



Penerapan Algoritma K-Nearest Neighbor (KNN) Dengan Fitur Relief-F Dalam Penentuan Status Stunting

Kemal Musthafa Rajabi^{1✉}, Wina Witanti², Rezki Yuniarti³

Fakultas Sains Dan Informatika, Universitas Jenderal Achmad Yani

Email: kemal.musthafa@student.unjani.ac.id^{1✉}

Abstrak

Stunting pada anak dapat diidentifikasi menggunakan beberapa indikator status gizi, seperti berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U), dan berat badan menurut tinggi badan (BB/TB), dengan menggunakan nilai skor-Z (Z-score) di bawah minus 2 pada indikator TB/U. Saat ini, metode pengukuran status stunting masih menggunakan pendekatan antropometri yang bersifat manual dan subjektif. Untuk mengatasi hal ini, penelitian ini mengembangkan model prediksi status stunting pada anak dengan menggunakan algoritma K-Nearest Neighbor (KNN) dan fitur Relief-F. Dan hasil penelitian menunjukkan bahwa seleksi fitur pada 5 Atribut yang berdasarkan iterasi dari Relief-F berhasil menaikkan nilai akurasi dengan bobot terbaik yang ditemukan pada iterasi ke-5 adalah pada atribut Usia Saat Ukur dan Tinggi. Nilai akurasi terbaik ada pada nilai K = 1, data test 50% (50:50) dengan seleksi fitur 2 atribut menghasilkan akurasi 98,16%.

Kata Kunci : *stunting, k-nearest neighbor, seleksi fitur relief-f*

Abstract

Stunting in children can be identified using various nutritional status indicators, such as weight-for-age (WFA), height-for-age (HFA), and weight-for-height (WFH), using Z-scores below minus 2 on the HFA indicator. Currently, the method of measuring stunting status still relies on manual and subjective anthropometric approaches. To address this issue, this research endeavors to develop a predictive model for stunting status in children utilizing the K-Nearest Neighbor (KNN) algorithm and the Relief-F feature selection technique. The research findings reveal that feature selection based on 5 attributes, determined through Relief-F iterations, successfully enhances accuracy, with the optimal attribute combination identified at the 5th iteration being Age at Measurement and Height. The highest accuracy is achieved at K = 1, employing a 50% test data distribution (50:50), and utilizing the feature selection of 2 attributes, resulting in an accuracy of 98.16%.

Keyword: Stunting, k-nearest neighbors, feature selection use relief-f method

PENDAHULUAN

Stunting adalah keadaan di mana seorang anak memiliki tinggi badan yang lebih pendek dari yang seharusnya sesuai dengan usianya, dan kondisi ini mempengaruhi sekitar 22% dari total jumlah anak-anak di dunia. Stunting pada anak dapat diidentifikasi melalui beberapa indikator status gizi, termasuk berat badan sesuai dengan usia (BB/U), tinggi badan sesuai dengan usia (TB/U), dan berat badan sesuai dengan tinggi badan (BB/TB). Salah satu indikator yang digunakan untuk mengukur pertumbuhan bayi dan anak adalah TB/U dengan menggunakan nilai skor-Z (Z-score). Jika nilai skor-Z kurang dari minus 2, hal ini menandakan adanya stunting. Saat ini, penentuan status stunting pada anak masih mengandalkan metode pengukuran antropometri secara manual dan subjektif, seperti pengukuran berat badan dan tinggi badan.

Salah satu cara untuk menentukan status stunting pada anak adalah dengan menggunakan algoritma K-Nearest Neighbor (KNN). KNN adalah metode klasifikasi terkenal dalam pengolahan data yang telah banyak digunakan dalam berbagai bidang, termasuk dalam konteks kesehatan.

