Volume 18 No. 2 (April 2026), pp. 85-98 | ISSN 2541-1760 (online) | ISSN 1907-3984 (print)
Jurnal Akademika
http://ojs.unh.ac.id/index.php/ | DOI: https://doi.org/10.53564/carvay33

Evaluasi SMOTE dan SMOTE+Tomek pada KNN dan Random Forest untuk Klasifikasi Status Stunting Balita

Marrylinteri Istoningtyas [1], M. Irwan Bustami [2], Irawan [3], Maria Rosario B [4], Sansan Rosita [5]
[1,3,5]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia
[2]Program Studi Sistem Komputer, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia
[4]Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia
Email: [1]marrylinteri@unama.ac.id , [2]irwan.sk05@gmail.com, [3]irawanirend@stikom-db.ac.id , [4]diamar_ros@yahoo.com , [5]snsnrst@gmail.com

Article Information

Article history

Received 06 Februari 2026
Revised 30 Maret 2026
Accepted 20 April 2026
Available 30 April 2026

Keywords

Stunting
Class Imbalance
K-Nearest Neighbours
Random Forest
SMOTE

Corresponding Author:

Marrylinteri Istoningtyas,
Universitas Dinamika Bangsa,
Jambi,
Indonesia,
Email: marrylinteri@unama.ac.id

Abstract

Class imbalance is a recurring obstacle in machine learning based screening of child nutritional status. This study evaluates and compares the effect of SMOTE and SMOTE+Tomek Links on the classification of toddler nutritional status using K-Nearest Neighbours (KNN) and Random Forest (RF). The data consist of 9,426 anthropometric records with four predictors, labelled into three classes: normal (7,212; 76.51%), moderate stunting (1,612; 17.10%) and severe stunting (602; 6.39%), a majority to minority ratio of about 12:1. Min-Max scaling and resampling were fitted on training data only, inside a pipeline, and the models were assessed on an independent 20% test set (n = 1,886) with stratified 5-fold cross validation. At baseline, RF reached 0.975 accuracy and 0.932 macro-F1, while KNN displayed an accuracy paradox: 0.913 accuracy but only 0.392 recall on severe stunting (47 of 120 cases). SMOTE raised KNN recall to 0.733 and macro-F1 from 0.753 to 0.816. For RF, SMOTE improved both criteria at once: recall rose from 0.825 to 0.908 (99 to 109 cases) and macro-F1 from 0.932 to 0.949. SMOTE+Tomek performed almost identically, removing only 30 of 17,307 training samples. RF with SMOTE is therefore recommended.

Keywords: *stunting, class imbalance, K-Nearest Neighbours, Random Forest, SMOTE*

Abstrak

Ketidakseimbangan kelas merupakan hambatan yang berulang pada penapisan status gizi anak berbasis pembelajaran mesin. Penelitian ini mengevaluasi dan membandingkan pengaruh SMOTE dan SMOTE+Tomek Links terhadap klasifikasi status gizi balita menggunakan *K-Nearest Neighbours* (KNN) dan *Random Forest* (RF). Data terdiri atas 9.426 rekaman antropometri dengan empat variabel prediktor dan label tiga kelas: normal (7.212; 76,51%), stunting sedang (1.612; 17,10%), dan stunting berat (602; 6,39%), dengan rasio kelas mayoritas terhadap minoritas sekitar 12:1. Penskalaan Min-Max dan penyeimbangan kelas hanya di-*fit* pada data latih di dalam *pipeline*, lalu model diuji pada data uji independen 20% (n = 1.886) dengan *stratified 5-fold cross validation*. Pada kondisi *baseline*, RF mencapai akurasi 0,975 dan macro-F1 0,932, sedangkan KNN menunjukkan *accuracy paradox*: akurasi 0,913 tetapi *recall* kelas stunting berat hanya 0,392 atau 47 dari 120 kasus. SMOTE menaikkan *recall* KNN menjadi 0,733 dan macro-F1 dari 0,753 menjadi 0,816. Pada RF, SMOTE memperbaiki kedua kriteria sekaligus: *recall* naik dari 0,825 menjadi 0,908 (99 menjadi 109 kasus) dan macro-F1 dari 0,932 menjadi 0,949. SMOTE+Tomek menghasilkan kinerja yang praktis identik karena hanya menghapus 30 dari 17.307 sampel data latih. RF dengan SMOTE karena itu direkomendasikan.

Kata Kunci: *stunting, ketidakseimbangan kelas, K-Nearest Neighbours, Random Forest, SMOTE*

1. Pendahuluan

Stunting adalah kondisi gangguan pertumbuhan tinggi badan pada balita akibat kekurangan gizi kronis yang menghambat potensi fisik dan kognitif anak [1]. Stunting merupakan isu global yang berdampak luas terhadap perkembangan anak serta berpotensi membatasi potensi ekonomi dan sosial di masa depan. Selain di Indonesia[2],[3],[4], kondisi stunting juga masih banyak ditemui di negara berkembang seperti Rwanda[5], Ghana[6], dan Bangladesh[7].

Berbagai upaya penurunan stunting telah dijalankan di Indonesia melalui perluasan akses layanan kesehatan dan edukasi gizi di posyandu. Prevalensi stunting nasional menurun dari 27,7% pada 2019 menjadi 24,4% pada 2021 dan 21,6% pada 2022 [8]. Penurunan agregat tersebut menyisakan konsekuensi metodologis yang penting bagi penapisan pada tingkat individu: semakin rendah prevalensi, semakin kecil pula proporsi kasus stunting—terutama stunting berat—di dalam data yang terkumpul. Fokus penelitian ini karena itu bukan pada estimasi prevalensi wilayah, melainkan pada klasifikasi status gizi setiap balita, yaitu tugas yang justru bertambah sulit ketika kelas sasaran semakin langka.

Dataset antropometri yang digunakan pada penelitian ini memperlihatkan ketidakseimbangan kelas yang nyata: 7.212 sampel (76,51%) berstatus normal, 1.612 sampel (17,10%) stunting sedang, dan hanya 602 sampel (6,39%) stunting berat, sehingga rasio kelas mayoritas terhadap kelas minoritas terkecil mencapai sekitar 12:1. Pada distribusi seperti ini, pengklasifikasi standar cenderung mengoptimalkan kelas mayoritas demi memaksimalkan akurasi. Akibatnya muncul *accuracy paradox*, yaitu akurasi keseluruhan yang tampak tinggi meskipun kemampuan mendeteksi kasus stunting rendah[9], [10]. Dalam konteks kesehatan masyarakat, kesalahan berupa *false negative* pada anak stunting berakibat tertundanya intervensi gizi sehingga lebih merugikan daripada *false positive*.

