

KLASIFIKASI STUNTING DAN SEBARAN ZONA DAERAH METODE EXTREME GRADIENT BOOST, RANDOM FOREST, DAN DECISION TREE

Tjut Awaliyah Zuraiyah
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
email

Bira Arya Setha
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
biraaryaset@gmail.com

Dinnar Nurhuda Hermawan
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
biraaryaset@gmail.com

Kriti Mauludin
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
email

Yamato
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
email

Andi Chairunnas
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
email

Nurdina Widanti
Jurusan Ilmu Komunikasi
Universitas Pakuan
Bogor, Indonesia
email

Abstract—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. ***CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.** (Abstract)

Keywords—component, formatting, style, styling, insert (key words)

I. PENDAHULUAN

Permasalahan utama tentang gizi yang dihadapi Indonesia adalah balita *stunting* atau biasa disebut balita kerdil. *Stunting* adalah suatu kondisi di mana anak mengalami kekurangan gizi kronis dalam jangka waktu lama dan tidak mendapat nutrisi yang cukup, sehingga menyebabkan pertumbuhan lambat dan tinggi badan tidak sesuai dengan usianya. Kegagalan pertumbuhan dapat dimulai pada masa perkembangan janin dan menjadi nyata setelah anak mencapai 2 tahun. [1]. *Stunting* berpengaruh terhadap kehidupan anak jika tidak mendapat penanganan yang cepat dan baik [2]

Stunting dapat terjadi karena beberapa faktor, yaitu pola asuh orang tua yang buruk, kurangnya pengetahuan ibu tentang kesehatan dan gizi sebelum atau sesudah melahirkan, terbatasnya pelayanan kesehatan, terbatasnya akses keluarga terhadap makanan bergizi, dan terbatasnya akses terhadap air bersih dan sanitasi. [1]

Kejadian *stunting* akan berdampak buruk pada balita jika tidak ditangani lebih lanjut. Dampak jangka pendek dari *stunting* antara lain terganggunya perkembangan fisik dan mental, berkurangnya kecerdasan, dan masalah metabolisme. Sedangkan dampak jangka panjang dari *stunting* tercermin pada menurunnya kemampuan kognitif dan menurunnya daya tahan tubuh sehingga membuat tubuh mudah terserang penyakit dan

berisiko terkena penyakit degeneratif seperti diabetes, penyakit kardiovaskular, kanker, stroke, dan menjadi tidak sehat. menjadi kompetitif di tempat kerja menyebabkan produktivitas rendah. [3]

Hasil dari Survei Status Gizi Indonesia (SSGI) yang diumumkan oleh Kementerian Kesehatan pada Rapat Kerja Nasional BKKBN tahun 2023 menunjukkan bahwa prevalensi *stunting* di Indonesia mengalami penurunan dari 17.8% pada tahun 2023. Menurut *website Indonesia.go.id* pemerintah menargetkan penurunan angka *stunting* menjadi 14% pada akhir 2024.

Mendeteksi *stunting* pada balita dapat dideteksi sejak dini dengan rutin memeriksa grafik tumbuh kembang anak di Posyandu terdekat. Kondisi balita diukur oleh pengurus Posyandu, kemudian pengukuran tersebut dikirimkan kepada ahli atau pakar untuk mengetahui apakah pengukuran tersebut menunjukkan adanya perlambatan pertumbuhan [4] Hasil identifikasi dini gizi buruk *stunting* memudahkan orang tua dan pengelola anak dalam menerapkan upaya pencegahan gizi buruk *stunting* pada tahap awal pertumbuhan.

Pemeriksaan *stunting* dilakukan dengan menghitung *z-score* berdasarkan indeks antropometri yang ditetapkan oleh Kementerian Kesehatan Republik Indonesia dengan ambang batas standar deviasi < -3 (Kemenkes, 2022), dimana pemeriksaan *stunting* pada balita masih cukup menyita waktu karena dilakukan secara manual dan rawan kesalahan. kemudian diperlukan suatu sistem yang dapat mengklasifikasikan data pemeriksaan balita sehingga dapat memprediksi dengan cepat dan akurat apakah anak mengalami *stunting* atau tidak dengan cepat dan akurat. Oleh karena itu, agar orang tua dan petugas posyandu dapat lebih cepat mengetahui kondisi anak, maka perlu dibangun suatu sistem yang menggunakan metode

data *Mining* untuk dapat mengklasifikasikan data yang dihasilkan. Agar dapat dengan cepat memprediksi balita tersebut masuk dalam kategori *stunting* atau tidak.

Data *Mining* adalah suatu proses penggalian informasi dan identifikasi pola serta hubungan dalam kumpulan data untuk memecahkan masalah tertentu. Misalnya, salah satu metode data *Mining* adalah klasifikasi. Teknik klasifikasi digunakan untuk mengelompokkan atau mengklasifikasikan data ke dalam kelas-kelas tertentu berdasarkan label yang ditentukan. [1]

Klasifikasi adalah proses yang digunakan untuk mengidentifikasi jenis data yang label kelasnya tidak diketahui. Memprediksi label kelas adalah tujuan utama klasifikasi. Proses klasifikasi mempunyai dua tahap, tahap pertama adalah pelatihan, yaitu menganalisis data pelatihan menggunakan algoritma klasifikasi. Dan yang kedua adalah klasifikasi, yaitu memprediksi keakuratan klasifikasi menggunakan data uji. [5]

