Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 4 Iss. 1 January 2024, pp: 257-265
P-ISSN: 2797-2313 | ISSN(E): 2775-8575

Comparative Analysis of Machine Learning Algorithms for Predicting Child Stunting

Analisis Perbandingan Algoritma *Machine Learning* untuk Prediksi Stunting pada Anak

Indah Pratiwi Putri[1], Terttiaavini[2*], Nur Arminarahmah[3]

[1,2]Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Sains,
Universitas Indo Global Mandiri, Indonesia
[3]Program Studi Teknik Informatika, FakultasTeknologi Informasi,
Universitas Islam Kalimantan MAB Banjarmasin, Indonesia

E-Mail: [1]wiwid@uigm.ac.id, [2]avini.saputra@uigm.ac.id, [3]nur.armina@gmail.com

Received Nov 04th 2023; Revised Dec 28th 2023; Accepted Jan 14th 2024
Corresponding Author: Terttiaavini

Abstract

This study highlights the serious issue of childhood stunting, particularly inconsistent data collection and the lack of accurate information in evaluating this condition. Its aim is to develop a Machine Learning (ML) model to predict stunting cases more effectively. The research methodology involves three ML algorithms: Naive Bayes, K-Nearest Neighbors, and Random Forest, evaluated based on Accuracy, Precision, and Recall. This research utilises the KNIME platform to help manage data more efficiently and accurately. The evaluation results indicate that Random Forest exhibits the highest accuracy (87.75%) and F1-score (0.922), demonstrating a good balance between Precision and Recall. However, K-Nearest Neighbors excel in identifying a majority of the actual stunting cases. Consequently, the Random Forest model might be the optimal choice for diagnosing stunting in children due to its high accuracy and superior ability to detect stunting cases compared to other models. This study provides insights into applying ML to support early detection of stunting, enabling more precise and prompt healthcare interventions for children requiring intensive attention.

Keyword: Childhood Stunting, K-Nearest Neighbors, KNIME, Machine Learning, Naive Baye, Random Forest

Abstrak

Penelitian ini menyoroti permasalahan serius stunting pada anak-anak, terutama dalam pendataan yang tidak konsisten dan kurangnya informasi akurat dalam evaluasi kondisi tersebut. Tujuannya adalah mengembangkan model Machine Learning (ML) untuk memprediksi kasus stunting dengan lebih baik. Metode penelitian melibatkan tiga algoritma ML: *Naive Bayes*, *K-Nearest Neighbors*, dan *Random Forest*, dievaluasi berdasarkan *Accuracy*, *Precision*, dan *recall*. Penelitian ini memanfaatkan platform KNIME untuk membantu pengelolaan data yang lebih efisien dan akurat. Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi (87.75%) dan *F1-score* (0.922), menunjukkan keseimbangan yang baik antara *Precision* dan *recall*. Meskipun demikian, K-Nearest Neighbors menonjol dalam menemukan sebagian besar kasus stunting yang sebenarnya. Kesimpulannya, model *Random Forest* mungkin menjadi pilihan terbaik untuk mendiagnosis stunting pada anak-anak, karena kombinasi akurasi tinggi dan kemampuan menemukan kasus stunting yang lebih baik dari model lainnya. Penelitian ini memberikan wawasan tentang penerapan ML dalam mendukung deteksi dini stunting, memungkinkan intervensi yang lebih tepat dan cepat bagi anak-anak yang membutuhkan perhatian kesehatan yang lebih intensif.

Kata Kunci: *KNIME, K-Nearest Neighbors, Naive Bayes, Random Forest*, Stunting pada Anak

1. PENDAHULUAN

Stunting pada anak-anak merupakan permasalahan kesehatan yang serius yang harus segera ditangani. Data dari PBB tahun 2020 menunjukkan bahwa lebih dari 149 juta atau sekitar 22% balita di seluruh dunia menderita stunting. Di Indonesia, sekitar 6,3 juta balita terkena dampak kondisi ini [1], [2]. Indonesia sendiri menargetkan penurunan stunting sebesar 14% di tahun 2024 [3]. Namun, upaya penanggulangan stunting menghadapi tantangan dalam pendataan stunting yang lengkap dan konsisten. Variabilitas dalam standar

pengukuran, perbedaan metode pengumpulan data, serta keterbatasan sumber daya manusia dan teknologi di beberapa wilayah Indonesia menjadi hambatan utama dalam upaya pendataan stunting.

Untuk menyimpulkan kondisi seorang balita atau anak terkait stunting, dibutuhkan pendekatan yang holistik dan data yang komprehensif. Evaluasi stunting pada seorang anak dapat melibatkan beberapa aspek, yaitu pengukuran antropometri, penilaian kesehatan secara keseluruhan, pemeriksaan gizi dan nutrisi, asesmen psikososial, riwayat perkembangan [4]. Namun, terkadang informasi yang diperoleh untuk melakukan evaluasi tidak akurat, hal ini dapat menghasilkan kesimpulan yang kurang tepat tentang kondisi stunting anak. Akibatnya, penanganan atau intervensi yang diperlukan untuk mengatasi masalah stunting pada anak tersebut menjadi lambat.