Pembelajaran mesin (*machine learning*) menawarkan pendekatan analitik yang adaptif untuk mempelajari pola pada data antropometri. Namun penelitian terdahulu pada domain stunting umumnya melaporkan kinerja dalam bentuk akurasi atau ukuran kecocokan model secara agregat. Pendekatan *Learning Vector Quantization* [11], *Support Vector Machine* (SVM) [12], *Geographically Weighted Regression-Kriging* (GWRK))[13], *Multi Ordinal Logistic Regression* [14], maupun *Decision Tree*[15]— belum menempatkan strategi penyeimbangan kelas sebagai bagian dari rancangan eksperimen, sehingga pengaruh dominasi kelas normal terhadap sensitivitas model tidak dapat ditelusuri dari hasil yang dilaporkan. Perbandingan langsung antara model berbasis jarak seperti KNN dan model *ensemble* seperti Random Forest dalam merespons teknik penyeimbangan data pada kasus stunting juga masih terbatas.

Penelitian ini bertujuan mengevaluasi dan membandingkan pengaruh SMOTE dan SMOTE+Tomek Links terhadap klasifikasi status gizi balita pada algoritma KNN dan Random Forest. Evaluasi difokuskan pada metrik yang tidak terdominasi kelas mayoritas, yaitu *precision*, *recall*, dan *F1-score* per kelas serta macro-F1, dan dilengkapi *Stratified 5-Fold Cross Validation*. Dengan demikian penilaian tidak bertumpu pada akurasi keseluruhan, melainkan pada kemampuan model mengenali kelas minoritas—khususnya stunting berat—beserta konsekuensi *trade-off precision–recall* yang menyertainya.

2. Kajian Terdahulu

2.1. Penelitian Terdahulu

Penerapan pembelajaran mesin (*machine learning*) dalam bidang kesehatan masyarakat, khususnya untuk pemodelan status gizi dan deteksi dini stunting pada anak, telah berkembang pesat dalam beberapa tahun terakhir. Berbagai pendekatan algoritma telah dieksplorasi untuk mengklasifikasikan risiko gizi buruk berdasarkan data antropometri maupun variabel sosial ekonomi. Tabel 1 merangkum beberapa penelitian relevan terkait pemodelan stunting beserta metode dan fokus kajian yang digunakan.

Tabel 1. Penelitian Terdahulu

Penelitian	Metode yang Digunakan	Fokus/Objek Kajian	Batasan/Catatan
Aziz dkk. [11]	*Learning Vector Quantization* (LVQ)	Klasifikasi keluarga berisiko *stunting*	Ketidakseimbangan kelas tidak diperlakukan sebagai variabel eksperimen; kinerja per kelas minoritas tidak dilaporkan.
Andriyani dkk. [12]	*Support Vector Machine* (SVM)	Klasifikasi data *stunting* balita	Fokus pada optimasi parameter SVM; tanpa penyeimbangan data, sensitivitas pada kelas minoritas tidak dapat dinilai.
Iriany dkk. [13]	*Geographically Weighted Regression-Kriging* (GWRK)	Klasifikasi prevalensi *stunting* spasial	Unit analisis berupa wilayah, bukan individu, sehingga tidak langsung dapat dipakai untuk penapisan per balita.
Asgedom dkk. [14]	*Multi Ordinal Logistic Regression*	Tingkat keparahan *stunting* di Ethiopia	Pendekatan statistik inferensial untuk faktor risiko, bukan membangun pengklasifikasi yang diuji pada data terpisah.
Wajgi & Wajgi [15]	*Decision Tree*	Deteksi malnutrisi pada bayi	Pohon keputusan tunggal rentan condong ke kelas mayoritas; strategi penyeimbangan tidak disertakan.

Meskipun penelitian terdahulu telah menunjukkan keandalan *machine learning* dalam memprediksi stunting, sebagian besar studi tersebut belum menempatkan ketidakseimbangan kelas sebagai variabel perlakuan yang dievaluasi. Padahal pada data antropometri lapangan proporsi balita bergizi normal jauh melampaui balita stunting sedang maupun berat. Ketika distribusi tersebut diabaikan, model dapat menampilkan *accuracy paradox*, yakni akurasi keseluruhan yang tinggi disertai kegagalan mengenali kasus minoritas [9], [10].

Untuk menangani isu *class imbalance*, beberapa penelitian di domain medis lainnya telah memanfaatkan teknik *oversampling* dan *hybrid sampling* [16] ,[17], [18],[19]. T*eknik Synthetic Minority Over-Sampling Technique* (SMOTE) terbukti efektif meningkatkan sensitivitas model pada klasifikasi penyakit stroke [16], deteksi spam [17], serta evaluasi kinerja [20]. Sementara itu, strategi hybrid seperti SMOTE+Tomek Links digunakan untuk membersihkan *overlap* antar-kelas sekaligus menambah data sintetis, sebagaimana diterapkan pada diagnosis tuberkulosis[18] dan prediksi stroke otak [19].

Meskipun teknik penyeimbangan data seperti SMOTE dan SMOTE+Tomek telah banyak diterapkan pada berbagai domain [16], [17], [20], [21], [18], [19]; evaluasi komparatif secara mendalam mengenai efektivitas kedua teknik ini pada kasus stunting balita—khususnya dengan membandingkan model berbasis jarak (*instance-based*) seperti K-Nearest Neighbours (KNN) dan model berbasis *ensemble learning* seperti Random Forest (RF)—masih sangat terbatas.

Kebaruan penelitian ini terletak pada evaluasi komparatif dua strategi penyeimbangan kelas—SMOTE dan SMOTE+Tomek—yang diuji secara berpasangan pada dua paradigma pembelajaran yang berbeda, yaitu *instance-based* (KNN) dan *ensemble* (Random Forest), dengan pelaporan kinerja pada tingkat kelas. Rancangan tersebut memungkinkan pemisahan dua

sumber perbaikan yang biasanya bercampur, yakni kontribusi *oversampling* dan kontribusi pembersihan batas kelas, sekaligus menampilkan *trade-off precision–recall* pada kelas stunting berat yang menjadi sasaran deteksi dini.

2.2. Landasan Teori

2.1.1 Ketidakseimbangan Kelas

Ketidakseimbangan kelas terjadi ketika distribusi jumlah sampel antarkelas pada dataset timpang, yaitu satu kelas memiliki jumlah sampel jauh lebih besar daripada kelas lain[9], [10]. Dalam pemodelan stunting, kelas berstatus gizi normal jauh lebih dominan dibanding kelas stunting sedang maupun berat. Kondisi ini menyebabkan algoritma standar cenderung mengoptimalkan prediksi pada kelas mayoritas demi mengejar nilai akurasi, sehingga menurunkan sensitivitas terhadap kelas minoritas yang justru krusial untuk deteksi dini kesehatan [16] ,[17].