Ada beberapa metode untuk klasifikasi, diantaranya adalah metode *eXtreme Gradient Boosting* (XGBoost), Random Forest, dan Decision Tree. Ketiga algoritme tersebut tergolong supervised learning karena bekerja menggunakan data yang telah memiliki label sebelumnya. Random Forest memiliki keunggulan diantaranya cocok diterapkan pada data berjumlah besar, dan bekerja dengan baik dengan data nonlinear. XGBoosting mampu melakukan pemrosesan paralel, yang dapat mempercepat perhitungan komputasi, memiliki fleksibilitas pengaturan objek yang besar, dan mengatasi split saat negative loss. Decision Tree menghasilkan suatu model yang dapat memprediksi kategori data dengan cara mempelajari aturan penentuan kategori berdasarkan fitur-fitur yang dimiliki oleh data. Komparasi performa ketiga algoritme menggunakan metrik evaluasi sebagai acuannya yakni nilai akurasi. [6]

Penelitian dengan metode tersebut pernah dilakukan oleh beberapa peneliti terdahulu dan penelitian terkait klasifikasi *stunting* juga dilakukan dengan menerapkan beberapa algoritma tertentu seperti penelitian yang dilakukan oleh Sabbir Ahmed Hemo dan Md. Israt Rayhan pada tahun 2021 yang berjudul *Classification tree and random forest model to predict under-five malnutrition in Bangladesh*. Random forest memiliki akurasi masing-masing 70,1% dan 72,4% untuk memprediksi *stunting* dan *underweight*. Classification tree telah memprediksi 68,7% dan 70,5% anak-anak dengan *stunting* dan *underweight* secara akurat. [7]

Penelitian kedua dilakukan oleh Juhidur Rahman Khan, Javed H Tomal dan Enayetur Raheem pada tahun 2021 dengan judul *Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh*. GBOOST memberikan hasil terbaik dalam memprediksi *stunting* diikuti oleh RF, SVM, TREE, dan Linear Regresion. Dalam studi ini, 5 metode teratas yang dipilih juga memiliki kinerja yang lebih baik (dalam hal akurasi klasifikasi) daripada metode dengan kinerja terbaik yang ditemukan dalam studi sebelumnya untuk memprediksi *stunting*. Akurasi metode dengan kinerja terbaik dalam studi tersebut (GBOOST) adalah $(100 - 31,529) = 68,47\%$. [8]

Penelitian ketiga dilakukan oleh Andi Nugroho, Harco Leslie Hendric Spits Warnars, Ford Lumban Gaol

dan Tokuro Matsuo pada tahun 2022 yang berjudul *Trend of Stunting Weight for Infants and Toddlers Using Decision Tree*. Decision Tree + Anova memberikan hasil terbaik dalam memprediksi *stunting* diikuti oleh KNN, Decision Tree, dan KNN + Anova. Akurasi metode dengan kinerja terbaik dalam studi tersebut (Decision Tree + Anova) adalah 0.96. [9]

Penelitian keempat dilakukan oleh M. Syaqui Haris, Mochammad Anshori dan Ahsanun Naseh Khudori pada tahun 2023 yang berjudul *Prediction Of Stunting Prevalence In East Java Province With Random Forest Algorithm*. Hasil penelitian ini menunjukkan bahwa prediksi Multi Linear Regression (MLR) masih lebih baik daripada prediksi Random Forest (RF) dengan nilai MAE 0,93 hingga 1,02 dan nilai MSE 1,34 hingga 1,64. [10]

Penelitian kelima dilakukan oleh Jonathan Mkungudza, Halima S. Twabi dan Samuel O. M. Manda pada tahun 2024 yang berjudul *Development of a diagnostic predictive model for determining child stunting in Malawi: a comparative analysis of variable selection approaches*. (AUC=0,67, 95% CI: 0,64, 0,69), yang menggunakan data dari satu fasilitas kesehatan merupakan prediktor *stunting* yang diidentifikasi oleh model berkinerja terbaik dan konsisten dengan pengetahuan yang ada tentang faktor penentu *stunting*, seperti demografi anak dan indikator kekayaan. [11]

Penelitian terakhir dilakukan oleh Hao Shen, Hang Zhao dan Yi Jiang pada tahun 2023 yang berjudul *Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea*. Penelitian ini menunjukkan bahwa kombinasi LASSO-XGBoost memiliki kinerja terbaik dalam memprediksi *stunting* pada anak di bawah lima tahun di PNG (AUC: 0,767, 95% CI: 0,714–0,819). [12]

Penelitian terakhir dilakukan oleh Obvious Nchimunya Chilyabanyama, Roma Chilengi, Michelo Simuyandi, Caroline C. Chisenga, Masuzyo Chirwa, Kalongo Hamusonde, Rakesh Kumar Saroj, Najeeha Talat Iqbal, Innocent Ngaruye, dan Samuel Bosomprah, pada tahun 2022 yang berjudul *Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia*. Penelitian ini membandingkan 4 metode yaitu logistic regression, SV classification, Random Forest, dan XGBoost. Dari beberapa model tersebut Random Forest merupakan model dengan hasil terbaik untuk nilai prediksi akurasinya [13]

II. LANDASAN TEORI

A. Klasifikasi

Klasifikasi merupakan pemahaman secara kritis terhadap obyek yang melibatkan pekerjaan membuat urutan, kelompok atau kelas, dan hal itu memberikan makna tertentu atas sebuah realitas. [14]

B. Penyebaran zona daerah

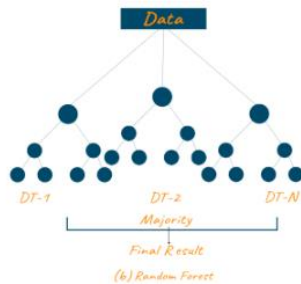
Penyebaran zona daerah adalah sekumpulan teknik yang dapat digunakan dalam pengolahan data SIG (sistem informasi dan geospasial). Teknik analisis ini digunakan untuk mengidentifikasi perubahan morfologi dan pola perkembangan suatu kawasan. [15]

