

Klasifikasi Status Penderita Gizi Stunting Pada Balita Menggunakan Metode *Random Forest* (Studi Kasus di Kelurahan Petamanan Kota Pasuruan)

Yunita Nur Aprilia¹, Dian Ahkam Sani^{2*}, Nanda Martyan Anggadimas³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Merdeka Pasuruan

Email: ¹nurapriliana@student.unmerpas.ac.id, ²dian.ahkam@unmerpas.ac.id,
³nandama@unmerpas.ac.id

Abstract. Child stunting, as in this case, is characterized by lower than average body growth. This is caused by a mismatch between long-term nutrient intake and body requirements. Possible impacts include delayed cognitive development, impaired learning ability, and increased risk of metabolic syndrome. To overcome these problems, a structured and data-based system is needed with one of the disagreements used, namely the Random Forest Method in the system using toddler stunting nutrition data as the basis for the classification process. In developing the system built aims to help track the state of health of young children, especially stunting by using several indicators to support innovation, provide a classification model for toddlers who suffer from stunting nutrition, and measure and evaluate the performance results of the Random Forest Method on the data variables used. From this research can show that the results of this study are that this system has successfully created a classification model and is very effective for measuring and evaluating the performance results of the Random Forest Method on Classification of Status of Patients with Stunting Nutrition in Toddlers using a dataset of 300 data so as to produce an average accuracy of 81%, the average precision result is 76%, the average recall result is 69%, and the average F1 score result is 72%.

Keywords: Classification, Stunting Nutrition, Data Mining, Random Forest..

Abstrak. Stunting pada anak, seperti dalam kasus ini, ditandai dengan pertumbuhan tubuh yang lebih rendah dari rata-rata. Hal ini disebabkan oleh ketidaksesuaian antara asupan nutrisi jangka panjang dan kebutuhan tubuh. Dampak yang mungkin termasuk perkembangan kognitif yang tertunda, gangguan dalam kemampuan belajar, serta peningkatan risiko terhadap sindrom metabolic. Untuk Mengatasi permasalahan tersebut diperlukan sistem yang terstruktur dan berbasis data dengan salah satu pendepakatan yang digunakan yaitu Metode Random Forest pada sistem dengan menggunakan data gizi stunting balita sebagai dasar proses klasifikasi. Dalam mengembangkan sistem yang dibangun bertujuan untuk membantu melacak keadaan kesehatan anak kecil, khususnya stunting dengan menggunakan beberapa indikator guna mendukung inovasi, memberikan model klasifikasi terhadap balita yang menderita gizi stunting, dan mengukur serta mengevaluasi hasil kinerja dari Metode Random Forest terhadap variabel data yang digunakan. Dari penelitian ini dapat menunjukkan bahwa Hasil dari penelitian ini adalah system ini berhasil dibuatnya model klasifikasi dan sangat efektif 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 %.

Kata Kunci : Klasifikasi, Gizi Stunting, Data Mining, Random Forest.

1. Pendahuluan

1.1. Latar Belakang

Stunting adalah masalah penyakit yang meluas di Indonesia dan beberapa negara berkembang lainnya. Stunting pada anak, seperti dalam kasus ini, ditandai dengan pertumbuhan tubuh yang lebih rendah dari rata-rata. Hal ini disebabkan oleh ketidaksesuaian antara asupan nutrisi jangka panjang dan

kebutuhan tubuh. Dampak yang mungkin termasuk perkembangan kognitif yang tertunda, gangguan dalam kemampuan belajar, serta peningkatan risiko terhadap sindrom metabolik, hipertensi, dan obesitas.

Keseimbangan gizi memiliki kaitan dengan tingkat kecerdasan pada anak. Ketika asupan gizi menurun, maka status gizi dan kesehatan anak cenderung menurun juga. (Menteri Kesehatan, 2014). Masalah pada gizi memiliki keterkaitan yang erat dengan masalah stunting pada anak balita. Stunting merujuk pada suatu keadaan kurang gizi jangka panjang yang muncul di Indonesia selama tahap penting pertumbuhan dan perkembangan, dimulai sejak janin. (Kementerian Kesehatan RI, 2018). Penting untuk diingat bahwa pertumbuhan yang lambat pada anak adalah masalah serius yang sering tidak disadari oleh masyarakat. Kejadian stunting yang tinggi pada balita biasanya terkait dengan tingkat sosial ekonomi yang rendah dan pendidikan orang tua yang terbatas, yang juga mempengaruhi pola asuh yang diberikan kepada balita.

Dalam Pemantauan dan pengumpulan data mengenai stunting di berbagai puskesmas di wilayah Kota Pasuruan mempunyai peranan penting dalam menilai pertumbuhan dan perkembangan yang baik pada janin maupun bayi yang baru lahir. Salah satu masalah yang sering timbul di Kelurahan Petamanan Kota Pasuruan adalah ketidakakuratan dan ketidaktepatan pengumpulan data stunting setiap bulannya, yang masih dilaksanakan secara manual memakai Microsoft Excel. Hal ini mengakibatkan ukuran data menjadi besar dan proses perhitungan menjadi lebih berat, sehingga masyarakat hanya dapat memeriksa.

Salah satu cara yang dapat diterapkan pada penelitian ini yaitu Klasifikasi status gizi balita dengan metode Random Forest dapat digunakan dalam pendekatan data mining. Keuntungan menggunakan algoritma random forest sebagai metode dalam klasifikasi yaitu dapat menangani masalah overfitting, dapat digunakan untuk klasifikasi dan regresi, untuk meningkatkan kinerja dan keakuratan prediksi secara keseluruhan.