Ketidaklengkapan atau ketidakakuratan informasi dalam evaluasi stunting dapat memperlambat identifikasi kasus-kasus stunting yang memerlukan perhatian khusus dan intervensi segera. Permasalahan ini membutuhkan penanganan dalam memprediksi untuk menangani kasus-kasus stunting secara lebih efisien. Diperlukan pendekatan yang lebih canggih dalam pengumpulan dan analisis data, serta perbaikan dalam infrastruktur pendataan dan penilaian stunting. Dalam menangani masalah evaluasi stunting pada anak-anak, diperlukan penerapan teknik evaluasi yang lebih inovatif dengan menggunakan *machine learning* (ML) untuk memprediksi kemungkinan anak teridentifikasi mengalami stunting, sehingga langkah-langkah preventif dapat dilakukan secara lebih tepat dan proaktif. Penerapan ML dalam klasifikasi melibatkan sejumlah algoritma yang dapat digunakan [5], [6]. Untuk mengembangkan model klasifikasi yang optimal, diperlukan eksperimen perbandingan beberapa algoritma menggunakan *dataset* yang spesifik agar mendapatkan gambaran yang komprehensif terkait performa algoritma klasifikasi dalam memprediksi kasus stunting dengan tingkat akurasi dan responsifitas yang lebih baik [7].

Tujuan dari penelitian ini adalah menentukan model klasifikasi yang terbaik dengan cara membandingkan kinerja dari tiga algoritma *machine learning* untuk memprediksi kasus stunting pada anak. Diharapkan hasil penelitian akan memberikan pemahaman yang lebih tentang performa algoritma klasifikasi, sehingga memungkinkan pengembangan model yang lebih efektif dalam mendiagnosis kasus stunting pada anak lebih dini. Penelitian ini memanfaatkan platform KNIME untuk mengelola, membersihkan, dan menganalisis data stunting pada anak-anak. Melalui alat ini, peneliti dapat melakukan berbagai langkah pemrosesan data seperti pembersihan, transformasi, pemilihan fitur, dan pemisahan data secara efisien [8]. Harapannya, model hasil penelitian ini mampu menjadi landasan yang baik untuk deteksi dini kasus stunting, memungkinkan intervensi yang lebih tepat dan cepat bagi anak-anak yang membutuhkan perhatian kesehatan yang lebih intensif. Dengan perbaikan dalam metode evaluasi, diharapkan dapat mengurangi dampak negatif jangka panjang dari kondisi stunting terhadap pertumbuhan dan perkembangan anak-anak.

2. LITERATURE REVIEW

2.1. Studi terdahulu tentang Stunting pada Anak-anak

Stunting merujuk pada kondisi terhambatnya pertumbuhan anak secara fisik yang umumnya disebabkan oleh kekurangan gizi kronis dan kondisi lingkungan yang tidak memadai selama periode kritis pertumbuhan, biasanya dari masa kehamilan hingga dua tahun pertama kehidupan [4]. Stunting ditandai dengan tinggi badan yang lebih pendek dari rata-rata usia dan jenis kelamin anak, dan bisa memiliki dampak jangka panjang terhadap kesehatan fisik, kognitif, dan perkembangan anak. Sejumlah studi terdahulu telah secara konsisten menyoroti stunting pada anak-anak sebagai hasil dari beragam faktor. Faktor risiko terkait stunting pada anak-anak meliputi kondisi gizi yang tidak memadai, lingkungan yang kurang higienis, akses yang terbatas terhadap pelayanan Kesehatan [9]–[11], serta faktor sosial ekonomi yang rendah [12]–[14].

Dengan pemahaman yang lebih pada faktor-faktor penyebab stunting, memungkinkan penyedia layanan kesehatan dan pemangku kepentingan terkait dapat mengimplementasikan strategi yang lebih efektif untuk meningkatkan pemahaman pada masyarakat, guna mengurangi prevalensi stunting.

2.2. Model Klasifikasi untuk Mendiagnosis Stunting

Model klasifikasi akan mencapai optimal, jika menggunakan algoritma yang sesuai dengan karakteristik data. Tidak ada satu metode yang menjadi pilihan terbaik untuk semua situasi. Pemilihan model yang optimal membutuhkan pemahaman mendalam terhadap data dan tujuan analisisnya. Algortima klasifikasi dapat diterapkan mendiagnosis Stunting adalah *Random Forest*, *Naive Bayes*, *K-Nearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Decision Tree* dan *Logistic Regression*. Beberapa penelitian yang bertujuan untuk membangun model klasifikasi stunting dilakukan oleh M. S. Haris, M. Anshori dan A. N. Khudori (2023) membahasa tentang prevalensi stunting menggunakan *Algoritma Random Forest* dengan menggunakan data stunting di Provinsi Jawa Timur, menghasilkan nilai akurasi = 93% [15], penelitian lain yang dilakukan oleh Harliana dan Anggraini (2023) menggunakan algoritma *naïve bayes* dengan sumber data stunting dari Posyandu Desa Kalitengah menghasilkan nilai akurasi 87,3% [16]. Penelitian yang dilakukan oleh S. Lonang and D. Normawati (2022) menggunakan algoritma *K-Nearest Neighbors* (KNN) pada seluruh data anak Indonesia dengan penambahan teknik *Backward Elimination* dan parameter k=8 mencapai akurasi sebesar 92,20% [17].