C. Extreme Gradient Boost

EXtreme Gradient Boosting atau XGBoost merupakan salah satu metode boosting yaitu kumpulan DT yang pembangunan pohon berikutnya akan bergantung pada pohon sebelumnya. Pohon pertama dalam XGBoost akan lemah dalam melakukan klasifikasi dengan inisialisasi probability yang ditentukan oleh peneliti dan kemudian akan dilakukan update bobot pada setiap pohon yang dibangun sehingga menghasilkan kumpulan pohon klasifikasi yang kuat. [16]

D. Random Forest

Random Forest adalah metode yang terdiri dari sekumpulan pohon keputusan yang dihasilkan secara acak dari sampel data pelatihan. Setiap pohon di hutan ditentukan berdasarkan subset acak dari variabel prediktor. Keunggulannya adalah kemampuan memenangkan kinerja yang lebih baik dan lebih baik dibandingkan dengan metode tradisional. Cara membangun pohon keputusan di Random Forest adalah dengan memilih fungsi root berdasarkan nilai gain yang dihitung dari entropi. Proses ini diulangi untuk setiap nilai atribut hingga semua sampel memiliki kelas yang sama. [17]



Gambar 1. Alur Random Forest

Alur dan Rumus untuk membuat pohon Keputusan:

$$Entropy(s) = \sum_{i=0}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan:

S: Himpunan kasus

A: Fitur

n: jumlah partisi S.

p_i : proporsi dari S_i terhadap S

Rumus Gain:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

S: himpunan kasus

A: atribut

N: jumlah partisi atribut A

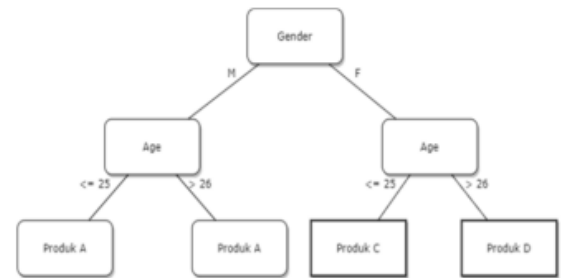
$|S_i|$: jumlah kasus pada partisi ke -i

$|S|$: jumlah kasus dalam S

E. Decision Tree

Algoritma Decision Tree merupakan suatu metode pengklasifikasian yang menggunakan contoh pohon, menyatakan node yang menggambarkan tiap atribut, yang mana daun menggambarkan nilai dari tiap kelas. Node akar menyatakan node yang berada paling atas dari pohon. Setiap node ini menggambarkan node pembagi, yang mana tiap node ini merupakan satu masukan dan memiliki sedikitnya dua keluaran. Leaf node adalah node terakhir, hanya mempunyai satu masukan, dan tidak mempunyai keluaran. Pohon keputusan pada tiap leaf node menyatakan label tiap kelas. Pohon keputusan pada tiap cabangnya menyatakan keadaan yang harus diisi dan tiap puncak pohonnya menggambarkan nilai kelas data. [18]

Pada metode ini suatu item dikelompokkan juga digambarkan berupa sebuah pohon keputusan, sehingga mudah untuk dimengerti. Metode decision tree digambarkan pada Gambar 2, menerangkan bagaimana suatu produk dipilih pembeli. [18]



Gambar 2. Alur Decision Tree

F. Confussion Matrix

Confusion Matrix adalah alat ukur berbentuk matrix yang digunakan untuk mendapatkan jumlah ketepatan klasifikasi terhadap kelas dengan algoritme yang dipakai [19]. Berikut akan disajikan bentuk Confusion Matrix pada Tabel 1.

Tabel 1. Confusion Matrix

Confusion Matrix		Nilai Sebenarnya	
		True	False
Nilai Prediksi	True	TP (True Positive) Correct Result	FP (False Positive) Unexpected result
	False	FN (False Negative) Missing result	TN (True Negative) Correct absence of result

Pada Tabel 1 nilai TP (true positive) dan TN (true negative) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, presisi, dan recall. Jika

label prediksi keluaran bernilai benar (*true*) dan nilai sebenarnya bernilai salah (*false*) disebut sebagai *false positive* (FP). Sedangkan jika prediksi label keluaran bernilai salah (*false*) dan nilai sebenarnya bernilai benar (*true*) maka hal ini disebut sebagai *false negative* (FN). Berikut formulasi untuk menghitung akurasi, presisi, dan *recall* pada pembentukan model klasifikasi ditunjukkan pada persamaan (3), persamaan (4), dan persamaan (5).

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

III. METODE

CROSS-INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM) adalah metode yang menyediakan proses standar dalam data mining yang lebih mudah diterapkan karena setiap tahapan atau fase didefinisikan dan terstruktur dengan jelas serta memiliki metodologi data Mining yang lengkap dan terdokumentasi dengan baik. (Lathifah, 2023)

A. Business Understanding

Business understanding ini tahap awal yang memiliki fokus pada pemahaman tujuan proyek dan persyaratandari perspektif bisnis, kemudian mengubah pengetahuan inimenjadi sebuah definisi masalah dan rencana awal data *mining* yang dirancang untuk mencapai tujuan.

B. Data Understanding

Tahap pemahaman data dimulai dengan pengumpulan data awal dan hasil kegiatan dalam rangka untuk membiasakan diri dengan data, untuk mengidentifikasi masalah kualitas data, untuk menemukan wawasan pertama ke data atau untuk mendeteksi subset menarik untuk membentuk hipotesis untuk informasi yang tersembunyi.

C. Data Preparation

Tahap persiapan data mencakup semua kegiatan untuk membangun dataset akhir (data yang akan dimasukkan ke dalam pemodelan) dari data mentah awal. Tugas data persiapan kemungkinan akan dilakukan beberapa kali dan tidak dapat diprediksi jumlahnya. Hasil tugas meliputi tabel, record dan pemilihan atribut serta transformasi dan pembersihan data untuk pemodelan.

