

**ANALISIS DETERMINAN PENYEBAB STUNTING PROVINSI
DI INDONESIA: APLIKASI MODEL *RANDOM FOREST* DAN
INTERPRETASI SHAP PADA DATA SSGI TAHUN 2024**

TUGAS AKHIR

Diajukan sebagai syarat menyelesaikan jenjang strata Satu (S-1) di
Program Studi Teknik Informatika, Fakultas Teknologi Industri, Institut
Teknologi Sumatera

Oleh:

Cindy Nadila Putri

122140002



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2026**

LEMBAR PENGESAHAN

Saya menyatakan bahwa Tugas Akhir berjudul “Analisis Determinan Penyebab *Stunting* Provinsi di Indonesia: Aplikasi Model *Random Forest* dan Interpretasi SHAP pada Data SSGI Tahun 2024” merupakan hasil karya saya sendiri dan belum pernah diajukan, baik sebagian maupun seluruhnya, di Institut Teknologi Sumatera atau institusi pendidikan lain oleh saya maupun pihak lain.

Lampung Selatan, 20 Januari 2026
Penulis,

Cindy Nadila Putri
NIM. 122140002

Foto 2x3

Diperiksa dan disetujui oleh,
Pembimbing

1. Martin Clinton Tosima Manullang, Ph.D.
NIP. 19930109 2019 03 1 017
.....
2. Martin Clinton Tosima Manullang, Ph.D.
NIP. 19930109 2019 03 1 017
.....

Pengaji

1. Dosen Pengaji I
NIP. 19900000 2000 00 0 000
.....
2. Dosen Pengaji II
NIP. 19900000 2000 00 0 000
.....

Disahkan oleh,
Koordinator Program Studi Teknik Informatika
Fakultas Teknologi Industri
Institut Teknologi Sumatera

Andika Setiawan, S.Kom., M.Cs.
NIP. 19911127 2022 03 1 007

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir dengan judul “Analisis Determinan Penyebab *Stunting* Provinsi di Indonesia: Aplikasi Model *Random Forest* dan Interpretasi SHAP pada Data SSGI Tahun 2024” adalah karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan benar.

Nama : Cindy Nadila Putri

NIM : 122140002

Tanda Tangan :

Tanggal :

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Institut Teknologi Sumatera, saya yang bertanda tangan di bawah ini:

Nama : Cindy Nadila Putri

NIM : 122140002

Program Studi : Teknik Informatika

Fakultas : Teknologi Industri

Jenis Karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Institut Teknologi Sumatera **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty Free Right)** atas karya ilmiah saya yang berjudul:

Analisis Determinan Penyebab Stunting Provinsi di Indonesia: Aplikasi Model Random Forest dan Interpretasi SHAP pada Data SSGI Tahun 2024

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Institut Teknologi Sumatera berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Lampung Selatan

Pada tanggal : 20 Januari 2026

Yang menyatakan

Cindy Nadila Putri

KATA PENGANTAR

Pada halaman ini mahasiswa berkesempatan untuk menyatakan terima kasih secara tertulis kepada pembimbing dan pihak lain yang telah memberi bimbingan, nasihat, saran dan kritik, kepada mereka yang telah membantu melakukan penelitian, kepada perorangan atau lembaga yang telah memberi bantuan keuangan, materi dan/atau sarana. Cara menulis kata pengantar beraneka ragam, tetapi hendaknya menggunakan kalimat yang baku. Ucapan terima kasih agar dibuat tidak berlebihan dan dibatasi pada pihak yang terkait secara ilmiah (berhubungan dengan subjek/materi penelitian).

Puji syukur kehadirat Allah SWT/Tuhan Yang Maha Esa atas limpahan rahmat, karunia, serta petunjuk-Nya sehingga penyusunan tugas akhir ini telah terselesaikan dengan baik. Dalam penyusunan tugas akhir ini penulis telah banyak mendapatkan arahan, bantuan, serta dukungan dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. [Rektor ITERA] selaku Rektor Institut Teknologi Sumatera.
2. [Dekan FTI] selaku Dekan Fakultas Teknologi Industri.
3. [Koor Prodi IF] selaku Ketua Program Studi Teknik Informatika.
4. [Dosen Pembimbing] selaku Dosen Pembimbing atas ide, waktu, tenaga, perhatian, dan masukan yang telah disumbangsihkan kepada penulis.
5. [Isi nama lainnya]

Akhir kata penulis berharap semoga tugas akhir ini dapat memberikan manfaat bagi kita semua.

RINGKASAN

Analisis Determinan Penyebab *Stunting* Provinsi di Indonesia: Aplikasi Model

Random Forest dan Interpretasi SHAP pada Data SSGI Tahun 2024

Cindy Nadila Putri

Halaman Ringkasan berisi uraian singkat tentang latar belakang masalah, rumusan masalah, tujuan, metodologi penelitian, hasil dan analisis data, serta kesimpulan dan saran. Isi ringkasan tidak lebih dari 1000 kata (sekitar maksimal 2 halaman).

 Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

 Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

ABSTRAK

Analisis Determinan Penyebab *Stunting* Provinsi di Indonesia: Aplikasi Model
Random Forest dan Interpretasi SHAP pada Data SSGI Tahun 2024

Cindy Nadila Putri

Halaman ABSTRAK berisi uraian tentang latar belakang, tujuan, metodologi penelitian, hasil / kesimpulan. Ditulis dalam BAHASA INDONESIA tidak lebih dari 250 kata, dengan jarak antar baris satu spasi. Pada akhir abstrak ditulis kata “Kata Kunci” yang dicetak tebal, diikuti tanda titik dua dan kata kunci yang tidak lebih dari 5 kata. Kata kunci terdiri dari kata-kata yang khusus menunjukkan dan berkaitan dengan bahan yang diteliti, metode/instrumen yang digunakan, topik penelitian. Kata kunci diketik pada jarak dua spasi dari baris akhir isi abstrak.

Kata Kunci: **kunci1, kunci2**

ABSTRACT

Determinant Analysis of Stunting Causes Across Indonesian Provinces:
Application of Random Forest Model and SHAP Interpretation on SSGI 2024
Data
Cindy Nadila Putri

Halaman ABSTRACT berisi uraian tentang latar belakang, tujuan, metodologi penelitian, hasil / kesimpulan. Ditulis dalam BAHASA INGGRIS tidak lebih dari 250 kata, dengan jarak antar baris satu spasi. Secara khusus, kata dan kalimat pada halaman ini tidak perlu ditulis dengan huruf miring meskipun menggunakan Bahasa Inggris, kecuali terdapat huruf asing lain yang ditulis dengan huruf miring (misalnya huruf Latin atau Greek, dll). Pada akhir abstract ditulis kata “Keywords” yang dicetak tebal, diikuti tanda titik dua dan kata kunci yang tidak lebih dari 5 kata. Keywords terdiri dari kata-kata yang khusus menunjukkan dan berkaitan dengan bahan yang diteliti, metode/instrumen yang digunakan, topik penelitian. Keywords diketik pada jarak dua spasi dari baris akhir isi abstrak.

Keywords: keywords1, keywords2

DAFTAR ISI

LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PERSETUJUAN PUBLIKASI	iv
KATA PENGANTAR	v
RINGKASAN	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiii
DAFTAR KODE	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 Tinjauan Pustaka	8
2.2 Dasar Teori	11

2.2.1	Teori 1	11
2.2.1.1	Subsubbab	12
2.2.2	Teori 2	12
BAB III METODE PENELITIAN	14
3.1	Alur Penelitian	14
3.2	Langkah Penelitian	15
3.2.1	Identifikasi Masalah	16
3.2.2	Studi Literatur	16
3.2.3	Penyiapan Data	17
3.2.3.1	Ekstraksi Data.....	18
3.2.3.2	Seleksi Fitur Determinan.....	19
3.2.3.3	Transformasi dan Agregasi Data	20
3.2.4	Pra-pemrosesan Data	20
3.2.5	Pembangunan Model <i>Random Forest</i>	21
3.2.5.1	Validasi Silang dan Optimasi Hiperparameter	22
3.2.5.2	Pelatihan Model Final.....	23
3.2.6	Evaluasi Performa Model	23
3.3	Interpretasi Hasil dengan SHAP	25
3.3.1	Interpretasi Global	26
3.3.2	Interpretasi Lokal	27
3.3.2.1	Pemilihan Sampel Provinsi	27
3.3.2.2	Visualisasi dan Analisis <i>Waterfall Plot</i>	28
3.4	Alat dan Bahan	28
3.4.1	Alat Penelitian	28
3.4.2	Bahan Penelitian	29
3.5	Metode Pengembangan	30
3.6	Ilustrasi Perhitungan Evaluasi.....	31
3.7	Rancangan Pengujian	32

BAB IV HASIL DAN PEMBAHASAN	33
4.1 Hasil Penelitian	33
4.2 Hasil Pengujian	33
4.3 Analisis Hasil Penelitian	34
BAB V KESIMPULAN DAN SARAN	36
5.1 Kesimpulan	36
5.2 Saran	36
DAFTAR PUSTAKA	37
LAMPIRAN	39
A Dataset	39
B Hasil Wawancara	39
C Rincian Kasus Uji	39

DAFTAR TABEL

Tabel 2.1 Literasi Penelitian Terdahulu.....	10
Tabel 2.2 Contoh Tabel.....	11
Tabel 3.1 Distribusi Variabel Determinan Stunting berdasarkan Kategori.....	30
Tabel 4.1 Data <i>dummy</i> Pengujian.....	34

DAFTAR GAMBAR

Gambar 2.1	Contoh gambar dan caption	12
Gambar 3.1	Diagram Alir Penelitian.....	14
Gambar 3.2	Alur Penyiapan Data.....	18
Gambar 3.3	Alur Pra-pemrosesan Data.....	21
Gambar 3.4	Alur Pembangunan Model <i>Random Forest Regression</i> ...	22
Gambar 3.5	Alur Interpretasi Model dengan SHAP	26
Gambar 4.1	Contoh Graf Pengujian	34

