



**KLASIFIKASI BALITA STUNTING BERDASARKAN  
KARAKTERISTIK BALITA DAN KELUARGA  
MENGUNAKAN *EXTREME GRADIENT BOOSTED TREES*  
*ALGORITHM*.**

**SKRIPSI**

Oleh

**Mohammad Nasrul Azis**

**182410101066**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS JEMBER**

**2023**



**KLASIFIKASI BALITA STUNTING BERDASARKAN  
KARAKTERISTIK BALITA DAN KELUARGA  
MENGUNAKAN *EXTREME GRADIENT BOOSTED TREES*  
*ALGORITHM*.**

diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana (S1)  
pada Program Studi Sistem Informasi Universitas Jember

**SKRIPSI**

**Oleh**

**Mohammad Nasrul Azis**

**182410101066**

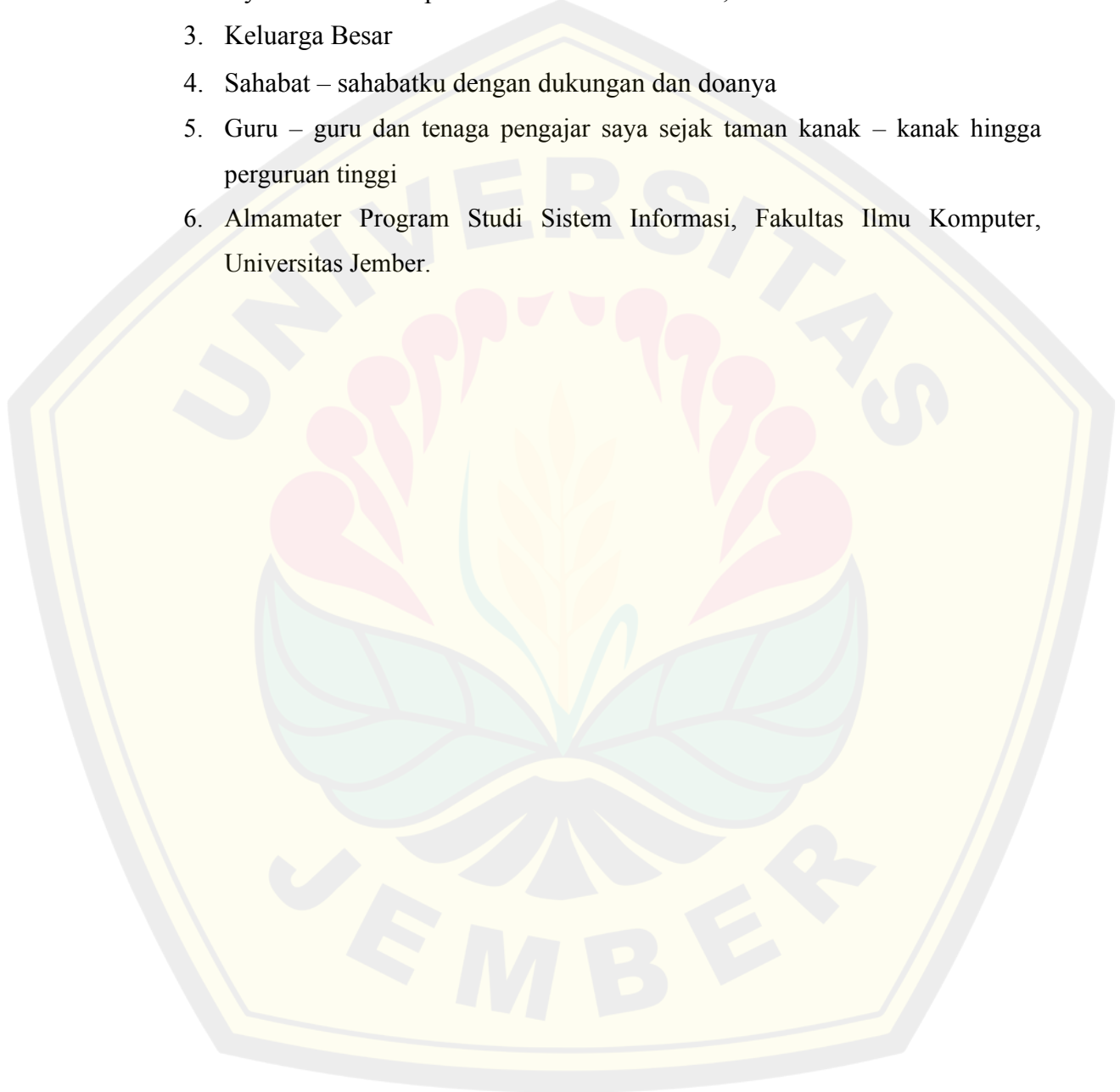
**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS JEMBER**

**2023**

**PERSEMBAHAN**

Skripsi ini saya persembahkan untuk:

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan dalam mengerjakan skripsi;
2. Ayahanda Hadi Jupri dan Ibunda Dewi Insani;
3. Keluarga Besar
4. Sahabat – sahabatku dengan dukungan dan doanya
5. Guru – guru dan tenaga pengajar saya sejak taman kanak – kanak hingga perguruan tinggi
6. Almamater Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Jember.



**MOTTO**

“For you is your religion, and for me is mine”

QS. Al-Kafirun 109:6



**PERNYATAAN ORISINALITAS**

Saya yang bertanda tangan di bawah ini:

Nama : Mohammad Nasrul Azis

NIM : 182410101066

Menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “KLASIFIKASI BALITA STUNTING BERDASARKAN KARAKTERISTIK BALITA DAN KELUARGA MENGGUNAKAN EXTREME GRADIENT BOOSTED TREES ALGORITHM”, adalah benar – benar hasil karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya, belum pernah diajukan pada instansi manapun, dan bukti karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika dikemudian hari pernyataan ini tidak benar.

Jember 17 Oktober 2023

Yang menyatakan,



Mohammad Nasrul Azis

182410101066

**HALAMAN PERSETUJUAN**

Skripsi berjudul “Klasifikasi Balita Stunting Berdasarkan Karakteristik Balita dan Keluarga Menggunakan Extreme Gradient Boosted Trees Algorithm” telah diuji dan disetujui oleh Fakultas Ilmu Komputer Universitas Jember pada :

Hari : Selasa

Tanggal : 19 September 2023

Tempat : Fakultas Ilmu Komputer Universitas Jember

Pembimbing

Tanda Tangan

1. Pembimbing Utama

Nama : Nelly Oktavia Adiwijaya, S.Si, MT

NIP : 198410242009122008



(.....)

2. Pembimbing Pendamping

Nama : Dr. Dwiretno Istiyadi S, ST., M.Kom

NIP : 197803302003121003



(.....)

Penguji

1. Penguji Utama

Nama : Nova El Maidah, S.Si.,M.Cs

NIP : 198411012015042001



(.....)

2. Penguji Pendamping

Nama : Gama Wisnu Fajarianto, S.Kom., M.Kom

NIP : 760015717



(.....)

**ABTRAK**

*Despite various efforts to address the challenges of toddler stunting, the prevalence of stunting in Tuban Regency remains high. This is not only caused by simple nutritional issues but also involves other crucial aspects such as the lack of access to balanced nutrition, inadequate access to healthcare services, and unresolved sanitation problems. Furthermore, socio-economic factors also play a significant role in determining the prevalence of stunting. In an effort to understand and tackle this issue, this research adopts a machine learning method with a primary focus on the application of the Extreme Gradient Boosted Trees (XGBoost) algorithm. This method is chosen for its ability to handle complex prediction issues and process large data, which can help identify patterns or important factors contributing to the occurrence of stunting in toddlers more efficiently and accurately. The research findings indicate that in addition to traditional factors such as the height/length of toddlers and age, environmental factors such as inadequate sanitation conditions, and socio-economic factors such as advanced maternal age, as well as exclusive breastfeeding practices, also play an essential role in determining the occurrence of stunting. In the model testing, it was found that the generated classification model could identify stunted toddlers with high accuracy, reaching 95.9%, with a precision of 94.7%, recall of 98.1%, and an F1-Score of 96.4%. These results demonstrate the strong potential of implementing machine learning methods using the XGBoost algorithm to support the early identification of stunting cases, providing a solid foundation for more effective health intervention efforts.*

*Keywords : Stunting; Machine Learning (ML); XGBoost; Hyperparameter Tuning*

**RINGKASAN**

**Klasifikasi Balita Stunting Berdasarkan Karakteristik Balita dan Keluarga Menggunakan Extreme Gradient Boosted Trees Algorithm;** Mohammad Nasrul Azis, 182410101066; 2023; 37 Halaman; Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Jember.

Meskipun berbagai upaya telah dilakukan untuk menangani tantangan balita stunting, angka prevalensi stunting di Kabupaten Tuban masih tinggi. Hal ini tidak hanya disebabkan oleh masalah gizi yang sederhana, melainkan melibatkan aspek-aspek penting lainnya seperti kurangnya akses terhadap nutrisi yang seimbang, kurangnya akses terhadap layanan kesehatan yang memadai, serta permasalahan sanitasi yang belum teratasi secara menyeluruh. Ditambah lagi, faktor sosial-ekonomi juga memainkan peran penting dalam menentukan prevalensi stunting. Dalam usaha untuk memahami dan menangani permasalahan ini, penelitian ini mengadopsi metode machine learning yang menjadi fokus utamanya adalah penerapan algoritma Extreme Gradient Boosted Trees (XGBoost). Metode ini dipilih karena kemampuannya dalam mengatasi masalah prediksi yang kompleks dan dalam mengolah data yang besar, yang dapat membantu dalam mengidentifikasi pola atau faktor-faktor penting yang berkontribusi terhadap kejadian stunting pada balita secara lebih efisien dan akurat.

Hasil penelitian menunjukkan bahwa selain faktor-faktor tradisional seperti panjang/tinggi badan balita dan usia, faktor lingkungan seperti kondisi sanitasi yang kurang layak, dan faktor sosial-ekonomi seperti usia ibu yang terlalu tua, serta praktik pemberian ASI eksklusif juga memiliki peran penting dalam menentukan kejadian stunting. Dalam pengujian model, ditemukan bahwa model klasifikasi yang dihasilkan mampu mengidentifikasi balita stunting dengan akurasi yang tinggi, mencapai 95.9%, dengan presisi sebesar 94.7%, recall sebesar 98.1%, dan F1-Score sebesar 96.4%. Hasil ini menunjukkan potensi yang kuat dari implementasi metode machine learning menggunakan algoritma XGBoost, dalam mendukung identifikasi dini kasus stunting, yang dapat memberikan landasan yang kokoh bagi upaya intervensi kesehatan yang lebih efektif.



## PRAKATA

Puji syukur kehadirat Allah SWT atasan limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “KLASIFIKASI BALITA STUNTING BERDASARKAN KARAKTERISTIK BALITA DAN KELUARGA MENGGUNAKAN EXTREME GRADIENT BOOSTED TREES ALGORITHM”. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan Pendidikan Strata Satu (S1) pada Program Studi Sistem Informasi Universitas Jember

Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

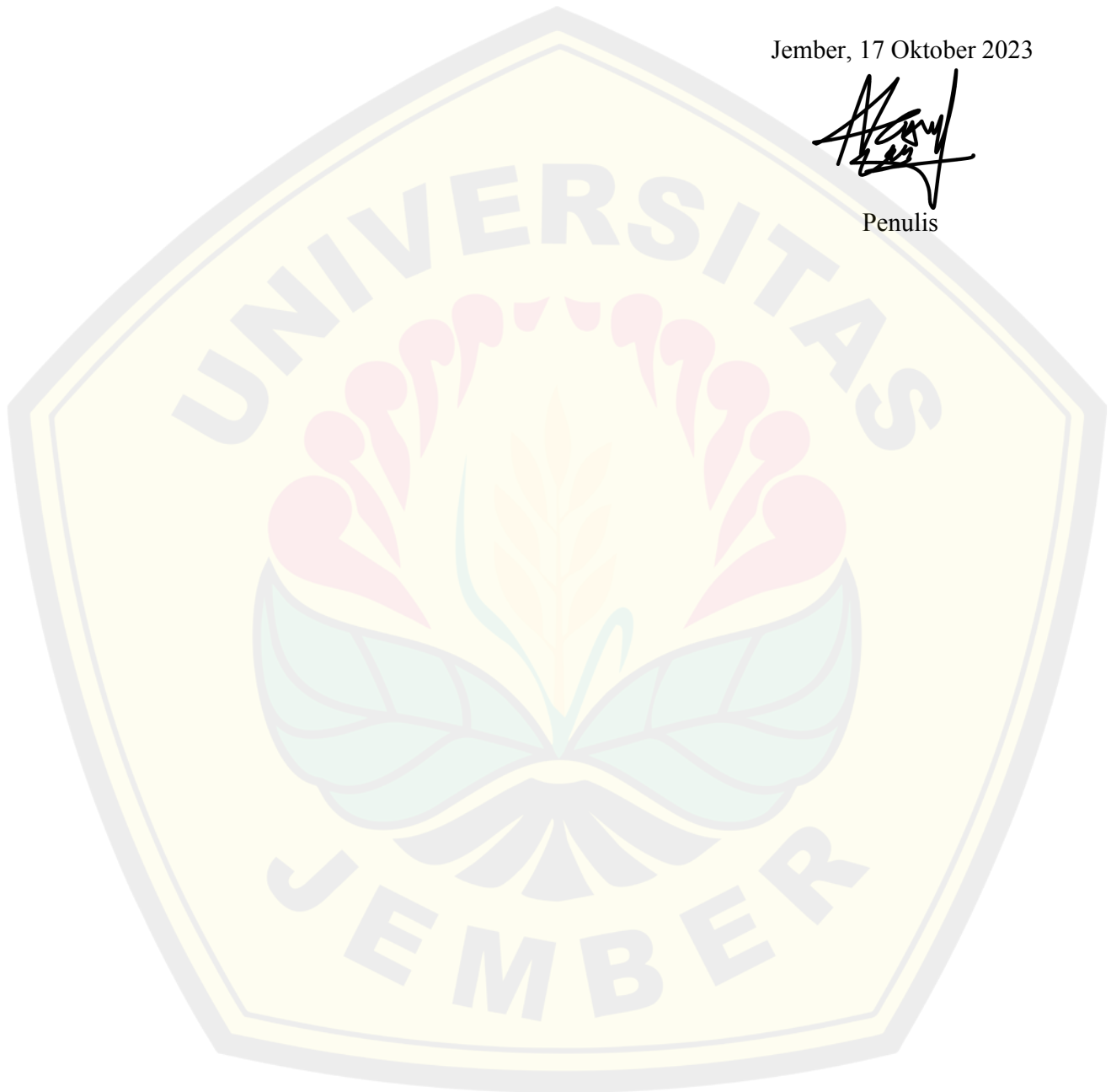
1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya untuk mempermudah dan melancarkan proses pengerjaan skripsi;
2. Oktalia Juwita selaku Dosen Pembimbing Akademik yang telah membimbing selama penulis menjadi mahasiswa;
3. Nelly Oktavia A, S.Si, MT selaku Dosen Pembimbing Utama dan Dr. Dwiretno Istiyadi S,ST.,M.Kom selaku Dosen Pembimbing Pendamping yang telah meluangkan waktu, pikiran dan perhatian dalam penulisan skripsi;
4. Seluruh Bapak dan Ibu dosen beserta staff karyawan di Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Jember;
5. Guru – guru dan tenaga pengajar Pendidikan formal maupun informal sejak taman kanak – kanak hingga perguruan tinggi;
6. Ayahanda Hadi Jupri dan Ibunda Dewi Insani yang selalu mendukung, mendoakan, memberi kekuatan dan menjadi tujuan hidup terbesar;
7. Maftukhatul Qomariyah Virati sebagai mentor yang telah membimbing, membantu dan mendampingi selama proses pengerjaan skripsi;
8. Sahabat terbaik Virli Galuh, Mei Tintri, Rosidah, Ardana Yuli dan Ahsan Hilmi yang selalu menemani, membantu dan memberi semangat;
9. Keluarga Besar Angkatan 2018 Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Jember (DELPHI);
10. Keluarga Besar UKM Olahraga Fakultas Ilmu Komputer (UKM-O MACO);

Dengan harapan bahwa penelitian ini nantinya terus berlanjut dan berkembang kelak, penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat bagi semua pihak.

Jember, 17 Oktober 2023



Penulis



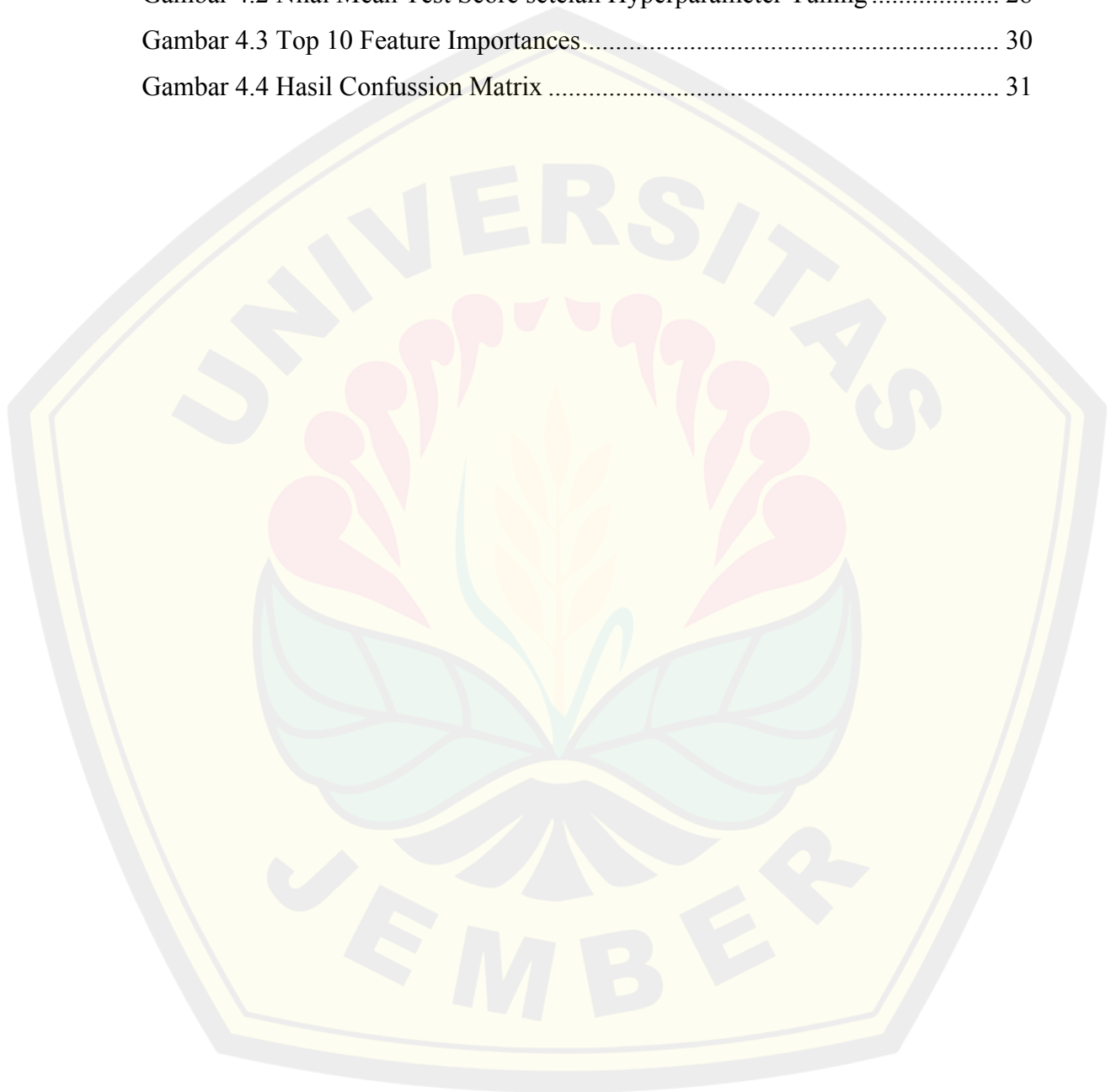
**DAFTAR ISI**

PERSEMBAHAN .....	ii
MOTTO .....	iii
PERNYATAAN ORISINALITAS .....	iv
HALAMAN PERSETUJUAN .....	v
ABTRAK .....	vi
RINGKASAN .....	vii
PRAKATA .....	viii
DAFTAR ISI .....	x
DAFTAR GAMBAR .....	xii
DAFTAR TABEL .....	xiii
BAB 1. PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan Penelitian .....	3
1.4. Manfaat Penelitian .....	4
1.5. Batasan Masalah .....	4
BAB 2. TINJAUAN PUSTAKA .....	5
2.1. Penelitian Terdahulu .....	5
2.2. Stunting .....	7
2.3. Pendataan Keluarga .....	8
2.4. Posyandu .....	9
2.5. Algoritma Extreme Gradient Boosted Trees (XGBoost) .....	9
2.6. Confusion Matrix .....	11
BAB 3. Metodologi Penelitian .....	13
3.1. Jenis Penelitian .....	13
3.2. Waktu dan Tempat Penelitian .....	13
3.3. Tahapan Penelitian .....	13
3.3.1. Perumusan Masalah .....	13
3.3.2. Studi Literatur .....	14

3.3.3.	Pengumpulan Data .....	14
3.3.4.	Perancangan dan Implementasi Model .....	16
3.3.5.	Penyusunan Laporan .....	18
BAB 4.	HASIL DAN PEMBAHASAN .....	19
4.1.	Pengumpulan Data.....	19
4.2.	Preprocessing Data .....	20
4.2.1.	Penggabungan Data Balita dan Keluarga.....	20
4.2.2.	Pelabelan Data Stunting .....	21
4.2.3.	Pemilihan Variabel yang Relevan.....	23
4.2.4.	Pembersihan Data.....	24
4.2.5.	Transformasi Data.....	24
4.3.	Pembagian Data.....	25
4.4.	Konfigurasi Hyperparameter Tuning .....	26
4.4.1.	Metode Hyperparameter Tuning.....	26
4.4.2.	Pemilihan Hyperparameter.....	26
4.4.3.	Rancangan Grid Search Tuning Hyperparameter .....	26
4.4.4.	Cross Validation.....	27
4.5.	Hasil Klasifikasi XGBoost.....	28
4.5.1.	Performa Klasifikasi.....	28
4.5.2.	Atribut paling berpengaruh .....	30
4.6.	Evaluasi (Confusion Matrix).....	31
BAB 5.	PENUTUP .....	33
5.1.	Kesimpulan.....	33
5.2.	Saran.....	34
DAFTAR PUSTAKA	.....	35
LAMPIRAN	.....	38

**DAFTAR GAMBAR**

Gambar 3.1. Tahapan Penelitian .....	13
Gambar 3.2. Implementasi dan perancangan model .....	16
Gambar 4.1 Grid Search untuk tuning Hyperparameter .....	27
Gambar 4.2 Nilai Mean Test Score setelah Hyperparameter Tuning .....	28
Gambar 4.3 Top 10 Feature Importances .....	30
Gambar 4.4 Hasil Confussion Matrix .....	31



**DAFTAR TABEL**

Tabel 2.1. Hyperparameter Tuning XGBoost .....	10
Tabel 2.2. Confusion Matrix .....	11
Tabel 3.1. Variabel independent dataset .....	14
Tabel 3.2. Tabel dependent dataset .....	15
Tabel 3.3. Indikator pada hyperparameter Tuning .....	17
Tabel 4.1 Standar Panjang/Tinggi Badan Menurut Umur PB/U Anak Laki-Laki	21
Tabel 4.2 Standar Panjang/Tinggi Badan Menurut Umur PB/U Anak Perempuan	22
.....	22
Tabel 4.3 Dataset setelah pelabelan .....	23
Tabel 4.4 Hasil GridSearchCV Split 80:20 .....	29
Tabel 4.5 Kombinasi Hyperparameter Terbaik .....	29
Tabel 4.6 Tabel Perbandingan Confussion Matrix .....	32

## BAB 1. PENDAHULUAN

### 1.1. Latar Belakang

Angka prevalensi balita stunting di Indonesia pada tahun 2020 menduduki peringkat ke-2 se Asia tenggara mencapai angka 31.8% (Asian Development Bank, 2021). Angka yang masih tinggi tersebut menandakan bahwa masih ada masalah kesehatan khususnya stunting di Indonesia yang tidak hanya mempengaruhi kesehatan tetapi juga mempengaruhi produktivitas ekonomi nasional (Wardani et al., 2021). Stunting adalah kegagalan mencapai potensi pertumbuhan linier yang ditunjukkan dengan *height-for-age Z score* (HAZ)  $< -2$  SD sesuai dengan referensi yang berlaku akibat status kesehatan dan/atau nutrisi yang tidak optimal (Endy P. Prawirohartono, 2021).

Pada tahun 2021 menurut Studi Status Gizi Indonesia (SSGI) prevalensi balita stunting di Indonesia menjadi 24.4% (Kementerian Kesehatan Republik Indonesia, 2021). Meski angka ini sudah turun, namun presiden Jokowi menargetkan pada 2024 angka prevalensi ini menjadi 14% sehingga masih jauh dari target (BPMI Sekretariat Presiden, 2022). Di Kabupaten Tuban sendiri angka prevalensi stunting masih di atas angka prevalensi nasional. Pada tahun 2019 angka prevalensi stunting di kabupaten Tuban mencapai 27.15% (Kementerian Kesehatan Republik Indonesia, 2019). Pada tahun 2021, angka prevalensi stunting di kabupaten Tuban mencapai 25.1% (Kementerian Kesehatan Republik Indonesia, 2021). Meskipun angka prevalensi stunting di kabupaten Tuban sudah menurun tiap tahunnya, masalah ini masih menjadi pekerjaan rumah untuk pemerintah daerah dan dinas terkait (Novia Puspitasari & Adi, 2020).

