarti model jarang memberikan prediksi positif yang salah (false positive rendah), presisi tinggi menunjukkan perkiraan kepadatan yang andal, mencerminkan lebih sedikit positif palsu[26]. *Recall* (Sensitivitas) mengukur seberapa banyak dari kasus positif yang berhasil dideteksi oleh model. Jika recall rendah, berarti ada banyak kasus positif yang terlewatkan (false negative tinggi). Dalam penelitian ini Recall lebih rendah daripada *Precision*, ini berarti model lebih konservatif dalam mengklasifikasikan kasus positif. Model yang dipilih hanya memprediksi status gizi balita sebagai berisiko, sehingga beberapa kasus stunting yang sebenarnya tidak terdeteksi (false negative tinggi) dalam dataset, hal ini pula dapat menjadikan pengembangan untuk penelitian selanjutnya, yaitu dengan menaikan nilai Recall dari Precision, untuk dapat mencakup lebih banyak lagi balita yang membutuhkan intervensi dalam penanganan gizi balita.

E. Perbandingan Kinerja

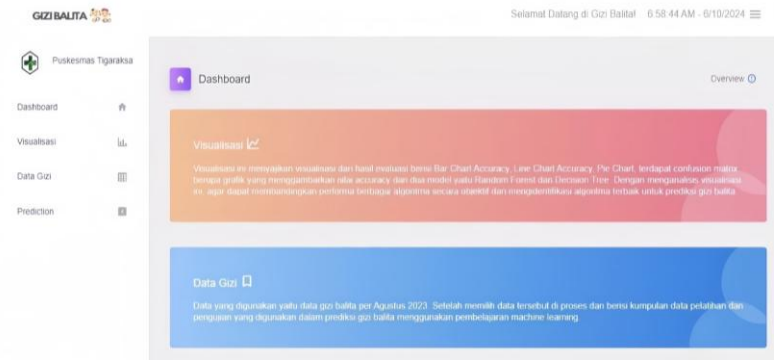
Peningkatan kinerja yang terjadi pada model *Random Forest* dan model *Decision Tree* telah mencapai nilai performa yang sangat baik. Dapat dilihat pada tabel 4. peran *GridSearch* memberikan kepada model untuk meningkatkan kinerja model tersebut, perbedaan dari penelitian sebelumnya bahwa penelitian ini memiliki nilai kinerja yang lebih tinggi dari pada penelitian sebelumnya (dijabarkan pada bagian pendahuluan), selain ada beberapa penelitian sebelumny tidak menggunakan keseimbnagan data dan tidak menggunakan *hyperparameter tuning*.

TABEL IV
PERBANDINGAN KINERJA

Model	Perbandingan Hasil Prediksi (%)	
	No GridSearchCV	*GridSearchCV*
Random Forest	0.92	0.93
Decision Tree	0.92	0.94

F. Hasil Impementasi

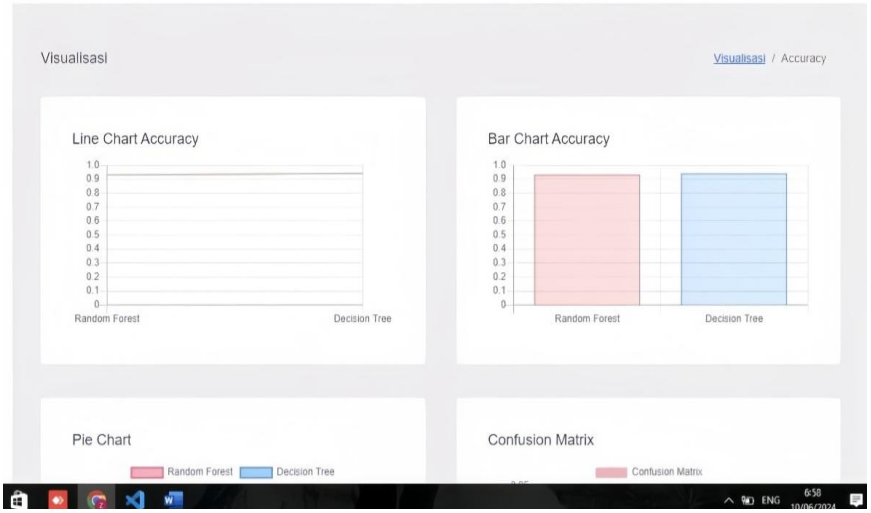
Penelitian ini mengusulkan rancang bangun prototype dengan tujuan untuk dapat diterapkan pada tenaga kesehatan dalam merekam. Mengolah dan menganalisis data, adapun prototype prediksi gizi balita menggunakan algoritma *Random Forest* dan *Decision Tree*, secara lengkap dapat dilihat pada gambar 12 sampai gambar 16.