DAFTAR RUMUS

Rumus 2.1 Rumus sederhana	12
Rumus 2.2 Mean Absolute Error (MAE)	12
Rumus 2.3 Distribusi Normal	13
Rumus 2.4 Sistem persamaan linier	13

DAFTAR KODE

Kode 4.1 Akuisisi Gambar	33
------------------------------------	----

BAB I

PENDAHULUAN

1.1 Latar Belakang

Manusia memerlukan kondisi tubuh yang sehat untuk dapat beraktivitas secara maksimal. Saat ini salah satu masalah kesehatan yang banyak terjadi di Indonesia adalah kendala *stunting*. Masalah ini telah menjadi perhatian serius karena dampaknya yang meluas, terutama pada anak-anak sebagai generasi penerus bangsa. Berbagai faktor, seperti pola hidup, akses terhadap makanan bergizi, serta kualitas pelayanan kesehatan, turut berkontribusi terhadap tingginya prevalensi masalah ini.

Penyakit *stunting* merupakan kondisi di mana tubuh mengalami kekurangan gizi secara berlebihan dan terjadi pada rentang waktu yang cukup lama [1]. Dampak dari masalah ini akan mengakibatkan kendala pertumbuhan pada anak sehingga tinggi badan anak cenderung menjadi lebih pendek. Tak hanya memengaruhi pertumbuhan fisik, penyakit *stunting* juga akan berpengaruh ke dalam aspek pertumbuhan lainnya seperti mental, intelektual, dan kognitif anak [2]. Oleh karena itu, penting untuk memahami faktor-faktor yang menyebabkan terjadinya *stunting* agar dapat mengambil langkah-langkah pencegahan yang efektif.

Di Indonesia, kasus *stunting* terjadi pada balita usia 0-5 tahun berada pada persentase sebesar 19,8% menurut Survei Status Gizi Indonesia (SSGI) [3]. Meskipun telah mengalami penurunan dari tahun sebelumnya yaitu angka 21,5%, hasil ini menunjukkan bahwa target pemerintah Indonesia yaitu menurunkan prevalensi *stunting* sampai 14,4% di tahun 2029 masih belum tercapai [4], [5]. *Stunting* dapat berasal dari faktor-faktor yang sangat kompleks dilihat dari aspek sosial, biologis, maupun lingkungan. Biasanya, penyebab utama pada *stunting* dapat terjadi karena kurangnya asupan gizi, buruknya sanitasi,

serta akses yang rendah terhadap pelayanan kesehatan [6], [7]. Terdapat faktor penting lain yang menyebabkan terjadinya stunting yaitu kurangnya edukasi ibu tentang betapa penting untuk menjaga gizi seimbang pada masa kehamilan, masa menyusui, dan masa pertumbuhan anak [7]. Pada beberapa kasus, ibu hamil yang melahirkan bayi dengan kondisi berat badan lahir rendah (BBLR) mengalami fase kekurangan gizi selama masa kehamilannya [7]. Hal ini dapat mengakibatkan terjadinya *stunting* di kemudian hari.

Tingkat *stunting* di Indonesia sendiri dinilai tidak merata di seluruh wilayah. Beberapa provinsi seperti Bali dan DI Yogyakarta menunjukkan angka prevalensi yang relatif rendah, berbanding terbalik dengan provinsi Nusa Tenggara Timur dan Sulawesi Barat yang mencatat angka jauh lebih tinggi [3]. Situasi yang timpang ini menunjukkan adanya variasi faktor determinan yang memengaruhi kejadian *stunting* di tiap daerah, baik dari segi sosial, ekonomi, pendidikan, maupun kondisi lingkungan [8]. Faktor lainnya seperti status gizi ibu, akses terhadap layanan kesehatan, dan sanitasi dasar memiliki kontribusi yang berbeda-beda pada setiap wilayah. Adanya variasi faktor risiko ini mengindikasikan bahwa pendekatan penanganan yang bersifat generalis memiliki keterbatasan dalam menjangkau akar masalah di tiap daerah, sehingga dibutuhkan landasan data yang spesifik per wilayah untuk mendukung perumusan strategi intervensi yang lebih presisi dan tepat sasaran. Maka dari itu, diperlukan analisis yang lebih mendalam untuk memahami faktor-faktor yang paling berpengaruh untuk situasi *stunting* tiap provinsi.

Penelitian sebelumnya menerapkan algoritma *machine learning* untuk memprediksi *stunting* pada kalangan remaja di Ethiopia dan menemukan bahwa metode konvensional kurang mampu menangkap interaksi kompleks antarvariabel [8]. Namun penelitian tersebut dilakukan secara terbatas pada kelompok usia remaja dan wilayah tertentu, sehingga belum secara langsung dapat digeneralisasi dan masih memerlukan kajian tersendiri untuk skala nasional Indonesia yang mencakup berbagai karakteristik. Dengan demikian,

penting untuk mengeksplorasi pendekatan serupa pada skala provinsi di Indonesia untuk memahami faktor determinan *stunting* berdasarkan kondisi lokal.