Relief-F adalah sebuah metode seleksi fitur yang dapat membantu dalam mengidentifikasi fitur-fitur yang memiliki pengaruh paling signifikan dalam proses seleksi fitur atau atribut. Adapun penelitian, dilakukan untuk meningkatkan akurasi dalam memprediksi jenis kelamin (Gender) dengan menggunakan metode seleksi fitur Relief-F. Berdasarkan hasil penelitian, setelah dilakukan penghapusan fitur yang kurang penting pada data uji, akurasi yang berhasil diperoleh mencapai 93,35%. Dan penelitian melakukan penelitian dengan mengkombinasikan metode SVM dan Relief-F dalam proses klasifikasi data. Berdasarkan hasil penelitian, setelah menggabungkan SVM dengan Relief-F, terjadi peningkatan akurasi klasifikasi sebesar 24,37%.

Berdasarkan latar belakang tersebut, dari penelitian sebelumnya dimana penggunaan K-Nearest Neighbor dalam berbagai kasus ada sedikit masalah dalam hasil akurasi. Contohnya, fitur atau atribut yang kurang relevan dan metode yang digunakan, dan Relief-F ini berguna untuk seleksi fitur dimana penelitian sebelumnya beda objek atau metode berhasil meningkatkan akurasi dengan seleksi fitur atau atribut yang digunakan.

Tujuan dari penelitian ini adalah menerapkan algoritma K-Nearest Neighbor dengan metode seleksi fitur Relief-F untuk menentukan atribut terbaik dan akurasi pada balita dalam penentuan status stunting atau normal.

Luaran yang diperoleh dari penelitian ini adalah pembuatan prototipe untuk menentukan status stunting pada balita dengan menggunakan algoritma KNN dan metode seleksi fitur Relief-F. Jika hasil akurasi yang didapat baik maka penentuannya tepat dan

akurat dan luaran lainnya yaitu tulisan ilmiah yang dipublikasi pada seminar informatika. Penelitian ini bermanfaat untuk Instansi Kesehatan dan hasil penelitian ini menjadi pendukung dalam penentuan status stunting.

TINJAUAN PUSTAKA

Stunting

Stunting pada anak dapat diidentifikasi melalui beberapa indikator status gizi, termasuk berat badan sesuai dengan usia (BB/U), tinggi badan sesuai dengan usia (TB/U), dan berat badan sesuai dengan tinggi badan (BB/TB). Salah satu indikator yang digunakan untuk mengukur pertumbuhan bayi dan anak adalah TB/U dengan menggunakan nilai skor-Z (Z-score). Jika nilai skor-Z kurang dari minus 2, hal ini menandakan adanya stunting.

Pre-processing

Preprocessing merupakan proses awal yang akan mentransformasikan data masukan menjadi data dengan format yang sesuai dan siap untuk diproses. Beberapa contoh hal yang dilakukan dalam preprocessing meliputi berbagai proses yang diperlukan antara lain : penggabungan, pengubahan bentuk, ataupun pentransformasian data sebagai cara untuk membersihkan, mengintegrasikan, mereduksi dan mendiskritisasi. Lebih lanjut proses yang ada dalam preprosesing dapat terdiri dari salah satu kegiatan proses ataupun gabungan dari beberapa proses diatas. Proses yang ada tergantung dari tujuan yang akan dicapai dalam preprocessing tersebut. Pemilihan proses yang tepat perlu dilakukan mengingat karena proses yang sesuai dalam tahapan preprocessing data akan meningkatkan performansi klasifikasi. Pada penelitian ini, normalisasi data dilakukan dengan menggunakan metode Min Max dengan rumus berikut.

$$Newdata = \frac{(Data - Min) * (NewMax - NewMin)}{(Max - Min)} + NewMin \dots \dots \dots (1)$$

Keterangan:

Newdata = Data hasil normalisasi

Data = Nilai data yang dipilih dari data perkolom

Min = Nilai minimum dari data perkolom

Max = Nilai maksimum dari data perkolom

Newmin = Batas nilai minimum yang kita berikan

Newmax = Batas nilai maksimum yang kita berikan

Relief-F

Relief-F adalah pengembangan dari algoritma Relief yang pertama kali diperkenalkan oleh Kira dan Rendell pada tahun 1992. Kemudian, konsep Relief diperluas dengan menggunakan skor ketergantungan (dependency score) yang dinamakan "Relief-F." Relief-F berfokus pada klasifikasi data yang bersifat biner (hanya memiliki dua kelas) dan multikelas (memiliki lebih dari dua kelas).