Beberapa studi membandingkan berbagai model untuk menentukan model yang paling efektif dalam melakukan klasifikasi [18]. Evaluasi terhadap keakuratan, kehandalan, dan kesesuaian model tersebut dalam mendiagnosis stunting memberikan wawasan penting dalam mengembangkan solusi prediktif yang lebih tepat dan responsif terhadap kondisi stunting pada anak-anak. Penelitian yang dilakukan oleh Lonang dkk (2023) melakukan evaluasi komparatif pada lima algoritma klasifikasi, yakni *Logistic Regression*, *Decision Tree* [19], *Random Forest*, *K-Nearest Neighbors* (KNN), dan *Support Vector Machine* (SVM) untuk mengklasifikasi balita stunting. Studi ini menggunakan *dataset* dari pencatatan dan pelaporan gizi balita berbasis masyarakat (EPPGBM) Puskesmas Ubung, Lombok Tengah, Indonesia. Hasil penelitian ini menyimpulkan bahwa model KNN dinilai sebagai yang terbaik, dengan tingkat akurasi mencapai 94,85% [20].

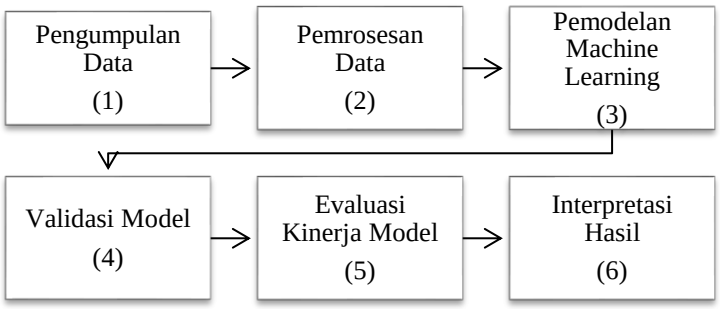
Penelitian berbeda dilakukan oleh Kusumaningrum dkk (2020) mengevaluasi secara komparatif kinerja empat algoritma, yaitu *Regresi logistik*, *Naive Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM) menggunakan data Indonesia. Metode SVM digabungkan dengan TF-IDF menghasilkan nilai akurasi tertinggi sebesar 0,98 dengan standar deviasi 0,03 [21].

Perbandingan antara model klasifikasi dapat melibatkan beberapa metrik evaluasi untuk menghasilkan nilai Akurasi (*Accuracy*), Presisi (*Precision*), *Recall* (*Sensitivity*), *F1-Score* dan *Confusion Matrix. hasil* Perbandingan model klasifikasi memberikan gambaran yang jelas tentang performa relatif dari setiap model dalam mendiagnosis stunting pada anak-anak. Oleh karena itu, dengan mempertimbangkan keseluruhan evaluasi metrik-metrik tersebut akan mendapatkan gambaran yang holistik dan memilih model yang paling sesuai dengan kebutuhan spesifik dalam penanganan stunting pada anak-anak.

3. METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Metodologi penelitian merujuk pada langkah-langkah sistematis yang digunakan untuk merencanakan, melaksanakan, dan menganalisis sebuah penelitian. Pada penelitian ini, metodologi yang digunakan terdiri dari beberapa tahap. Gambar 1 menjelaskan tentang tahapan penelitian dalam bentuk diagram.



Gambar 1. Tahapan penelitian

Tahapan penelitian secara rinci melibatkan proses yang terstruktur untuk mengarahkan dan melaksanakan penelitian ini. Berikut adalah penjelasan dari tahapan penelitian tersebut:

3.2. Pengumpulan Data

Data yang digunakan tersebut terdiri dari jenis kelamin (*gender*), umur (*age*), berat lahir (*birth weight*), panjang lahir (*birth length*), berat badan (*body weight*), panjang badan (*body length*), menyusui (*breastfeeding*). Sumber data berasal dari data stunting di Indonesia tahun 2022 dengan jumlah data = 10.000 Jumlah data dari masing -masing atribut dijelaskan pada tabel 1.

Tabel 1. Jumlah data pada masing-masing atribut

Nama Variabel	Katagori dan jumlah data
Jenis kelamin	Male = 6204
	Female = 3792
Umur	0-5 bulan = 0
	6-11 bulan = 3827
	12-23 bulan = 5430
	24-35 bulan = 248
	36-47 bulan = 358
	48-59 bulan = 137
Berat Lahir	2.0 Kg = 678
	2.3 Kg = 1130

Nama Variabel	Katagori dan jumlah data
	2.7 Kg = 483
	2.8 Kg = 3409
	2.9 Kg = 1807
	3.0 Kg = 1969
	3.1 Kg = 524
Panjang Lahir	48 cm = 449
	49 cm = 7928
	50 cm = 1623
berat badan	< 6.0 Kg = 610
	6.0 – 6.9 Kg = 1931
	7.0 – 7.9 Kg = 3000
	8.0 – 8.9 Kg = 1722
	9.0 – 9.9 Kg = 1520
	10.0 – 10.9 Kg = 1217
Panjang Badan	< 50 cm =702
	50 – 59 cm = 0
	60 – 69 cm = 5729
	70 – 79 cm = 2492
	80 – 89 cm = 492
	90 – 99 cm = 585
Stunting	Yes = 7955
	No = 2045

Sumber data: data stunting anak di Indonesia tahun 2023

Dataset stunting menggunakan tipe data numerik untuk mewakili katagori yang relevan dengan pertumbuhan dan kondisi kesehatan anak. Data numerik ini memungkinkan model *machine learning* dapat bekerja dengan baik pada dataset tersebut.