Pemerintah nasional telah mengeluarkan kebijakan dan regulasi untuk menangani angka prevalensi stunting yang diharapkan bisa berkontribusi untuk menurunkan angka prevalensi stunting. Selain program nasional, program intervensi gizi spesifik yang dilakukan oleh Kementerian Kesehatan (Kemenkes) melalui Pos Pelayanan Terpadu (Posyandu) dan Pusat Kesehatan Masyarakat (Puskesmas) melalui gerakan 1000 Hari Pertama Kehidupan (HPK) (Sekretariat

Wakil Presiden Republik Indonesia, 2017). Dalam upaya untuk mengurangi angka prevalensi stunting, intervensi dapat dilakukan dengan prediksi yang akurat sejak dini. Saat ini *machine learning* menjadi metode populer dalam memprediksi berbagai jenis data medis dan biomedis karena kemampuannya yang membuat prediksi yang akurat dan efektif (Rahman et al., 2021). Beberapa penelitian sudah dilakukan untuk mengembangkan sistem untuk prediksi stunting sejak dini menggunakan berbagai metode *machine learning* seperti *Support Vector Machine*, *Random Forest*, *Linear Regression*, *Gradient Boosted Tree*, *Neural Network*, Teknik *Hybrid* dan metode-metode lainnya (Bitew et al., 2022; Khan et al., 2021; Talukder et al. 2020).

Ashis Talukder M.Sc dan Benojir Ahammed M.Sc melakukan penelitian menggunakan dataset *Bangladesh Demographic and Health Survey* (BDHS) mengemukakan dari lima algoritma yang digunakan, hasil terbaik didapatkan pada algoritma *Random Forest* (RF) dengan akurasi 68.51%, sensitivitas 94.66% dan derajat keandalan 69.76% (Talukder et al. 2020). Penelitian lain menggunakan dataset *Ethiopian Demographic and Health Survey* yang dilakukan oleh Fikrewold H Bitew, Corey S Sparks dan Samuel H Nyarko pada tahun 2022 menemukan bahwa dari 5 algoritma populer yang digunakan yaitu *random forest* (RF), *neural network*, *extreme gradient boosting* (XGBoost), *k-nearest neighbours* (k-NN) dan *generalised linear model*, algoritma *XGBoost* menunjukkan hasil prediksi yang lebih baik daripada algoritma lainnya (Bitew et al., 2022). Penelitian lainnya menggunakan dataset *Bangladesh Demographic and Health Survey* (BDHS) yang dilakukan oleh Jahidur Rahman Khan, Javed H. Tomal dan Enayetur Raheem pada tahun 2021 menunjukkan metode *gradient boosting* menghasilkan kesalahan klasifikasi paling sedikit dalam prediksi stunting, diikuti oleh metode *random forest* (RF), *support vector machine* (SVM), *classification tree* dan *logistic regression* dengan seleksi *forward-stepwise* (Khan et al., 2021).

Berdasarkan pembahasan sebelumnya, penelitian ini akan melakukan klasifikasi pada balita stunting menggunakan metode *gradient boosted tree* (XGBoost) untuk prediksi stunting sehingga dapat dilakukan intervensi dini. Objek



penelitian di Kabupaten Tuban merupakan representasi dari wilayah dengan angka prevalensi stunting yang signifikan melebihi target nasional yang telah ditentukan. Dataset yang digunakan pada penelitian ini didapatkan dari pendataan keluarga 2021 (PK21) yang telah dilakukan oleh dinas kesehatan Kabupaten Tuban dan data posyandu yang dikumpulkan oleh Dinas Kesehatan Kabupaten Tuban. Dalam upaya peningkatan performa model beberapa hyper parameter dapat disesuaikan oleh analis antara lain : *n\_estimators*, *max\_depth*, *max\_delta\_step*, *subsample*, *colsample\_bytree*, *reg\_lambda*, *min\_child\_weight* dan *learning\_rate* (Fadhilurrahman, 2023). Model machine learning yang memiliki parameter terbaik yang akan digunakan sebagai dasar evaluasi model secara kuantitatif dalam prediksi balita stunting yang akurat.

### 1.2. Rumusan Masalah

1. Bagaimana Pengaruh Karakteristik Balita dan Keluarga Terhadap Klasifikasi Balita Stunting Menggunakan Algoritma Extreme Gradient Boosted Trees (XGBoost)?
2. Bagaimana Hasil Akurasi, Presisi, *Recall* dan *F1-score* Prediksi Balita Stunting Berdasarkan Hasil Klasifikasi Menggunakan Algoritma *Extreme Gradient Boosted Trees* (XGBoost) untuk Strategi Intervensi Kesehatan Keluarga?

### 1.3. Tujuan Penelitian

Dari rumusan masalah yang ada, maka tujuan penelitian yang ingin dicapai adalah :

1. Menganalisis Pengaruh Karakteristik Balita dan Keluarga Terhadap Klasifikasi Balita Stunting Menggunakan Algoritma Extreme Gradient Boosted Trees (XGBoost).
2. Memberikan Hasil Akurasi, Presisi, *Recall* dan *F1-Score* Prediksi Balita Stunting Berdasarkan Hasil Klasifikasi Balita Stunting Menggunakan Algoritma *Extreme Gradient Boosted Trees* (Xgboost) untuk Strategi Intervensi Kesehatan Keluarga.

#### 1.4. Manfaat Penelitian

Manfaat dari penelitian ini adalah :

##### 1. Bagi Akademis

Bagi Akademis diharapkan dapat membantu menyajikan informasi terkait judul penelitian kepada pembaca khususnya mahasiswa Fakultas Ilmu Komputer Universitas Jember.

##### 2. Bagi Peneliti

Bagi Peneliti diharapkan dapat mengembangkan dan memahami mengenai klasifikasi balita stunting menggunakan algoritma *Extreme Gradient Boosted Trees* (XGBoost)

##### 3. Bagi Objek Penelitian

Bagi Objek Penelitian diharapkan dapat mengetahui dan memahami mengenai pengaruh karakteristik balita dan keluarga pada prediksi balita stunting berdasarkan hasil klasifikasi balita stunting menggunakan algoritma *Extreme Gradient Boosted Trees* (XGBoost) untuk strategi intervensi kesehatan keluarga.

#### 1.5. Batasan Masalah

Batasan masalah yang dapat dijadikan sebagai acuan atau parameter dalam penelitian ini, antara lain:

1. Data didapatkan dari data Pendataan Keluarga 2021 yang diverifikasi ulang pada tahun 2022 oleh Dinas Kesehatan Kabupaten Tuban dan data kunjungan Posyandu yang ada di Dinas Kesehatan Kabupaten Tuban.
2. Penelitian ini menggunakan atribut untuk klasifikasi berdasarkan gabungan data kondisi keluarga dan data kondisi balita.
3. Penelitian ini hanya akan memberikan saran intervensi kesehatan keluarga berdasarkan model, tanpa evaluasi efektivitas intervensi secara langsung.

## BAB 2. TINJAUAN PUSTAKA

Tinjauan pustaka menjelaskan tentang kajian literatur atau referensi yang berkaitan dengan topik penelitian yang bersumber dari jurnal, prosiding, website, buku dan penelitian terdahulu.

### 2.1. Penelitian Terdahulu

Penelitian sebelumnya telah membahas beberapa metode *machine learning* yang digunakan dalam klasifikasi dan prediksi stunting. Penelitian terdahulu ini akan digunakan sebagai dasar untuk klasifikasi balita stunting khususnya menggunakan metode XGBoost yang digunakan dalam penelitian ini.

1. Penelitian dilakukan oleh Ashis Talukder M.Sc dan Benojir Ahammed M.Sc pada tahun 2020 dengan judul *Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh*. Penelitian ini menggunakan data pada tahun 2014 dari *Bangladesh Demographic and Health Survey* (BDHS). Atribut yang dianggap sebagai kemungkinan faktor resiko untuk malnutrisi anak di Bangladesh yaitu umur anak, pendidikan ibu, indeks kekayaan, tempat tinggal, BMI ibu dan jarak lahir. 5 metode *machine learning* digunakan pada penelitian ini, *linear dicriminant analysis* (LDA), *k-nearest neighbors* (k-NN), *suport vector machines* (SVM), *random forest* (RF) dan *logistic refression* (LR). Dari kelima algoritma di atas, hasil terbaik didapatkan pada algoritma RF, dengan akurasi 68.51%, sensitivitas 94.66%, dan derajat keandalan 69.76% (Talukder et al. 2020). Penelitian ini relevan dengan penelitian yang akan dilakukan karena mengeksplorasi objek yang serupa dan beberapa atribut yang sama, dengan menerapkan beberapa metode *machine learning*.
2. Penelitian dilakukan oleh Fikrewold H Bitew, Corey S Sparks dan Samuel H Nyarko pada tahun 2022 dengan judul *Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia*. Penelitian ini menggunakan data *Ethiopian Demographic and Health Survey* pada tahun 2016 yang merupakan data terbaru dan bagian dari survei demografik dan

kesehatan dunia yang dilakukan setiap 5 tahun. Jumlah sampel yang diambil adalah 9471 anak di bawah lima tahun. Atribut yang digunakan pada survei yang dianggap sebagai faktor risiko kekurangan nutrisi pada anak di Ethiopia adalah umur anak (dalam bulan), jenis kelamin anak, jarak kelahiran (dalam bulan), ukuran saat lahir, pendampingan ibu, bmi ibu, tempat tinggal, agama, edukasi ibu dan kelompok kekayaan. Hasil dari data tersebut adalah anak stunting, wasting dan underweight yang ditentukan oleh *Z-scores*. Berdasarkan WHO indikator kekurangan nutrisi anak bisa dilihat mengikuti standard berikut: *stunting: height-for-age* < -2 SD; *wasting: weight-for-height* < -2 SD dan *underweight: weight-for-age* < -2 SD. Penelitian ini menggunakan 5 Algoritma *machine learning* yang populer yaitu *extreme gradient boosting*, *k-nearest neighbours* (k-NN), *random forest*, *neural network* dan *generalised linear model*. Diantara 5 algoritma yang digunakan pada studi ini, algoritma *xgbTree* menunjukkan hasil prediksi yang lebih baik daripada algoritma lainnya (Bitew et al., 2022). Penelitian ini relevan dengan penelitian yang akan dilakukan karena mengeksplorasi objek yang serupa dan beberapa atribut yang sama, dengan menerapkan beberapa metode *machine learning*. Hasil terbaik diperoleh dengan menggunakan algoritma XGBTree.

3. Penelitian dilakukan oleh Jahidur Rahman Khan, Jabed 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*. Penelitian ini menggunakan data *Bangladesh Demographic and Health Survey* (BDHS) pada tahun 2014. Variabel yang digunakan dalam studi ini adalah, umur anak (dalam bulan), jenis kelamin anak, status menyusui anak, pemberian vitamin A dalam 6 bulan terakhir, anak ke berapa, diare pada anak 2 minggu terakhir, anak panas 2 minggu terakhir, umur kelahiran, umur suami, umur saat kelahiran pertama, edukasi kehamilan, media edukasi kehamilan, status pekerjaan ibu, bmi ibu, jarak kelahiran, jumlah keluarga, jumlah anak di bawah 5 tahun, jumlah anak, agama, kategori kekayaan, fasilitas toilet, fasilitas toilet bersama, tempat tinggal dan kabupaten. Metode *machine learning* digunakan dalam penelitian ini adalah

*neural networks* (NN), *support vector machines* (SVM), *random forest* (RF), *classification and regression trees* (CART) dan *gradient boosting* (GBOOST). Hasil dari penelitian ini menunjukkan pendidikan orang tua, nutrisi dan karakteristik rumah tangga adalah faktor yang penting untuk memprediksi stunting pada anak di Bangladesh. Dari beberapa model klasifikasi yang digunakan, *gradient boosting* menghasilkan kesalahan klasifikasi paling sedikit dalam memprediksi stunting, diikuti dengan *RF*, *SVM*, *classification tree* dan *logistic regression* dengan seleksi *forward-stepwise* (Khan et al., 2021). Penelitian ini relevan dengan penelitian yang akan dilakukan karena mengeksplorasi objek yang serupa dan beberapa atribut yang sama, dengan menerapkan beberapa metode machine learning. Hasil terbaik didapatkan menggunakan metode gradient boosting. Penelitian ini juga mengukur relevansi variabel yang digunakan terkait dengan kejadian stunting pada balita.

## 2.2. Stunting

Stunting pada anak usia di bawah lima tahun merupakan indikasi pertumbuhan linear yang tidak baik selama masa kritis 1000 hari pertama kehidupan (HPK) dan dikenali sebagai tinggi badan untuk usia kurang dari 2 standar deviasi dari median standar pertumbuhan anak yang ditentukan oleh World Health Organization (WHO) akibat status kesehatan dan/atau nutrisi yang tidak optimal (Beal et al., 2018; Endy P. Prawirohartono, 2021). Stunting disebabkan oleh dua hal, yaitu faktor langsung dan tidak langsung. Faktor langsung meliputi kekurangan asupan nutrisi dan adanya penyakit, khususnya penyakit infeksi. Sementara penyebab tidak langsung meliputi faktor ketahanan pangan keluarga, gaya hidup dan pola makan, kondisi lingkungan dan kesehatan serta akses terhadap pelayanan kesehatan (Trihono et al., 2015).

Dampak yang akan ditimbulkan stunting tidak hanya jangka pendek, namun juga akan dirasakan dalam jangka panjang. Stunting pada 2 tahun pertama kehidupan berpeluang untuk anak memiliki IQ di bawah 89 dan lebih rendah 4,57 kali dibandingkan IQ anak yang tidak stunting (Daracantika, 2021). Kerugian

negara di masa mendatang menjadi salah satu dampaknya karena anak pendek memiliki potensi tumbuh menjadi dewasa yang kurang pendidikan, miskin, kurang sehat dan rentan terkena penyakit tidak menular (obesitas, penyakit kardiovaskuler, dll) .