Langkah-langkah untuk mencari nilai bobot fitur dengan menggunakan metode Relief-F:

1. Mulai dengan mengatur nilai awal semua bobot fitur menjadi 0 dan menentukan jumlah iterasi yang akan dilakukan.
2. Pilih satu data secara acak sebagai titik pusat atau titik acuan.
3. Temukan data dengan kelas yang sama yang memiliki jarak terdekat dengan titik pusat. Jarak terdekat antara titik pusat dan data dengan kelas yang sama disebut hit, sedangkan jarak terdekat antara titik pusat dan data dengan kelas yang berbeda disebut miss.
4. Perbarui bobot untuk setiap fitur. Untuk fitur dengan data kategorikal, hitung bobot menggunakan persamaan (2) berikut.

$$diff(A, Ri, HM) = \begin{cases} 0; & \text{value}(A, Ri) = \text{value}(A, H, M) \\ 1; & \text{otherwise} \end{cases} \dots\dots\dots(2)$$

Keterangan :

A = Attributes, atau atribut dari dataset.

Ri = Random Instance, data secara acak sebagai titik pusat atau titik acuan

H = Hit, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang sama.

M = Miss, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang beda.

5. Sementara itu, fitur dengan data numerik dihitung menggunakan persamaan (3) berikut,

$$diff(A, Ri, HM) = \frac{|\text{value}(A, Ri) - \text{value}(A, H, M)|}{\max(A) - \min(A)} \dots\dots\dots(3)$$

Keterangan :

A = Attributes, atau atribut dari dataset.

Ri = Random Instance, data secara acak sebagai titik pusat atau titik acuan

H = Hit, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang sama.

M = Miss, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang beda.

Max(A) = Nilai tertinggi pada dataset.

Min(A) = Nilai terendah pada dataset.

6. Kemudian rumus untuk bobot perbaruan dapat dihitung menggunakan persamaan (4) berikut.

$$W[A] = W[A] - \frac{diff(A,Ri,H)}{m} + \frac{diff(A,Ri,M)}{m} \dots\dots\dots(4)$$

Keterangan :

W[A]= Weight, nilai bobot dari atribut yang akan dicari.

A = Attributes, atau atribut dari dataset.

Ri = Random Instance, data secara acak sebagai titik pusat atau titik acuan

H = Hit, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang sama.

M =Miss, data yang memiliki jarak terdekat dengan titik pusat dan mempunyai kelas yang beda.

m= Banyaknya iterasi yang akan dilakukan.

7. Langkah selanjutnya adalah melanjutkan iterasi berikutnya yang dimulai dari langkah 1 hingga bobot fitur yang baru berhasil diperoleh.

K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah salah satu algoritma dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. K-NN adalah algoritma yang sederhana tetapi kuat dalam mengklasifikasikan objek atau contoh data berdasarkan kelas mayoritas tetangga terdekatnya. Algoritma ini termasuk dalam kategori algoritma pembelajaran berbasis instan (instance-based learning). Selain itu, metode K-Nearest Neighbor memiliki kelebihan yakni kemudahan dalam implementasi dan kemampuan yang baik dalam menangani dataset yang besar.

Langkah-langkah metode K-Nearest Neighbor :

- Pilih jumlah k yang akan digunakan.
- Hitunglah jarak antara data training dan data testing dengan menggunakan persamaan yang sesuai, ditunjukkan pada persamaan (5).