3.3. Pemrosesan Data

Penelitian ini menggunakan KNIME dalam pemrosesan data. KNIME merupakan salah satu platform analisis data yang memungkinkan pemrosesan data yang efisien untuk pembangunan model data [22]. Dalam konteks penelitian ini, KNIME digunakan untuk berbagai langkah pemrosesan data, termasuk membersihkan data, transformasi, pemilihan fitur, dan pemisahan data. Kelebihan KNIME adalah antarmuka yang intuitif dengan pendekatan berbasis simpul (*node-based*) yang memungkinkan peneliti untuk mengatur alur kerja (*workflow*) secara visual. Dengan memanfaatkan node-node yang tersedia, para peneliti dapat membangun dan mengatur proses pemrosesan data secara sistematis dan efektif serta memberikan fleksibilitas dalam menyesuaikan alur kerja (*workflow*) sesuai dengan kebutuhan analisis yang spesifik [23].

3.4. Pemodelan *Machine Learning* (ML)

Pemodelan ML pada penelitian ini menggunakan tiga model klasifikasi yaitu *Naive Bayes*, *Random Forest* dan *K-Nearest Neighbors* (KNN).

3.4.1 Naive Bayes

Naive Bayes adalah algoritma klasifikasi menggunakan teorema Bayes yang digunakan untuk memperbarui probabilitas suatu kejadian berdasarkan informasi baru. Secara umum, rumus Bayes menyatakan bahwa probabilitas suatu kejadian yang terjadi (*posterior*) berdasarkan informasi yang telah diamati (*prior*) dapat dihitung dari probabilitas kejadian tersebut terjadi (*likelihood*) dan probabilitas kejadian lain yang mungkin terjadi (*evidence*).

Dalam rumus ini, *posterior probability* (probabilitas kelas setelah observasi fitur) dapat dihitung dari *likelihood probability* (probabilitas fitur terjadi pada kelas tertentu) dan *prior probability* (probabilitas kelas sebelum melihat fitur). Langkah-langkah algoritma *Naïve Bayes* adalah sebagai berikut :

1. Persiapan Data: Siapkan dataset dengan fitur-fitur yang menggambarkan entitas dan label kelas yang ingin diprediksi. Bagi data menjadi data latih dan data uji.
2. Perhitungan Probabilitas kelas: Hitung probabilitas masing-masing kelas dari data latih. Rumus Probabilitas Kelas (Priors), yaitu:

$$P(C_k) = \frac{jumlah\ sampel\ dengan\ kelas\ C_k}{total\ jumlah\ sampel} \quad (1)$$

3. Perhitungan Probabilitas Fitur Terhadap Kelas: Hitung probabilitas masing-masing fitur untuk setiap kelas. Rumus Probabilitas Fitur Terhadap Kelas (*Likelihood*), yaitu

$$P(x_i\ C_k) = \frac{jumlah\ sampel\ dengan\ fitur\ x_i\ dalam\ kelas\ C_k}{jumlah\ sampel\ dalam\ kelas\ C_k} \quad (2)$$

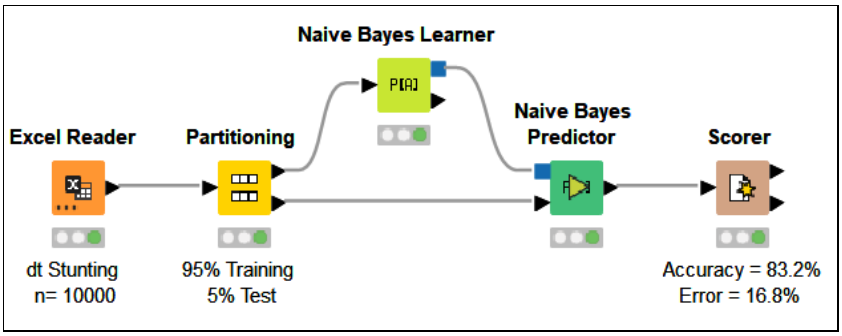
4. Prediksi Kelas Baru: Dengan menggunakan probabilitas yang telah dihitung, untuk setiap data uji, hitung probabilitas kelas untuk setiap kemungkinan kelas berdasarkan fitur-fiturnya. Ini melibatkan perkalian probabilitas fitur dalam kelas tertentu. Kemudian, prediksi kelas dari data baru adalah kelas dengan probabilitas tertinggi. Rumus Probabilitas Posterior (*Predictive*), yaitu

$$P\ (C_k|\ x_1, x_2, \dots, x_n) = P\ (C_k)\ \text{x}\ \prod_{i=1}^{n} P(x_1|C_k) \quad (3)$$

   Dimana $P\ (C_k)$ adalah probabilitas kelas C_k sebelum melihat data, $P(x_1|C_k)$ adalah probabilitas fitur x_i terjadi dalam kelas (C_k), $P\ (C_k|\ x_1, x_2, \dots, x_n)$ adalah probabilitas posterior kelas C_k setelah melihat data dengan fitur $x_1, x_2, \dots, x_n$, n adalah jumlah fitur dalam data.
5. Evaluasi dan Validasi: Evaluasi performa model Naïve Bayes menggunakan data uji atau teknik validasi silang untuk mengevaluasi seberapa baik model tersebut dapat memprediksi kelas-kelas yang benar.