Dalam penelitian sebelumnya dari beberapa metode yang bisa digunakan, terdapat beberapa pilihan untuk kasus dengan Metode Random Forest, salah satunya dibahas dalam penelitian “Prediksi Stunting Pada Balita Dengan Menggunakan algoritma Random Forest”, oleh Aditya Yudha., Roswan Latuconsina, Ashri Dinimaharawati (2021). Membahas mengenai prediksi keadaan stunting pada balita dan melakukan pengembangan sebuah sistem berupa perangkat lunak yang berbasis website dan android yang mampu melakukan pengukuran keadaan Stunting pada balita. mampu menghasilkan nilai akurasi rata-rata sebesar 97.8721%. (Perdana, A. Y., Latuconsina, R., & Dinimaharawati, A., 2021).

Pada penelitian lainnya yang membahas tentang Metode Random Forest menunjukkan bahwa, menerapkan metode random forest yang dioptimasi dengan algoritma genetika untuk mengklasifikasi status gizi balita secara akurat dan cepat. Setelah dilakukan pengujian, maka diperoleh rata-rata akurasi oleh metode random forest yang dioptimasi dengan algoritma genetika sebesar 89,58%, rata-rata precision 74,34%, rata-rata recall 58,68% dan f1-score 65,54% dengan parameter ukuran populasi bernilai 20, iterasi sebanyak 3, nilai crossover rate sebesar 0,7, nilai mutation rate sebesar 0,3, dan jumlah feature 4. (Candra, E. N., Cholissodin, I., & Wihandika, R. C., 2022).

Ditahun berikutnya Juwariyem, J., & Sriyanto, S., (2023) Melakukan penelitian yang berjudul “Prediksi Stunting Pada Balita Menggunakan Algoritma Random Forest”. Dimana pada penelitian tersebut menjelaskan bahwa Penelitian ini menggunakan metode Random Forest untuk melihat tingkat akurasi prediksi stunting pada balita yang menggunakan dataset dengan jumlah sebanyak 10001 record data, 7 atribut dan 1 kelas atribut. Berdasarkan hasil pengujian Random Forest untuk melihat akurasi dari prediksi keberhasilan data yang diujikan, diperoleh hasil akurasi yaitu 85,86 %. Dari hasil yang diperoleh terhadap pengujian yang telah dilakukan, diketahui bahwa metode Random Forest merupakan metode yang lebih baik dalam memprediksi akurasi pada stunting. (Juwariyem, J., & Sriyanto, S., (2023))

Pada tahun berikutnya ditahun 2024 oleh Reza, A. A. R., & Rohman, M. S. (2024). Yang berjudul “Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization”. dimana pada penelitian tersebut menjelaskan bahwa penelitian ini merupakan Kasus stunting di Indonesia tergolong tinggi, mencapai 21,6% pada tahun 2022, mengindikasikan prevalensi stunting yang cukup signifikan. Identifikasi stunting dilakukan melalui pendekatan data mining yang dinilai lebih efisien. Namun, algoritma klasifikasi dalam data mining sering kali mengalami

ketidakseimbangan data, yang menyebabkan rendahnya akurasi dan hasil prediksi yang tidak akurat. Untuk mengatasi hal tersebut, penelitian ini menggunakan algoritma Random Forest dengan optimasi menggunakan metode random search. Hasil pengujian menunjukkan bahwa Random Forest mencapai akurasi yang relatif tinggi, yaitu 90,7%. Setelah dilakukan optimasi menggunakan pencarian acak, akurasi semakin meningkat menjadi 96,33%. Kombinasi algoritma dan optimasi terbukti sangat efektif, menghasilkan peningkatan akurasi sebesar 5,63%.

Berdasarkan permasalahan dan hasil penelitian terdahulu yang telah diuraikan sebelumnya bahwa metode Random Forest memiliki kemampuan dan mengukur, serta mengevaluasi dengan mengklasifikasikan status gizi balita secara akurat dan cepat. Oleh karena itu, penelitian ini “Mengklasifikasikan Status Gizi Stunting pada Balita Menggunakan Pendekatan Random Forest”. Manfaat penelitian ini menjadi bahan referensi untuk penelitian berikutnya yang berkaitan dengan Klasifikasi Status Penderita Gizi Stunting Pada Balita ataupun penelitian yang berkaitan tentang metode *Random Forest*, dan digunakan dalam memudahkan tenaga medis posyandu dalam mengetahui perkembangan pertumbuhan balita.

2. Tinjauan Pustaka

Dalam penelitian ini, membahas beberapa teori untuk memahami penulisan pada skripsi, diantaranya sebagai berikut :

2.1 Stunting

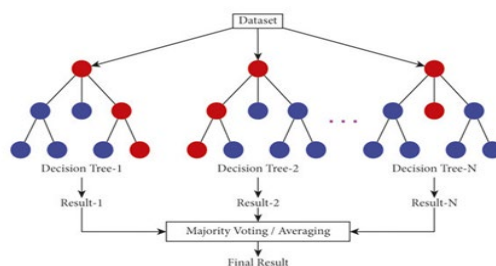
Stunting adalah masalah gizi jangka panjang yang muncul karena kurangnya asupan nutrisi selama periode yang berlangsung cukup lama, yang disebabkan oleh pemberian makanan yang tidak memenuhi kebutuhan gizi yang diperlukan. Ada beberapa indikator yang dapat digunakan untuk menilai kondisi gizi pada bayi. Evaluasi status gizi bisa dilakukan melalui pemeriksaan laboratorium atau melalui pengukuran antropometri. Penggunaan pengukuran tubuh atau antropometri dipilih karena metode ini dianggap sebagai cara yang paling sederhana dan ekonomis untuk menilai status gizi. Di Indonesia, rumus yang digunakan untuk mengukur stunting pada anak adalah menerapkan z-score atau Z-indeks. Z score adalah hasil yang menunjukkan pengukuran dari median. (Organization, W. H. (2006)).

$$Z - score = \frac{TB - TB \text{ Standar}}{SD TB \text{ Standar}} \quad (1)$$

Dengan TB adalah Tinggi Badan dari balita, U adalah Umur balita, dan SD TB Standar adalah Standar deviasi tinggi badan standar pada balita.