2.1.2 K-Nearest Neighbours (KNN)

kKNN merupakan algoritma non-parametrik berbasis *instance-based learning* yang mengklasifikasikan data baru berdasarkan mayoritas kelas dari k tetangga terdekatnya menurut jarak Euclidean. Kinerjanya sangat bergantung pada kepadatan lokal data, sehingga ketika sampel kelas minoritas sedikit, batas keputusan cenderung dimenangkan kelas mayoritas.

2.1.3 Random Forest (RF)

Random Forest merupakan algoritma *ensemble learning* yang membangun banyak pohon keputusan melalui *bagging* dan pemilihan fitur acak, lalu menggabungkan hasilnya melalui pemungutan suara mayoritas. Agregasi tersebut menurunkan varians dan membuat model lebih tahan terhadap dominasi kelas mayoritas dibandingkan pengklasifikasi tunggal.

2.1.4 Synthetic Minority Over-Sampling Technique (SMOTE)

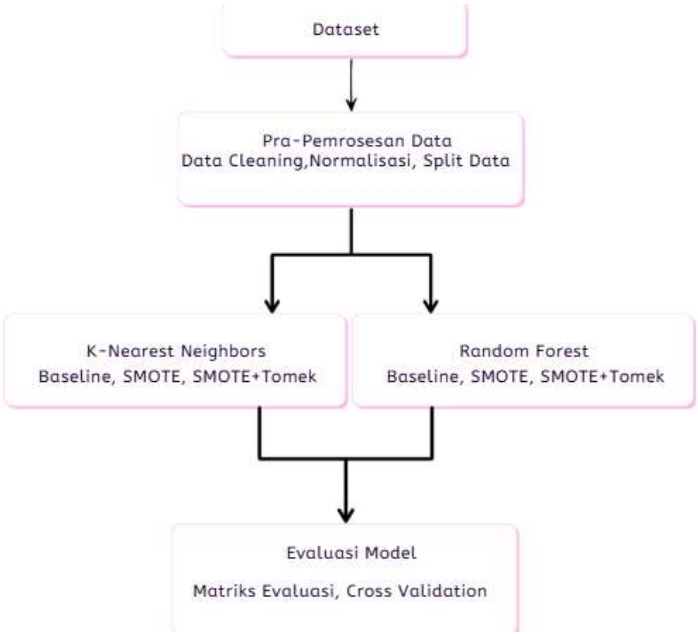
SMOTE merupakan teknik *oversampling* yang bekerja dengan cara membangkitkan data sintetis baru pada kelas minoritas melalui proses interpolasi, bukan sekadar menduplikasi sampel yang ada [20], [21]. Sampel sintetis baru x_{baru} dibuat di sepanjang segmen garis antara sampel minoritas x_i dan salah satu tetangga terdekatnya x_{zi} sesuai Persamaan (2).

2.1.5 SMOTE + Tomek Links

SMOTE+Tomek merupakan pendekatan *hybrid* yang mengombinasikan *oversampling* (SMOTE) dan *undersampling* (Tomek Links) [18],[19].Setelah SMOTE membangkitkan data sintetis, teknik Tomek Links mengidentifikasi pasangan sampel terdekat dari dua kelas yang berbeda (kelas mayoritas dan minoritas). Pasangan (x_i, y_j) dikatakan sebagai Tomek Link jika tidak ada titik x_k yang memiliki jarak lebih dekat ke x_i atau x_j dibandingkan jarak $d(x_i, y_j)$. Sampel kelas mayoritas dari pasangan Tomek Link kemudian dihapus untuk membersihkan area *overlapping* sehingga decision boundary antar-kelas menjadi lebih jelas [18],[19].

3. Metodologi Penelitian

Penelitian ini merupakan studi eksperimental kuantitatif yang membandingkan kinerja KNN dan Random Forest pada tiga strategi data, yaitu tanpa penyeimbangan, dengan SMOTE, dan dengan SMOTE+Tomek Links. Tahapan penelitian disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1 Sumber dan Deskripsi Data

Penelitian ini menggunakan data sekunder publik berupa data antropometri balita sebanyak 9.999 sampel awal dengan empat variabel prediktor, yaitu jenis kelamin, umur (bulan), berat badan (kg), dan tinggi badan (cm), pada rentang umur 0 sampai 24 bulan. Variabel status gizi pada data asal memuat empat kategori, yaitu normal, stunting sedang, stunting berat, dan tinggi (*tall*), yang diturunkan dari ambang *height-for-age z-score* (HAZ) menurut WHO Child Growth Standards: kurang dari −3 SD untuk stunting berat, −3 SD hingga kurang dari −2 SD untuk stunting sedang, −2 SD hingga +3 SD untuk normal, dan di atas +3 SD untuk kategori tinggi.

Setelah pembersihan data (*data cleaning*), tersisa 9.426 sampel yang siap diproses. Distribusi kelas pada data akhir memperlihatkan ketidakseimbangan yang nyata sebagaimana dirangkum pada Tabel 2. Rasio kelas normal terhadap kelas stunting berat mencapai sekitar 12:1, sedangkan terhadap kelas stunting sedang sekitar 4,5:1.

Tabel 2. Sebaran Sampel Data Stunting

Kelas Target	Kategori Status Gizi	Jumlah Sampel (∑)	Persentase (%)
0	Normal	7.212	76,51
1	Stunting Sedang	1.612	17,10
2	Stunting Berat	602	6,36
Total		**9.426**	**100**

3.2 Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk meningkatkan kualitas data mentah dan meminimalkan bias pada pemodelan [22], [23]. Tahapan pra-pemrosesan yang dilakukan pada penelitan ini adalah sebagai berikut:

a. Pembersihan Data (*Data Cleaning*)

Dua langkah pembersihan dilakukan pada tahap ini. Pertama, atribut *wasting* dihapus untuk mencegah kebocoran informasi (*data leakage*) karena atribut tersebut

merupakan turunan penilaian status gizi dan berkorelasi tinggi dengan label target. Kedua, seluruh sampel berkategori tinggi (*tall*) dikeluarkan dari analisis, yaitu 573 sampel atau 5,73% dari 9.999 sampel awal. Pembatasan ini dilakukan karena fokus penelitian adalah membedakan derajat keparahan stunting, sedangkan kategori tinggi berada pada arah sebaran yang berlawanan dan tidak relevan bagi penapisan kekurangan gizi kronis. Tugas klasifikasi dengan demikian disederhanakan menjadi tiga kelas berjenjang, yaitu normal, stunting sedang, dan stunting berat. Tidak ada penyaringan nilai ekstrem lain yang diterapkan karena seluruh rekaman berada dalam rentang yang wajar secara biologis menurut *fixed exclusion range* WHO untuk indikator tinggi badan menurut umur, yaitu HAZ di dalam rentang −6 SD sampai +6 SD[24].