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \dots\dots\dots(5)$$

Keterangan :

X_{1i} = Data latih atau training dari dataset yang akan diklasifikasi.

X_{2i} = Data uji atau testing dari dataset yang akan diklasifikasi.

$\sum_{i=0}^n (x_{1i} - x_{2i})^2$ = Jumlah hasil jarak perhitungan Euclidean Distance.

- Urutkan jarak secara menaik dan identifikasi k jarak terdekat.
- Pilih kelas yang cocok untuk data yang memiliki jarak terdekat tersebut..

- e. Cari jumlah tetangga terdekat yang paling banyak, kemudian tentukan kategori data tersebut sebagai kategori yang sedang dicari.

METODE PENELITIAN

Prosedur penelitian yang dilakukan dalam klasifikasi status gizi balita melibatkan penggunaan metode KNN dengan fitur Relief-F yang akan diimplementasikan untuk menentukan status stunting pada balita. Berikut adalah tahapan dalam penelitian tersebut.

Penelitian ini melibatkan beberapa tahapan, pertama adalah persiapan data, di mana data balita dipersiapkan sebagai objek penelitian untuk menentukan status stunting, kedua melibatkan implementasi proses data mining dengan menggunakan algoritma K-Nearest Neighbor dengan Fitur Relief-F.

Pada penelitian ini, langkah pertama yang dilakukan ialah studi pustaka untuk mengumpulkan berbagai informasi dan sebagai referensi terkait dengan topik yang dicari yaitu K-Nearest Neighbor, Data mining, dan Relief-F. Selanjutnya pengumpulan data untuk perolehan data, pada tahap ini melakukan observasi pengumpulan data balita pada Puskesmas Tanjungsari.

Kemudian di peroleh dataset balita pada Puskesmas Tanjungsari dari tahun 2020 sampai dengan 2022 sebanyak 6.913 record data. Data pada tahun 2020 sebanyak 2.252 data, pada tahun 2021 sebanyak 2.247 data, dan pada tahun 2022 sebanyak 2.414 data.

Untuk penjelasannya ada pada Gambar 1. berikut.



HASIL DAN PEMBAHASAN

Gambar 1. Diagram metode penelitian

Analisis Dan Perancangan

Pengumpulan data pada tahap ini diawali dengan melakukan wawancara ke bagian petugas Kesehatan gizi Puskesmas Tanjungsari, hasil dari wawancara tersebut memperoleh data balita dari tahun 2020 sampai 2022 sebanyak 6.913 record. Data pada tahun 2020 sebanyak 2.252 data, pada tahun 2021 sebanyak 2.247 data, dan pada tahun 2022 sebanyak 2.414 data.

Selanjutnya, dilakukan tahap pre-processing yang terdiri dari dua proses, yaitu pembersihan data (data cleaning) dan seleksi data (data selection). Pada tahap ini dilakukannya pembersihan jika ada data yang kosong atau hilang, serta data yang terdapat kesalahan. Dataset balita akan mengalami proses seleksi terlebih dahulu. Seleksi data dilakukan dengan menggunakan dataset yang terdiri dari 6.913 record dari tahun 2020 hingga 2022, dengan 32 kolom atau atribut yang mencakup informasi seperti NIK, Nama, Jenis Kelamin, Tanggal Lahir, Berat Badan Lahir, Tinggi Badan Lahir, Nama Orang Tua, Provinsi, Kabupaten/Kota, Kecamatan, Puskesmas, Desa/Kelurahan, Posyandu, RT, RW, Alamat, Usia saat Pengukuran, Tanggal Pengukuran, Berat Badan, Tinggi Badan, Lingkar Lengan Atas (LiLA), Berat Badan Menurut Usia (BB/U), Z-Score BB/U, Tinggi Badan Menurut Usia (TB/U), Z-Score TB/U, Berat Badan Menurut Tinggi Badan (BB/TB), Z-Score BB/TB, Kenaikan Berat Badan, Jumlah Vitamin A, KPSP, KIA.