Data dalam penelitian ini diolah menggunakan platform KNIME sebagai alat utama untuk analisis. Data disimpan dalam file Excel yang diakses oleh KNIME melalui *node Excel Reader*. Total dataset yang digunakan adalah n=10,000, seperti yang dijelaskan dalam Tabel 1. Dataset stunting tidak memiliki nilai *null*, sehingga tidak memerlukan penghapusan atau penggantian nilai kosong. Kemudian, dataset stunting dipartisi menggunakan *node partitioning* dengan 95% data training dan 5% data testing. Perbandingan antara jumlah data testing dan training ditentukan melalui serangkaian percobaan untuk memastikan diperolehnya nilai akurasi yang optimal.

Model klasifikasi Naïve Bayes dibangun menggunakan *node Naïve Bayes Learner*. Model yang terbentuk diuji menggunakan *node Naïve Bayes Predictor*. Hasil uji coba ditampilkan melalui *node Scorer*. Workflow KNIME untuk algoritma Naïve Bayes ditampilkan pada Gambar 2.



Gambar 2. Workflow KNIME untuk algoritma Naïve Bayes

3.4.2 *K-Nearest Neighbors* (KNN)

KNN merupakan teknik yang memungkinkan klasifikasi objek dengan menggunakan data latih yang memiliki jarak paling dekat dengan objek yang akan diklasifikasikan. Cara kerja KNN adalah dengan mencari jarak terpendek antara data yang akan diprediksi dengan *k* tetangga terdekatnya dalam kumpulan data pembelajaran.

KNN memiliki keunggulan dalam kemudahan pemahaman dan implementasi, serta fleksibilitasnya dalam menangani data non-linier dan dinamis tanpa memerlukan proses pelatihan yang kompleks. Algoritma ini mampu menyesuaikan diri dengan data baru tanpa perlu pelatihan ulang, cocok untuk data yang tidak memiliki struktur yang jelas. Namun, KNN rentan terhadap perubahan skala fitur yang dapat memengaruhi kinerjanya, rentan terhadap nilai pencilan (*outliers*), dan membutuhkan komputasi yang tinggi karena perlu menghitung jarak dari setiap data baru ke semua data latih saat melakukan prediksi.

Prinsip utama KKN adalah dengan mencari *k* tetangga terdekat dari data yang akan diprediksi di dalam dataset latih, menggunakan pengukuran jarak dengan *Euclidean distance* [24]. KNN tidak melakukan proses pelatihan yang kompleks; ia hanya menyimpan data latih untuk digunakan dalam pengklasifikasian data baru. Dalam proses ini, pemilihan nilai *k* yang optimal sangat penting karena akan memengaruhi akurasi prediksi: nilai *k* yang terlalu kecil bisa sensitif terhadap *noise*, sementara nilai *k* yang terlalu besar bisa memperkenalkan bias yang tidak diinginkan.

Langkah-langkah algoritma *Euclidean distance* adalah:

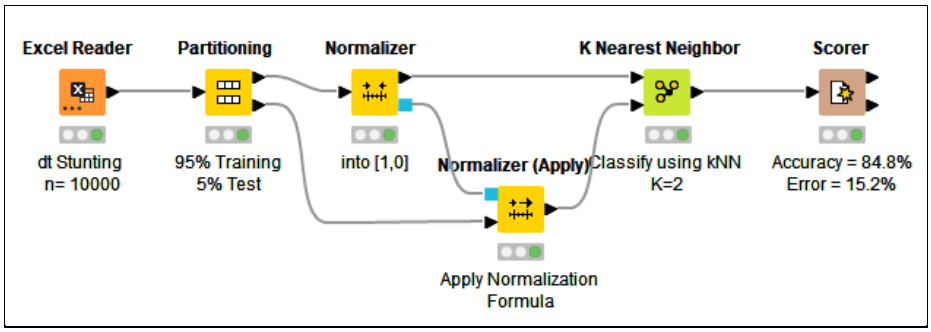
1. Pilih jumlah tetangga (K) yang akan digunakan untuk menentukan kelas,
2. Hitung jarak antara data baru dengan setiap data poin dalam dataset,

3. Pilih K data poin dengan jarak terdekat, lalu identifikasi kelas dari data baru berdasarkan mayoritas kelas dari tetangga-tetangga tersebut. Persamaan *Euclidean distance* adalah sebagai berikut:

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (4)$$

Dimana x_{i1} adalah sampel data, y_i adalah data uji dan $d_{(x,y)}$ adalah jarak antara x_i dan y_i

Pada KNIME, pembentukan model klasifikasi menggunakan algoritma KNN dimulai dengan *node Excel Reader*. Data kemudian dipartisi menjadi 95% untuk data training dan 5% untuk data testing. Proses KNN melibatkan normalisasi data menjadi rentang [1,0]. Model klasifikasi KNN diimplementasikan menggunakan *node K-Nearest Neighbor*, dan hasil pengujian ditampilkan melalui *node Scorer*. Workflow KNIME untuk algoritma *K-Nearest Neighbors* dapat dilihat pada Gambar 3.