### 2.3. Pendataan Keluarga

Pendataan Keluarga 2021 (PK-21) adalah kegiatan pengumpulan data-data terkait pembangunan keluarga, kependudukan, keluarga berencana, dan anggota keluarga. Kegiatan ini dilakukan oleh BKKBN secara bersamaan setiap lima tahun sekali secara rutin. Program tersebut dilaksanakan sesuai dengan UU 52 tahun 2009 tentang Perkembangan Kependudukan dan Pembinaan Keluarga dan PP nomor 87 tentang Perkembangan Kependudukan, Pembinaan Keluarga, KB, dan Sistem Informasi Keluarga. (Ratna Nuraini, 2021). Pada tahun 2022 Badan Kependudukan dan Keluarga Berencana Nasional melakukan pemutakhiran PK-21 yang dilakukan pada bulan September 2022 hingga November 2022. Dari hasil pemutakhiran, terdapat 35.309.446 dari 68.487.139 data keluarga Indonesia yang berhasil di mutakhirkan (Fitang Budhi Adhitia, 2023). Data PK-21 yang digunakan pada penelitian ini adalah :

- a. **Kelayakan air minum**, sumber air minum utama yang digunakan oleh keluarga.
- b. **Kelayakan Jamban**, kepemilikan jamban pada keluarga masih menjadi masalah terutama di daerah pedesaan.
- c. **Istri terlalu muda**, umur istri yang dikategorikan terlalu muda yaitu  $< 20$  tahun.
- d. **Istri terlalu tua**, umur istri yang dikategorikan terlalu tua adalah  $> 35$  tahun.
- e. **Anak terlalu dekat**, jarak kelahiran anak yang disarankan adalah lebih dari 2 tahun, jika di bawah angka tersebut maka jarak kelahiran anak terlalu dekat.
- f. **Anak terlalu banyak**, jumlah anak yang disarankan tidak lebih dari 2 anak, jika lebih dari angka tersebut maka jumlah anak terlalu banyak.
- g. **Memiliki baduta**, kondisi keluarga pada saat ini memiliki bayi di bawah dua tahun.

**h. Memiliki balita**, kondisi keluarga pada saat ini memiliki bayi di bawah lima tahun.

#### 2.4. Posyandu

Posyandu adalah salah satu jenis UKBM yang dikelola oleh masyarakat, untuk masyarakat, dari masyarakat, dan bersama masyarakat, untuk memberikan pemberdayaan kepada masyarakat dan memberikan kemudahan kepada masyarakat dalam mendapatkan layanan kesehatan untuk ibu, bayi dan anak usia dini.. Data posyandu yang digunakan pada penelitian ini adalah :

- a. **Jenis kelamin**, jenis kelamin bayi yaitu laki-laki atau perempuan.
- b. **Berat lahir bayi**, berat bayi dihitung saat melahirkan dengan satuan gram (g)
- c. **Panjang lahir bayi**, panjang bayi dihitung pada saat melahirkan dengan satuan centimeter (cm)
- d. **Asfiksia**, penyakit ini adalah gagal nafas spontan pada bayi saat awal kelahiran.
- e. **Usia bayi**, usia bayi saat ini.
- f. **Umur bayi dalam kandungan**, umur bayi dalam kandungan dibagi 2, yaitu mature (37 – 42 minggu) dan premature (< 37 minggu).
- g. **Panjang bayi**, panjang bayi saat ini dihitung saat pendataan di posyandu
- h. **Berat bayi**, berat bayi saat ini dihitung saat pendataan di posyandu
- i. **Pemberian asi eksklusif**, pemberian asi eksklusif oleh ibu selama 6 bulan pertama kelahiran.

#### 2.5. Algoritma Extreme Gradient Boosted Trees (XGBoost)

XGBoost adalah salah satu contoh dari algoritma boosting. Istilah boosting merujuk pada kelas algoritma belajar yang menggabungkan berbagai "model basis" yang lebih sederhana untuk mendapatkan hasil belajar terbaik. Model-model basis tersebut masing-masing memiliki kemampuan prediktif yang rendah, namun ketika dipilih dengan cermat dengan boosting, dapat membentuk model yang relatif lebih akurat.

Pada algoritma XGBoost, model basis yang digunakan adalah pohon keputusan (decision tree). Pohon keputusan secara matematis menghasilkan pembagian variabel penjelas  $x \in \mathbb{R}^k$  kedalam  $T$  partisi, sebutlah  $R_1, \dots, R_T$  dan memberikan bobot kepada masing-masing partisi, sebutlah  $\omega_1, \dots, \omega_T$ . Dengan demikian, sebuah pohon keputusan  $\phi(x)$  dapat ditulis menggunakan persamaan (1).

$$\phi(x) = \sum_{t=1}^T \omega_t I(x \in R_t) \quad (1)$$

Dengan kata lain, XGBoost berusaha mempelajari fungsi  $p(x)$  sebagai ansambel aditif dari  $S$  buah pohon yang dapat dilihat pada persamaan (2):

$$p(x) = \theta_0 + \sum_{s=1}^S \theta_s \phi_s(x) \quad (2)$$

dimana  $\theta_0$  adalah bias model dan  $\theta_s$  adalah bobot untuk masing-masing pohon.

Dari data, fungsi log-loss dan bentuk ansambel aditif di atas, XGBoost akan mencari  $p(x)$  terbaik. Pencarian tersebut dilakukan dengan metode greedy berbasis gradien, dimana pohon-pohon  $\phi_1, \dots, \phi_s$  ditambahkan satu per satu, hingga mencapai fungsi log-loss minimum (Fadhlurrahman, 2023). Penjelasan hyperparameter dapat dilihat pada Tabel 2.1.

Tabel 2.1. Hyperparameter Tuning XGBoost

Nama hyperparameter	Notasi	Keterangan
n_estimators	$S$	Jumlah pohon dalam XGBoost
max_depth	$T$ atau $T_s$	Kedalaman pohon maksimum untuk model basis XGBoost.
max_delta_step	$\delta_s$	Nilai bobot maksimum untuk pohon yang diperbolehkan.
subsample	$n/N$	Rasio subsampel yang digunakan untuk melatih pohon.
colsample_bytree	$k/K$	Rasio variabel penjelas yang digunakan untuk melatih pohon.
reg_lambda	$\lambda$	Regularisasi L2 untuk bobot pohon.



min_child_weight	$w_s$	Jumlah bobot terkecil untuk pembuatan partisi pada pohon.
learning_rate	$\eta$	Laju pembelajaran untuk menghindari <i>overfitting</i> .

## 2.6. Confusion Matrix

Confusion *matrix* adalah sebuah tabel yang menunjukkan jumlah data uji yang telah diklasifikasi benar dan jumlah data uji yang telah diklasifikasi salah (Arifin, 2018). Metode ini digunakan untuk mengukur kinerja model menggunakan akurasi, presisi, recall dan nilai-F1. Bentuk dari confusion matrix dapat dilihat pada Tabel 2.2.

Tabel 2.2. Confusion Matrix

Nilai prediksi klasifikasi	Nilai sebenarnya	
	True	False
True	TP (True Positive) Correct Result	FP (False Positive) Unexpected Result
False	FN (False Negative) Missing Result	TN (True Negative) Correct Absence of Result

Akurasi adalah ukuran seberapa akurat model dalam memprediksi kelas. Akurasi juga dapat digunakan untuk kelas dengan distribusi dan jumlah kelas yang tidak seimbang. Akurasi dihitung dengan membagi rasio prediksi yang benar untuk kelas positif dan negatif terhadap keseluruhan data. Perhitungan akurasi dapat dilihat pada persamaan (3).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

*Precision* adalah rasio prediksi benar positif yang dibandingkan dengan seluruh hasil prediksi positif. *Precision* mengukur seberapa akurat model yang dibuat dari seluruh kelas positif hasil prediksi. Perhitungan *precision* dapat dilihat pada persamaan (4).

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

*Recall* adalah rasio prediksi benar positif yang dibandingkan dengan seluruh data benar positif. Perhitungan *recall* dapat dilihat pada persamaan (5).

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

*F1 Score* didapatkan dari rata - rata *precision* dan *recall* yang dibobotkan. *F1 Score* dibutuhkan supaya mendapatkan nilai yang seimbang antara nilai *precision* dan *recall*. Nilai *F1 Score* berkisar antara 0 dan 1, semakin nilai mendekati 1 kinerja model semakin baik, sedangkan nilai yang semakin mendekati 0 maka kinerja model semakin buruk. Perhitungan *F1 Score* dapat dilihat pada persamaan (6).

$$F1 \text{ score} = \frac{2(\text{recall} \times \text{precision})}{\text{recall} + \text{precision}} \quad (6)$$



### BAB 3. METODOLOGI PENELITIAN

#### 3.1. Jenis Penelitian

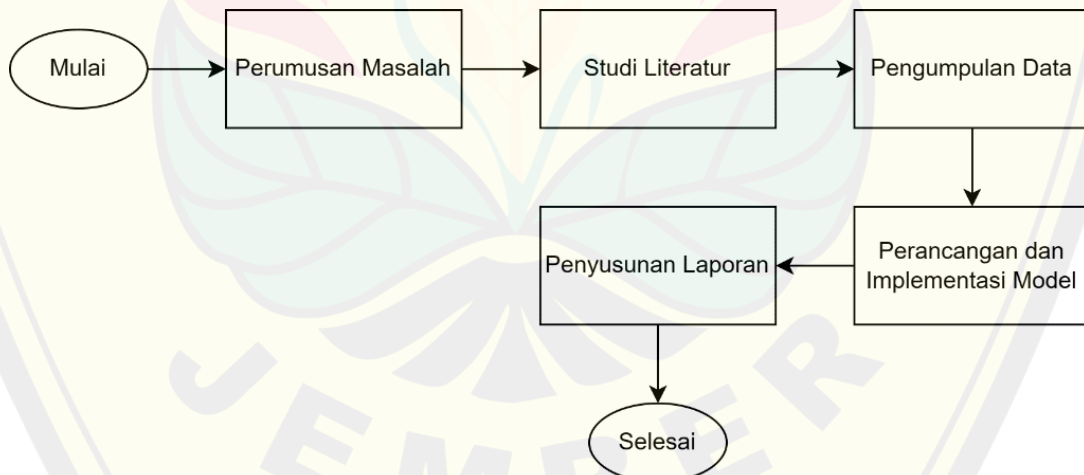
Pada penelitian ini jenis penelitian yang diterapkan adalah jenis penelitian terapan (applied research). Penelitian ini menggunakan metode Extreme Gradient Boosted Tree (XGBoost) dengan penyesuaian hyperparameter untuk mendapatkan hasil akurasi yang terbaik.

#### 3.2. Waktu dan Tempat Penelitian

Penelitian ini dilakukan di Fakultas Ilmu Komputer, Universitas Jember. Penelitian berlangsung selama 6 bulan berawal pada bulan November 2022 hingga bulan Mei 2023.

#### 3.3. Tahapan Penelitian

Dalam melakukan penelitian, penting untuk menentukan tahapan penelitian yang akan diikuti. Tahapan penelitian merupakan urutan bagaimana sebuah penelitian akan dilakukan. Urutan tahapan penelitian dapat dilihat pada Gambar 3.1.



Gambar 3.1. Tahapan Penelitian

##### 3.3.1. Perumusan Masalah

Pada tahap ini perumusan masalah dilakukan oleh peneliti dan mencari solusi masalah yang ada berdasarkan literatur, informasi dan data yang telah diperoleh.

### 3.3.2. Studi Literatur

Pada tahap ini, studi literatur yang berkaitan terhadap penelitian ini dilakukan oleh peneliti dengan mengumpulkan bahan literatur. Sumber yang didapat berasal dari laporan ilmiah, buku, jurnal, prosiding, website dan karya ilmiah dari penelitian sebelumnya.

### 3.3.3. Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data PK-21 yang telah diverifikasi pada tahun 2022 dan data posyandu dari dinas kesehatan yang berjumlah 1000 sampel di antaranya 348 sampel balita stunting dan 652 sampel balita tidak stunting. Dataset ini memiliki 17 variabel independen Tabel 3.1 dan 1 variabel dependen Tabel 3.2.