D. Modeling

Dalam fase ini, berbagai teknik pemodelan dapat dipilih dan diterapkan dan parameter model dikalibrasi dengan nilai-nilai yang optimal. Biasanya, ada beberapa teknik data *mining* yang dapat menyelesaikan masalah yang sama. Beberapa teknik memiliki persyaratan tertentu pada formulir data. Oleh

karena itu, melangkah kembali ke tahap persiapan data seringkali dilakukan.

E. Evaluation

Pada tahap ini dalam proyek *mining* telah membangun sebuah model yang tampaknya memiliki kualitas tinggi dari perspektif analisis data. Sebelum melanjutkan ke distribusi final model, penting untuk lebih teliti mengevaluasi model dan meninjau langkah-langkah yang telah dilakukan untuk membangun model. Hal ini dilakukan untuk meyakinkan bahwa memang benar model yang dihasilkan mencapai tujuan bisnis. Salah satu tujuan utama adalah untuk menentukan apakah ada beberapa masalah bisnis penting yang belum terselesaikan telah cukup dipertimbangkan. Pada akhir tahap ini, keputusan tentang penggunaan data dan hasil *Mining* harus dicapai

F. Deployment

Penciptaan model umumnya bukan akhir dari proyek. Bahkan jika tujuan dari model adalah untuk meningkatkan pengetahuan dari data, pengetahuan yang diperoleh akan perlu diorganisasikan dan disajikan dengan cara yang pelanggan dapat menggunakannya. Hal ini sering melibatkan dan menerapkan "Hidup" model dalam proses pembuatan keputusan organisasi, misalnya secara real-time personalisasi halaman Web atau scoring berulang database marketing. Namun, tergantung pada persyaratan, sehingga pada fase deployment dapat sesederhana mungkin dalam menghasilkan laporan atau mempermudah menerapkan proses data *mining* di seluruh perusahaan. Dalam banyak kasus itu adalah terjadi masalah pada pelanggan saat penerapan model bukan analisis data. Namun, bahkan fase *deployment* penting bagi pelanggan untuk memahami di kegiatan apa yang perlu dilakukan dalam rangka untuk benar-benar memanfaatkan model yang dihasilkan.

IV. HASIL DAN PEMBAHASAN

A. Deskripsi Data

1. Data Selection

Data selection merupakan proses pengambilan data yang berhubungan dengan analisis dari basis data. Data yang dipilih pada data stunting ini yaitu usia saat ukur (bulan), berat, tinggi, BB/U, ZS BB/U, TB/U, ZS TB/U, BB/TB, ZS BB/TB.

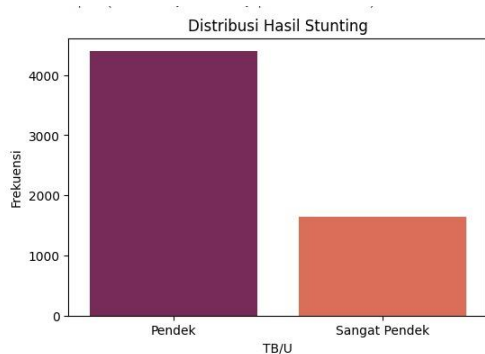
2. Data Deskripsi

	Usia Saat Ukur	Berat	Tinggi	ZS BB/U	ZS TB/U	ZS BB/TB
count	6032.000000	6032.000000	6032.000000	3654.000000	6032.000000	3654.000000
mean	34.028846	10.434090	83.35630	-2.296483	-2.789736	-1.015651
std	13.613181	2.124166	9.25064	0.800268	0.638939	1.073543
min	0.000000	2.000000	44.70000	-5.760000	-6.000000	-5.490000
25%	24.000000	9.100000	78.00000	-2.800000	-3.060000	-1.700000
50%	35.000000	10.500000	84.50000	-2.310000	-2.640000	-1.040000
75%	45.000000	11.900000	90.00000	-1.830000	-2.330000	-0.400000
max	59.000000	20.000000	100.10000	1.320000	-2.010000	5.810000

Gambar 3. Data Deskripsi

Berdasarkan Gambar 3 diketahui bahwa data terdiri dari 6032 sampel dengan jumlah variabel sebanyak 9 dan terdiri dari beberapa kelas.

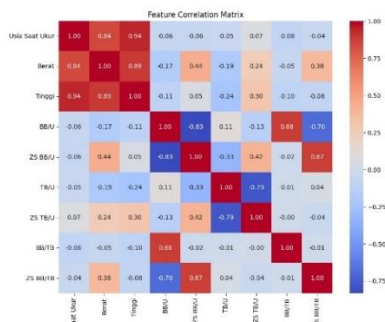
3. BarChart



Gambar 4. Barchart

Pada gambar barchart tersebut menampilkan bahwa kasus dengan indeks pendek (*stunting*) lebih banyak dari pada sangat pendek (*severe stunting*)

4. Feature Correlations



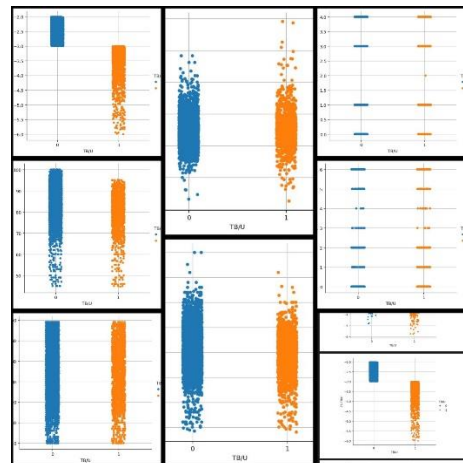
Gambar 5. Feature Correlations

Hasil akan dikorelasikan dengan Feature correlations. Feature correlations digunakan untuk mengetahui kekuatan features (variabel) terhadap target (kelas), dimana jika variabel usia saat ukur (bulan), berat, tinggi memiliki nilai yang tinggi terhadap stunting, maka variabel tersebut sangat berpengaruh, dan sebaliknya, jika semakin rendah maka semakin tidak berpengaruh seperti nilai BB/U dan ZS BB/U. Feature correlations dapat dilihat pada gambar 5.