Gambar 3. *Workflow* KNIME untuk algoritma *K-Nearest Neighbors*

3.4.3 *Random Forest* (RF)

RF merupakan sebuah algoritma yang termasuk dalam kategori *ensemble learning*. RF memanfaatkan sejumlah besar pohon keputusan (*decision trees*) yang bekerja secara bersamaan untuk melakukan prediksi atau klasifikasi. Konsep RF adalah membuat sejumlah besar pohon keputusan yang masing-masing dihasilkan dari sampel acak dari *dataset* yang sama, namun bervariasi. Setiap pohon dalam RF melakukan prediksi atau klasifikasi secara independen, dan output dari keseluruhan model diambil dari mayoritas hasil prediksi dari setiap pohon individual. Keunggulan utama dari RF adalah

kemampuannya dalam menangani *dataset* yang besar, menangani fitur-fitur yang tidak terlalu penting, serta kemampuannya untuk mengurangi *overfitting* yang umumnya terjadi pada pohon keputusan tunggal. Namun meskipun *Random Forest* sangat efektif dalam mengatasi *overfitting* dan menghasilkan prediksi yang kuat, kekurangannya terletak pada proses pembuatan model yang melibatkan banyak pohon keputusan menjadi komputasi yang memakan waktu, terutama pada *dataset* besar, sementara interpretasi model kompleks dengan banyak pohon bisa sulit. Meskipun algoritma ini secara alami mengurangi *overfitting*, keberadaan *data noise* atau fitur yang tidak relevan tetap bisa menyebabkan *overfitting*, serta pemilihan *hyperparameter* yang tidak tepat dapat memengaruhi kinerja secara keseluruhan, sementara penggunaan memori yang signifikan menjadi pertimbangan terutama pada model dengan jumlah pohon yang besar. Rumus yang digunakan untuk membangun pohon keputusan adalah menggunkan metode C.45

Langkah-langkah algoritma informasi gain dalam metode C.45, adalah

1. Hitung Entropi Data Awal menggunakan: Hitung nilai entropi dari kelas target sebelum membagi data berdasarkan atribut apa pun. Entropi mengukur tingkat ketidakpastian atau keacakan dalam data. Rumusnya adalah:

$$Entropy\ (s) - \sum_{i=1}^{c} p_i\, log_2\ (pi) \quad (5)$$

Di mana p_i adalah proporsi data yang termasuk dalam kelas tertentu dari total kelas (c).

2. Hitung Entropi Setiap Atribut: Hitung entropi dari setiap atribut dengan menggunakan nilai atribut tersebut sebagai kriteria pemisahan. Entropi dihitung untuk setiap nilai unik pada atribut tersebut.
3. Hitung Informasi Gain: Setelah mendapatkan entropi dari setiap atribut, hitung informasi gain untuk setiap atribut. Informasi gain mengukur seberapa baik atribut tersebut dalam memisahkan data. Rumusnya adalah:

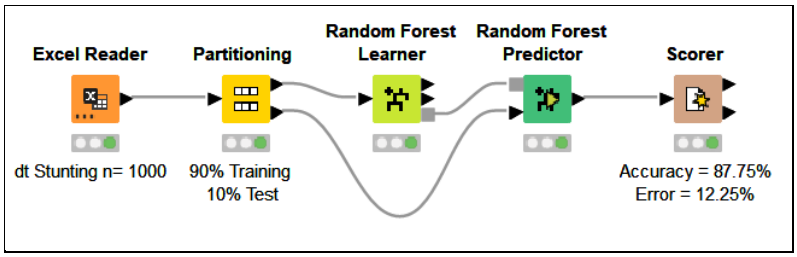
$$Gain\ (S, A) = |S_v|\ \ |S|Entropy\ (S) - \sum_{v \epsilon Values(A)} \left|\frac{S_v}{S}\right| * Entropy\ (S_v) \quad (6)$$

Dimana S adalah dataset awal, A adalah atribut yang sedang dievaluasi, $Values(A)$ adalah nilai-nilai unik yang dimiliki oleh atribut A, $|Sv|$ adalah jumlah data pada subset S dengan nilai $A = v$, dan $|S|$ adalah total jumlah data pada S.

4. Pilih Atribut dengan Informasi Gain Tertinggi: Pilih atribut dengan nilai informasi gain tertinggi sebagai kandidat pemisahan node pada pohon keputusan.

Langkah-langkah ini membantu C4.5 dalam memilih atribut yang paling informatif untuk membuat keputusan pemisahan yang optimal dalam pembentukan pohon keputusan.

Pembentukan model klasifikasi menggunakan *algoritma* RF pada KNIME dimulai dengan *node Excel Reader*. Data kemudian dipartisi menjadi 90% untuk data training dan 10% untuk data testing. Model klasifikasi *Random Forest* diimplementasikan menggunakan *node Random Forest learner*, Model yang terbentuk kemudian diuji menggunakan *node Random Forest learner Predictor*. Hasil uji coba ditampilkan melalui *node Scorer*. Workflow KNIME untuk algoritma *Random Forest* ditampilkan pada Gambar 4.