*Tabel 3.1. Variabel independent dataset*

Variabel	Keterangan	Value
Kelayakan air minum	Sumber air minum utama keluarga	1. Air kemasan/isi ulang 2. Ledeng/pam 3. Sumur bor 4. Sumur terlindung 5. Sumur tidak terlindung 6. Air permukaan (Sungai, Danau, dll) 7. Air Hujan 8. Lainnya
Kelayakan Jamban	Kepemilikan jamban pada keluarga	1. Ya, dengan septic tank 2. Ya, tanpa septic tank 3. Tidak, jamban umum/bersama 4. Lainnya
Istri terlalu muda	Umur istri < 20 tahun	1. Ya 2. Tidak
Istri terlalu tua	Umur istri > 35 tahun	1. Ya 2. Tidak
Anak terlalu dekat	Jarak kelahiran anak < 2 tahun	1. Ya 2. Tidak
Anak terlalu banyak	Jumlah anak > 2 orang	1. Ya 2. Tidak

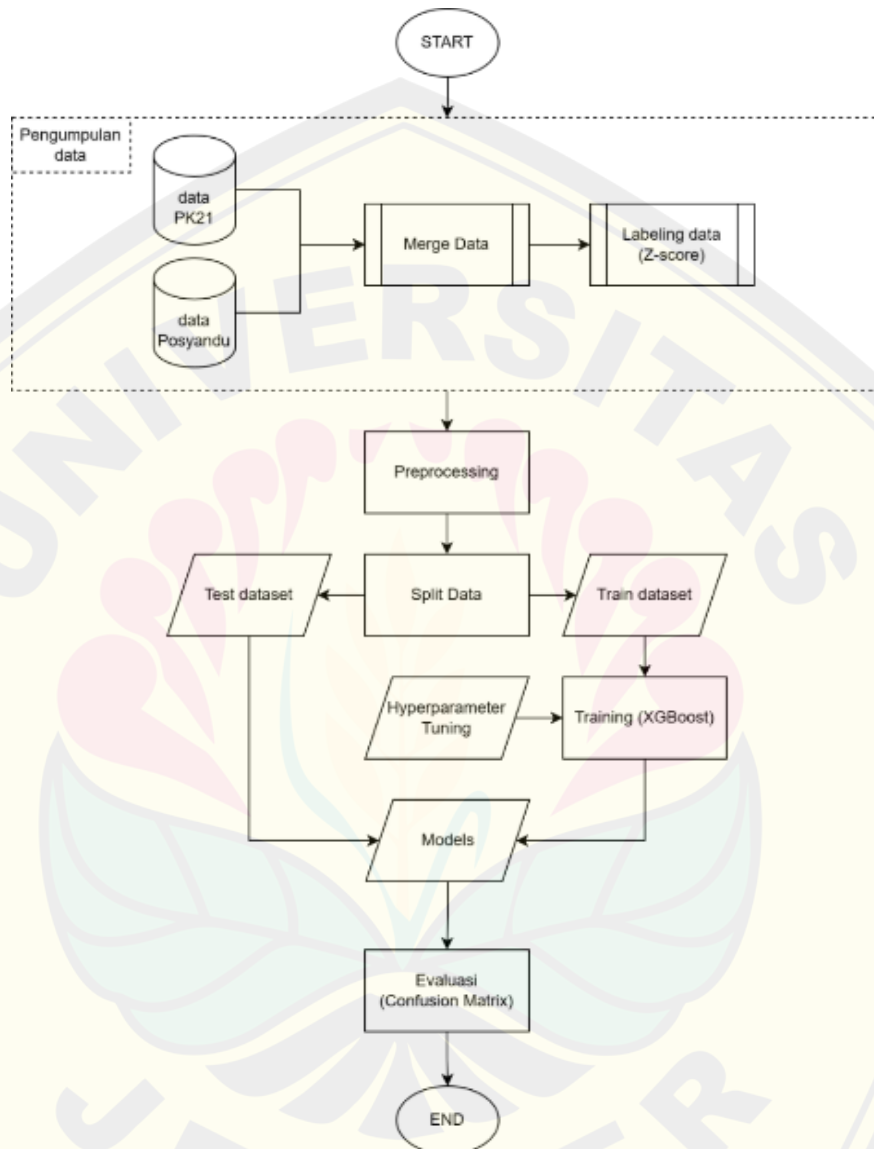
Variabel	Keterangan	Value
Memiliki baduta	Kondisi keluarga saat ini memiliki baduta	1. Ya 2. Tidak
Memiliki balita	Kondisi keluarga saat ini memiliki balita	1. Ya 2. Tidak
Jenis kelamin	Jenisi kelamin bayi	1. Laki laki 2. Perempuan
Berat lahir bayi	Berat lahir bayi dengan satuan gram (g)	Float
Panjang lahir bayi	Panjang lahir bayi dengan satuan centimeter (cm)	Float
Asfiksia	Penyakit gagal nafas spontan pada bayi saat awal kelahiran	1. Ya 2. Tidak
Usia bayi	Usia bayi saat ini (dalam bulan)	Integer
Umur bayi dalam kandungan	Umur bayi dalam kandungan	1. Mature (37-42 minggu) 2. Premature (< 37 minggu)
Panjang bayi	Panjang bayi saat ini	Float
Berat bayi	Berat bayi saat ini	Float
Pemberian asi eksklusif	Pemberian asi eksklusif oleh ibu selama 6 bulan pertama kelahiran	1. Ya 2. Tidak

Tabel 3.2. Tabel dependent dataset

Variabel	Keterangan	Value
Stunting	Status stunting anak dihitung menggunakan Z-score height-to-age	1. Stunting 2. Tidak Stunting

### 3.3.4. Perancangan dan Implementasi Model

Pada tahap ini, terdapat beberapa tahapan yang harus dilakukan. Urutan tahap perancangan dan implementasi model dapat dilihat pada Gambar 3.2.



Gambar 3.2. Implementasi dan perancangan model

#### 1. Pengumpulan data

Langkah pertama adalah melakukan merge data pada data PK21 dan data posyandu yang diperoleh dari Dinas Kesehatan Kabupaten Tuban. Kemudian melakukan pelabelan status gizi anak secara manual dengan menghitung Z-score.

## 2. Preprocessing

Pada tahap ini dilakukan pembersihan data dengan menghapus data yang hilang, mendeteksi outlier, menghapus data duplikat dan normalisasi data dengan memberikan bobot pada tiap atribut.

## 3. Split data

Pada tahap ini data dibagi menjadi dua menggunakan train/test split menjadi train set dan test set menggunakan k-fold cross validation, kemudian dibagi menjadi k subset yang berukuran sama, dimana k-1 digunakan untuk menguji melatih model dan sisanya digunakan untuk menguji model, proses ini diulang sebanyak k kali dengan memvariasikan subset yang digunakan untuk melatih dan menguji model. Hasil k iterasi digabungkan untuk memberikan estimasi kinerja model yang lebih stabil dan akurat.

## 4. Training dataset

Implementasi *hyperparameter tuning* menggunakan metode XGBoost pada data training dilakukan dengan teknik grid yang dilakukan perulangan dengan parameter yang dapat dilihat pada Tabel 3.3.

Tabel 3.3. Indikator pada hyperparameter Tuning

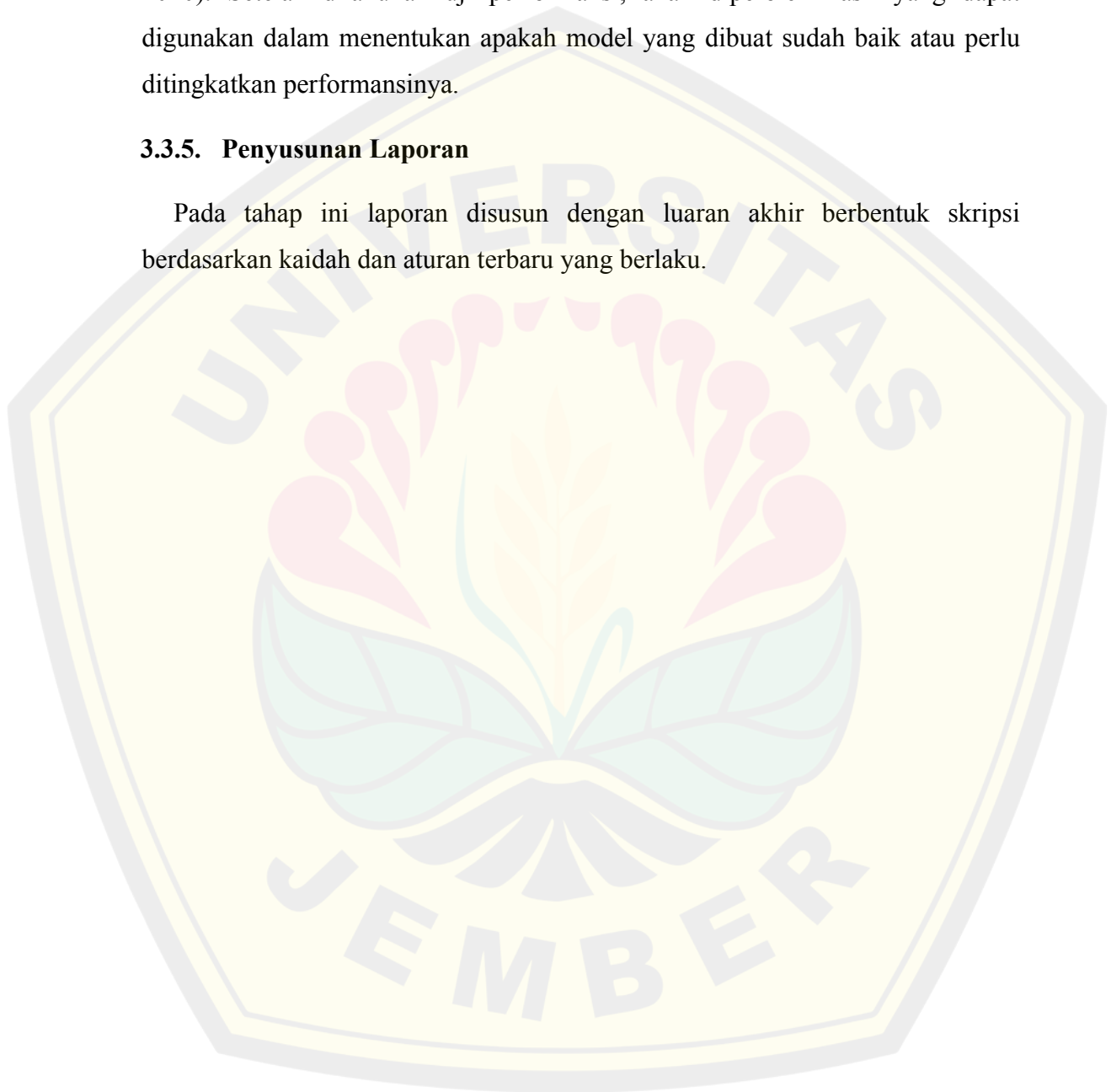
Parameter	Indikator
n_estimator	[50,100,150,200,250]
colsample_bytree	[0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]
max_depth	[3,4,5,6,7]
reg_lambda	[0.2,0.4,0.6,0.8,1]
min_child_weight	[0,1]
learning_rate	[0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0]

## 5. *Evaluasi*

Tahap ini dilakukan dengan menggunakan data testing sehingga kemampuan model dalam memprediksi data baru dapat diukur. Evaluasi performa model dilakukan dengan menguji akurasi, presisi, *recall* dan *f1-score* (Harianto et al., 2020). Setelah dilakukan uji performansi, akan diperoleh hasil yang dapat digunakan dalam menentukan apakah model yang dibuat sudah baik atau perlu ditingkatkan performansinya.

### 3.3.5. Penyusunan Laporan

Pada tahap ini laporan disusun dengan luaran akhir berbentuk skripsi berdasarkan kaidah dan aturan terbaru yang berlaku.





## BAB 4. HASIL DAN PEMBAHASAN

### 4.1. Pengumpulan Data

Dalam penelitian ini, sumber data utama berasal dari dua sumber yang berbeda, yaitu data balita yang diperoleh dari posyandu dan data Pendataan Keluarga 2021 (PK21) yang telah dikumpulkan oleh Dinas Kesehatan Kabupaten Tuban. Kombinasi keduanya memberikan pandangan yang komprehensif mengenai kondisi balita dan faktor-faktor keluarga yang berkaitan dengan kejadian stunting pada tahap awal kehidupan. Jumlah data yang digunakan pada penelitian ini berjumlah 1000 data. Data ini sangat penting untuk dipahami sehingga dapat mengidentifikasi ciri-ciri balita dan keluarga yang menyebabkan kejadian stunting sejak dini. Data yang digunakan dalam penelitian ini diperoleh dari dua sumber utama, yaitu :

#### 1. Data Balita dari Posyandu:

Data balita di posyandu dikumpulkan oleh petugas kesehatan posyandu secara berkala saat balita datang untuk mendapatkan pelayanan kesehatan rutin. Setiap informasi yang relevan dengan penelitian ini dicatat dalam formulir khusus. Data ini mencakup berbagai informasi tentang setiap balita, seperti usia, jenis kelamin, berat badan, tinggi badan, umur bayi dalam kandungan, serta status pemberian asi eksklusif bayi. Data balita yang didapatkan dari posyandu dapat dilihat pada Tabel 4.1. Data balita secara detail ditampilkan dalam Lampiran 4.1.

*Tabel 4.1 Data Balita*

No	NIK Ibu	Nama Ibu	Usia (Tahun)	...	Usia saat ini (bulan)	Panjang /Tinggi Badan Saat ini (cm)	Berat Badan saat ini (kg)	Pemberian ASI Eksklusif (Bayi usia 0-6 bulan)
1	3523185209970001	Isniminawati	25	...	3	56	5.3	Ya
2	3523186701830002	Patri	39	...	21	81	12	Ya
3	3523185189100006	Asmaul Husna	31	...	17	80	8.9	Ya
...	...	...	...	...	...	...	...	...
1010	3523054202000003	Lilik nur indah sari	23	...	8	68	7.5	Ya

## 2. Data Keluarga dari Dinas Kesehatan:

Data keluarga dikumpulkan oleh petugas kesehatan dari dinas kesehatan melalui program PK21. Survei ini dilakukan untuk mendapatkan informasi tentang kondisi sosial-ekonomi keluarga dan lingkungan tempat tinggal balita. Informasi dalam data keluarga mencakup karakteristik dan kondisi keluarga di mana balita tersebut tumbuh dan berkembang. Data keluarga yang didapatkan dari dinas kesehatan Kabupaten Tuban dapat dilihat pada Tabel 4.2. Data keluarga secara detail ditampilkan dalam Lampiran 4.2.