Jika hasil korelasi mendekati nilai 1, maka korelasi tersebut dikatakan baik, namun jika mendekati nilai -1 maka korelasi dikatakan buruk. Dapat dilihat dari hasil visualisasi di atas, yang memiliki nilai korelasi baik terdapat pada variabel "Usia Saat Ukur Bulan", "Berat" dan "Tinggi" yaitu

memiliki nilai hasil korelasinya 0,84, 0.89 dan 0.94, adapun nilai korelasi yang buruk terdapat pada variabel "BB/TB" dan ZS BB/TB, yaitu memiliki nilai -0.83, jika hasil korelasi yang berbentuk diagonal menyamping tersebut itu merupakan korelasi variabel dengan dirinya sendiri.

5. Categorical Plot



Gambar 6. Categorical Plot

Pada *categorical plot*, plot dilakukan pada setiap variabel berdasarkan TB/U sehingga membentuk suatu pola persebaran data pada masing-masing variabel. Pesebaran data balita yang terdiagnosis stunting rata-rata memiliki nilai yang tinggi pada variabel-variabel tersebut. Hal tersebut selaras dengan *correlations features* pada gambar 5, dimana variabel tersebut memiliki nilai yang besar dalam mempengaruhi balita berpotensi stunting.

B. Preparation Data

1. Data Cleaning

Dalam proses ini dimana data sebelumnya yang sudah didapatkan akan dilakukan tahap data preprocessing pada Python. Dimana untuk proses preprocessing data ini melakukan data cleaning (mengisi data yang kosong atau data bernilai null). Pada gambar 7 dapat dilihat terdapat 2378 data dengan nilai null pada variabel ZS BB/U dan ZS BB/TB. Untuk mengisi data yang kosong pada variabel tersebut, dilakukannya perhitungan nilai modus pada variabel ZS BB/U dan perhitungan mean pada variabel ZS BB/TB. Untuk proses preprocessing tersebut bisa dilihat pada gambar 7.

Usia Saat Ukur	0	Usia Saat Ukur	0
Berat	0	Berat	0
Tinggi	0	Tinggi	0
BB/U	0	BB/U	0
ZS BB/U	2378	ZS BB/U	0
TB/U	0	TB/U	0
ZS TB/U	0	ZS TB/U	0
BB/TB	0	BB/TB	0
ZS BB/TB	2378	ZS BB/TB	0

dtype: int64 dtype: int64

Gambar 7. Data sebelum dan sesudah *Cleaning*

2. Mencari Nilai Outlier

Tahap selanjutnya mendeteksi nilai outliers pada masing-masing variabel dengan menggunakan metode Z-Score untuk mendapatkan data outlier. Hasil dari mencari outlier tersebut dapat dilihat pada gambar 8.

	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB
0	25	9.1	78.7	4	-2.31	0	-2.47	6	-1.015651
1	24	10.0	78.5	4	-2.31	0	-2.24	6	-1.015651
2	17	9.0	75.0	4	-2.31	0	-2.20	6	-1.015651
3	35	12.5	86.2	4	-2.31	0	-2.21	6	-1.015651
4	15	7.5	73.5	4	-2.31	0	-2.28	6	-1.015651

Gambar 8. Nilai Outlier

3. Data Transformation

Tahap awal yang dilakukan pada data transformation ini yaitu mengubah nilai categorical menjadi numeric agar dapat diproses untuk kedepannya. Hasil dapat dilihat pada gambar 9.

	Usia Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB
0	25	9.1	78.7	4	NaN	0	-2.47	6	NaN
1	24	10.0	78.5	4	NaN	0	-2.24	6	NaN
2	17	9.0	75.0	4	NaN	0	-2.20	6	NaN
3	35	12.5	86.2	4	NaN	0	-2.21	6	NaN
4	15	7.5	73.5	4	NaN	0	-2.28	6	NaN
...
6027	57	14.1	99.2	0	-1.83	0	-2.07	0	-0.88
6028	57	12.0	95.2	3	-3.09	0	-2.95	2	-2.07
6029	57	12.0	97.0	3	-3.09	0	-2.57	2	-2.46
6030	57	11.6	95.6	3	-3.16	0	-2.72	2	-2.21
6031	57	14.5	98.6	0	-1.43	0	-2.01	0	-0.25

6032 rows x 9 columns

Gambar 9. Nilai *Categorical* menjadi Nilai *Numeric*

Pada Gambar 9. Nilai Categorical menjadi Nilai Numeric dimana pada kelas TB/U nilai 0 = Pendek dan nilai 1 = Sangat Pendek.