Gambar 4. Workflow KNIME untuk Algoritma *Random Forest*

3.4.4 Evaluasi Kinerja Model

Hasil evaluasi model klasifikasi dari ketiga *algoritma machine learning* yakini *Naive Bayes*, *K-Nearest Neighbors* dan *Random Forest* dapat dilihat pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Klasifikasi Stunting

Nilai Akurasi	*Naive Bayes*	*K-Nearest Neighbors*	*Random Forest*
Accuracy	83.2	84.8	87.75
Recall	0.897	0.967	0.956
Precision	0.892	0.869	0.891
F1-Score	0.895	0.916	0.922

Dari hasil evaluasi model klasifikasi stunting menggunakan tiga algoritma yang berbeda tersebut, maka dapat diambil kesimpulan, yaitu:

1. Performa Algoritma
   Dalam hal akurasi, *Random Forest* menunjukkan hasil tertinggi dengan 87.75%, diikuti oleh *K-Nearest Neighbors* dengan 84.8%, dan *Naive Bayes* dengan 83.2%. Namun, ketika melihat *Recall*, yang mengukur kemampuan model dalam menemukan seluruh kasus *stunted* yang sebenarnya, *K-Nearest Neighbors* memiliki nilai tertinggi (0.967), diikuti oleh *Random Forest* (0.956) dan *Naive Bayes* (0.897). *Recall* yang tinggi menunjukkan bahwa K-Nearest Neighbors cenderung lebih baik dalam menemukan kasus stunting dengan lebih baik daripada algoritma lainnya.

2. Presisi dan *F1-Score*
   Meskipun *Recall K-Nearest Neighbors* yang tinggi, perlu diperhatikan bahwa *presisi Naive Bayes* (0.892) dan *Random Forest* (0.891) hampir setara dengan nilai *Recall*-nya. *F1-Score* (rata-rata harmonis dari presisi dan recall) *Random Forest* (0.922) memiliki nilai tertinggi, menunjukkan keseimbangan yang baik antara presisi dan recall dalam mengidentifikasi kasus stunting.

3. Pertimbangan Holistik
   Dalam pemilihan model untuk mendiagnosis stunting, sementara *Random Forest* memiliki akurasi yang baik, *K-Nearest Neighbors* menonjol dalam kemampuan menemukan sebagian besar kasus stunting yang sebenarnya. Namun, *Random Forest* memiliki keseimbangan yang baik antara presisi dan *recall* (*F1-score* tinggi), menandakan bahwa model ini dapat menjadi pilihan yang baik untuk mendiagnosis stunting, menggabungkan akurasi yang baik dengan kemampuan menemukan kasus stunting yang lebih baik dari model lainnya dalam evaluasi ini.

4. KESIMPULAN

Hasil evaluasi menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi dengan 87.75%, diikuti oleh *K-Nearest Neighbors* dengan 84.8%, dan *Naive Bayes* dengan 83.2%. Meskipun *Random Forest* memiliki akurasi yang baik, *K-Nearest Neighbors* menonjol dalam kemampuan menemukan sebagian besar kasus stunting yang sebenarnya (nilai *Recall* tertinggi). Namun, *Random Forest* memiliki keseimbangan yang baik antara *precision* dan *recall* (*F1-score* tinggi), menandakan bahwa model ini dapat menjadi pilihan yang baik untuk mendiagnosis stunting, menggabungkan akurasi yang baik dengan kemampuan menemukan kasus stunting yang lebih baik dari model lainnya dalam evaluasi ini.

Penelitian ini menyoroti pentingnya penerapan ML dalam mendiagnosis stunting pada anak-anak dan menyediakan pemahaman lebih lanjut tentang performa berbagai algoritma dalam hal prediksi stunting. Diharapkan bahwa model hasil penelitian ini dapat menjadi landasan yang baik untuk deteksi dini kasus stunting, memungkinkan intervensi yang lebih tepat dan cepat bagi anak-anak yang membutuhkan perhatian kesehatan yang lebih intensif. Dengan perbaikan dalam metode evaluasi, diharapkan dapat mengurangi dampak negatif jangka panjang dari kondisi stunting terhadap pertumbuhan dan perkembangan anak-anak.

REFERENSI

[1] Eko, "149 Juta Anak di Dunia Alami Stunting Sebanyak 6,3 Juta di Indonesia, Wapres Minta Keluarga Prioritaskan Kebutuhan Gizi," *Direktorat Pendidikan Anak Usia Dini*, 2023. https://paudpedia.kemdikbud.go.id/berita/149-juta-anak-di-dunia-alami-stunting-sebanyak-63-juta-di-indonesia-wapres-minta-keluarga-prioritaskan-kebutuhan-gizi?do=MTY2NC01YjRhOGZkNA==&ix=MTEtYmJkNjQ3YzA=

[2] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A review of child stunting determinants in Indonesia," *Matern. Child Nutr.*, vol. 14, no. 4, pp. 1–10, 2018, doi: 10.1111/mcn.12617.

[3] M. Wahid and Mujib Rahman, "Rakornas 2023: Pastikan Prevalensi Stunting Turun Menjadi 14% Pada Tahun 2024," *Kementerian Sekretariat Negara RI*, 2023. https://stunting.go.id/rakornas-2023-pastikan-prevalensi-stunting-turun-menjadi-14-pada-tahun-2024/

[4] M. Rosyidah, Y. L. R. Dewi, and I. Qadrijati, "Effects of Stunting on Child Development: A Meta-Analysis," *J. Matern. Child Heal.*, vol. 6, no. 1, pp. 25–34, 2021, doi: 10.26911/thejmch.2021.06.01.03.

[5] A. Heryati, Erduandi, and Terttiaavini, "Penerapan Jaringan Saraf Tiruan Untuk Memprediksi Pencapaian Prestasi Mahasiswa," in *Konferensi Nasional Sistem Informasi 2018 STMIK Atma Luhur Pangkalpinang, 8 – 9 Maret 2018*, 2018, pp. 8–9.