2.2 Random Forest

Random forest didefinisikan sebagai kelompok klasifikasi dari pohon regresi, dilatih dari data pelatihan menggunakan pilihan fitur acak dalam proses generate tree. Cara kerja metode random forest adalah dengan membuat pohon keputusan dari pemilihan data secara acak dan juga pemilihan fitur secara acak. Pohon keputusan yang terbentuk banyak, oleh karena itu dinamakan forest atau hutan. Kemudian metode pengujian data pada random forest adalah dengan memasukkan data tersebut ke dalam semua pohon keputusan yang terbentuk dan hasilnya akan ditentukan sesuai dengan jumlah jawaban terbanyak yang muncul. (Juwariyem, J., & Sriyanto, S. (2023)).



Gambar 1. Gambaran umum dari algoritma random forest

Pada gambar 1 merupakan gambaran umum dari *Algoritma Random Forest* yang Dimana pada gambar tersebut merupakan pohon Keputusan pada algoritma random fores. *Algoritma random forest* ini dapat menghitung entropy yang mampu menentukan information gain ataupun keuntungan informasi dengan persamaan. Rumus Random Forest :

$$\text{Entropy (Y)} = \sum_{i=1}^n - p_i \log_2 p_i \quad (2)$$

Dengan Y adalah himpunan dari sebuah kasus. n adalah jumlah anggota dari Y. dan P_i adalah proporsi dari S_i terhadap S.

$$\text{Information Gain (Y, } \alpha) = \text{Entropy (Y)} - \sum (Y/X) \quad (3)$$

Dengan E (Y) adalah entropi dari atribut target. E(Y/X) adalah entropi rata-rata dari atribut klasifikasi terhadap atribut target. X adalah atribut dari klasifikasi dan Y adalah atribut target. P_i merupakan kemungkinan dari class I, sedangkan IG (information gain) yang merupakan hasil pengurangan dari entropy parent dengan entropy leaf node (child). Pada random forest akan dibentuk beberapa decision tree, dengan setiap hasilnya akan diambil untuk dijadikan voting (hasil dengan jumlah terbanyak) akan dijadikan keluaran dari algoritma random forest.

2.3. Evaluasi

Metode yang digunakan untuk evaluasi dari ensemble model yang digunakan dari algoritma yaitu *Confusion matrix* adalah teknik yang digunakan untuk mengidentifikasi penyebaran prediksi benar dan salah yang untuk penyesuaian dua kelas yaitu kelas positif dan kelas negative. Model Ini memberikan perbandingan antara output yang dihasilkan oleh sistem dengan nilai yang seharusnya. Keakuratan, presisi, dan daya ingat hanyalah beberapa metrik yang menggunakan banyak matriks kinerja yang disediakan oleh *Confusion matrix*. Keakuratan prediksi adalah tingkat kesesuaiannya dengan kenyataan. (Muhammad Al-Husaini, Irani Hoeronis, Hen Hen Lumana, Luh Desi Puspareni, (2023)).

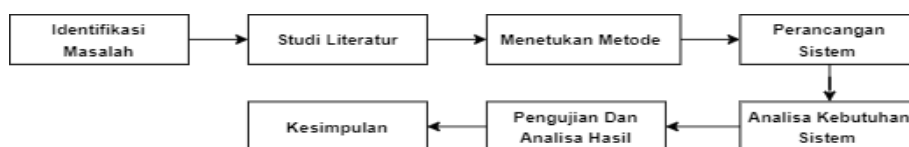
Tabel 1. Confusion Matrix

	Positive	Negative
Predicted Positive	TP	FP
Predicted Negative	FN	TN

Keterangan:

TP (True Positive) : nilai prediksi sistem positif, nilai sebenarnya positif
 TN (True Negative) : nilai prediksi sistem negatif, nilai sebenarnya negatif
 FP (False Positive) : nilai prediksi sistem positif, nilai sebenarnya negatif
 FN (False Negative) : nilai prediksi sistem negatif, nilai sebenarnya positif

3. Metode Penelitian



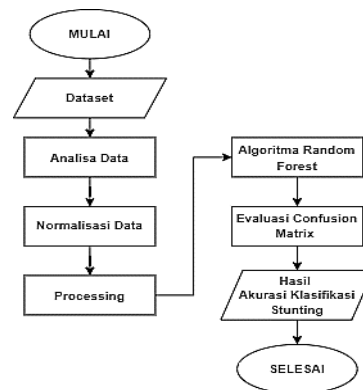
Gambar. 2 Alur Penelitian

3.1 Studi Literatur

Tahap kedua yaitu melakukan studi literatur berdasarkan topik penelitian yang telah ditentukan. Studi Literatur yaitu pencarian bahan bacaan melalui buku, jurnal penelitian terdahulu, artikel atau website yang ada di internet merupakan bentuk tindakan yang dilakukan peneliti untuk membantu tercapainya tujuan penelitian. Pada penelitian ini, studi literatur yang digunakan berkaitan dengan hasil klasifikasi yang diperoleh dari Metode Random Forest terhadap suatu hasil melalui penelitian terdahulu.

3.2 Perancangan Sistem

Tahap kedua yaitu tentang perancangan system yang akan digunakan sebagai acuan dalam penelitian. Yang dimana pada rancangan ini berisikan proses apa saja yang akan dijalankan terhadap system yang dibuat, dan termasuk kebutuhan dalam system tersebut.