Dalam beberapa tahun terakhir, pendekatan *machine learning* mulai banyak digunakan dalam bidang kesehatan dan gizi untuk mengidentifikasi pola tersembunyi dalam data yang kompleks [9]. Salah satu algoritma yang populer adalah *Random Forest*, yang dikenal memiliki kemampuan tinggi dalam menangani variabel dalam jumlah besar serta menghasilkan prediksi yang akurat. Keunggulan lain dari *Random Forest* adalah kemampuannya dalam mengukur tingkat kepentingan setiap fitur (*feature importance*), sehingga dapat membantu memahami faktor-faktor mana yang paling berpengaruh terhadap suatu fenomena. Namun demikian, model ini sering dianggap sebagai *black box* karena sulit dijelaskan secara langsung oleh pengambil kebijakan atau peneliti non-teknis.

Untuk mengatasi keterbatasan interpretasi tersebut, digunakan metode interpretabilitas seperti SHAP (*SHapley Additive exPlanations*) yang mampu menjelaskan kontribusi masing-masing fitur terhadap hasil prediksi model [9]. Melalui pendekatan ini, setiap faktor dapat dinilai apakah ia meningkatkan atau menurunkan kemungkinan terjadinya *stunting* pada suatu wilayah. Dengan demikian, kombinasi antara *Random Forest* dan SHAP tidak hanya memberikan hasil prediksi yang akurat, tetapi juga penjelasan yang dapat dipahami secara intuitif oleh pengambil keputusan. Pendekatan ini memberikan nilai tambah dalam analisis kebijakan berbasis data, terutama untuk menentukan prioritas intervensi di daerah dengan tingkat *stunting* tinggi.

Menjawab permasalahan tersebut, penelitian ini diarahkan untuk mengidentifikasi faktor-faktor determinan utama yang memengaruhi prevalensi *stunting* antarprovinsi di Indonesia menggunakan data SSGI tahun 2024 melalui pendekatan *Random Forest* yang diinterpretasikan dengan SHAP. Penelitian ini berfokus pada dua aspek utama, yaitu menilai performa model dalam

memprediksi tingkat *stunting* serta memahami peran masing-masing variabel terhadap hasil prediksi. Hasil penelitian diharapkan dapat memberikan dasar empiris bagi pemerintah dan pemangku kebijakan dalam merancang strategi intervensi gizi yang lebih efektif dan berbasis bukti ilmiah.

1.2 Rumusan Masalah

Berdasarkan pemaparan latar belakang, dihasilkan rumusan masalah pada penelitian yang akan dijabarkan pada poin berikut:

1. Apa saja faktor-faktor yang menjadi determinan utama penyebab *stunting* antarprovinsi di Indonesia berdasarkan data SSGI tahun 2024?
2. Bagaimana performa model *Random Forest* dalam memprediksi prevalensi *stunting* antarprovinsi di Indonesia?
3. Bagaimana interpretasi SHAP digunakan untuk menjelaskan kontribusi masing-masing faktor terhadap hasil prediksi model *Random Forest*?

1.3 Tujuan Penelitian

Setelah penjabaran rumusan masalah, pada penelitian ini dibuat tujuan penelitian yaitu sebagai berikut:

1. Mengidentifikasi faktor-faktor determinan utama penyebab *stunting* antarprovinsi di Indonesia berdasarkan data SSGI tahun 2024.
2. Menganalisis performa model *Random Forest* dalam memprediksi prevalensi *stunting* antarprovinsi di Indonesia.
3. Menerapkan interpretasi SHAP untuk menjelaskan kontribusi masing-masing faktor terhadap hasil prediksi model *Random Forest*.

1.4 Batasan Masalah

Agar penelitian tidak terlalu luas dan tidak keluar dari pokok permasalahan, maka ditentukan batasan oleh beberapa poin berikut:

1. Data yang digunakan untuk penelitian ini terbatas pada Survei Status Gizi Indonesia (SSGI) tahun 2024 yang telah dipublikasikan oleh Kementerian

Kesehatan RI. Perlu diketahui bahwa penelitian ini tida menggunakan data tambahan dari sumber lain. Sehingga penelitian yang dilakukan bukan analisis *time-series* antartahun dan memiliki ruang lingkup spasial terbatas pada 34 provinsi di Indonesia sesuai ketersediaan data.

2. Populasi yang diambil untuk menjadi fokus penelitian ini adalah anak balita usia 0-59 bulan (5 tahun), sesuai dengan sasaran pengukuran prevalensi *stunting* pada SSGI.
3. Variabel yang digunakan untuk penelitian ini mencakup berbagai faktor yang memiliki potensi untuk memengaruhi prevalensi penyebab *stunting*, seperti karakteristik balita, karakteristik ibu, serta kondisi rumah tangga yang meliputi status gizi, pendidikan, ekonomi, dan akses terhadap fasilitas kesehatan, air bersih, serta sanitasi. variabel-variabel ini merupakan sebagian dari indikator yang tersedia dalam dataset SSGI tahun 2024 dan dipilih berdasarkan relevansi terhadap topik penelitian.
4. Analisis yang dilakukan pada penelitian ini difokuskan pada tingkat provinsi di Indonesia, sehingga hasilnya mencerminkan kondisi agregat dan tidak mempertimbangkan variasi di tingkat kabupaten/kota atau individu.
5. Model yang digunakan untuk penelitian ini adalah algoritma *Random Forest* untuk memprediksi prevalensi *stunting*, dengan interpretasi hasil menggunakan metode SHAP (*SHapley Additive exPlanations*) untuk memahami kontribusi setiap variabel terhadap prediksi prevalensi *stunting*.
6. Penelitian ini berfokus pada analisis faktor determinan penyebab *stunting*, bukan pada pengembangan sistem prediksi *stunting* secara *real-time* atau aplikasi praktis lainnya.