Gambar 12. Halaman *Dashboard*

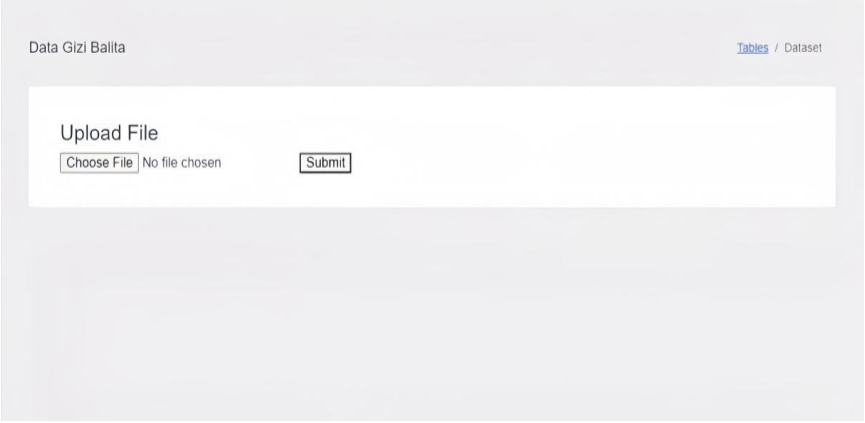
JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58

Pada gambar 12 menampilkan dashboard dari prototype yang dibangun untuk sebagai halaman depan, yang mana terdapat 4 menu yang disajikan.

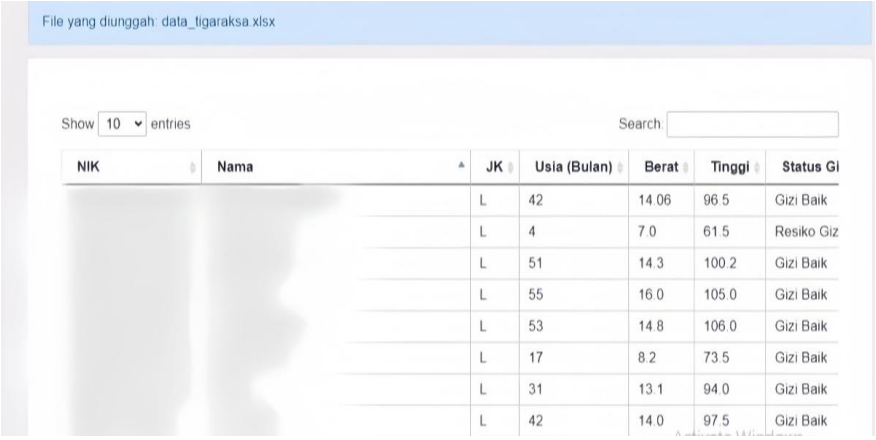


Gambar 13. Halaman Visualisasi

Pada gambar 13 menampilkan halaman visualisasi berguna untuk mengetahui hasil data yang telah diolah ditampilkan dalam berupa grafik untuk lebih mudah dalam membaca informasi.

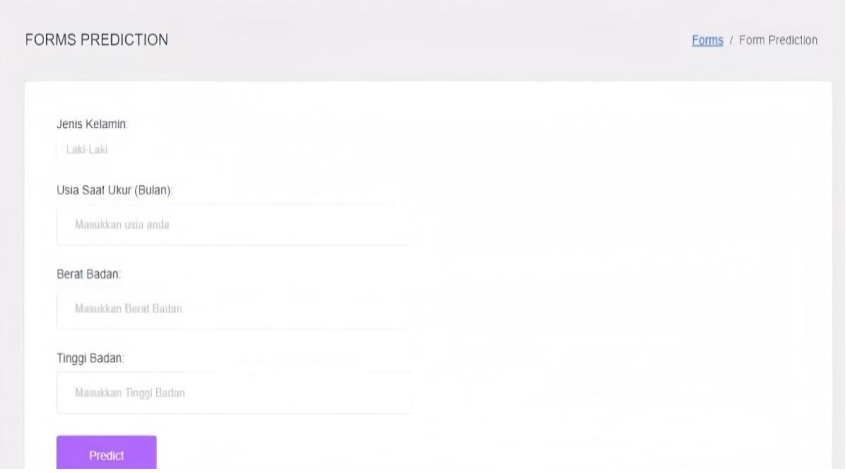


Gambar 14. *Upload File*



Gambar 15. Halaman Data Gizi

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 11, No. 1, March 2026, Pp.47-58