Selanjutnya dilakukan transformasi data dengan menggunakan z-score. Dapat dilihat bahwa terdapat perbedaan nilai pada data sebelum dilakukannya preprocessing dan setelah dilakukannya proses preprocessing. Perbedaan yang sangat kontras terlihat perbedaannya adalah nilai median pada data sebelum

dan setelah preprocessing. Proses data transformation bisa dilihat pada gambar 10.

										proportion
Usia	Saat Ukur	Berat	Tinggi	BB/U	ZS BB/U	TB/U	ZS TB/U	BB/TB	ZS BB/TB	
45	12.3	84.2	4	-2.31	1	-4.00	6	-1.015651		0.000332
13	8.0	70.0	4	-2.31	0	-2.94	6	-1.015651		0.000332
14	8.0	70.0	0	-1.36	0	-2.19	0	-0.420000		0.000332
28	10.0	81.0	4	-2.31	0	-2.86	6	-1.015651		0.000332
52	10.5	84.1	3	-3.54	1	-4.74	0	-0.700000		0.000332
...
28	9.2	84.0	4	-2.31	0	-2.13	6	-1.015651		0.000166
		81.5	1	-2.45	0	-2.25	0	-1.590000		0.000166
		81.4	1	-2.48	0	-2.31	0	-1.560000		0.000166
		81.0	4	-2.31	0	-2.66	6	-1.015651		0.000166
59	18.0	96.0	0	-0.04	0	-2.91	3	2.690000		0.000166

6015 rows x 11 columns

Gambar 10. *Data Transformation*

4. Data Splitting

Dataset pada penelitian ini dibagi menjadi dua, yaitu data train dan data test dengan presentase splitting 70% data train dan 30% data test, Setelah pembagian data didapatkan 4222 data latih dan 1810 data uji. Pembagian data dapat dilihat pada gambar 11.

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=2)

[ ] print(X.shape, X_train.shape, X_test.shape)

(6032, 8) (4222, 8) (1810, 8)
```

Gambar 11. *Data Splitting*

C. Modeling

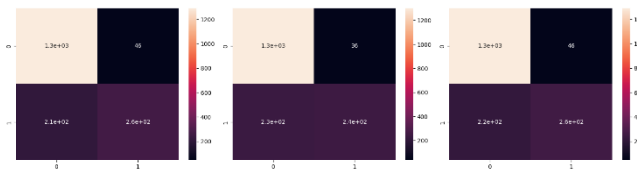
Dalam tahapan ini, data sebelumnya yang telah diproses dimasukan ke dalam model algoritma XGBoost, Random Forest, dan Decision Tree lalu diproses untuk mendapatkan hasil klasifikasi terbaik diantara algoritma XGBoost, Random Forest, dan Decision Tree. Pemodelan dapat dilihat pada gambar 12.