*Tabel 4.2 Data Keluarga*

No	Kode keluarga	nama kepala keluarga	kode provinsi	...	istri terlalu muda	istri terlalu tua	anak terlalu dekat	anak terlalu banyak
1	35231820190204002028033	PERI PRADA EKO S	35	...	0	0	0	0
2	35231820190204001031041	SULIKAN	35	...	0	1	0	0
3	35231820190101001032033	SAKIDI	35	...	0	0	0	0
...	...	...	...	...	...	...	...	...
1010	35230520120203002024026	EKO YULIONO	35	...	0	0	0	0

## 4.2. Preprocessing Data

Data yang akan digunakan dilakukan preprocessing sebelum dapat digunakan untuk melatih model klasifikasi. Mempersiapkan data untuk analisis diperlukan penggabungan data balita dan keluarga, pemilihan variabel yang relevan, pelabelan data, pembersihan data, dan transformasi data agar sesuai dengan kebutuhan analisis.

### 4.2.1. Penggabungan Data Balita dan Keluarga

Setelah data balita dari posyandu dan data keluarga dari dinas kesehatan terkumpul, dilakukan penggabungan data untuk mengintegrasikan informasi dari kedua sumber data ini. Integrasi data ini dilakukan secara manual dengan memanfaatkan fitur pencarian keluarga di situs PK21. Dalam tahap ini, nomor keluarga dan kepala keluarga dicocokkan dengan NIK ibu sebagai acuan, sehingga memungkinkan untuk menggabungkan data keluarga dan data balita yang relevan.

Hasil dari penggabungan data ini akan mencakup informasi karakteristik balita dan keluarga secara bersamaan, seperti yang terlihat pada Tabel 4.3.

Tabel 4.3 Hasil Penggabungan Data Keluarga dan Data Balita

No	Kode keluarga	nama kepala keluarga	...	anak terlalu dekat	anak terlalu banyak	NIK Ibu	Nama Ibu	...	Berat Badan saat ini (kg)	Pemberian ASI Eksklusif (Bayi usia 0-6 bulan)
1	3523182019020 4002028033	PERI PRADA EKO S	...	0	0	352318520 9970001	Isnimina wati	...	5.3	Ya
2	3523182019020 4001031041	SULIKAN	...	0	0	352318670 1830002	Patri	...	12	Ya
3	3523182019010 1001032033	SAKIDI	...	0	0	352318518 9100006	Asmaul Husna	...	8.9	Ya
...	...	...	...	...	...	...	...	...	...	...
1010	3523052012020 3002024026	EKO YULIONO	...	0	0	352305420 2000003	Lilik nur indah sari	...	7.5	Ya

Tabel ini akan menjadi dasar untuk dilakukan analisis lanjutan dalam upaya pemahaman yang lebih mendalam terkait pengaruh karakteristik balita dan keluarga terhadap balita stunting.

#### 4.2.2. Pelabelan Data Stunting

Data yang tersedia dari posyandu belum mencakup informasi mengenai status stunting pada balita. Untuk mengatasi hal ini, diperlukan pelabelan status stunting pada data balita dengan menggunakan variabel panjang/tinggi badan dan umur balita. Proses pelabelan ini mengacu pada pedoman yang telah ditetapkan oleh Peraturan Menteri Kesehatan No 2 Tahun (Menteri Kesehatan, 2020) yang dapat dilihat pada Tabel 4.4 untuk anak laki-laki

Tabel 4.4 Standar Panjang/Tinggi Badan Menurut Umur PB/U Anak Laki-Laki

Umur (bulan)	Panjang Badan (cm)						
	-3 SD	-2 SD	-1 SD	Median	+1 SD	+2 SD	+3 SD
0	44.2	46.1	48.0	49.9	51.8	53.7	55.6
1	48.9	50.8	52.8	54.7	56.7	58.6	60.6
2	52.4	54.4	56.4	58.4	60.4	62.4	64.4
...	...	...	...	...	...	...	...
60	96.1	100.7	105.3	110.0	114.6	119.2	123.9

Dan Tabel 4.5 untuk anak perempuan.

Tabel 4.5 Standar Panjang/Tinggi Badan Menurut Umur PB/U Anak Perempuan

Umur (bulan)	Panjang Badan (cm)						
	-3 SD	-2 SD	-1 SD	Median	+1 SD	+2 SD	+3 SD
0	43.6	45.4	47.3	49.1	51.0	52.9	54.7
1	47.8	49.8	51.7	53.7	55.6	57.6	59.5
2	51.0	53.0	55.0	57.1	59.1	61.1	63.2
...	...	...	...	...	...	...	...
60	95.2	99.9	104.7	109.4	114.2	118.9	123.7

Tabel tersebut menguraikan standar deviasi antropometri anak yang dapat digunakan untuk menghitung status stunting pada anak dengan menggunakan rumus pada persamaan (10) jika TB/PB anak kurang dari median,

$$(TB/U) = \frac{TP/PB_{anak} - TP/PB_{median}}{TP/PB_{median} - (-1SD)} \quad (10)$$

dan menggunakan rumus pada persamaan (11) jika TP/PB anak lebih dari median.

$$(TB/U) = \frac{TP/PB_{anak} - TP/PB_{median}}{(-1SD) - TP/PB_{median}} \quad (11)$$

Keterangan:

TB : Tinggi Badan

U : Umur

TP/TB<sub>anak</sub> : Tinggi badan / Panjang Badan Anak

TP/PB<sub>median</sub> : Nilai median standar deviasi anak

-1SD : Nilai -1 standar deviasi anak

Setelah dilakukan perhitungan dan nilai standar deviasi anak kurang dari -2, maka kondisi anak yang dinyatakan mengalami stunting ditandai dengan angka 1 dan anak yang dinyatakan tidak mengalami stunting ditandai dengan angka 0 dan dimasukkan ke dalam dataset yang dapat dilihat pada Tabel 4.6. Data gabungan dataset secara detail ditampilkan dalam Lampiran 4.3.

Tabel 4.6 Dataset setelah pelabelan

No	Kode keluarga	nama kepala keluarga	...	anak terlalu dekat	NIK Ibu	Nama Ibu	...	Berat Badan saat ini (kg)	Pemberian ASI Eksklusif (Bayi usia 0-6 bulan)	Status
1	35231820190204002028033	PERI PRADA EKO S	...	0	3523185209970001	Isniminawati	...	5.3	Ya	1
2	35231820190204001031041	SULIKAN	...	0	3523186701830002	Patri	...	12	Ya	0
3	35231820190101001032033	SAKIDI	...	0	3523185189100006	Asmaul Husna	...	8.9	Ya	0
...	...	...	...	...	...	...	...	...	...	...
1010	35230520120203002024026	EKO YULIONO	...	0	3523054202000003	Lilik nur indah sari	...	7.5	Ya	0

#### 4.2.3. Pemilihan Variabel yang Relevan

Penelitian ini mengutamakan variabel-variabel yang memiliki keterkaitan langsung dengan tujuan penelitian. Pemilihan variabel ini didasarkan pada kriteria berikut:

##### 1. Relevansi Variabel dengan Tujuan Penelitian:

Pemilihan variabel dilakukan dengan mempertimbangkan hubungannya secara langsung dengan status gizi balita serta karakteristik keluarga yang diduga memiliki pengaruh terhadap kasus stunting pada balita. Dalam konteks ini, variabel-variabel yang dipilih harus memiliki relevansi yang kuat dalam konteks pemahaman terhadap stunting.

##### 2. Hubungan Variabel dengan Teori atau Penemuan Sebelumnya:

Pemilihan variabel juga didasarkan pada teori atau penelitian sebelumnya yang dilakukan Jahidur Rahman Khan, Javed H. Tomal & Enayetur Raheem tentang pemilihan model dan variabel menggunakan model machine learning, dan penelitian yang dilakukan Fikrewold H Bitew, Corey S Sparks & Samuel H Nyarko tentang algoritma machine learning untuk prediksi balita kurang gizi di Ethiopia. Dari beberapa penelitian tersebut menunjukkan hubungan antara variabel yang berpengaruh terhadap stunting pada balita.

### 3. Penghapusan Variabel yang Tidak Relevan:

Dalam tahap praproses data, dilakukan evaluasi terhadap beberapa variabel yang tidak relevan atau tidak memberikan kontribusi signifikan terhadap pemahaman stunting pada balita. Variabel-variabel tersebut akan dihapus dari dataset untuk mengurangi dimensi data dan menghindari kompleksitas yang tidak perlu.

#### 4.2.4. Pembersihan Data

Data yang diperoleh dari posyandu dan dinas kesehatan sering kali mengandung kesalahan atau nilai yang hilang. Oleh karena itu, langkah pertama dalam praproses data adalah pembersihan data untuk mengatasi masalah-masalah berikut:

##### 1. Penanganan Missing Data:

Dataset yang digunakan dapat mengandung nilai yang kosong atau tidak lengkap. Oleh karena itu, dilakukan analisis untuk mengidentifikasi data yang hilang dan kemudian memutuskan apakah akan menghapusnya atau mengisi dengan nilai rata-rata. Jika data yang hilang cukup penting maka baris data tersebut akan dihapus.

##### 2. Penanganan Outlier:

Outlier dapat mempengaruhi performa algoritma klasifikasi, oleh karena itu dilakukan identifikasi dan penanganan terhadap nilai-nilai ekstrim yang tidak wajar dalam data secara manual. Jika terdapat outlier yang signifikan, nilai outlier tersebut dihapus atau disesuaikan.

#### 4.2.5. Transformasi Data

Selanjutnya, data perlu diubah atau ditransformasi agar sesuai dengan asumsi dan kebutuhan algoritma Extreme Gradient Boosted Trees. Beberapa transformasi yang dilakukan antara lain:

#### 1. Normalisasi atau Standarisasi:

Jika variabel-variabel dalam data memiliki skala yang berbeda, peneliti akan melakukan normalisasi atau standarisasi untuk mengubah nilai-nilai variabel ke dalam rentang tertentu, seperti  $[0, 1]$  atau nilai rata-rata 0 dan simpangan baku 1.

#### 2. Encoding Variabel Kategorikal:

Algoritma Extreme Gradient Boosted Trees hanya dapat bekerja dengan variabel numerik, oleh karena itu perlu mengubah variabel kategorikal menjadi representasi numerik. Teknik encoding seperti one-hot encoding digunakan untuk merubah variabel kategorikal seperti variabel jenis kelamin.

### 4.3. Pembagian Data

Pada bagian ini, data dibagi menjadi data training dan data testing sebelum melakukan pelatihan model klasifikasi menggunakan algoritma Extreme Gradient Boosted Trees (XGBoost) untuk mengklasifikasikan balita menjadi dua kategori, yaitu "Stunting" dan "Tidak Stunting".

Sebelum melakukan pelatihan model, data tersebut dibagi menjadi dua kelompok, yaitu data training dan data testing. Pembagian proporsi data yang digunakan sebanyak 3 kali yaitu, 90% data training : 10% data testing, 80% data training : 20% data testing dan 70% data training : 30% data testing. Pembagian data dilakukan secara acak untuk menghindari bias dalam proses pelatihan dan evaluasi model.

Proses pembagian data dilakukan menggunakan teknik `train_test_split` dari library `scikit-learn`. Proses ini juga dilakukan secara stratified untuk memastikan distribusi kelas yang seimbang dalam kedua kelompok data. Hal ini penting untuk memastikan bahwa data testing mencerminkan karakteristik data secara keseluruhan dan mewakili proporsi stunting dan tidak stunting yang ada dalam populasi balita secara proporsional.

#### **4.4. Konfigurasi Hyperparameter Tuning**

Proses tuning hyperparameter dilakukan dengan menggunakan metode Grid Search Cross-Validation (CV). Metode ini secara sistematis mengeksplorasi berbagai kombinasi hyperparameter dengan tujuan untuk mengidentifikasi konfigurasi hyperparameter optimal yang dapat memaksimalkan kinerja model kami dalam tugas tertentu. Proses ini melibatkan pengujian model menggunakan berbagai kombinasi nilai hyperparameter untuk memastikan hasilnya konsisten dan dapat diandalkan..

##### **4.4.1. Metode Hyperparameter Tuning**

Metode Grid Search CV digunakan untuk mengeksplorasi berbagai tingkat kompleksitas model dan memahami bagaimana pengaruhnya terhadap kinerja model, dengan mempertimbangkan trade-off antara kinerja dan waktu pelatihan dalam upaya untuk meningkatkan kinerja model. Teknik Grid Search CV secara sistematis mencoba semua kombinasi nilai hyperparameter yang tercantum dalam grid search yang telah ditentukan sebelumnya.

##### **4.4.2. Pemilihan Hyperparameter**

Hyperparameter yang digunakan adalah hyperparameter yang memiliki pengaruh yang signifikan terhadap kinerja model menurut dokumentasi XGBoost untuk Tree Booster, yaitu 'n\_estimators', 'max\_depth', 'colsample\_bytree', 'reg\_lambda', 'min\_child\_weight', dan 'learning\_rate'.