b. Normalisasi Data

Penskalaan fitur numerik diperlukan agar algoritma berbasis jarak seperti KNN tidak terbias oleh fitur dengan rentang nilai besar [25], [26], [27], [28], [29]. Untuk mencegah kebocoran informasi, penskalaan tidak dilakukan pada keseluruhan data sebelum pembagian. Objek *scaler* di-*fit* hanya pada data latih dan, pada validasi silang, hanya pada data latih setiap *fold*; parameter hasil *fit* tersebut kemudian dipakai untuk mentransformasi data validasi dan data uji melalui mekanisme *Pipeline*.

c. Pembagian Data (*Data Split*)

Data dibagi menggunakan *Stratified Split* dengan rasio 80% data latih (7.540 sampel) dan 20% data uji (1.886 sampel) [30], [31]. Stratifikasi menjaga proporsi ketiga kelas tetap sama pada kedua subset, sehingga data uji memuat 1.443 sampel kelas normal, 323 sampel stunting sedang, dan 120 sampel stunting berat. Data uji tersebut disisihkan sejak awal dan tidak pernah digunakan pada tahap pelatihan maupun validasi silang.

3.3 Strategi Penyeimbang Kelas

Dua teknik penyeimbangan diterapkan pada data latih, yaitu SMOTE dan SMOTE+Tomek Links, yang mekanismenya telah diuraikan pada Bab 2.

Kedua teknik hanya diterapkan pada data latih. Pada validasi silang, penyeimbangan ditempatkan sebagai tahap di dalam *Pipeline* sehingga *resampling* dijalankan ulang pada data latih setiap *fold*, sedangkan data validasi hanya ditransformasi. Penempatan ini penting karena menerapkan SMOTE pada seluruh data latih sebelum validasi silang akan menyebarkan sampel sintetis beserta tetangga asalnya ke *fold* berbeda sehingga skor validasi menjadi optimistis. Penskalaan dilakukan sebelum penyeimbangan agar pembangkitan sampel sintetis, yang bertumpu pada jarak Euclidean, tidak didominasi fitur berskala besar. Pada data latih (7.540 sampel), SMOTE menyetarakan ketiga kelas menjadi 17.307 sampel, lalu Tomek Links menghapus 30 sampel (0,17%).

3.4 Konfigurasi Parameter Model

Seluruh eksperimen dijalankan dengan Python menggunakan pustaka scikit-learn dan imbalanced-learn. Konfigurasi parameter setiap komponen dirangkum pada Tabel 3. Nilai *random_state* disamakan pada seluruh komponen agar hasil dapat direproduksi.

Tabel 3. Konfigurasi Parameter Model dan Prosedur Validasi

Komponen	Konfigurasi
Penskalaan fitur	Normalisasi Min-Max ke rentang [0, 1]; di-fit hanya pada data latih di dalam Pipeline; dilakukan sebelum penyeimbangan kelas
K-Nearest Neighbours	n_neighbors = 5; weights = uniform; metric = minkowski, p = 2 (Euclidean)
Random Forest	n_estimators = 100; criterion = gini; max_depth = None; min_samples_split / min_samples_leaf = 2 / 1; class_weight = None
SMOTE	sampling_strategy = auto (kelas minoritas disamakan dengan kelas mayoritas); k_neighbors = 5
SMOTE+Tomek	parameter smote sama dengan baris SMOTE; tomek sampling_strategy = all
Validasi silang	StratifiedKFold; n_splits = 5; shuffle = True
Umum	random_state = 42 (pembagian data, SMOTE, Random Forest, dan validasi silang)

3.5 Arsitektur Pemodelan dan Prosedur Algoritma

Kedua algoritma dipasang sebagai tahap akhir pada *Pipeline* yang sama sehingga urutan penskalaan, penyeimbangan, dan pelatihan identik untuk seluruh kombinasi perlakuan. Prosedur eksperimen dirangkum pada Pseudocode 1, yang sekaligus menegaskan letak pencegahan kebocoran informasi.

```
============================================================
Pseudocode 1. Prosedur eksperimen dan evaluasi model
============================================================
Input  : D = {X, y}, N = 9.426; S = {Baseline, SMOTE, SMOTE+Tomek};
         A = {KNN, RF}
Output : P/R/F1 per kelas, macro-F1, confusion matrix, skor CV
1.  D_clean <- DataCleaning(D)   // buang wasting & kelas tinggi
2.  X_tr, X_te, y_tr, y_te <- StratifiedSplit(X, y, test_size=0.2,
        stratify=y, random_state=42)   // belum ada penskalaan
3.  FOR EACH s IN S, a IN A DO
4.    res  <- (s = Baseline) ? passthrough : Resampler(s)
5.    pipe <- Pipeline([MinMaxScaler(), res, Classifier(a)])
      // scaler & res di-fit ULANG pada data latih tiap fold;
      // data validasi tiap fold tidak pernah di-resample
6.    cv_scores <- CrossValidate(pipe, X_tr, y_tr,
        cv=StratifiedKFold(5, shuffle=True, random_state=42))
7.    pipe.fit(X_tr, y_tr);  y_hat <- pipe.predict(X_te)
8.    Save(a, s, Evaluate(y_te, y_hat), cv_scores)
9.  END FOR
============================================================
```

3.6 Metrik Evaluasi dan Strategi Validasi

Kinerja model dinilai menggunakan *confusion matrix* beserta metrik turunannya, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* per kelas, serta macro-F1 dan weighted-F1 sebagai ukuran agregat. Macro-F1 dipakai sebagai metrik utama karena merata-ratakan kinerja seluruh kelas tanpa membobot jumlah sampel, sehingga tidak terdominasi kelas mayoritas.