1.5 Manfaat Penelitian

Penelitian ini dilaksanakan dengan harapan agar nantinya dapat memiliki manfaat sebagai berikut:

1. Bagi masyarakat, penelitian ini diharapkan dapat memberi gambaran empiris mengenai faktor-faktor utama yang memiliki kontribusi terhadap *stunting* di berbagai provinsi di Indonesia terutama di Sumatera. Hasil penelitian diharapkan dapat menjadi dasar bagi tenaga kesehatan dan pemerintah dalam menyusun strategi intervensi gizi yang lebih terarah dan sesuai dengan karakteristik wilayah masing-masing.
2. Bagi peneliti, penelitian ini bermanfaat untuk mengembangkan pemahaman mengenai penerapan algoritma *machine learning*, khususnya *Random Forest* dengan interpretasi SHAP dalam menganalisis data kesehatan masyarakat.
3. Bagi akademisi, dapat dijadikan referensi mahasiswa lain yang ingin meneliti terkait faktor-faktor penyebab *stunting* maupun penerapan model *machine learning* untuk analisis data survei kesehatan.

1.6 Sistematika Penulisan

Secara keseluruhan, struktur penulisan laporan ini disusun untuk memberikan gambaran umum setiap tahapan penelitian mengenai analisis determinan penyebab *stunting* di provinsi-provinsi di Indonesia. Setiap bab disusun secara sistematis agar pembahasan mengenai penerapan model *Random Forest* dan interpretasi SHAP pada data SSGI tahun 2024 dapat diikuti dengan runtut dan mudah dipahami.

Bab I

Bab I berisi pendahuluan. Pendahuluan menguraikan latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah dan sistematika penulisan dari penelitian.

Bab II

Bab II berisi tinjauan pustaka. Pada bab ini akan dibahas teori-teori mengenai *stunting*, faktor-faktor determinan penyebab *stunting*, Survei Status

Gizi Indonesia (SSGI), serta penerapan algoritma *machine learning* khususnya *Random Forest* pada model prediksi risiko *stunting*. Selain itu, bab ini juga membahas hasil penelitian terdahulu yang relevan sebagai landasan konseptual dan pembeda penelitian ini dari studi yang telah dilakukan sebelumnya.

Bab III

Bab ini menjelaskan metode yang akan digunakan dalam penelitian, termasuk sumber dan jenis data, variabel penelitian, tahapan pengolahan data, serta rancangan model analisis menggunakan *Random Forest* dan SHAP. Selain itu, bab ini juga menjelaskan mengenai teknik evaluasi model serta prosedur interpretasi hasil untuk mengidentifikasi faktor determinan utama penyebab *stunting*.

Bab IV

Bab ini menguraikan hasil analisis data, performa model *Random Forest*, hasil interpretasi SHAP, serta pembahasan mengenai faktor-faktor determinan yang memiliki pengaruh prevalensi *stunting* antarprovinsi di Indonesia.

Bab V

Bab V berisi Kesimpulan dan saran. Pada bab ini akan diberikan kesimpulan dari hasil penelitian yang menjawab rumusan masalah berdasarkan hasil analisis yang telah dilakukan sebelumnya. Saran-saran juga diberikan sebagai masukan bagi pihak terkait dan peneliti selanjutnya untuk pengembangan kebijakan dan penelitian lanjutan di bidang *stunting*.

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Berisi penelitian terdahulu yang menjadi konsep / pendukung penelitian yang dilakukan. Lakukan pembahasan secara sistematis dengan menjelaskan masalah apa yang diangkat di penelitian terdahulu, metode yang digunakan, kontribusi yang diberikan, serta analisis penulis terkait dengan keunggulan atau keterbatasannya. Tuangkan perbandingan penelitian terdahulu dengan penelitian yang akan dikerjakan, minimal 5 jurnal pembanding (4-5 tahun terakhir). Kemudian penulis sebaiknya melakukan rangkuman terutama terkait dengan peluang pengembangan atau tugas akhir yang akan dilakukan

Perujukan literatur dapat dilakukan dengan menambahkan entri baru dalam file `references.bib`. Cara merujuk sitasi menggunakan `\cite{nama_label_sitasi}`. Hasil sitasi seperti ini: **knuth2001art, Vogels2006Am**, atau **Suryanto2019MAEIVanny2014Banding**. Daftar Pustaka hanya akan memunculkan sitasi yang direferensikan menggunakan command `\cite{}`.

Tuliskan **judul jurnal, penulis jurnal, tahun jurnal terbit, penjelasan isi jurnal, metode penelitian, hasil penelitian & pengujian**.