Pada penelitian ini diperoleh data balita dengan tipe data tidak numerik. Sehingga, perlu dilakukan transformasi data dengan bentuk kualitatif menjadi kuantitatif. Tahap ini diperlukan karena data yang ada bukan data tipe numerik. Dalam tahap transformasi ini menggunakan teknik preprocessing untuk merepresentasikan data kategori agar dapat diproses dengan algoritma K-Nearest Neighbors (KNN), perlu mengubah variabel kategori menjadi bentuk numerik dengan Label Encoding atau Ordinal Encoding. Metode Label Encoding ini mengubah setiap kategori yang menggantikan nilai-nilai kategori dengan bilangan bulat berurutan. Sedangkan dalam Ordinal Encoding disini diurutkan berdasarkan tingkatnya seperti kriteria "Tinggi", "Normal", "Pendek", dan "Sangat Pendek" diubah kedalam numerik menjadi "0", "1", "2", dan "3".

Pada tahap ini Proses preprocessing data dilakukan sebelum memulai tahap data mining. Teknik praprocessing diperlukan karena jika kualitas data tidak baik, akan menyebabkan kualitas data mining yang buruk, sehingga pembersihan data dapat dilakukan untuk memastikan pemulihan data yang baik.

Dalam tahap ini, analisis data menggunakan metode data mining yang diterapkan dengan menggunakan algoritma klasifikasi K-Nearest Neighbor. Pengumpulan data dilakukan menggunakan dataset balita dari Puskesmas Tanjungsari, yang mencakup periode

tahun 2020 hingga 2022. Total dataset terdiri dari 6.913 record, dengan 2.252 data pada tahun 2020, 2.247 data pada tahun 2021, dan 2.414 data pada tahun 2022. Data tersebut disimpan dalam format dokumen excel.

Selanjutnya, dilakukan tahap pre-processing yang terdiri dari dua proses, yaitu pembersihan data (data cleaning) dan seleksi data (data selection). Setelah proses pre-processing selesai, dilakukan tahap transformasi data untuk mengubah data kategori pada nilai atribut dengan inisialisasi numerik dan mengubah format data excel dari sebelumnya xlsx menjadi format csv, hal ini dilakukan untuk mempermudah saat proses penginputan.

Selain itu, data juga dibagi menjadi data uji dan data latih. Data latih digunakan untuk melatih model algoritma K-Nearest Neighbor, sedangkan data uji digunakan untuk mengevaluasi performa model yang telah dilatih.

Proses pengolahan data dilakukan pada keseluruhan proses mengolah data set untuk nantinya diinputkan ke dalam sistem. Tahapan pada pengolahan data ini dimulai dari menyeleksi data hingga tahap hasil.

Perhitungan Relief-F bertujuan untuk pencarian bobot terbaik. Pada perhitungan ini data yang diambil untuk perhitungan manual sebanyak 20 baris. Pada tahap perhitungan hasil akurasi klasifikasi status stunting melalui beberapa tahap. Sebelum sistem diimplementasikan, dilakukan pengujian menggunakan matriks kebingungan (Confusion Matrix) untuk mendapatkan nilai akurasi yang dihasilkan. Hal ini bertujuan untuk memastikan bahwa pola yang dibentuk oleh sistem sesuai dengan model klasifikasi KNN.

Pada perhitungan iterasi ke-5 nilai awal semua bobot fitur didapat dari hasil bobot pada iterasi ke-4. Kemudian nilai nearest hit adalah -4,67 dan untuk nilai nearest miss adalah -4,94. Lalu kemudian bisa dilihat untuk nilai bobot nya. Berikut penjelasannya ada pada Tabel 1.