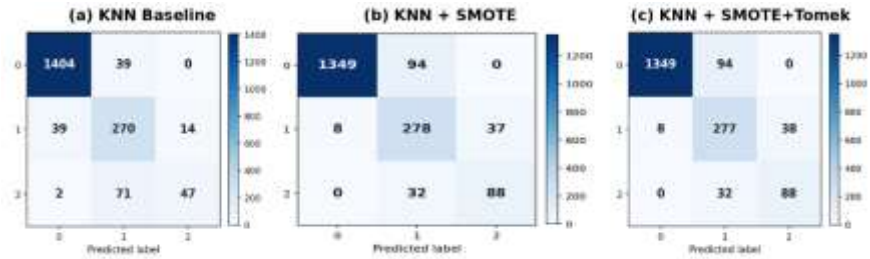
Validasi dilakukan dalam dua tahap, yaitu pengujian pada data uji independen (n = 1.886) dan *Stratified 5-Fold Cross Validation* pada data latih. Perbedaan antar strategi dilaporkan sebagai selisih absolut metrik dan selisih jumlah kasus terdeteksi, disertai selang kepercayaan 95% Wilson untuk *recall* kelas minoritas. Perbedaan yang selang kepercayaannya masih bertumpang tindih dinyatakan secara eksplisit sebagai belum teruji.

4. Hasil dan Pembahasan

Bagian ini memaparkan hasil enam kombinasi model pada data uji independen yang sama, yaitu 1.886 sampel yang terdiri atas 1.443 sampel normal, 323 stunting sedang, dan 120 stunting berat.

4.1 Hasil Eksperimen

Gambar 2 menyajikan *confusion matrix* ketiga konfigurasi KNN, sedangkan metrik per kelasnya dirinci pada Tabel 4. Pada kondisi *baseline*, akurasi mencapai 0,913 tetapi macro-F1 hanya 0,753: kelas normal memperoleh *recall* 0,973, sedangkan kelas stunting berat hanya 0,392 atau 47 dari 120 kasus, dengan 71 kasus bergeser ke kelas stunting sedang. Selisih 16 poin antara akurasi dan macro-F1 itulah wujud *accuracy paradox*, yaitu metrik agregat yang tertutup dominasi kelas normal penyusun 76,5% data.

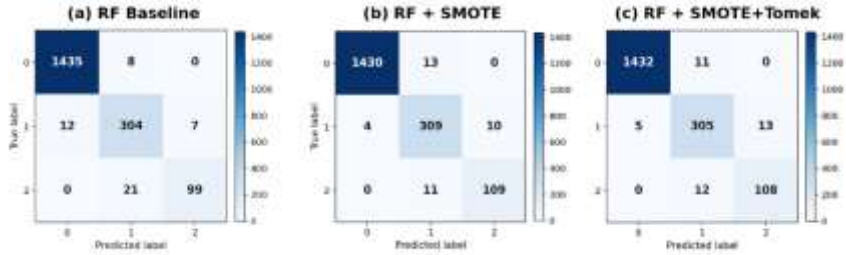


Gambar 2. Confusion Matrix Model KNN pada Data Uji

Penerapan SMOTE mengubah komposisi kesalahan tanpa mengubah akurasi secara berarti (0,909). *Recall* kelas stunting berat naik dari 0,392 menjadi 0,733, yaitu dari 47 menjadi 88 kasus, dan tidak ada lagi kasus stunting berat yang diprediksi normal. Kenaikan tersebut ditebus turunnya *precision* kelas itu dari 0,771 menjadi 0,704 dan *recall* kelas normal dari 0,973 menjadi 0,935. Secara agregat macro-F1 naik menjadi 0,816. Pola ini sesuai cara kerja SMOTE: penambahan sampel sintetis memperluas wilayah kelas minoritas sehingga batas keputusan bergeser ke arah kelas mayoritas, yakni pertukaran yang menguntungkan bagi penapisan. Selang kepercayaan 95% Wilson untuk *recall* kelas stunting berat bergerak dari [0,309; 0,481] menjadi [0,648; 0,804] dan keduanya tidak bertumpang tindih.

Penambahan tahap Tomek Links hampir tidak berpengaruh. Dari 1.886 sampel uji hanya satu yang berpindah prediksi, sehingga macro-F1 bergerak tipis dari 0,816 menjadi 0,814 dan *recall* ketiga kelas praktis tetap. Hal ini sejalan dengan kecilnya intervensi tahap tersebut, yang hanya menghapus 30 dari 17.307 sampel data latih (0,17%); penjelasannya diuraikan pada Subbab 4.3.

Gambar 3 menyajikan *confusion matrix* ketiga konfigurasi Random Forest. Pada kondisi *baseline*, model mencapai akurasi 0,975 dan macro-F1 0,932. Kelas stunting berat memperoleh *precision* 0,934 tetapi *recall* 0,825, yaitu 99 dari 120 kasus, dengan 21 sisanya diprediksi sebagai stunting sedang dan tidak satu pun jatuh ke kelas normal. Jarak kinerja antarkelas jauh lebih sempit daripada KNN, namun arah biasnya sama: kelas paling jarang memperoleh *recall* terendah.



Gambar 3. Confusion Matrix Model Random Forest pada Data Uji

Berbeda dengan pola pada KNN yang menukar *precision* demi *recall*, pada Random Forest kedua kriteria membaik bersamaan setelah penerapan SMOTE. Akurasi naik menjadi 0,980 dan macro-F1 menjadi 0,949, sementara *recall* kelas stunting berat naik dari 0,825 menjadi 0,908 atau dari 99 menjadi 109 kasus. *Precision* kelas tersebut hanya turun 1,8 poin menjadi 0,916, jauh lebih kecil daripada kenaikan *recall* sebesar 8,3 poin, sehingga F1-nya naik dari 0,876 menjadi 0,912. *Recall* kelas stunting sedang pun naik menjadi 0,957. Dengan demikian SMOTE tidak memindahkan kesalahan dari satu batas kelas ke batas lain, melainkan menguranginya pada kedua batas sekaligus.

Konfigurasi RF+SMOTE+Tomek menghasilkan akurasi 0,978 dan macro-F1 0,943, keduanya sedikit di bawah RF+SMOTE, dengan *recall* kelas stunting berat 0,900 atau 108 kasus. Selisihnya terhadap RF+SMOTE hanya menyangkut empat sampel, sehingga sejalan dengan temuan pada KNN bahwa tahap Tomek Links tidak memberikan perbaikan berarti.

4.2 Perbandingan Kinerja Model

Rangkuman kinerja seluruh model disajikan pada Tabel 4 untuk metrik per kelas dan Tabel 5 untuk metrik agregat, keduanya dihitung dari data uji yang sama.