1. Sistem Informasi Pendaftaran Haji dan Umroh Di PT Multazam Sriwijaya Barakah Palembang Menggunakan Metode Rapid Application Development. *Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer*

adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

2. Sistem Informasi Umroh Di PT XYZ Lampung Menggunakan Metode Rapid Application Development. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Tabel ringkasan tinjauan pustaka ditulis setelah penjelasan masing-masing jurnal.

Tabel 2.1 Literasi Penelitian Terdahulu

No.	Judul	Masalah	Metode	Hasil
1.	Sistem Informasi Pendaftaran Haji dan Umroh Di PT Multazam Sriwijaya Barakah Palembang Menggunakan Metode Rapid Application Development	Belum adanya sistem untuk pendaftaran haji & umrah	Rapid Application Development	Sistem Informasi Pendaftaran Haji dan Umroh di PT Multazam Sriwijaya Barakah Palembang
2.	Sistem Informasi Umroh Di PT XYZ Lampung Menggunakan Metode Rapid Application Development	Belum adanya sistem untuk pendaftaran haji & umrah	Rapid Application Development	Sistem Informasi Pendaftaran Umroh di PT XYZ Lampung

No.	Judul	Masalah	Metode	Hasil
3.	Sistem Informasi Umroh Di PT XYZ Lampung Menggunakan Metode Rapid Application Development	Belum adanya sistem untuk pendaftaran haji & umrah	Rapid Application Development	Sistem Informasi Pendaftaran Umroh di PT XYZ Lampung

2.2 Dasar Teori

Berisi teori/konsep yang berkaitan/digunakan dalam tugas akhir yang dikerjakan. Gunakanlah data melalui buku/jurnal referensi, publikasi tugas akhir, penelitian, buku, dan informasi web yang dapat dipertanggungjawabkan, hindari penggunaan dasar teori melalui tautan Wikipedia, surat kabar, atau portal berita, yang dapat memiliki isi yang tidak bersifat fakta.

2.2.1 Teori 1

Berikut adalah contoh penyisipan tabel menggunakan \begin{longtable}{}:

Tabel 2.2 Contoh Tabel

Col1	Col2	Col2	Col3
1	6	87837	787
2	7	78	5415
3	545	778	7507
4	545	18744	7560
5	88	788	6344

2.2.1.1 Subsubbab

Berikut adalah contoh subsubbab. Ini adalah level subbab maksimal dalam laporan Tugas Akhir, dan tidak boleh lebih dalam.

Gambar 2.1 adalah contoh Gambar yang diambil dari internet yang harus dicantumkan sumbernya dan memiliki lisensi Creative Common. Jika gambar adalah milik peneliti lain atau tidak dibuat atau diambil sendiri maka peneliti wajib meminta izin kepada peneliti lain tersebut untuk mencantumkan gambar. Gunakan `\begin{figure}` untuk memasukkan gambar. Gunakan `\caption{[nama caption]}` untuk memberikan caption gambar. Nomor caption akan diurutkan secara otomatis. Jangan lupa untuk melabeli setiap gambar dengan `\label{[nama label]}`, agar bisa direferensi menggunakan `\ref{[nama label]}`



Gambar 2.1 Contoh gambar dan caption

Sumber: Contoh

2.2.2 Teori 2

Untuk membuat sebuah rumus persamaan, gunakan kode `\begin{equationcaptioned}` seperti dibawah:

$$x + 1 = 2$$

(Rumus 2.1)

Teks caption rumus tidak akan muncul di teks, tetapi akan muncul di Daftar Rumus.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

(Rumus 2.2)

Berikut adalah contoh penulisan persamaan yang lebih kompleks, yaitu persamaan distribusi normal.

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

(Rumus 2.3)

Jika menuliskan banyak persamaan secara berurutan, gunakan `\begin{split}`:

$$2x - 5y = 8$$

$$3x + 9y = -12$$

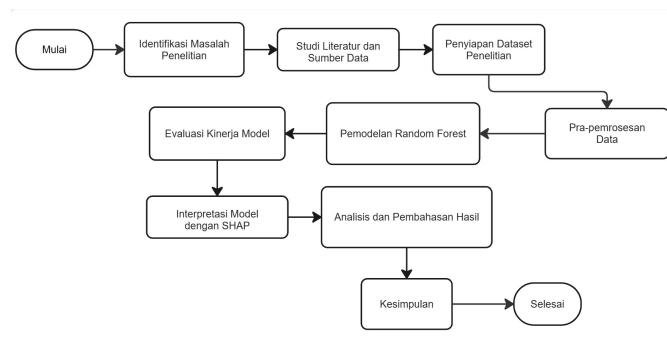
(Rumus 2.4)

BAB III

METODE PENELITIAN

3.1 Alur Penelitian

Penelitian ini dirancang untuk menganalisis faktor-faktor determinan penyebab *stunting* pada tingkat provinsi di Indonesia dengan memanfaatkan data sekunder *Annual Report Survei Status Gizi Indonesia (SSGI)* tahun 2024. Pendekatan yang digunakan mengombinasikan pemodelan *machine Learning* regresi menggunakan *Random Forest* dan metode interpretasi model SHAP (*SHapley Additive exPlanations*). Model ini dibangun menggunakan dataset agregat tingkat provinsi yang memuat berbagai indikator status gizi balita serta faktor determinan yang mencakup aspek kesehatan ibu dan anak, pola asuh, layanan kesehatan, penyakit, perlindungan sosial, dan kondisi sanitasi. Penelitian ini bertujuan untuk mengidentifikasi faktor determinan utama yang berkontribusi terhadap tingginya prevalensi *stunting* antar provinsi, serta memberikan pemahaman yang bersifat interpretatif sebagai dasar perumusan kebijakan dan intervensi gizi yang lebih tepat sasaran. Secara garis besar, tahapan penelitian digambarkan dalam diagram alir (*flowchart*) pada Gambar 3.1.