ITERASI KE-5		W[A] = HASIL ITERASI KE-4. Ri=16								
		16 .	1	53	13,8	95,4	1	-1,49	2	-2,25
NEAREST HIT		15 .	1	54	16,3	95,6	1	-0,39	2	-2,38
NEAREST MISS		18 .	1	52	14,6	100	1	-1,03	1	-1,17
HIT	-4,67		JK	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U
MISS	-4,94		0,2	-0,07895	0,067568	-0,05435	-0,4	0,258101	-0,2	0,139279

Tabel 1 Iterasi ke-5 1

Pada Tabel di atas dapat dilihat bahwa nilai bobot dapat diidentifikasi setiap atributnya yang nanti akan dipakai dalam seleksi atribut.

Dalam perancangan perangkat lunak untuk sistem klasifikasi data balita, diperlukan

analisis kebutuhan perangkat yang dijelaskan melalui kebutuhan fungsional. Perancangan perangkat lunak dalam penelitian ini menggunakan model Unified Modeling Language (UML) dengan beberapa diagram yaitu, use case diagram, activity diagram sebagai representasi pengembangan dari use case, dan Sequence Diagram.

Implementasi Dan Pengujian

Penelitian ini melibatkan pengembangan perangkat lunak berbasis web yang sesuai dengan kebutuhan menggunakan bahasa pemrograman PHP. Dalam pengembangan tersebut, digunakan alat bantu Visual Studio dan database MySQL yang terintegrasi dengan aplikasi XAMPP. Sebagai media untuk menjalankan perangkat lunak, digunakan web browser Google Chrome. Database diimplementasikan sesuai dengan perancangan database sebelumnya menggunakan MySQL.

Pada tahap ini pengujian dilakukan untuk melakukan perbandingan dan implementasi seleksi fitur pada data balita dalam penentuan status stunting dengan semua 8 atribut yang dipakai. Dan untuk mengukur jaraknya menggunakan Euclidean Distance agar bisa memproses dan mencari nilai K atau data terdekat yang akan dicari pada proses perhitungan.

Perhitungan Confusion Matrix pada K-Nearest Neighbour dengan menghitung Accuracy, Precision, dan Recall sebagai berikut:

Precision

a. $\text{Precision} : \text{TP} / (\text{TP} + \text{FP}) \times 100\%$

-TP : True Positif

-FP : False Positif

1. Precision pada kelas Stunting

$$= 161 / (161+5)$$

$$= 0,97 \times 100\%$$

$$= 97\%$$

2. Precision pada kelas Normal

$$= 480 / (480 + 28)$$

$$= 0,945 \times 100\%$$

$$= 94,5\%$$

Hasil rata-rata precision : 95,75%

b. $\text{Recall} : \text{TP} / (\text{TP} + \text{FN}) \times 100\%$

-TP : True Positif

-FN : False Negatif

1. Recall pada kelas Stunting

$$= 161 / (161 + 28)$$

$$= 0,852 \times 100\%$$

$$= 85,2\%$$

2. Recall pada kelas Normal

$$= 480 / (480 + 5)$$

$$= 0,99 \times 100\%$$

$$= 99\%$$

Hasil rata-rata recall : 92,1%

c. Akurasi : $\frac{TP+TN}{TP+TN+FP+FN} \times 100\%$

-TP : True Positif

-FN : False Negatif

-FP : False Positif

1. Akurasi pada kelas Stunting

$$= 161 + 480 / (161 + 480 + 5 + 28) \times 100\%$$

$$= 0,951 \times 100\%$$

$$= 95,1\%$$

2. Akurasi pada kelas Normal

$$= 480 / (480 + 161 + 28 + 5)$$

$$= 0,951 \times 100\%$$

$$= 95,1\%$$

Hasil rata-rata akurasi : 95,1%

Pengujian Model klasifikasi K-Nearest Neighbor

Pada tahap ini dilakukan pengujian prediksi pada data baru. Dengan klasifikasi data balita 2020 – 2021. Berikut data yang digunakan :

a. Pengujian pada data imbalance

Data baru yang akan diinput yaitu :