```
models = [
    ('DT', DecisionTreeClassifier()),
    ('RF', RandomForestClassifier()),
    ('XGBM', xgb.XGBClassifier()),
]
```

Gambar 12. *Modeling*

D. Evaluation

Visualisasi confusion matrix untuk prediksi terhadap data testing dengan model Decision Tree didapatkan 1292 untuk true positive (TP), kemudian 46 untuk data false positive (FP), 215 untuk data false negative (FN), dan 257 untuk data true negative (TN). Untuk model Random Forest didapatkan 1302 untuk true positive (TP), kemudian 36 untuk data false positive (FP), 232 untuk data false negative (FN), dan 240 untuk data true negative (TN). Untuk model XGBoost didapatkan 1296 untuk true positive (TP), kemudian 46 untuk data false positive (FP), 208 untuk data false negative (FN), dan 264 untuk data true negative (TN).



Gambar 13. Confusion Matrix Decision Tree, Random Forest, & XGBoost

Berdasarkan pengujian dataset stunting menggunakan 3 model algoritma yaitu Decision Tree, Random Forest, dan XGBoost, diperoleh hasil penelitian eksperimen yang dapat disimpulkan ke dalam tabel 2.

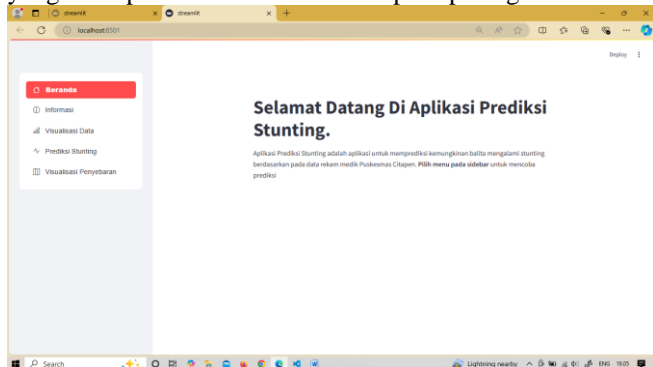
Tabel 2. Perbandingan Model

Rasio	70:30		
Perbandingan	Decision Tree	Random Forest	XGBoost
Akurasi	85.58%	84.86%	85.97%

Dari hasil confusion matrix dan accuracy XGBoost dipilih untuk digunakan ke tahap deployment dikarenakan lebih unggul dibandingkan dengan Decision Tree dan Random Forest pada splitting data 70:30 memiliki akurasi 85.97%.

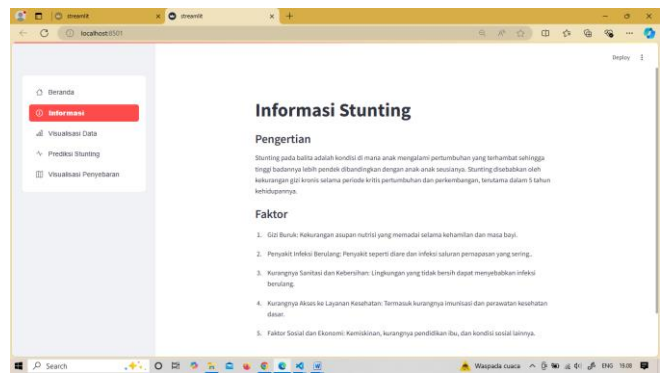
E. Deployment

Setelah model XGBoost dan pembagian data dengan skenario 70% data latih dan 30% data uji terpilih sebagai model terbaik, selanjutnya dibuatkan aplikasi website dengan menggunakan framework Streamlit dengan teks editor Visual Studio Code agar pemodelan prediksi penyakit diabetes dapat diakses lebih mudah. Dalam aplikasi ini terdapat halaman beranda yang merujuk pada halaman awal yang ditampilkan di browser web seperti pada gambar 14.



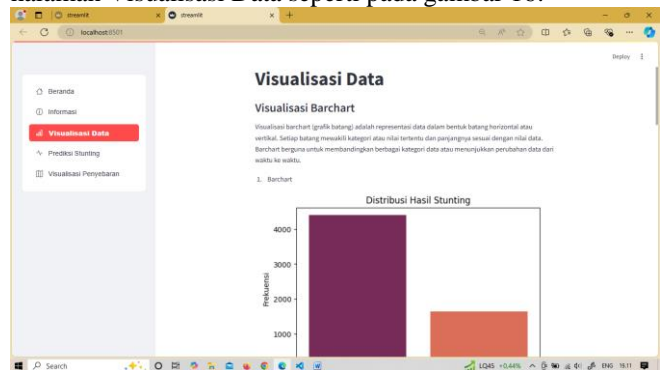
Gambar 14. Tampilan Halaman Beranda

Pada halaman beranda terdapat informasi mengenai penjelasan aplikasi website yang telah dibuat dan pilihan menu yang berada pada sidebar. Pada pilihan halaman terdapat 4 halaman selain beranda, yaitu halaman Informasi, halaman Visualisasi Data, halaman Prediksi *Stunting*, dan halaman Visualisasi Penyebaran. Tampilan halaman informasi seperti pada gambar 15.



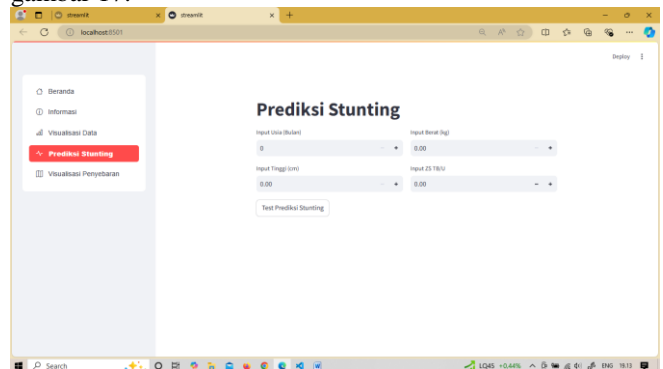
Gambar 15. Halaman Informasi

Halaman Informasi menjelaskan pengertian *stunting* dan faktor-faktor apa saja yang mempengaruhi kemungkinan balita terkena *stunting*. Kemudian tampilan halaman Visualisasi Data seperti pada gambar 16.



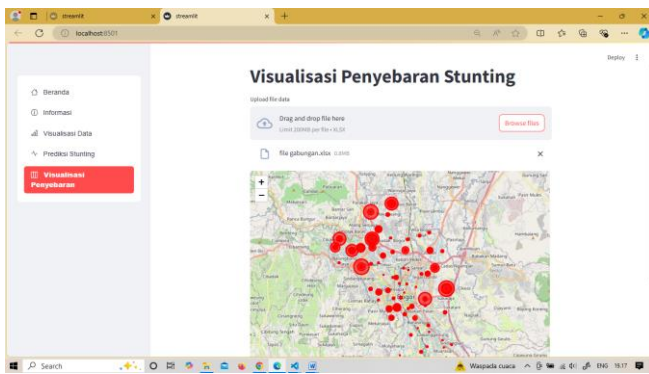
Gambar 16. Halaman Visualisasi Data

Pada halaman ini terdapat visualisasi data dari model yang telah dibuat sebelumnya yaitu visualisasi *barchart*, *correlations* features, dan *categorical plot*. Selanjutnya tampilan halaman Prediksi *Stunting* seperti pada gambar 17.



Gambar 17. Halaman Prediksi Stunting

Dapat dilihat pada gambar 17 halaman Prediksi berisi form-form yang dapat diisi untuk melakukan prediksi *stunting*, jika semua form telah diisi sesuai data dan mengklik tombol prediksi, maka akan muncul diagnosis balita terkena *severe stunting* (Sangat Pendek) dan *Stunting* (Pendek). Dan tampilan akhir yaitu tampilan visualisasi penyebaran seperti gambar 18.



Gambar 18. Halaman Visualisasi Penyebaran

Halaman visualisasi penyebaran merupakan halaman yang menampilkan sebaran zona daerah balita yang terkena diagnosis stunting. Tampilan zona daerah yang berwarna merah menandakan bahwa zona daerah tersebut merupakan daerah yang terkena *stunting*.

V. KESIMPULAN

Pada Featured Correlations dapat diketahui, bahwa variabel yang sangat mempengaruhi yaitu "Usia Saat Ukur (Bulan)", "Berat" dan "Tinggi" yaitu memiliki nilai hasil korelasinya 0.84, 0.89 dan 0.94, adapun nilai korelasi yang buruk terdapat pada variabel "BB/TB" dan "ZS BB/TB", yaitu memiliki nilai -0.83, yang menandakan bahwa variabel tersebut sangat tidak berpengaruh terhadap hasil stunting.

Klasifikasi dilakukan menggunakan algoritma Decision Tree, Random Forest dan XGBoost. Tahap pertama dilakukan yaitu tahap cleaning dan transformation data, setelah itu dilakukan pembagian data latih dan data uji yaitu 70:30.