Gambar 3.1 Diagram Alir Penelitian

Berdasarkan Gambar 3.1, alur penelitian disusun secara sistematis yang diawali dengan tahap identifikasi masalah mengenai urgensi penanganan *stunting* di Indonesia dan kompleksitas faktor determinan yang mempengaruhinya di tingkat provinsi. Tahap ini didukung oleh studi literatur mendalam untuk memahami variabel-variabel determinan multidimensi yang mencakup berbagai aspek, serta metode pengolahan data dan pemodelan *machine learning* yang relevan. Berlandaskan pemahaman tersebut, selanjutnya dilakukan pengumpulan dataset yang bersumber dari Laporan Tahunan Survei Status Gizi Indonesia (SSGI) tahun 2024. Laporan Tahunan SSGI 2024 kemudian digunakan untuk menyiapkan dataset agregat tingkat provinsi yang memuat variabel target (prevalensi *stunting*) dan variabel fitur (faktor determinan). Data mentah yang diperoleh selanjutnya diproses melalui tahapan pra-pemrosesan data yang meliputi pembersihan data, penyaringan data, dan pemisahan variabel target dan fitur.

Setelah data siap, tahap berikutnya adalah pembangunan model regresi menggunakan algoritma *Random Forest*. Proses ini mencakup penentuan parameter terbaik melalui *Grid Search* dan validasi silang (*Cross Validation*) untuk memastikan performa model yang optimal. Model yang telah dilatih kemudian dievaluasi kinerjanya menggunakan metrik statistik seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), serta koefisien determinasi (R^2 Score). Langkah terakhir adalah menginterpretasikan hasil prediksi model *Random Forest* menggunakan SHAP untuk mengetahui kontribusi setiap fitur determinan terhadap angka *stunting*, baik secara global skala nasional maupun lokal per provinsi.

3.2 Langkah Penelitian

Tahapan penelitian disusun secara sistematis untuk menjawab rumusan masalah yang telah ditetapkan. Secara garis besar, alur penelitian terdiri

dari identifikasi masalah, studi literatur, penyiapan data, pra-pemrosesan, pembangunan model, hingga evaluasi dan interpretasi hasil. Penjelasan rinci mengenai masing-masing tahapan diuraikan sebagai berikut.

3.2.1 Identifikasi Masalah

Tahap awal penelitian ini dilakukan dengan mengidentifikasi permasalahan mengenai prevalensi *stunting* di Indonesia serta kompleksitas faktor determinan yang mempengaruhinya. Identifikasi masalah dalam penelitian ini dilakukan melalui tahapan observasi statistik awal terhadap distribusi dataset SSGI 2024 yang menyajikan data *stunting* dalam berbagai aspek determinan, seperti kesehatan ibu, pelayanan kesehatan, pola asuh, konsumsi pangan, lingkungan, serta faktor sosial ekonomi. Keberagaman indikator tersebut memperlihatkan bahwa permasalahan *stunting* bersifat multidimensional dan tidak hanya dipengaruhi oleh faktor tunggal. Temuan ini mendorong munculnya pertanyaan penelitian mengenai kesesuaian strategi intervensi gizi yang bersifat seragam dalam menurunkan prevalensi *stunting* di seluruh provinsi di Indonesia. Berdasarkan kondisi tersebut, penelitian ini ditujukan untuk mengidentifikasi faktor determinan utama penyebab *stunting* pada tingkat provinsi menggunakan pendekatan analisis data interpretasi berbasis *machine learning*.

3.2.2 Studi Literatur

Pada tahap studi literatur, dilakukan identifikasi mengenai pendekatan, metode, dan jenis data yang digunakan dalam penelitian terkait *stunting* terdahulu. Hal ini dilakukan untuk menemukan celah penelitian yang dapat diisi melalui pendekatan baru yang diusulkan. Hasil tinjauan literatur menunjukkan bahwa sebagian besar studi menerapkan metode prediksi *stunting* dengan algoritma *machine learning*, terutama algoritma *Random Forest*, *Gradient Boosting*, dan *XGBoost* dengan fokus utama pada peningkatan akurasi prediksi. Selain itu, sebagian besar penelitian *machine learning* mengenai *stunting* dilakukan

menggunakan data individu atau wilayah yang terbatas dengan fokus utama pada peningkatan akurasi prediksi. Penelitian yang menggunakan data agregat tingkat provinsi masih sangat minim, begitu pula dengan penggunaan metode interpretasi model untuk memahami kontribusi faktor determinan penyebab *stunting*.