Gambar 3. Perancangan Sistem

Tahapan-tahapan yang dilakukan dalam penelitian ini dijelaskan seperti pada Gambar 3. Berikut ini penjelasan dari perancangan sistem tersebut :

a. Pengumpulan Data

Dalam penelitian ini, pengumpulan dataset yang berisi informasi tentang balita, termasuk atribut-atribut yang relevan seperti jenis kelamin, umur, berat, tinggi, tb/u, dan lainnya, Sumber data yang dipakai pada riset ini berasal dari Kelurahan Petamanan. Dan Berbicara dengan staf medis yang bertanggung jawab atas layanan Kesehatan untuk memperoleh informasi tambahan. Data ini diperoleh dari petugas gizi berupa file excel, yang mencakup jumlah kasus stunting pada balita di tahun 2022-2023 di Kelurahan Petaman Kota Pasuruan.

b. Analisa Data

Analisa data adalah proses sistematis untuk memahami, menyelidiki, dan mengekstrak informasi yang terkandung dalam kumpulan data. Pada penelitian ini tahap Analisa data terdapat 2 tahap yaitu :

1. Korelasi Data

Dalam proses ini, data balita stunting yang telah dikumpulkan kemudian dianalisis untuk melihat apakah terdapat korelasi antar variabel. Hasil analisis menunjukkan bahwa hanya variabel Umur/bulan, Berat, Tinggi, Lingkar lengan atas, Status Pendidikan, dan BB/TB yang memiliki keterkaitan satu sama lain. Variabel-variabel ini menunjukkan hubungan yang signifikan. Sementara itu, variabel lainnya tidak memiliki korelasi karena nilai yang diperoleh bersifat negatif dan tidak menunjukkan adanya keterkaitan.

2. Setelah Data Dikorelasikan, dalam proses analisis data, terdapat beberapa variabel yang dikategorikan. Pada penelitian ini, variabel data dibagi menjadi dua jenis: variabel bebas Dalam metode *Random Forest*, terdapat lima variabel yaitu: Umur (dalam bulan), Berat, Tinggi, Lingkar lengan atas, dan Status Pendidikan. Dan pada variabel terikat dalam penelitian ini adalah hasil akhir dari perhitungan menggunakan metode *Random Forest* yang berupa status dengan label BB/TB (Status Gizi). Variabel ini dikategorikan ke dalam dua kelas, yaitu "Stunting" dan "Normal".

c. Normalisasi Data

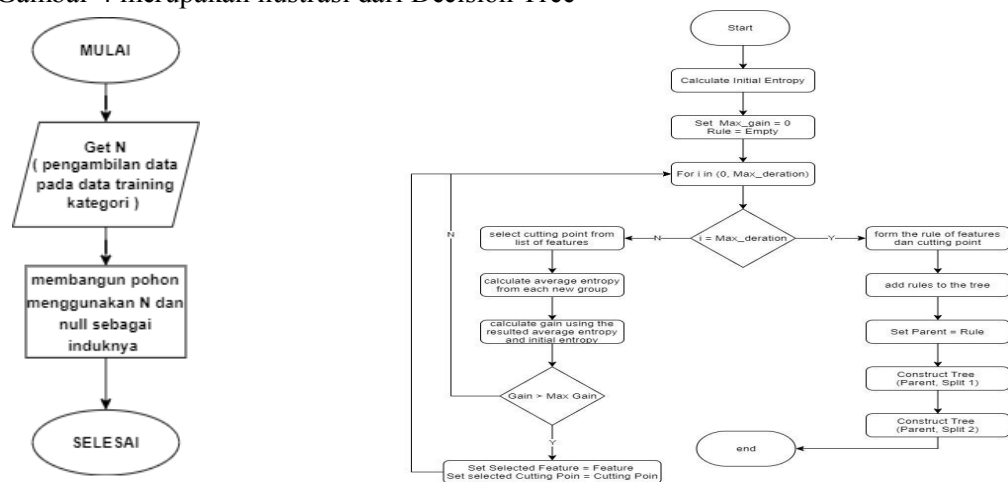
Proses Normalisasi data merupakan salah satu teknik yang penting untuk dilakukan pada tahapan preprocessing. Setelah dilakukannya Pada penelitian ini, menggunakan cara pra-pemrosesan data dalam analisis data dan pembelajaran mesin yang digunakan untuk mengubah nilai-nilai fitur dalam dataset sehingga terdistribusi diberi nilai dalam rentang 0 hingga 1 untuk melakukan normalisasi data. Alasannya hal ini dikarenakan seringkali sebuah data memiliki rentang nilai antar variable yang sangat jauh. (Aji, Priyo Wahyu Setiyo, S. Suprianto, and Rohman Dijaya, 2023).

d. Processing

Proses ini menjelaskan tentang proses pengolahan data yaitu Pembagian dataset menjadi subset pelatihan (data training) dan subset pengujian (data testing). Pada penelitian ini data yang digunakan sebagai data training untuk menguji kinerja model yang telah dilatih mencapai tingkat akurasi untuk data training sebesar 70%, Dan untuk data testing 30% dari data training. Dari Hasil penyaringan yang dilakukan untuk penelitian ini memberikan informasi tentang reliabilitas hasil. Tujuan dari penelitian ini adalah untuk mengevaluasi keefektifan algoritma Random Forest dalam melakukan prediksi tentang kondisi gizi balita yang berisiko mengalami stunting.