Nilai K = 5

Usia Saat Ukur = 50

Berat = 76

Tinggi = 96

ZS BB/U = -1.25

ZS TB/U = -2.5

Nomor	Usia Saat Ukur	Berat	Tinggi	ZS BB/U	ZS TB/U	Status Stunting	Jarak
5766	0.917	1	0.961	1	0.883	Normal	1.533
4805	0.867	0.955	0.955	0.995	0.892	Normal	1.57
4607	0.983	0.843	0.955	0.773	0.863	Normal	1.606
5186	0.85	0.843	0.902	0.898	0.851	Normal	1.631
2625	0.767	0.846	0.902	0.904	0.877	Normal	1.633

Tabel 2. Hasil uji klasifikasi data imbalance 5 atribut

]

Berdasarkan perhitungan, dengan Usia Saat Ukur: 50, Berat: 76, Tinggi: 96, ZS BB/U: -1.25, ZS TB/U: -2.5, maka hasilnya: Normal.

b. Pengujian pada data balance

Data baru yang akan diinput yaitu :

Nilai K = 5

Usia Saat Ukur = 50

Berat = 76

Tinggi = 96

ZS BB/U = -1.25

ZS TB/U = -2.5

Nomor	Usia Saat Ukur	Berat	Tinggi	ZS BB/U	ZS TB/U	Status Stunting	Jarak
125	0.867	1	0.955	1	0.892	Normal	1.611
2097	0.983	0.883	0.955	0.796	0.863	Normal	1.667
1754	0.767	0.886	0.902	0.916	0.877	Normal	1.687
2304	0.95	0.842	0.929	0.788	0.851	Normal	1.699
228	0.767	0.819	0.886	0.907	0.861	Normal	1.747

Tabel 3. Hasil uji klasifikasi data balance 5 atribut

Berdasarkan perhitungan, dengan Usia Saat Ukur: 50, Berat: 76, Tinggi: 96, ZS BB/U: -1.25, ZS TB/U: -2.5, maka hasilnya: Normal.

Setelah melakukan pengujian Confusion Matrix dan black box testing, maka dapat disimpulkan bahwa didapatkan hasil untuk mengukur performa dari metode K-Nearest Neighbor ini dengan menggunakan 3 metode evaluasi pada confusion matrix yaitu Accuracy, Precision, Recall. Berikut adalah kesimpulan dari pengujian metode dengan confusion matrix dengan mengambil rata-rata dari 5 Atribut K=1 tidak balance dan data balance.

Tabel 4. 1 Confusion Matrix rata-rata pada Data tidak balance

Klasifikasi	Accuracy	Precision	Recall
5 Atribut	95.17%	96.85%	88.65%
4 Atribut	96.62%	97.85%	92%
3 Atribut	97.18%	98.15%	93.35%
2 Atribut	98.16%	98.8%	95.65%

Tabel 4. 2 Confusion Matrix rata-rata pada Data balance

Klasifikasi	Accuracy	Precision	Recall
5 Atribut	90.23%	90.45%	90.25%
4 Atribut	90.99%	91.2%	91%
3 Atribut	93.88%	94.1%	93.9%

Berdasarkan pada hasil pengujian yang dilakukan seperti yang tercantum dalam tabel, dapat dihitung persentase kesesuaian sistem dengan menggunakan fungsi kesesuaian sebagai berikut :

Jumlah dari Kode	= 6
Pengujian	
Kode pengujian sesuai	= 6
Kode pengujian tidak sesuai	= 0

$$\begin{aligned}\text{Persentase} &= \frac{(\text{jumlah kode uji sesuai} - \text{jumlah kode uji tidak sesuai})}{\text{jumlah kode uji}} \\ &\quad \times 100 \\ \text{Persentase} &= \frac{(6 - 0)}{6} \times 100 = 100\%\end{aligned}$$

Hasil perhitungan fungsi kesesuaian sistem dapat disimpulkan bahwa pengujian terhadap Perangkat Lunak Sistem menggunakan black box testing telah berjalan sesuai dengan spesifikasi yang telah ditetapkan dengan persentase 100%.