Gambar 16. Halaman *Prediction*

IV. KESIMPULAN

Pada dasarnya, beberapa model yang dibuat seperti *Random Forest* dan *Decision Tree* memiliki nilai yang berbeda. Prediksi kinerja model dengan menggunakan *Decision Tree* memiliki nilai *accuracy* yang lebih optimal dibandingkan *Random Forest* terlebih dengan yang menggunakan *GridSearch.* Fokus penelitian ini adalah hasil pengukuran gizi balita. Data dikumpulkan dari 4976 balita dengan 5 atribut data yang diimbangi dengan *SMOTE dan SMOTE – Tomek Link* sebagai *imbalanced data* serta menghasilkan hasil yang spesifik. Penelitian ini dapat dikembangkan dengan menggunakan beberapa algoritma data machine learning yang berbeda. Dengan menggabungkan metode - metode tersebut, akan diperoleh hasil yang lebih variatif. Serta menggunakan atribut data yang lebih banyak untuk mengolah data untuk prediksi. Karena semakin banyak atribut data yang digunakan semakin bagus juga hasil prediksi yang didapatkan.

UCAPAN TERIMA KASIH

Kami mengucapkan penghargaan setinggi-tingginya kepada semua orang yang telah membantu menyelesaikan penelitian ini. Kami ingin mengucapkan terima kasih kepada mentor kami atas nasihat mereka, pengetahuan mereka, dan bantuan mereka yang tak tergoyahkan selama upaya penelitian.

Saya ingin mengucapkan terima kasih kepada Dinas Kesehatan Kabupaten Tangerang dan Puskesmas Tigaraksa karena telah memberikan informasi dan pengetahuan yang berharga untuk penelitian ini. Kami berterima kasih kepada para pengulas yang telah membantu kami membuat penelitian ini lebih baik.

DAFTAR PUSTAKA

[1] D. Prayogi and R. Rahani, "Determinants Analysis of Stunting Prevalence among Toddlers in Central Java Province in 2022," *J. Stat. dan Apl.*, vol. 8, no. 2, pp. 152–163, 2024, doi: https://doi.org/10.21009/JSA.08202.

[2] D. P. Deswinta and A. Prasetyo, "Analisis Prevalensi Kasus Balita Wasting di Kawasan Timur Indonesia tahun 2022 dengan Pendekatan Spasial," *Gizi Indones. (Journal Indones. Nutr. Assoc.*, vol. 47, no. 2, pp. 185–194, 2024, doi: https://doi.org/10.36457/gizindo.v47i2.1023.

[3] T. Hidayat *et al.*, "Performance Prediction Using Cross Validation (GridSearchCV) for Stunting Prevalence," in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, 2024, pp. 1–6. doi: https://doi.org/10.1109/AIMS61812.2024.10512657.

[4] T. Hidayat, M. Ridwan, M. F. Iqbal, R. Rizky, and W. E. Manongga, "Determining Toddler ' s Nutritional Status with Machine Learning Classification Analysis Approach," *Matrik J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 24, no. 2, pp. 235–246, 2025, doi: https://doi.org/10.30812/matrik.v24i2.4092.

[5] A. Mallikarjun and H. H. Bharadi, "Impact of Malnutrition Among Children on Child Mortality in Developing Countries," *Int. J. Innov. Res. Multidiscip. Educ.*, vol. 03, no. 01, pp. 127–130, 2024, doi: https://doi.org/10.58806/ijirme.2024.v3i1n17.

[6] K. Neighbor, "Klasifikasi soal menggunakan multi-label problem transformation dengan metode random forest dan k-nearest neighbor," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 367–380, 2025, doi: http://dx.doi.org/10.29100/jipi.v10i1.5910.

[7] G. Sailasya and G. L. A. Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 539–545, 2021, doi: https://dx.doi.org/10.14569/IJACSA.2021.0120662.

[8] W. Li, "Applied Mathematics and Nonlinear Sciences," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, pp. 1–15, 2024, doi: https://doi.org/10.2478/amns-2024-0340.

[9] 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: https://doi.org/10.36040/jati.v7i6.7822.

[10] I. P. Putri, T. Terttiaavini, and N. Arminarahmah, "Analisis Perbandingan Algoritma Machine Learning untuk Prediksi Stunting pada Anak," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 1, pp. 257–265, 2024, doi: 10.57152/malcom.v4i1.1078.

[11] Y. N. Aprilia *et al.*, "Klasifikasi Status Penderita Gizi Stunting Pada Balita Menggunakan Metode Random Forest (Studi Kasus di Kelurahan Petamanan Kota Pasuruan)," *INTEGER J. Inf. Technol.*, vol. 9, no. 2, pp. 143–154, 2024, doi: https://doi.org/10.31284/j.integer.2024.v9i2.6080.