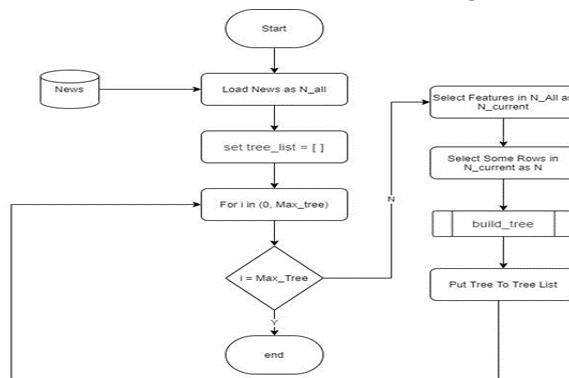
3.3 Algoritma Random Forest

Pembentukan Algoritma *Random Forest* yang bekerja berdasarkan mekanisme dataset bagging dan pelatihan sekumpulan *Decision Tree*, maka dari itu untuk menjelaskan proses pelatihan yang terjadi dalam algoritma RF terlebih dahulu akan dijelaskan proses pelatihan/ pembentukan sebuah data. Dalam proses perhitungan metode Random Forest dari data yang telah diperoleh sebelumnya bertujuan untuk menguji keakurasian dan kinerja dalam pembuatan sebuah program dengan metode Random Forest . yang diawali dengan get N yaitu proses mengambil dataset pada sub-kategori berita yang akan dijadikan Decision Tree setelah itu proses pelatihan decision tree mula-mula dilakukan berdasarkan sebuah data input N dan null sebagai induknya. Lalu dalam tahap pembuatan pohon keputusan, langkah pertama adalah menghitung entropi awal. Setelah itu, kita menentukan nilai awal untuk $\text{max_gain} = 0$ dan $\text{rule} = 0$. Setelah inisialisasi, kami memulai proses iterasi dengan perulangan "For i in (0, max_deration)." Perulangan ini digunakan untuk memeriksa nilai i pada setiap iterasi. Jika $i = \text{max_deration}$, kita memeriksa kondisi "yes," yang mengarah pada pembentukan aturan berdasarkan fitur dan titik potong, dan kemudian aturan tersebut ditambahkan ke pohon. Selanjutnya, kami menentukan $\text{parent} = \text{rule}$ dan melanjutkan dengan proses konstruksi pohon dengan menentukan (parent , split 1), yang kemudian diikuti dengan konstruksi pohon menggunakan (parent , split 2). Proses ini terus berlanjut jika kondisi perulangan "yes" terpenuhi. Namun, jika kondisi perulangan "no" terpenuhi, kita akan memasuki proses pemilihan fitur dan titik potong, dan kemudian melanjutkan dengan menghitung rata-rata entropi dari setiap kelompok baru. Setelah menghitung rata-rata, langkah selanjutnya adalah menghitung gain menggunakan entropi rata-rata dan entropi awal. Kemudian, kita memeriksa apakah $\text{gain} > \text{max_gain}$. Jika hasil pemeriksaan adalah "yes," maka kita menetapkan $\text{selected_feature} = \text{feature}$ dan $\text{selected_cutting_point} = \text{Cutting poin}$. Setelah semua proses ini selesai, kita kembali ke awal perulangan dan memeriksa kembali apakah kondisi $i = \text{max_entropy}$. Jika kondisi ini terpenuhi, kita melanjutkan dengan proses yang sesuai dengan "yes" pada $i = \text{max_entropy}$. Jika hasil pemeriksaan adalah "no," kita melanjutkan dengan proses yang sesuai dengan "no" pada tahap periksa $i = \text{max_entropy}$. Proses ini terus berlanjut hingga semua iterasi selesai. Setelah selesai dengan proses pembuatan pohon keputusan, kita dapat melanjutkan dengan Algoritma Random Forest Classifier untuk mengklasifikasikan data yang telah diproses. Gambar 4 merupakan ilustrasi dari Decision Tree



Gambar 4. Decision Tree

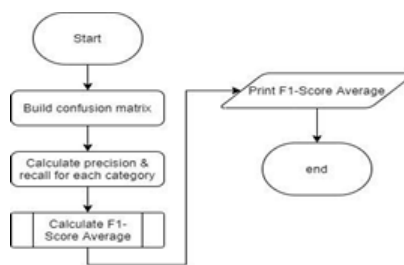
Setelah itu proses Algoritma Random Forest dimulai dengan mengambil data dari dataset dan menginisialisasi N_{all} . Kemudian, langkah berikutnya adalah menentukan pohon (tree) menggunakan sebuah array kosong. Selanjutnya, proses memasuki pengulangan for dengan variabel i berjalan dari 0 hingga max_tree . Dalam pengulangan ini, dilakukan pengecekan apakah i sama dengan Max_Tree . Jika kondisi ini terpenuhi, proses akan selesai. Namun, jika hasil pengecekan menunjukkan bahwa kondisinya tidak terpenuhi, maka proses akan melanjutkan dengan pemilihan fitur pada N_{All} untuk digunakan sebagai $N_{current}$. Setelah langkah ini selesai, proses berlanjut dengan pemilihan baris dalam $N_{current}$ untuk dijadikan data atau N . Kemudian, dilakukan proses $build_tree$. Setelah proses ini selesai, pohon (tree) akan dimasukkan ke dalam daftar pohon atau tree. Setelah data dimasukkan ke dalam daftar pohon, proses akan kembali ke dalam pengulangan for dengan variabel i berjalan dari 0 hingga max_tree . Kembali dilakukan pengecekan apakah i sama dengan Max_Tree . Jika kondisi ini terpenuhi, proses akan berakhir. Namun, jika kondisi tersebut tidak terpenuhi, proses akan kembali ke langkah sebelumnya di mana i tidak sama dengan Max_Tree . Setelah dilakukan beberapa tahap perhitungan, Pada Tahapan ini model yang telah dibuat dan telah melalui proses pengujian untuk menguji kebenaran data yang digunakan, apakah metode klasifikasi Random Forest dapat diterapkan pada status gizi balita atau tidak. Gambar 5. adalah ilustrasi dari algoritma *Random Forest*.



Gambar 5. Algoritma Random Forest

3.4 Evaluasi

Dalam tahapan evaluasi data diawali dengan membuat confusion matrix untuk model evaluasi, setelah itu dilanjutkan dengan proses perhitungan Accuration, precision dan recall. Proses selanjutnya adalah proses perhitungan rata-rata dari hasil perhitungan precision dan recall tiap sub kategori berita dengan menggunakan matrik perhitungan F1-Score, selanjutnya adalah mencetak hasil perhitungan F1-Score yaitu ukuran statistik yang digunakan dalam pengenalan pola dan pembelajaran mesin untuk mengevaluasi kinerja model klasifikasi yang digunakan untuk melihat hasil rata-rata dari perhitungan yang sudah dilakukan untuk menentukan hasil terbaik dari rata-rata tersebut. Gambar 6. Adalah ilustrasi dari Flowchart Evaluasi.