[6] D. A. Safitri, D. Fitriani, L. Hertati, T. Terttiavini, A. Heryati, and Asmawati, "PKM Mahasiswa Indo Global Mandiri Pada E-Commerce Marketplace Era Pandemi Covid Meningkat Tajam," *J. Sustain. Community Serv.*, vol. 1, no. 4, pp. 192–208, 2021.

[7] Hartatik *et al.*, *Data Science - Data Science*, no. September 2016. 2023. [Online]. Available: https://www.data-science.ruhr/about_us/

[8] D. Marcelina, A. Kurnia, and T. Terttiaavini, "Analisis Klaster Kinerja Usaha Kecil dan Menengah Menggunakan Algoritma K-Means Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. October, pp. 293–301, 2023.

[9] M. Tahangnacca, R. Amiruddin, Ansariadi, and A. Syam, "Model of stunting determinants: A systematic review," *Enferm. Clin.*, vol. 30, pp. 241–245, 2020, doi: 10.1016/j.enfcli.2019.10.076.

[10] M. S. Islam, A. N. Zafar Ullah, S. Mainali, M. A. Imam, and M. I. Hasan, "Determinants of stunting during the first 1,000 days of life in Bangladesh: A review," *Food Sci. Nutr.*, vol. 8, no. 9, pp. 4685–4695, 2020, doi: 10.1002/fsn3.1795.

[11] T. R. Chowdhury, S. Chakrabarty, M. Rakib, S. Afrin, S. Saltmarsh, and S. Winn, "Factors associated with stunting and wasting in children under 2 years in Bangladesh," *Heliyon*, vol. 6, no. 9, 2020, doi: 10.1016/j.heliyon.2020.e04849.

[12] C. Scheffler and M. Hermanussen, "Stunting is the natural condition of human height," *Am. J. Hum. Biol.*, vol. 34, no. 5, pp. 1–13, 2022, doi: 10.1002/ajhb.23693.

[13] T. Mulyaningsih, I. Mohanty, V. Widyaningsih, T. A. Gebremedhin, R. Miranti, and V. H. Wiyono, "Beyond personal factors: Multilevel determinants of childhood stunting in Indonesia," *PLoS One*, vol. 16, no. 11 November, pp. 1–19, 2021, doi: 10.1371/journal.pone.0260265.

[14] T. Huriah and N. Nurjannah, "Risk factors of stunting in developing countries: A scoping review," *Open Access Maced. J. Med. Sci.*, vol. 8, no. F, pp. 155–160, 2020, doi: 10.3889/oamjms.2020.4466.

[15] M. S. Haris, M. Anshori, and A. N. Khudori, "Prediction of Stunting Prevalence in East Java Province With Random Forest Algorithm," *J. Tek. Inform.*, vol. 4, no. 1, pp. 11–13, 2023, doi: 10.52436/1.jutif.2023.4.1.614.

[16] Harliana and D. Anggraini, "Penerapan Algoritma Naïve Bayes Pada Klasifikasi Status Gizi Balita di Posyandu Desa Kalitengah (Harliana, Dewi Anggraini)," *FAHMA - J. Inform. Komputer, Bisnis dan Manaj.*, vol. 21, no. 2, pp. 38–45, 2023.

[17] S. Lonang and D. Normawati, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 49, 2022, doi: 10.30865/mib.v6i1.3312.

[18] T. Terttiaavini, Y. Hartono, E. Ermatita, and D. P. Rini, "Comparison of Simple Additive Weighting Method and Weighted Performance Indicator Method for Lecturer Performance Assessment," *Mod. Educ. Comput. Sci.*, vol. 15, no. 2, pp. 1–11, 2023, doi: 10.5815/ijmecs.2023.02.01.

[19] T. Terttiaavini, S. amariena Hamim, and S. Agustri, "Aplikasi sistem pakar penentu bidang studi ditingkat perguruan tinggi berbasis web," *J. Ilm. Inform. ...*, vol. 7, no. 1, pp. 67–72, 2016, [Online]. Available: http://ejournal.uigm.ac.id/index.php/IG/article/view/188

[20] S. Lonang, A. Yudhana, and M. Kunta Biddinika, "Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting," *J. Media Inform. Budidarma*, vol. 7, pp. 2109–2117, 2023, doi: 10.30865/mib.v7i4.6553.

[21] R. Kusumaningrum, T. A. Indihatmoko, S. R. Juwita, A. F. Hanifah, K. Khadijah, and B. Surarso, "Benchmarking of multi-class algorithms for classifying documents related to stunting," *Appl. Sci.*, vol. 10, no. 23, pp. 1–13, 2020, doi: 10.3390/app10238621.

[22] A. Naik and L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 662–668, 2016, doi: 10.1016/j.procs.2016.05.251.

[23] KNIME Official, "KNIME Analytics Platform," 2023. https://www.knime.com/knime-analytics-platform

[24] N. Arminarahmah, A. D. GS, G. W. Bhawika, M. P. Dewi, and A. Wanto, "Mapping the Spread of Covid-19 in Asia Using Data Mining X-Means Algorithms," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1071, no. 1, p. 012018, 2021, doi: 10.1088/1757-899x/1071/1/012018.