Tabel 4. Perbandingan Performa Per Kelas pada Data Uji (nilai ditulis precision / recall / F1-score)

Algoritma	Strategi	Kelas 0 (Normal)	Kelas 1 (Sedang)	Kelas 2 (Berat)	Akurasi
KNN	Baseline	0,972/0,973/0,972	0,711/0,836/0,768	0,771/0,392/0,519	0,913
KNN	SMOTE	0,994/0,935/0,964	0,688/0,861/0,765	0,704/0,733/0,718	0,909
KNN	SMOTE+Tomek	0,994/0,935/0,964	0,687/0,858/0,763	0,698/0,733/0,715	0,909
RF	Baseline	0,992/0,994/0,993	0,913/0,941/0,927	0,934/0,825/0,876	0,975
RF	SMOTE	0,997/0,991/0,994	0,928/0,957/0,942	0,916/0,908/0,912	0,980
RF	SMOTE+Tomek	0,997/0,992/0,994	0,930/0,944/0,937	0,893/0,900/0,896	0,978

Tabel 4 memperlihatkan bahwa akurasi keseluruhan justru menyembunyikan perbedaan yang paling penting. KNN *baseline* dan KNN+SMOTE memiliki akurasi yang praktis sama (0,913 berbanding 0,909), padahal macro-F1 keduanya berselisih enam poin dan jumlah kasus stunting berat yang terdeteksi berbeda hampir dua kali lipat, yaitu 47 berbanding 88. Pada Random Forest, selisih akurasi antara *baseline* dan SMOTE hanya 0,5 poin, sedangkan selisih *recall* kelas stunting berat mencapai 8,3 poin atau setara sepuluh anak. Penilaian yang bertumpu pada akurasi akan menyimpulkan keempat konfigurasi tersebut setara, padahal perbedaannya justru terletak pada kemampuan mengenali kelas yang menjadi sasaran penapisan. Metrik yang menangkap perbedaan tersebut dirangkum pada Tabel 5.

Tabel 5. Ringkasan Metrik Agregat, Deteksi Kelas Stunting Berat, dan Skor Validasi Silang

Model	Akurasi	Macro-F1	Recall kelas 2	Precision kelas 2	Deteksi kelas 2	Macro-F1 CV
KNN Baseline	0,913	0,753	0,392	0,771	47	0,697 ± 0,019
KNN + SMOTE	0,909	0,816	0,733	0,704	88	0,773 ± 0,016
KNN + SMOTE+Tomek	0,909	0,814	0,733	0,698	88	0,772 ± 0,015
RF Baseline	0,975	0,932	0,825	0,934	99	0,910 ± 0,010
RF + SMOTE	**0,980**	**0,949**	**0,908**	0,916	**109**	0,918 ± 0,015
RF + SMOTE+Tomek	0,978	0,943	0,900	0,893	108	**0,921 ± 0,011**

Random Forest unggul atas KNN pada seluruh strategi data, dan selisihnya paling lebar pada kelas minoritas: dengan strategi yang sama (SMOTE), *recall* kelas stunting berat RF mencapai 0,908 sedangkan KNN 0,733. Keunggulan ini konsisten dengan sifat *ensemble* RF yang mengagregasi banyak pohon sehingga kurang sensitif terhadap kepadatan lokal sampel.

Pengaruh penyeimbangan kelas berbeda menurut algoritmanya, tetapi arahnya sama-sama positif. Pada KNN, perbaikan besar pada kelas minoritas ditebus turunnya *precision*; pada Random Forest, perbaikan terjadi tanpa imbalan berarti. Perbedaan besaran ini mencerminkan besarnya bias kelas yang masih tersisa pada model dasar: KNN menyisakan bias besar, sedangkan RF sudah relatif seimbang sejak awal sehingga ruang perbaikannya lebih sempit meskipun tetap nyata.

Kontribusi tahap Tomek Links kecil pada kedua algoritma, yaitu satu keputusan klasifikasi yang berubah pada KNN dan empat sampel pada RF. Pola yang konsisten pada kedua paradigma pembelajaran menunjukkan penyebabnya berasal dari struktur data, bukan dari karakteristik algoritma tertentu.

Kolom terakhir Tabel 5 menyajikan macro-F1 dari *stratified 5-fold cross validation* sebagai pemeriksaan kestabilan. Simpangan bakunya kecil pada seluruh model (0,010 sampai 0,019) dan urutan peringkat antarmodel sama dengan urutan pada data uji, sehingga keunggulan Random Forest maupun manfaat SMOTE bukan artefak satu pembagian data. Nilai validasi silang konsisten sedikit lebih rendah daripada nilai pada data uji, misalnya 0,918 berbanding 0,949 pada RF+SMOTE. Selisih itu justru konsekuensi yang diharapkan dari penempatan *resampling* di dalam *Pipeline*: data validasi setiap *fold* tidak pernah di-*resample* sehingga tetap timpang, berbeda dengan skema yang menyeimbangkan sebelum pembagian *fold* dan menghasilkan skor optimistis.

Berbeda dengan pola yang lazim dilaporkan pada studi ketidakseimbangan kelas, pemilihan model terbaik di sini tidak menuntut kompromi antara keseimbangan kinerja dan sensitivitas deteksi. RF+SMOTE unggul pada kedua kriteria sekaligus, yaitu macro-F1 tertinggi (0,949) dan deteksi kasus stunting berat terbanyak (109 dari 120), dengan akurasi 0,980 yang juga tertinggi. Untuk konteks deteksi dini stunting, ketika biaya *false negative* lebih besar daripada biaya *false positive*, konfigurasi ini yang direkomendasikan.

4.3 Diskusi Temuan Penelitian

Temuan ini menempatkan ketidakseimbangan kelas sebagai faktor yang besar pengaruhnya bergantung pada jenis pengklasifikasi. Pada KNN, keputusan ditentukan komposisi k tetangga terdekat sehingga wilayah dengan kepadatan sampel minoritas rendah hampir selalu dimenangkan kelas mayoritas; SMOTE bekerja tepat pada penyebab tersebut dengan menaikkan kepadatan lokal kelas minoritas.

Pada Random Forest, *bagging* dan pemilihan fitur acak sudah meredam sebagian besar bias kelas sehingga ruang perbaikan lebih sempit, tetapi arah perbaikannya tetap sama dan tidak disertai kemunduran pada kelas lain. Besarnya manfaat *oversampling* karena itu sebanding dengan

besarnya bias yang masih tersisa pada model dasar, sedangkan bentuk pertukaran yang menyertainya bergantung pada seberapa tegas model dasar tersebut sudah memisahkan kelas.

Kecilnya kontribusi Tomek Links dapat ditelusuri dari struktur pemisahan kelas. Karena label status gizi diturunkan dari ambang HAZ menurut WHO, batas antarkelas sepenuhnya ditentukan oleh jenis kelamin, umur, dan tinggi badan. Pemeriksaan pada 50 strata jenis kelamin x umur menunjukkan rentang tinggi badan ketiga kelas tidak bertumpang tindih pada satu strata pun; seluruh 100 pemeriksaan batas menghasilkan celah positif antara 0,1 dan 0,7 cm.