Gambar 6. Flowchart Evaluasi

4. Hasil dan Pembahasan

Pada penelitian ini, proses klasifikasi dengan menggunakan *Metode Random Forest* pada data stunting balita dibuat dengan menggunakan Bahasa pemrograman PHP Native dengan tampilan menggunakan Bahasa javascript bootstrap. Berikut ini merupakan pembahasan dan hasil dari penelitian yang dilakukan :

4.1 Hasil Pengujian

Tahap hasil pengujian akan digunakan sebagai acuan dalam penelitian. Yang dimana pada hasil pengujian ini berisikan hasil apa saja yang akan sudah dijalankan terhadap system. Dari alur dapat dijelaskan mengenai Langkah-langkah dalam perancangan sistem sebagai berikut :

a. Dataset.

Sebelum data tersebut diproses, dataset tersebut dikumpulkan terlebih dahulu. dalam penelitian ini data tersebut berisikan informasi tentang balita, termasuk atribut-atribut yang relevan seperti jenis kelamin, umur, berat, tinggi, lingkar lengan atas, dan lainnya, Sumber data yang dipakai pada riset ini berasal dari Kelurahan Petamanan. Data ini diperoleh dari petugas gizi berupa file excel, yang mencakup jumlah kasus stunting pada balita di tahun 2022-2023 di Kelurahan Petamanan Kota Pasuruan.

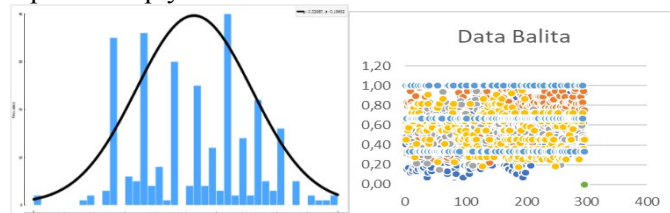
Tabel 2. Data Stunting Balita tahun 2022-2023

No.	...	Berat	Tinggi	LiLa	Status Pendidikan	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB
1	...	12.4	90.3	12	S1	Normal	-1,86	Pendek	-2,6	gizi baik	-0,55
2	...	8.2	79	14	SMP	sangat kurang	-3,2	Pendek	-2,61	gizi kurang	-2,34
3	...	7.5	68	13	SMA	Normal	-1,82	Pendek	-2,3	gizi baik	-0,74
4	...	10.3	84.5	13	SMK	Kurang	-2,12	Pendek	-2,36	gizi baik	-1,03
5	...	9.6	78	14	S3	Normal	-1,69	Pendek	-2,58	gizi baik	-0,59
....											
300		15,6	78	12	SMA	Kurang	-2,33	Pendek	-2,67	gizik kurang	-2,58

Pada tabel 2. yaitu data stunting balita tahun 2022-2023 dengan total 300 data yang terdiri dari 16 kolom. Dataset ini harus mencakup informasi tentang apakah balita mengalami stunting atau tidak.

b. Analisa Data

Setelah data dikumpulkan, proses selanjutnya adalah Analisa data yaitu proses sistematis untuk memahami, menyelidiki, dan mengekstrak informasi yang terkandung dalam kumpulan data. Pada tahap Analisa data terdapat 2 tahap yaitu Korelasi Data dan variabel data.



Gambar 7. Distribusi Data

Berdasarkan gambar 7. Dalam proses ini, data balita stunting yang telah dikumpulkan kemudian dianalisis untuk melihat apakah terdapat korelasi antar variabel. Hasil analisis menunjukkan bahwa hanya variabel Umur/bulan, Berat, Tinggi, Lingkar lengan atas, Status Pendidikan, dan BB/TB yang memiliki keterkaitan satu sama lain. Variabel-variabel ini menunjukkan hubungan yang signifikan. Sementara itu, variabel lainnya tidak memiliki korelasi karena nilai yang diperoleh bersifat negatif dan tidak menunjukkan adanya keterkaitan.

3. Normalisasi Data

Tabel 3. Tabel Setelah normalisasi

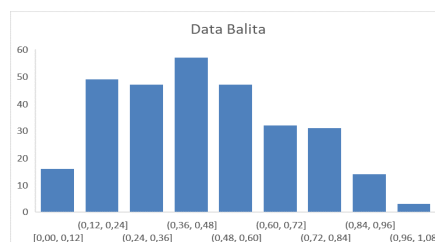
No	Umur/ bulan	Berat	Tinggi	Lingkar Lengan Atas	status pendidikan	Status Gizi
1	1	1	0	1	3	0
2	1	1	1	0	2	1
3	0	0	1	0	2	0
4	1	0	1	0	1	0
...	...	...	...	...	...	...
299	1	0	1	1	1	1
300	1	1	1	0	1	1

Proses Normalisasi data merupakan salah satu teknik yang penting untuk dilakukan pada tahapan preprocessing. Pada tabel 3 data yang setelah dilakukan analisa data, proses selanjutnya adalah Normalisasi data yaitu dengan cara pra-pemrosesan data dalam analisis data dan pembelajaran mesin yang digunakan untuk mengubah nilai-nilai fitur dalam dataset sehingga terdistribusi diberi nilai dalam rentang 0 hingga 1 untuk melakukan normalisasi data.

Tabel 4. Data Bobot

Atribut	Nama	Keterangan	Nilai
1 Umur	< 24 bulan	Balita	0
	> 24 bulan	Anak	1
2 Berat	2,5 – 16 kg	Kurang	1
	3,3 - 18,3 kg	Normal	0
Tinggi	< 85 cm	Rendah	1
	85-110 cm	Normal	0
3 Lila	< 12 cm	Kurang	1
	> 12 cm	Normal	0
4 Pendidikan	-	SD	0
	-	MA,SMA,SMK	1
	-	SMP	2
	-	S1,S2,S3	3

Pada tabel 4. data yang sudah dinormalisasikan terdapat masing- masing bobot yang memiliki nilai yang digunakan sebagai acuan perhitungan yaitu dilihat pada tabel. Dan untuk bobot target yaitu terbagi menjadi 2 kelas adalah Status Gizi yang memiliki nilai 0 adalah normal dan nilai 1 adalah stunting. Pada gambar 8 merupakan hasil normalisasi data. Yang Dimana diagram tersebut berisikan hasil dari variabel data stunting.