##### **4.4.3. Rancangan Grid Search Tuning Hyperparameter**

Grid search untuk hyperparameter tuning yang digunakan pada tahap ini dapat dilihat pada Gambar 4.1.



```
param_grid = {  
    'n_estimators': [50, 100, 150, 200, 250],  
    'max_depth': [3, 4, 5, 6, 7],  
    'colsample_bytree': [0.2, 0.4, 0.6, 0.8, 1],  
    'reg_lambda': [0.2, 0.4, 0.6, 0.8],  
    'min_child_weight': [0,1],  
    'learning_rate': [0.1, 0.2, 0.3, 0.4, 0.5]  
}
```

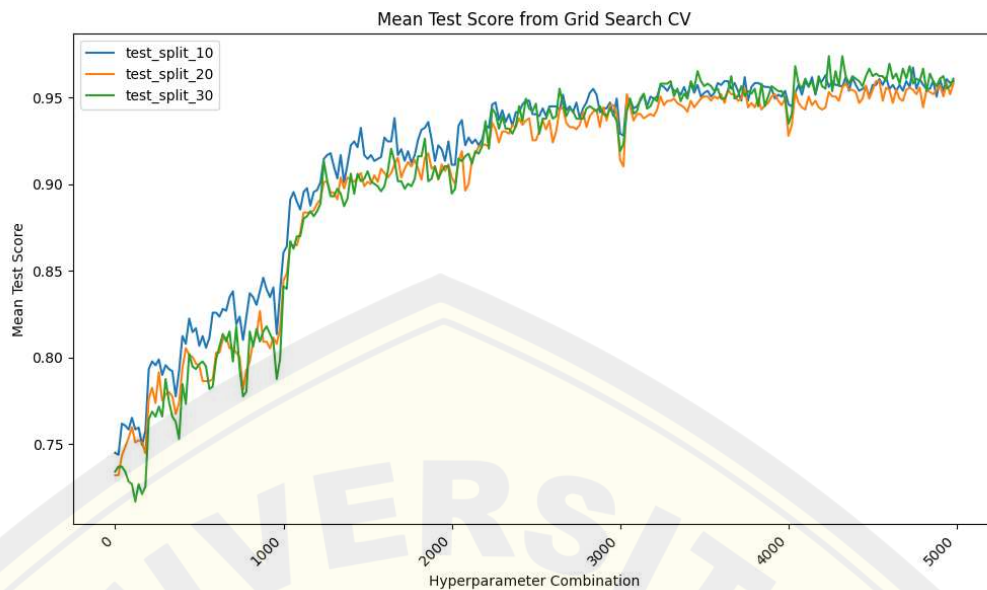
*Gambar 4.1 Grid Search untuk tuning Hyperparameter*

Pemilihan `n_estimators` dari 50 hingga 250 memungkinkan eksplorasi kompleksitas model dengan memeriksa trade-off antara kinerja dan waktu pelatihan. Kedalaman pohon (`max_depth`) dari 3 hingga 7 digunakan untuk mengeksplorasi tingkat kompleksitas pohon keputusan, sementara fraksi fitur (`colsample_bytree`) dari 0.2 hingga 1 mengatur sejauh mana setiap pohon dapat menggunakan fitur. Parameter lainnya, seperti `reg_lambda`, `min_child_weight`, dan `learning_rate`, dipilih berdasarkan upaya mengendalikan overfitting dan mengoptimalkan kecepatan konvergensi model. Keseluruhan pemilihan nilai-nilai ini dirancang untuk memahami bagaimana pengaruhnya terhadap kinerja model kami dalam mengatasi tugas kami yang spesifik.

#### **4.4.4. Cross Validation**

Validasi silang dilakukan dengan skema lipatan (`folds`) sebanyak 5 untuk mengevaluasi setiap kombinasi hyperparameter. Metode ini merupakan pendekatan yang kuat untuk mengukur kualitas prediksi model dengan menghindari overfitting pada data latih dan memberikan estimasi yang lebih akurat terhadap kinerja model pada data yang belum pernah dilihat sebelumnya. Setiap skor validasi silang dihitung berdasarkan rata-rata performa pada kelima lipatan.

Hasil Grid Search, seperti yang terlihat pada Gambar 4.2, menggambarkan performa model pada berbagai kombinasi hyperparameter untuk memberikan visualisasi yang lebih jelas.



*Gambar 4.2 Nilai Mean Test Score setelah Hyperparameter Tuning*

Dari grafik tersebut, dapat dilihat bahwa nilai Mean test score (Rata rata akurasi) terus meningkat seiring dengan variasi kombinasi hyperparameter, hingga akhirnya mencapai nilai yang stabil yang mengindikasikan bahwa model tidak akan mengalami peningkatan performa yang signifikan dengan mencoba lebih banyak kombinasi hyperparameter, sehingga kombinasi tersebut dapat digunakan dengan baik.

#### **4.5. Hasil Klasifikasi XGBoost**

Pada bagian ini, hasil klasifikasi dilihat dari performa klasifikasi dan atribut yang paling berpengaruh yang diperoleh dari model yang diimplementasikan menggunakan algoritma XGBoost setelah dilakukan konfigurasi hyperparameter tuning.

##### **4.5.1. Performa Klasifikasi**

Pada proses hyperparameter tuning menggunakan metode GridSearch CV untuk mengoptimalkan performa model, kombinasi hyperparameter diuji pada data train yang dapat dilihat pada Tabel 4.7

Tabel 4.7 Hasil GridSearchCV Split 80:20

no	colsample bytree	learning rate	max depth	min child weight	n esti mat ors	req lamb da	split0 test score	split1 test score	split2 test score	split3 test score	split4 test score	mean test score	std test score	rank test score
0	0.2	0.1	3	0	50	0.2	0.716981	0.753165	0.772152	0.664557	0.753165	0.732004	0.038163	4994
1	0.2	0.1	3	0	50	0.4	0.710692	0.759494	0.765823	0.696203	0.753165	0.737075	0.028123	4993
2	0.2	0.1	3	0	50	0.6	0.704403	0.753165	0.759494	0.664557	0.753165	0.726956	0.036972	4998
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4043	1	0.1	4	0	50	0.8	0.943396	0.936709	0.949367	0.924051	0.987342	0.948173	0.021315	889
<b>4044</b>	<b>1</b>	<b>0.1</b>	<b>4</b>	<b>0</b>	<b>100</b>	<b>0.2</b>	<b>0.962264</b>	<b>0.962025</b>	<b>0.955696</b>	<b>0.943038</b>	<b>0.981013</b>	<b>0.960807</b>	<b>0.012278</b>	<b>1</b>
4045	1	0.1	4	0	100	0.4	0.955975	0.955596	0.962025	0.93038	0.962025	0.95322	0.011751	253
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Dari tabel berikut dapat dilihat bahwa pada kombinasi ke 4044 pada split 80% train:20% test, ditemukan kombinasi hyperparameter terbaik dengan kombinasi yang dapat dilihat pada Tabel 4.8

Tabel 4.8 Kombinasi Hyperparameter Terbaik

'colsample_bytree'	1
'learning_rate'	0.1
'max_depth'	4
'min_child_weight'	0
'n_estimators'	100
'reg_lambda'	0.2

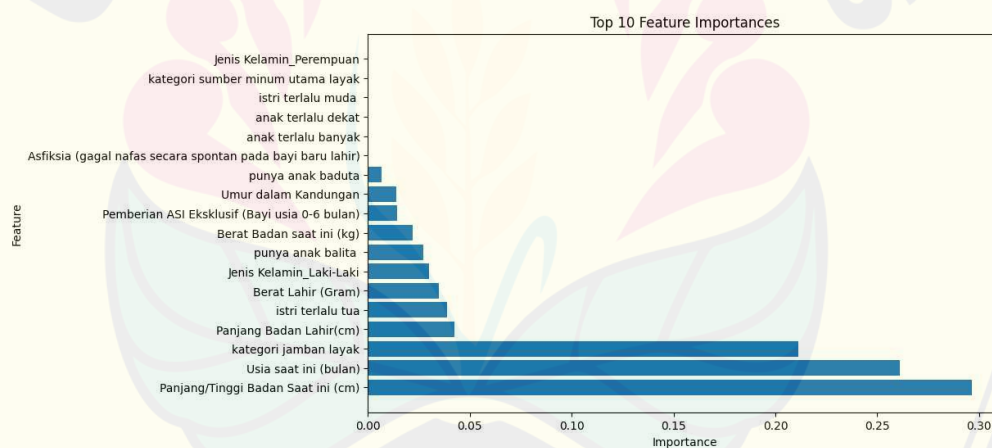
Dengan menggunakan kombinasi hyperparameter tersebut, model mencapai skor validasi silang (cross-validation score) sebesar 96%. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam menggeneralisasi pada data yang belum pernah dilihat sebelumnya. Kombinasi hyperparameter yang dihasilkan dari proses tuning ini membantu meningkatkan performa model dan mengurangi risiko overfitting. Skor cross-validation yang cukup tinggi juga menegaskan kehandalan model dalam mengklasifikasikan balita berisiko stunting dengan akurat.

Setelah menerapkan model XGBoost dengan konfigurasi hyperparameter terbaik yang ditemukan melalui proses tuning, model tersebut diuji pada data pengujian yang belum pernah dilihat sebelumnya. Sebelumnya telah dilakukan pembagian tiga dataset dengan perbandingan yang berbeda, yaitu 70% data latih dan 30% data uji, 80% data latih dan 20% data uji, serta 90% data latih dan 10% data uji. Dari ketiga pembagian tersebut, hasil terbaik diperoleh pada pembagian

data dengan perbandingan 80% data latih dan 20% data uji. Pada pembagian ini, model mencapai tingkat akurasi sebesar 95.9%, presisi sebesar 94.7%, recall sebesar 98.1%, dan nilai F1-score sebesar 96.4%.

#### 4.5.2. Atribut paling berpengaruh

Analisis fitur dilakukan untuk mengidentifikasi tiap atribut pada model klasifikasi yang memiliki pengaruh paling signifikan dalam prediksi risiko stunting pada balita. Analisis ini memberikan wawasan yang berharga tentang faktor-faktor yang paling mempengaruhi hasil klasifikasi. Property `feature_importances_` default pada model yang menggunakan `XGBClassifier` digunakan untuk mendapatkan atribut yang paling berpengaruh. Hasil tersebut kemudian dibuat grafik menggunakan `pyplot` pada library `matplotlib`. Grafik ini memberikan gambaran visual tentang skala kontribusi setiap atribut dalam mempengaruhi keputusan klasifikasi yang dapat dilihat pada Gambar 4.3.

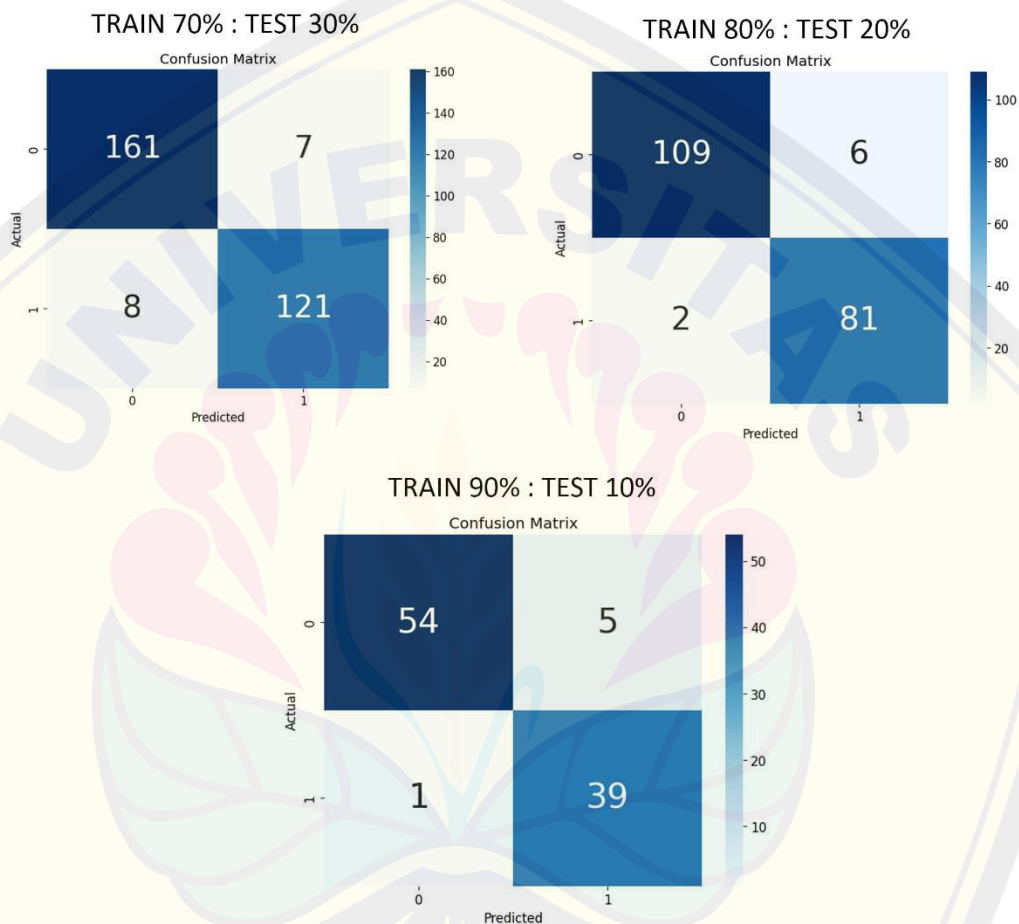


Gambar 4.3 Top 10 Feature Importances

Dari grafik tersebut, dapat dilihat bahwa atribut Panjang/Tinggi Badan Saat Ini memiliki kontribusi paling signifikan dalam prediksi stunting pada balita, diikuti oleh Usia saat ini, Namun, penelitian ini lebih fokus pada faktor lain yang juga memiliki pengaruh tinggi, seperti kondisi jamban, usia ibu, dan pemberian ASI eksklusif. Oleh karena itu, variabel-variabel tersebut perlu mendapatkan perhatian lebih lanjut dalam pencegahan stunting pada balita.