Model pada penelitian ini dievaluasi dengan confusion matrix. Sehingga diketahui bahwa algoritma Decision Tree yaitu 85.58%, Random Forest yaitu 84.86%, dan XGBoost yaitu 85.97%. Komparasi tersebut menunjukkan bahwa algoritma XGBoost menghasilkan nilai klasifikasi yang lebih akurat dibanding algoritma Decision Tree dan Random Forest yang selanjutnya dipilih untuk dibuatkan sistem aplikasi web menggunakan framework Streamlit. Berdasarkan serangkaian proses yang telah dilakukan dalam penelitian ini, maka dapat disimpulkan bahwa penelitian ini berhasil membangun sebuah model terbaik dan aplikasi web untuk prediksi stunting menggunakan algoritma XGBoost.

Pada aplikasi web memakai metode XGBoost yang dimana merupakan metode terbaik dibandingkan Decision Tree dan Random Forest untuk melakukan prediksi. Semua halaman berfungsi dengan sesuai yang diharapkan. Halaman Prediksi Stunting sudah diuji coba dengan nilai sesuai dengan data yang dimiliki. Dan untuk halaman visualisasi penyebaran berfungsi dengan sesuai dimana daerah yang terkena zona merah merupakan daerah yang terkena *stunting*.

REFERENSI

- [1] S. Lonang, "Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination," no. March, 2022, doi: 10.30865/mib.v5i1.2293.
- [2] M. R. Nugroho, R. N. Sasongko, and M. Kristiawan, "Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia," *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, vol. 5, no. 2, Mar. 2021, doi: 10.31004/obsesi.v5i2.1169.
- [3] R. Fitri, N. Huljannah, and T. N. Rochmah, "PROGRAM PENCEGAHAN STUNTING DI INDONESIA :," vol. 17, no. 3, pp. 281–292, 2022.
- [4] J. C. Servizio, E. Sormin, and C. Siagian, "Pelatihan Pengukuran Antropometri dan Edukasi Gizi Seimbang sebagai Upaya Revitalisasi Posyandu dalam Rangka Menurunkan Angka Stunting di Kelurahan Cawang/Jakarta Timur," 2022.
- [5] 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.
- [6] M. Erkamim, S. Suswadi, M. Z. Subarkah, and E. Widarti, "Komparasi Algoritme Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 127–134, Oct. 2023, doi: 10.21456/vol13iss2pp127-134.
- [7] S. A. Hemo and Md. I. Rayhan, "Classification tree and random forest model to predict under-five malnutrition in Bangladesh," *Biom Biostat Int J*, vol. 10, no. 3, pp. 116–123, Sep. 2021, doi: 10.15406/bbij.2021.10.00337.
- [8] J. R. Khan, J. H. Tomal, and E. Raheem, "Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh," *Inform Health Soc Care*, vol. 46, no. 4, pp. 425–442, 2021, doi: 10.1080/17538157.2021.1904938.
- [9] A. Nugroho, H. L. Hendric, S. Warnars, F. L. Gaol, and T. Matsuo, "Trend of Stunting Weight for Infants and Toddlers Using Decision Tree," 2022.
- [10] M. S. Haris, M. Anshori, and A. N. Khudori, "PREDICTION OF STUNTING PREVALENCE IN EAST JAVA PROVINCE WITH RANDOM FOREST ALGORITHM PREDIKSI NILAI PREVALENSI STUNTING DI PROVINSI JAWA TIMUR MENGGUNAKAN ALGORITMA RANDOM FOREST," vol. 4, no. 1, pp. 11–13, 2023, doi: 10.20884/1.jutif.2023.4.1.614.
- [11] J. Mkungudza, H. S. Twabi, and S. O. M. Manda, "Development of a diagnostic predictive model for determining child stunting in Malawi: a comparative analysis of variable selection approaches," *BMC Med Res Methodol*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12874-024-02283-6.
- [12] H. Shen, H. Zhao, and Y. Jiang, "Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea," *Children*, vol. 10, no. 10, Oct. 2023, doi: 10.3390/children10101638.

- [13] O. N. Chilyabanyama *et al.*, "Performance of Machine Learning Classifiers in Classifying Stunting among Under-Five Children in Zambia," *Children*, vol. 9, no. 7, Jul. 2022, doi: 10.3390/children9071082.
- [14] F. Syarifudin, "KLASIFIKASI ARTIKEL-ARTIKEL JURNAL PUSTAKALOKA BERDASARKAN SKEMA JITA," *Fihris: Jurnal Ilmu Perpustakaan dan Informasi*, vol. 17, no. 1, p. 20, Sep. 2022, doi: 10.14421/fhrs.2022.171.20-37.
- [15] V. Tumbelaka, J. Kindangen, and J. Rengkung, "MORFOLOGI KAWASAN PERMUKIMAN AKIBAT KEBERADAAN KAWASAN KAMPUS UNIVERSITAS SAM RATULANGI DI KELURAHAN BAHU DAN KLEAK," *Jurnal Spasial*, vol. 6, no. 1, 2019.
- [16] Y. Rombe, "PENGUNAAN METODE XGBOOST UNTUK KLASIFIKASI STATUS OBESITAS DI INDONESIA," Universitas Hasanuddin, Makassar, 2021.
- [17] A. Ariyandi, S. Lestanti, and S. Kirom, "KLASIFIKASI BALITA STUNTING MENGGUNAKAN RANDOM FOREST CLASSIFIER DI KABUPATEN BLITAR," 2023.
- [18] Robianto, S. Sitorus, and U. Ristian, "PENERAPAN METODE DECISION TREE UNTUK MENGLASIFIKASIKAN MUTU BUAH JERUK BERDASARKAN FITUR WARNA DAN UKURAN," 2021.
- [19] L. Qadrini, A. Seppewali, and A. Aina, "DECISION TREE DAN ADABOOST PADA KLASIFIKASI PENERIMA PROGRAM BANTUAN SOSIAL," vol. 2, no. 7, 2021.