Tumpang tindih yang teramati pada ruang fitur lengkap karena itu berasal dari variabel yang tidak menentukan label, yaitu berat badan, yang rata-ratanya hampir sama pada ketiga kelas (9,27; 9,69; dan 8,95 kg). Pasangan tetangga terdekat lintas kelas yang menjadi sasaran Tomek Links sebagian besar terbentuk karena kedekatan berat badan, bukan karena keraguan pada batas tinggi badan, sehingga penghapusannya tidak menggeser batas keputusan yang sesungguhnya. Temuan ini menunjukkan bahwa manfaat strategi *hybrid* tidak ditentukan oleh ada atau tidaknya tumpang tindih dalam ruang fitur secara keseluruhan, melainkan oleh apakah tumpang tindih tersebut terjadi pada dimensi yang benar-benar membawa informasi kelas.

Dari perspektif kesehatan masyarakat, konsekuensi kesalahan klasifikasi bersifat asimetris: satu kasus stunting berat yang tidak terdeteksi berarti tertundanya intervensi gizi, sedangkan satu kasus normal yang salah ditandai hanya berujung pada pemeriksaan lanjutan. RF *baseline* meloloskan 21 dari 120 kasus stunting berat sedangkan RF+SMOTE meloloskan 11 kasus. Selisih sepuluh kasus itu hanya 0,5% dari 1.886 sampel uji, tetapi menyangkut hampir separuh kasus yang sebelumnya tidak terdeteksi. Pelaporan kinerja model prediksi stunting karena itu sebaiknya selalu menyertakan *recall* dan F1 per kelas serta macro-F1.

5. Kesimpulan

Penelitian ini mengevaluasi pengaruh SMOTE dan SMOTE+Tomek Links terhadap klasifikasi status gizi balita menggunakan KNN dan Random Forest pada 9.426 sampel antropometri dengan rasio kelas sekitar 12:1. Berdasarkan macro-F1 dan *recall* kelas stunting berat pada data uji independen (n = 1.886), Random Forest unggul atas KNN pada seluruh strategi data. KNN *baseline* menampilkan *accuracy paradox* dengan akurasi 0,913 tetapi macro-F1 0,753 dan *recall* kelas stunting berat 0,392; SMOTE menaikkan keduanya menjadi 0,816 dan 0,733. Pada Random Forest, SMOTE memperbaiki seluruh metrik utama sekaligus, yaitu akurasi 0,975 menjadi 0,980, macro-F1 0,932 menjadi 0,949, dan *recall* kelas stunting berat 0,825 menjadi 0,908 atau 99 menjadi 109 dari 120 kasus, dengan penurunan *precision* hanya 1,8 poin. Tahap Tomek Links tidak memberikan perbaikan berarti pada kedua algoritma karena hanya menghapus 30 dari 17.307 sampel data latih. RF+SMOTE karena itu direkomendasikan karena unggul pada kriteria keseimbangan antarkelas sekaligus sensitivitas deteksi.

Penelitian ini memiliki sejumlah keterbatasan. Prediktor terbatas pada empat variabel antropometri sehingga determinan sosial ekonomi, pola asuh, dan sanitasi belum terwakili. Rentang umur data hanya 0 sampai 24 bulan, sehingga temuan berlaku bagi dua tahun pertama kehidupan. Evaluasi dilakukan pada satu dataset dan satu pembagian data uji; selang kepercayaan 95% untuk *recall* kelas stunting berat pada ketiga strategi RF masih bertumpang tindih sehingga perbedaannya belum teruji secara statistik. Pada data akhir masih terdapat 54 baris bernilai identik (0,57%) yang tidak dihapus karena kombinasi antropometri yang sama mungkin berasal dari dua anak berbeda. Terakhir, karena label diturunkan dari ambang HAZ, batas antarkelas bersifat tegas; pada dataset yang labelnya ditentukan penilaian klinis, kontribusi tahap pembersihan batas berpeluang lebih besar.

Penelitian selanjutnya disarankan melakukan validasi eksternal pada dataset posyandu dari wilayah lain, menggunakan *repeated stratified cross validation* beserta uji berpasangan seperti

McNemar atau Wilcoxon *signed-rank*, memperluas prediktor ke faktor non-antropometri, serta membandingkan Borderline-SMOTE, ADASYN, dan *cost-sensitive learning*.

6. Pernyataan Penulis

Penulis menyatakan bahwa tidak ada konflik kepentingan terkait publikasi artikel ini. Penulis menyatakan bahwa data dan makalah bebas dari plagiarisme serta penulis bertanggung jawab secara penuh atas keaslian artikel.

Daftar Pustaka

[1] WHO, "Malnutrition," World Health Organization. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/malnutrition

[2] M. Y. E. Soekatri, S. Sandjaja, and A. Syauqy, "Stunting Was Associated with Reported Morbidity, Parental Education and Socioeconomic Status in 0.5–12-Year-Old Indonesian Children," *IJERPH*, vol. 17, no. 17, p. 6204, Aug. 2020, doi: 10.3390/ijerph17176204.

[3] S. Zaleha and H. Idris, "IMPLEMENTATION OF STUNTING PROGRAM IN INDONESIA: A NARRATIVE REVIEW," *JAKI*, vol. 10, no. 1, pp. 143–151, Jun. 2022, doi: 10.20473/jaki.v10i1.2022.143-151.

[4] J. H. Rah, S. Sukotjo, N. Badgaiyan, A. A. Cronin, and H. Torlesse, "Improved sanitation is associated with reduced child stunting amongst Indonesian children under 3 years of age," *Maternal & Child Nutrition*, vol. 16, no. S2, p. e12741, Oct. 2020, doi: 10.1111/mcn.12741.

[5] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques," *J Prev Med Public Health*, vol. 56, no. 1, pp. 41–49, Jan. 2023, doi: 10.3961/jpmph.22.388.

[6] E. K. Anku and H. O. Duah, "Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms," *PLoS ONE*, vol. 19, no. 2, p. e0296625, Feb. 2024, doi: 10.1371/journal.pone.0296625.

[7] S. M. J. Rahman *et al.*, "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," *PLoS ONE*, vol. 16, no. 6, p. e0253172, Jun. 2021, doi: 10.1371/journal.pone.0253172.

[8] KEMENTERIAN KESEHATAN RI, *Buku Saku SSGI 2022*. KEMENTERIAN KESEHATAN RI, 2022.

[9] V. Patel and H. Bhavsar, "Review on Data Balancing Approaches for Skewed Data Classification," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India: IEEE, Sep. 2024, pp. 235–241. doi: 10.1109/IC3I61595.2024.10828620.