#### 4.6. Evaluasi (Confusion Matrix)

Evaluasi performa model klasifikasi lebih lanjut dilakukan melalui analisis confusion matrix pada tiga pembagian data yang berbeda, yaitu 70% data latih dan 30% data uji, 80% data latih dan 20% data uji, serta 90% data latih dan 10% data uji yang dapat dilihat pada Gambar 4.4 .



Gambar 4.4 Hasil Confussion Matrix

Confusion matrix tersebut memberikan wawasan mendalam tentang tingkat akurasi, presisi, recall dan f1-score yang dibuat oleh model dalam klasifikasi. Perbandingan matrix evaluasi untuk tiap bagian data training dan data test pada confusion matrix dapat dilihat pada Tabel 4.9.

*Tabel 4.9 Tabel Perbandingan Confusion Matrix*

	Data train : Data Test		
	90%:10%	80%:20%	70%:30%
Accuracy	0.939	0.959	0.949
Precision	0.915	0.947	0.958
Recall	0.981	0.981	0.952
F1-Score	0.947	0.964	0.955

Dari ketiga pembagian data yang ada memberikan informasi yang berharga tentang performa model dalam konteks yang berbeda. Pembagian data dengan perbandingan 80% data train:20% data test menghasilkan performa terbaik dengan akurasi 95.9%, presisi 94.7%, recall 98.1%, dan F1-score 96.4%. Hal ini menunjukkan bahwa model memiliki kemampuan yang tinggi dalam mengklasifikasikan balita berisiko stunting secara akurat.

## BAB 5. PENUTUP

Berdasarkan hasil dari penelitian terkait pengerjaan tugas akhir ini. Maka pada bab ini akan dibahas kesimpulan dan saran.

### 5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, hasil kesimpulan penelitian adalah sebagai berikut :

1. Dalam penelitian ini, implementasi model klasifikasi balita stunting menggunakan algoritma XGBoost dengan melakukan konfigurasi hyperparameter tuning dilakukan. Dari model tersebut dihasilkan *feature importances* menggunakan library XGBoost yang menunjukkan pengaruh karakteristik balita dan keluarga terhadap model. Panjang dan usia menjadi variabel yang paling berpengaruh karena menurut who dari dua variabel tersebut dapat ditentukan anak tersebut stunting atau tidak. Namun penelitian ini lebih memfokuskan terhadap variabel lain yang berpengaruh terhadap stunting seperti kondisi jamban yang kurang layak dan istri terlalu tua serta pemberian ASI eksklusif menjadi variabel yang cukup berpengaruh terhadap stunting pada balita sehingga perlu adanya perhatian lebih terhadap variabel tersebut.
2. Model yang dihasilkan mampu mengklasifikasikan balita stunting dengan akurasi yang mencapai 95.9%, presisi 94.7%, recall 98.1%, dan F1-Score 96.4%. Hasil tersebut menunjukkan performa yang baik dalam mengidentifikasi kemungkinan kasus balita stunting.

Penelitian ini memberikan kontribusi tentang bagaimana model klasifikasi dapat diterapkan untuk mengidentifikasi balita stunting dengan akurasi yang tinggi. Hasil ini memiliki implikasi penting dalam upaya mencegah stunting pada balita melalui identifikasi dini serta dilakukan tindakan pencegahan yang tepat.

## 5.2. Saran

Berdasarkan hasil penelitian ini, penelitian ini memberikan beberapa saran untuk penelitian selanjutnya:

1. Melakukan penelitian lebih lanjut untuk menganalisis faktor-faktor lain yang berpotensi mempengaruhi status gizi balita stunting, seperti faktor lingkungan, sosial, dan ekonomi.
2. Menggabungkan berbagai jenis algoritma klasifikasi dan teknik ensemble untuk membandingkan dan meningkatkan performa model.
3. Memperluas cakupan penelitian untuk mengklasifikasikan status gizi lainnya, seperti balita wasting atau overweight, guna memberikan informasi lebih komprehensif mengenai status gizi balita.



## DAFTAR PUSTAKA

- Arifin, O. (2018). SISTEM KLASIFIKASI BERITA DARING FAKTOR KEJAHATAN PENYALAHGUNAAN NARKOTIKA BERBASIS ALGORITMA NAIVE BAYES. *Telematika*, 11(2), 27. <https://doi.org/10.35671/telematika.v11i2.713>
- Asian Development Bank. (2021). *KEY INDICATORS FOR ASIA AND THE PACIFIC 2021 52nd EDITION*. Asian Development Bank.
- Beal, T., Tumilowicz, A., Sutrisna, A., Izwardy, D., & Neufeld, L. M. (2018). A review of child stunting determinants in Indonesia. In *Maternal and Child Nutrition* (Vol. 14, Issue 4). <https://doi.org/10.1111/mcn.12617>
- Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutrition*, 25(2), 269–280. <https://doi.org/10.1017/S1368980021004262>
- BPMI Sekretariat Presiden. (2022, March 22). *PEMERINTAH TARGETKAN ANGKA PREVALENSI STUNTING DI BAWAH 14 PERSEN PADA 2024*. <https://www.presidentri.go.id/siaran-pers/pemerintah-targetkan-angka-prevalensi-stunting-di-bawah-14-persen-pada-2024/>
- Daracantika, A. (2021). Systematic Literature Review: Pengaruh Negatif Stunting terhadap Perkembangan Kognitif Anak. *Jurnal BIKFOKES*.
- Endy P. Prawirohartono. (2021). *STUNTING: Dari Teori dan Bukti ke Implementasi di Lapangan*. UGM PRESS.
- Fadhlurrahman, A. (2023). *Memahami XGBoost dari Sudut Pandang Matematika pada Kasus Klasifikasi Pengaturan Masalah*.
- Fitang Budhi Adhitia. (2023). *Pemutakhiran Data Keluarga BKKBN, Dalam Setahun Jumlah Keluarga Bertambah 2,2 Juta*. <https://www.bkkbn.go.id/berita-pemutakhiran-data-keluarga-bkkbn-dalam-setahun-jumlah-keluarga-bertambah-22-juta>

Harianto, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Anomaly dengan Univariate Fitur Selection. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 40–49. <https://doi.org/10.29408/edumatic.v4i2.2433>

Kementerian Kesehatan Republik Indonesia. (2019). *LAPORAN PELAKSANAAN INTEGRASI SUSENAS MARET 2019 DAN SSGBI TAHUN 2019*. Badan Pusat Statistik.

Kementerian Kesehatan Republik Indonesia. (2021). *STUDI STATUS GIZI INDONESIA (SSGI)*. Humas Litbangkes.

Khan, J. R., Tomal, J. H., & Raheem, E. (2021). Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh. *Informatics for Health and Social Care*, 46(4), 425–442. <https://doi.org/10.1080/17538157.2021.1904938>

Novia Puspitasari, P., & Adi, A. C. (2020). Risk Factors of Stunting in Children Aged 1-5 Years at Wire Primary Health Care, Tuban Regency, East Java. *Journal of Maternal and Child Health*, 04, 387–395. [www.thejmch.com](http://www.thejmch.com)

Menteri Kesehatan. (2020). *PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR 2 TAHUN 2020 TENTANG STANDAR ANTROPOMETRI ANAK*.

Rahman, S. M. J., Ahmed, N. A. M. F., Abedin, M. M., Ahammed, B., Ali, M., Rahman, M. J., & Maniruzzaman, M. (2021). Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. *PLoS ONE*, 16(6 June 2021). <https://doi.org/10.1371/journal.pone.0253172>

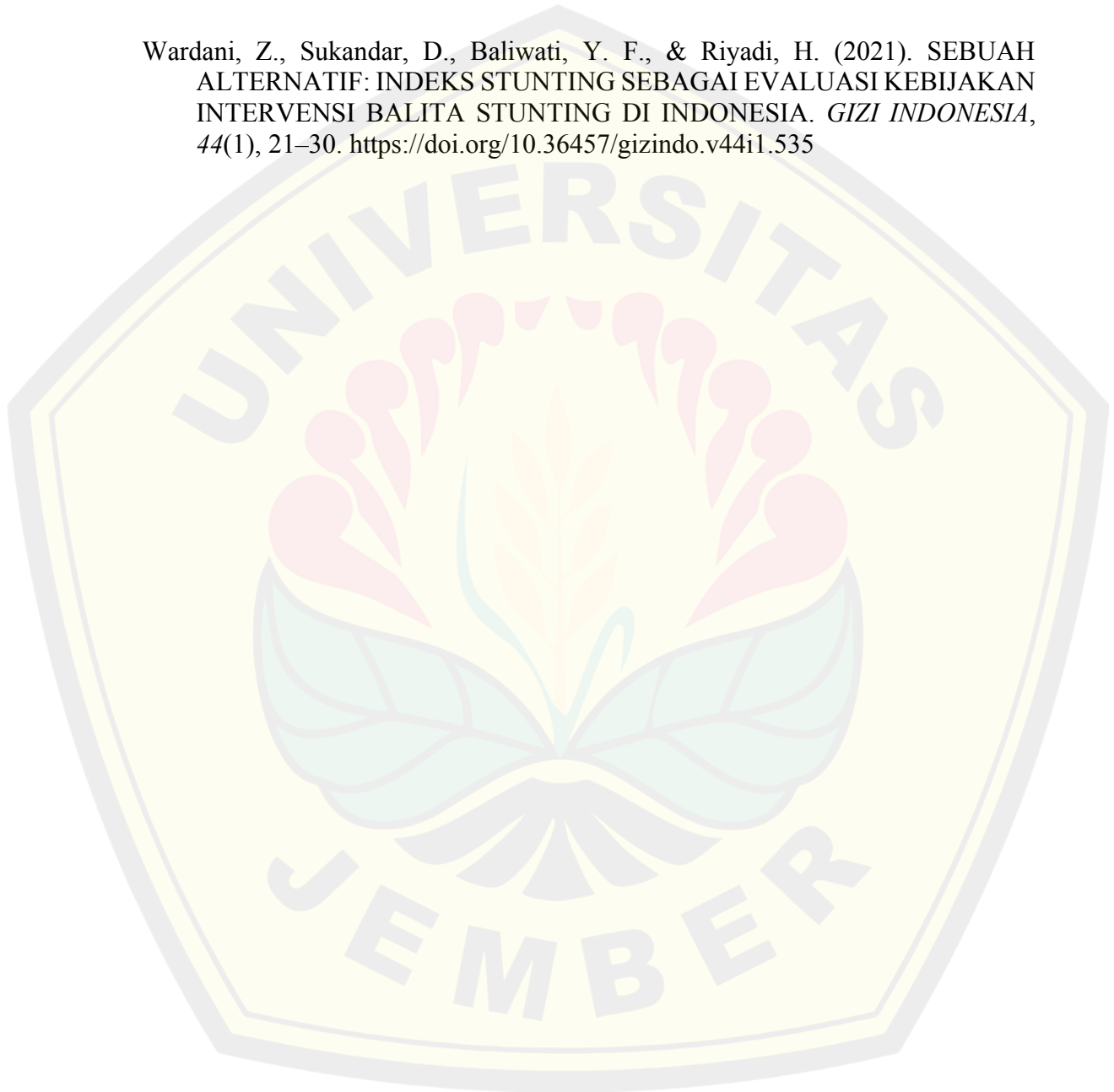
Ratna Nuraini. (2021). *Pendataan Keluarga, Upaya Melindungi dan Memenuhi Hak Setiap Warga Negara*. <https://indonesia.go.id/kategori/editorial/2665/pendataan-keluarga-upaya-melindungi-dan-memenuhi-hak-setiap-warga-negara>

Sekretariat Wakil Presiden Republik Indonesia. (2017). *100 Kabupaten/Kota Prioritas untuk Intervensi Anak Kerdil (Stunting) VOLUME 1*.

Talukder, A., & Ahammed, B. (2020). Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, 78. <https://doi.org/10.1016/j.nut.2020.110861>

Trihono, Sudomo, M., & Indonesia. Badan Penelitian dan Pengembangan Kesehatan. (2015). *Pendek (stunting) di Indonesia : masalah dan solusinya*.

Wardani, Z., Sukandar, D., Baliwati, Y. F., & Riyadi, H. (2021). SEBUAH ALTERNATIF: INDEKS STUNTING SEBAGAI EVALUASI KEBIJAKAN INTERVENSI BALITA STUNTING DI INDONESIA. *GIZI INDONESIA*, 44(1), 21–30. <https://doi.org/10.36457/gizindo.v44i1.535>



LAMPIRAN

Lampiran 1 Data Balita

<https://unej.id/databalita>



Lampiran 2 Data Keluarga

<https://unej.id/datakeluarga>



Lampiran 3 Data Gabungan

<https://unej.id/datagabungankeluargabalita>