Gambar 8. Diagram Hasil Normalisasi

4. Processing

Pada proses ini yaitu pembagian dataset menjadi subset pelatihan (data training) dan subset pengujian (data testing).

Tabel 5. Data Training

No	Umur/ bulan	Berat	Tinggi	Lingkaran Lengan Atas	status pendidikan	Status Gizi
1	1	1	0	1	3	0
2	1	1	1	0	2	1
3	0	0	1	0	2	0
4	1	0	1	0	1	0
5	0	0	1	0	3	0
...	...	...	...	...	...	...
209	1	1	0	1	1	1
210	0	1	1	1	1	1

Pada tabel 5 data ini digunakan sebagai data training untuk menguji kinerja model yang telah dilatih mencapai tingkat akurasi yaitu sebesar 70%, Dimana dengan jumlah data seluruhnya 300 x 0,7 menjadi 210 data. Sementara sebagian kecil pada tabel 6 digunakan sebagai data testing. Dan untuk data testing 30% dari data training yaitu sebesar 90 data.

Tabel 6. Data Testing

No	Umur/ bulan	Berat	Tinggi	Lingkaran Lengan Atas	status pendidikan	Status Gizi
1	0	0	0	1	2	0
2	1	1	0	0	1	1
3	1	1	1	0	3	1
4	0	0	0	1	2	0
5	0	0	0	0	2	0
...	...	...	...	...	...	...
89	1	0	1	1	2	1
90	1	1	1	0	1	1

5. Algoritma Random Forest

Dalam proses perhitungan metode Random Forest dari data yang telah diperoleh sebelumnya, pada proses ini bertujuan untuk menguji keakurasian dan kinerja dalam pembuatan sebuah program dengan metode Random Forest tabel 7. Merupakan hasil dari perhitungan random forest

Tabel 7. Perhitungan Random Forest

Label		Value		Total Data		Bobot	
Status Gizi Thd 0		0		300		0.3267	
Status Gizi Thd 1		1				0.3733	

No	Umur		Berat		Tinggi		Lingkaran Lengan Atas		Status Pendidikan		Status Gizi	Bobot Prediksi		Prediksi
	1	0	1	0	1	0	1	0	1	0		1	0	
1	0.64	0.29	0.35	0.6	0.46	0.46	0.54	0.29	0.19	0.21	1	0.003	0.002	1
2	0.64	0.29	0.65	0.4	0.54	0.54	0.46	0.71	0.25	0.27	1	0.008	0.004	1
3	0.64	0.29	0.65	0.4	0.54	0.54	0.54	0.29	0.19	0.21	1	0.008	0.001	1
4	0.64	0.29	0.35	0.6	0.54	0.54	0.54	0.29	0.24	0.41	1	0.005	0.004	1
5	0.36	0.71	0.35	0.6	0.46	0.46	0.46	0.71	0.19	0.21	0	0.002	0.011	0
6	0.36	0.71	0.35	0.6	0.46	0.46	0.54	0.29	0.24	0.41	1	0.002	0.009	0
7	0.64	0.29	0.65	0.4	0.54	0.54	0.54	0.29	0.25	0.27	1	0.010	0.002	1
8	0.64	0.29	0.35	0.6	0.46	0.46	0.54	0.29	0.32	0.11	1	0.006	0.001	1
9	0.36	0.71	0.35	0.6	0.54	0.54	0.54	0.29	0.24	0.41	0	0.003	0.010	0
10	0.64	0.29	0.35	0.6	0.46	0.46	0.46	0.71	0.19	0.21	1	0.003	0.004	0

4.2 Evaluasi

Proses selanjutnya yaitu dilakukannya proses pengujian dengan menggunakan *Confusion Matrix* untuk mengetahui nilai *Recall*, *Precision*, *Accuracy* dan *F1-score* yang diperoleh dari *Metode Random Forest*.

Tabel 8. Confusion Matrix

		Prediksi	
		1	0
Actual	1	51	10
	0	7	22

Dari tabel 8. data yang sudah diuji berjumlah 210 data menunjukkan bahwa nilai "True Positif sebanyak 51 , nilai False Negatif sebanyak 7, nilai False Positif sebanyak 10, dan nilai True Negatif sebanyak 22". Dengan perhitungan dapat dijabarkan sebagai berikut :

a.) Accuracy

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% = \frac{51+22}{51+10+7+22} \times 100\% = 81\%$$

b.) Precision

$$Precision = \frac{TP}{TP+FP} \times 100\% = \frac{51}{51+10} \times 100\% = 76\%$$

c.) Recall

$$Recall = \frac{TP}{TP+FN} \times 100\% = \frac{51}{51+7} \times 100\% = 69\%$$

d.) F1-Score

$$F1 - score = \frac{2 \times (Presisi \times Recall)}{Presisi + Recall} = \frac{2 \times (76 \times 69)}{76 + 69} = 72\%$$

Setelah dilakukannya perhitungan dengan Model *Confusion Matrix* dapat menghasilkan Accuracy, Presicion, Recall dan F1-Score. berikut ini adalah tabel hasil dari Confusion Matrix :

Tabel 9. Hasil Confusion Matrix

Rumus	Variable	Hasil
TP/Jumlah Data	accuracy	0.81
TP/(TP+FP)	Precision	0.76
TP/(TP+FN)	Recall	0.69
2*recall*presicion/recall+presicion	F1-Score	0.72

Berdasarkan tabel 9, bahwa Hasil dari nilai Accuracy sebesar 81% , nilai Precision sebesar 76%, nilai recall sebesar 69% dan terakhir pada nilai F1-Score sebesar 72%. Hal ini membuktikan bahwa metode Random Forest dapat digunakan untuk mengukur dan mengevaluasi serta menilai status gizi stunting balita.