[10] K. U. Apu and M. Ali, "A Systematic Literature Review on AI Approaches To Address Data Imbalance In Machine Learning," *FAET*, vol. 2, no. 01, pp. 58–77, Jan. 2025, doi: 10.70937/faet.v2i01.57.

[11] A. Aziz, F. Insani, J. Jasril, and F. Syafria, "Implementasi Metode Learning Vector Quantization (LVQ) Untuk Klasifikasi Keluarga Beresiko Stunting," *bits*, vol. 5, no. 1, Jun. 2023, doi: 10.47065/bits.v5i1.3478.

[12] S. Y. Andriyani, M. S. Lydia, and S. Efendi, "Optimization of Support Vector Machine Algorithm Using Stunting Data Classification," *J. Prisma. Sains*, vol. 11, no. 1, p. 164, Jan. 2023, doi: 10.33394/j-ps.v11i1.6619.

[13] A. Iriany, W. Ngabu, D. Arianto, and A. Putra, "CLASSIFICATION OF STUNTING USING GEOGRAPHICALLY WEIGHTED REGRESSION-KRIGING CASE STUDY: STUNTING IN EAST JAVA," *BAREKENG: J. Math. & App.*, vol. 17, no. 1, pp. 0495–0504, Apr. 2023, doi: 10.30598/barekengvol17iss1pp0495-0504.

[14] Y. S. Asgedom *et al.*, "Levels of stunting associated factors among under-five children in Ethiopia: A multi-level ordinal logistic regression analysis," *PLoS ONE*, vol. 19, no. 1, p. e0296451, Jan. 2024, doi: 10.1371/journal.pone.0296451.

[15] R. Wajgi and D. Wajgi, "Malnutrition detection in infants using machine learning approach," presented at the PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND COMPUTING APPLICATIONS-21 (ICCICA-21), Nagpur, India, 2022, p. 040006. doi: 10.1063/5.0076876.

[16] J. Wang, H. Gu, and H. Gu, "Optimizing Stroke Prediction in Machine Learning by Addressing Data Imbalance," in *2023 3rd International Signal Processing, Communications and Engineering Management Conference (ISPCEM)*, Montreal, QC, Canada: IEEE, Nov. 2023, pp. 665–669. doi: 10.1109/ISPCEM60569.2023.00125.

[17] S. He, "Addressing data imbalance in neural network spam detection with insights from SMS spam collection," *TNS*, vol. 39, no. 1, pp. 212–218, Jul. 2024, doi: 10.54254/2753-8818/39/20240636.

[18] N. Faulina, K. Nisa, and W. Warsono, "Enhancing Tuberculosis Diagnosis: Effective Naive Bayes Classification using SMOTE and Tomek Links for Imbalanced Data," *InPrime:Ind.Jour.Pure.Applied.Math*, vol. 6, no. 2, pp. 98–111, Oct. 2024, doi: 10.15408/inprime.v6i2.41463.

[19] M. M. Hasan Bhuiyan, S. A. Poly, and A. Kumar Acharyan, "Comparative Study on Brainstroke Prediction Using Data Balancing Techniques," in *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, Bhubaneswar, India: IEEE, Jan. 2024, pp. 1–5. doi: 10.1109/ASSIC60049.2024.10507962.

[20] W. Rahayu *et al.*, "Synthetic Minority Oversampling Technique (SMOTE) for Boosting the Accuracy of C4.5 Algorithm Model," *j. of artif. intell. and eng. appl.*, vol. 3, no. 3, pp. 624–630, Jun. 2024, doi: 10.59934/jaiea.v3i3.469.

[21] F. Y. A'la, N. Firdaus, Hartatik, and M. A. Safi'Ie, "SMOTE on Numeric Breast Cancer Dataset to Overcome Imbalance Class," in *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, Lombok, Indonesia: IEEE, Sep. 2023, pp. 335–339. doi: 10.1109/IC2IE60547.2023.10331221.

[22] C. El Morr, M. Jammal, H. Ali-Hassan, and W. El-Hallak, "Data Preprocessing," in *Machine Learning for Practical Decision Making*, vol. 334, in International Series in Operations Research & Management Science, vol. 334. , Cham: Springer International Publishing, 2022, pp. 117–163. doi: 10.1007/978-3-031-16990-8_4.

[23] G. Y. Lee, L. Alzamil, B. Doskenov, and A. Termehchy, "A Survey on Data Cleaning Methods for Improved Machine Learning Model Performance," Sep. 15, 2021, *arXiv*: arXiv:2109.07127. doi: 10.48550/arXiv.2109.07127.

[24] World Health Organization, *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*. Geneva: World Health Organization, 2006. [Online]. Available: https://www.who.int/publications/i/item/924154693X

[25] M. D. V. Prasad and S. T, "Enhancing K-Means Clustering Accuracy Through Modified Robust Scaling Technique," Nov. 18, 2024, *Computer Science and Mathematics*. doi: 10.20944/preprints202411.1245.v1.

[26] H. A. Ahmed, P. J. Muhammad Ali, A. K. Faeq, and S. M. Abdullah, "An Investigation on Disparity Responds of Machine Learning Algorithms to Data Normalization Method," *ARO*, vol. 10, no. 2, pp. 29–37, Sep. 2022, doi: 10.14500/aro.10970.

[27] A. Pajankar and A. Joshi, "Preparing Data for Machine Learning," in *Hands-on Machine Learning with Python*, Berkeley, CA: Apress, 2022, pp. 79–97. doi: 10.1007/978-1-4842-7921-2_6.

[28] V. V. Starovoitov and Yu. I. Golub, "Data normalization in machine learning," *Informatika (Minsk)*, vol. 18, no. 3, pp. 83–96, Sep. 2021, doi: 10.37661/1816-0301-2021-18-3-83-96.

[29] A. Asesh, "Normalization and Bias in Time Series Data," in *Digital Interaction and Machine Intelligence*, vol. 440, C. Biele, J. Kacprzyk, W. Kopeć, J. W. Owsiński, A. Romanowski, and

M. Sikorski, Eds., in Lecture Notes in Networks and Systems, vol. 440. , Cham: Springer International Publishing, 2022, pp. 88–97. doi: 10.1007/978-3-031-11432-8_8.

[30] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *IJACSA*, vol. 15, no. 2, 2024, doi: 10.14569/IJACSA.2024.0150235.

[31] B. Supri, Rudianto, Abdurohim, Badriatul Mawadah, and Helmi Ali, "Asian Stock Index Price Prediction Analysis Using Comparison of Split Data Training and Data Testing," *jemsi*, vol. 9, no. 4, pp. 1403–1408, Aug. 2023, doi: 10.35870/jemsi.v9i4.1339.