SIMPULAN

Berdasarkan pada hasil penelitian yang dilakukan dalam penerapan algoritma k-nearest neighbor (knn) dengan fitur relief-f dalam penentuan status stunting. Dataset yang digunakan dari tahun 2020 sampai 2022 sebanyak 6.913 record data. Dapat disimpulkan dari eksperimen dataset yang tidak balance dan yang balance di setiap kelasnya, eksperimen nilai K pada splitting data 10% (90 : 10) , splitting data 20 % (80 : 20), splitting data 30% (70 : 30), splitting data 40% (60 : 40) , dan splitting data 50% (50 : 50), secara umum menunjukkan bahwa peningkatan akurasi terhadap K-Nearest Neighbor pada kasus data balita dalam penentuan stunting tidak terlalu signifikan. Dan hasil penelitian menunjukkan bahwa seleksi fitur pada 5 Atribut yang berdasarkan iterasi dari Relief-F berhasil menaikkan nilai akurasi dengan bobot terbaik yang ditemukan pada iterasi ke-5 adalah pada atribut Usia Saat Ukur dan Tinggi. Nilai akurasi terbaik ada pada nilai K = 1, data test 50% (50:50) dengan seleksi fitur 2 atribut menghasilkan akurasi 98,16%.

Berdasarkan hasil penelitian yang telah dilakukan, terdapat peluang untuk melakukan pengembangan dalam identifikasi atribut relevan untuk data kategorial yang dapat mempengaruhi nilai akurasi yang tidak stabil. Dan juga dapat dikembangkan lebih lanjut mengenai pemilihan atribut yang dipakai dalam penentuan status stunting atau normal pada balita.

DAFTAR PUSTAKA

- Amaha, N.D. (2021) 'Maternal factors associated with moderate and severe stunting in Ethiopian children: analysis of some environmental factors based on 2016 demographic health survey', 6, pp. 1–9.
- Bachri, O.S., Mohamad, R. and Bhakti, H. (2021) 'PENENTUAN STATUS STUNTING PADA ANAK', 3(2), pp. 130–137.
- Kononenko, I. (no date) 'Estimating Attributes : Analysis and Extensions of RELIEF'.

- Mahasiswa, B. and Masyarakat, K. (no date) *STUDY GUIDE - STUNTING DAN UPAYA PENCEGAHANNYA STUDY GUIDE - STUNTING DAN UPAYA*.
- Neighbor, K. and Ariani, A. (2019) 'Klasifikasi Penyakit Ginjal Kronis menggunakan', 5(1), pp. 978–979.
- W. C. Wahyudin and F. Selection, "KLASIFIKASI STUNTING BALITA MENGGUNAKAN NAIVE BAYES," vol. 1, pp. 71–74, 2020.
- S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," vol. 6, pp. 49–56, 2022, doi: 10.30865/mib.v6i1.3312.
- N. Pudjihartono, T. Fadason, and A. W. Kempa-liehr, "A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction," vol. 2, no. June, pp. 1–17, 2022, doi: 10.3389/fbinf.2022.927312.
- T. R. Reddy, B. V. Vardhan, M. Gopichand, and K. Karunakar, *Gender Prediction in Author Pro fi ling Using ReliefF Feature Selection Algorithm*. Springer Singapore.
- J. H. J. Z. L. Zheng, "Support Vector Machine Classification Algorithm Based on Relief-F Feature Weighting," pp. 547–553, 2020, doi: 10.1109/ICCEA50009.2020.00121.
- keputusan M. K. R. nomor 1995/MENKES/SK/XII/2010, "STANDAR ANTROPOMETRI PENILAIAN STATUS GIZI ANAK." .
- D. T. Larose and C. D. Larose, *Knowledge in data wiley series on methods and applications*. 2014.