5 Penutup

5.1 Kesimpulan

Dari hasil penelitian pada Klasifikasi Status Penderitaan Gizi Stunting Pada Balita Menggunakan Metode Random Forest , maka dapat ditarik kesimpulan yaitu: Model klasifikasi status penderita gizi stunting pada balita dengan menggunakan Metode Random Forest untuk mengukur serta mengevaluasi hasil kinerja dari proses klasifikasi pada balita yang menderita gizi stunting terhadap variabel data yang digunakan, serta mengetahui bahwa balita tersebut mengalami stunting atau tidaknya telah berhasil dilakukan.. Hasil pengujian klasifikasi menggunakan metode Random Forest dengan menggunakan dataset sebanyak 300 data menghasilkan rata-rata accuracy sebesar 81%, hasil rata-rata precision sebesar 76%, hasil rata-rata recall sebesar 69%, dan hasil rata-rata F1 score sebesar 72%.

5.2 Saran

Beberapa saran dari penulis sebagai bahan untuk pengembangan penelitian selanjutnya yaitu: Untuk penelitian selanjutnya agar dapat dilakukan penambahan jumlah data dan tahun penelitian guna mendapatkan nilai akurasi yang lebih baik, Pada penelitian selanjutnya dapat menggunakan metode klasifikasi lainnya untuk dapat dibandingkan dengan hasil nilai akurasi pada penelitian ini, Pada penelitian Selanjutkan agar dapat dilakukan pengembangan sistem, dengan versi mobile agar memudahkan pengguna untuk mengakses sistem kapan saja dan di mana saja melalui perangkat seluler pengguna.

Referensi

- Aji, Priyo Wahyu Setiyo, S. Suprianto, and Rohman Dijaya. "Prediksi Penyakit Stroke Menggunakan Metode Random Forest." *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)* 4.4 (2023): 916-924.
- Al-Husaini, M., Hoeronis, I., Lumana, H. H., & Puspareni, L. D. Deteksi Dini Stunting Anak Balita Berbasis Ensemble Machine Learning di Purbaratu Kota Tasikmalaya. *learning*, 9, 10.
- Arisandi, Riza Rizqi Robbi, Budi Warsito, and Arief Rachman Hakim. "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation." *Jurnal Gaussian* 11.1 (2022): 130-139.
- Atikah, R., Yulidasari, F., Ocktaviana, A., & Anggraini, L. (2018). *Stunting dan Upaya Pencegahannya* (Hadianor (ed.)). CV. Mine.
- Drajana, I. C. R., & Bode, A. (2022). Prediksi Status Penderita Stunting Pada Balita Provinsi Gorontalo Menggunakan K-Nearest Neighbor Berbasis Seleksi Fitur Chi Square. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 5(2).
- F. K. Rahim, "Faktor risiko underweight balita umur 7-59 bulan," *KEMAS J.Kesehat. Masy.*, vol. 9, no. 2, pp. 115– 121, 2014.
- F. O. Aridiyah, N. Rohmawati, and M. Ririanty, "Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Balita di Wilayah Pedesaan dan Perkotaan (The Factors Affecting Stunting on Toddlers in Rural and Urban Areas)," *Pustaka Kesehat.*, vol. 3, no. 1, pp. 163– 170, 2015.
- Juwariyem, J., & Sriyanto, S. (2023). PREDIKSI STUNTING PADA BALITA MENGGUNAKAN ALGORITMA RANDOM FOREST. *IndraTech*, 4(1), 29-37.
- Kemenkes RI. (2011). *KEPMENKES RI Tentang Standar Antropometri Penilaian Status Gizi Anak*. In *Jornal de Pediatria* (Vol. 95, Issue 4, p. 41).
- Kemenkes, R. (2020). *Peraturan Kementrian Kesehatan Republik Indonesia Nomor 2 Tahun 2020 Tentang Standar Antropometri Anak*. 68(1), 1– 12
- M. R. Nugroho, R. N. Sasongko, and M. Kristiawan, "Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia," *J. Obs. J. Pendidik. Anak Usia Dini*, vol. 5, no. 2, pp. 2269– 2276, 2021.
- Organization, W. H. (2006). *WHO child growth standards: length/height-for-age, weight for-age, weight-for-length, weight-for-height and body mass index-for-age: methods and development*. World Health Organization.
- Reza, A. A. R., & Rohman, M. S. (2024). Prediction Stunting Analysis Using Random Forest Algorithm and Random Search Optimization. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 7(2), 534-544.
- Siti, M. (2022). *Perancangan Sistem Informasi Aplikasi Gizi Balita Menggunakan Naïve Bayes Classification Dan K-Means* (Doctoral dissertation, UNSADA).
- Tim Riskesdas 2018, (2019). *Laporan Provinsi Jawa Timur Riskesdas 2018*. Surabaya: Lembaga Penerbit Bidang Litbang Kesehatan
- Wahyudi, R. (2018). pertumbuhan dan perkembangan balita stunting (The Growth and Development Overview of the Stunting). *Jurnal Keperawatan*, IV(1)
- Wahyudin, W. C., Hana, F. M., & Prihandono, A. (2023). Prediksi Stunting Pada Balita Di Rumah Sakit Kota Semarang Menggunakan Naive Bayes. *Jurnal Ilmu Komputer Dan Matematika*, 4 (1), 32-36.
- Welassih, B. D. & Wirjatmadi, R. B., 2012. Beberapa Faktor Yang Berhubungan Dengan Status Gizi Balita Stunting. *The Indonesian Journal of Public Health*, VIII(3), pp. 99-104), 56–62.