[12] N. Azizah, E. Budiarti, A. Azis, M. Arsan, M. Nur, and H. Al, "Klasifikasi Status Gizi Anak dengan Decision Tree Berdasarkan Data Stunting di Kabupaten Gowa Child Nutrition Status Classification with Decision Tree Based on Stunting Data in Gowa Regency," *JUSTIN J. Sist. dan Teknol. Inf.*, vol. 12, no. 4, pp. 609–615, 2024, doi: https://doi.org/10.26418/justin.v12i4.79636.

[13] M. Ula, A. F. Ulva, M. Mauliza, M. A. Ali, and Y. R. Said, "Application of Machine Learning in Determining the Classification of Children'S Nutrition With Decision Tree," *J. Tek. Inform.*, vol. 3, no. 5, pp. 1457–1465, 2022, doi: https://doi.org/10.20884/1.jutif.2022.3.5.599.

[14] M. I. Kamil and A. P. Wibowo, "Implementation of the Naive Bayes Classifier Algorithm for Classifying Toddler Nutritional Status," *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 421–427, 2024, doi: https://doi.org/10.30871/jaic.v8i2.8669.
[15] O. N. Chilyabanyama *et al.*, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, 2022, [Online]. Available: https://www.mdpi.com/2227-9067/9/7/1082
[16] J. Wang, "Research on Stock Price Rise and Fall Prediction Based on Optimization Random Forest," in *Applied and Computational Engineering*, 2025, vol. 0, pp. 61–67. doi: https://doi.org/10.54254/2755-2721/2025.20601.
[17] L. González-Castro, M. Chávez, P. Duflot, V. Bleret, G. Del Fiol, and M. López-Nores, "Impact of Hyperparameter Optimization to Enhance Machine Learning Performance: A Case Study on Breast Cancer Recurrence Prediction," *Appl. Sci.*, vol. 14, no. 13, 2024, doi: 10.3390/app14135909.
[18] S. L. Setyowati *et al.*, "Optimizing Random Forest Parameters with Hyperparameter Tuning for Classifying School-Age KIP Eligibility in West Java Optimizing Random Forest Parameters with Hyperparameter Tuning for Classifying School-Age KIP Eligibility in West Java," vol. 7, no. 1, pp. 40–48, 2025, doi: http://dx.doi.org/10.37905/jjom.v7i1.28736.
[19] T. E. Putri, R. T. Subagio, Kusnadi, and P. Sobiki, "Classification System of Toddler Nutrition Status using Naïve Bayes Classifier Based on Z- Score Value and Anthropometry Index," in *Journal of Physics: Conference Series*, 2020, vol. 1641, no. 1. doi: 10.1088/1742-6596/1641/1/012005.
[20] S. Strasser and M. Klettke, "Transparent Data Preprocessing for Machine Learning," in *HILDA 2024 - Workshop on Human-In-the-Loop Data Analytics Co-located with SIGMOD 2024*, 2024. doi: 10.1145/3665939.3665960.
[21] S. Luftensteiner and G. C. Chasparis, "Data Preprocessing and Expert Knowledge Integration for Feature Selection in Process Industry," *Data Preprocessing Expert Knowl. Integr. Featur. Sel. Process Ind.*, p. 17, 2023, doi: https://dx.doi.org/10.2139/ssrn.4552873.
[22] K. Gohari *et al.*, "A Bayesian latent class extension of naive Bayesian classifier and its application to the classification of gastric cancer patients," *BMC Med. Res. Methodol.*, vol. 23, no. 1, pp. 1–15, 2023, doi: https://doi.org/10.1186/s12874-023-02013-4.
[23] P. Karkare, "Handling Missing Values in Data," in *Medium.Com*, 2019, no. October, pp. 750–756. doi: https://doi.org/10.5753/sbbd.2024.243102.
[24] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, and A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 393–399, 2021, doi: https://doi.org/10.29207/resti.v5i2.3008.
[25] Y. El Touati, J. Ben Slimane, and T. Saidani, "Adaptive Method for Feature Selection in the Machine Learning Context," *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 3, pp. 14295–14300, 2024, doi: https://doi.org/10.48084/etasr.7401.
[26] J. Juodakis, I. Castro, and S. Marsland, "Precision as a metric for acoustic survey design using occupancy or spatial capture-recapture," *Environ. Ecol. Stat.*, vol. 28, no. 3, pp. 587–608, 2021, doi: 10.1007/s10651-021-00513-4.