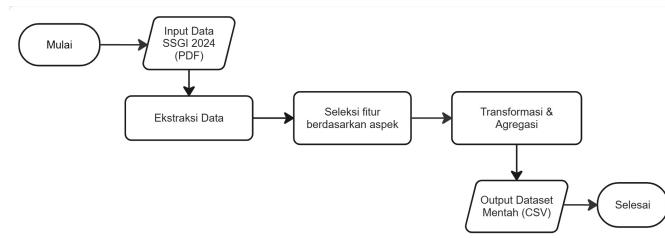
Berdasarkan temuan tersebut, penelitian ini mengadopsi pendekatan pemodelan *machine learning* regresi menggunakan algoritma *Random Forest* yang dipadukan dengan metode interpretasi SHAP untuk mengisi celah penelitian yang ada. Pendekatan ini diharapkan dapat memberikan wawasan baru mengenai faktor-faktor determinan utama penyebab *stunting* pada tingkat agregat provinsi menggunakan data SSGI 2024 yang masih terbatas penggunaannya. Tahap studi literatur juga dilakukan sebagai dasar pemilihan teori dan metode yang akan diterapkan dalam penelitian.

3.2.3 Penyiapan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Laporan Publikasi Survei Status Gizi Indonesia (SSGI) tahun 2024 yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia [3]. Laporan ini disajikan dalam bentuk dokumen PDF dan memuat informasi statistik yang komprehensif, mencakup ±5.189 tabel statistik yang menggambarkan indikator kesehatan ibu dan anak, pola asuh, akses pelayanan kesehatan, kondisi sanitasi dan lingkungan, serta karakteristik demografi rumah tangga. Secara spesifik, populasi target dalam survei ini difokuskan pada rumah tangga kelompok balita usia 0-59 bulan, dengan data yang disajikan menggunakan pendekatan *cross-sectional* untuk menggambarkan kondisi kesehatan pada satu titik waktu tertentu. Seluruh informasi statistik tersebut disajikan dalam bentuk numerik atau persentase yang merepresentasikan prevalensi kejadian di wilayah terkait.

Informasi dalam laporan SSGI 2024 disajikan dalam berbagai tingkat agregasi wilayah, mulai dari data nasional, provinsi, hingga rincian

kabupaten/kota. Dari berbagai tingkatan tersebut, penelitian ini membatasi unit analisis pada skala provinsi guna menganalisis variabilitas determinan *stunting* antarwilayah secara makro. Namun, kompleksitas struktur data ini menyebabkan data tidak dapat langsung digunakan sebagai *dataset* penelitian kuantitatif tanpa melalui proses penyiapan data terlebih dahulu. Oleh karena itu, diperlukan suatu tahapan penyiapan data untuk mengekstraksi, menyeleksi, dan menyusun kembali informasi yang relevan agar diperoleh *dataset* terstruktur yang sesuai dengan tujuan penelitian. Alur penyiapan data yang dilakukan dalam penelitian ini ditunjukkan pada Gambar 3.2.



Gambar 3.2 Alur Penyiapan Data

Alur penyiapan data pada Gambar 3.2 menggambarkan tahapan yang dilakukan untuk mengonversi data mentah SSGI 2024 dalam bentuk PDF menjadi *dataset* terstruktur yang siap digunakan pada tahap pra-pemrosesan dan pemodelan. Penjelasan lebih lanjut mengenai masing-masing tahapan penyiapan data diuraikan sebagai berikut:

3.2.3.1 Ekstraksi Data

Tahap pertama difokuskan pada pengumpulan data mentah dari laporan SSGI 2024 yang tersedia dalam bentuk dokumen PDF dengan ribuan tabel statistik yang kompleks. Mengingat data tersebar dalam berbagai sub-bab laporan, dilakukan identifikasi manual terlebih dahulu untuk memetakan nomor halaman tabel yang memuat indikator determinan *stunting* dan prevalensi gizi balita. Informasi dari tabel-tabel tersebut kemudian diekstraksi dan dikonversi

dari bentuk dokumen statis menjadi lembar kerja digital (*spreadsheet*) yang lebih fleksibel. Karena data asli menggunakan pendekatan *cross-sectional* dengan banyak kategori, proses ekstraksi dilakukan dengan memastikan angka persentase yang diambil sesuai dengan label baris dan kolom yang benar.

Pada tahap ini juga ditetapkan skala analisis yang digunakan dalam penelitian, yaitu skala provinsi sebagai representasi kondisi nasional. Pemilihan skala provinsi dilakukan dengan mempertimbangkan tujuan penelitian yang berfokus pada pemetaan faktor determinan *stunting* secara makro. Berdasarkan keputusan tersebut, seluruh tabel yang dianalisis dibatasi pada tabel tingkat provinsi yang tersedia dalam laporan SSGI 2024.

3.2.3.2 Seleksi Fitur Determinan

Tahap seleksi fitur dilakukan untuk menentukan variabel determinan *stunting* yang relevan dari 100 tabel tingkat provinsi hasil ekstraksi data SSGI 2024. Proses seleksi ini mempertimbangkan kesesuaian variabel dengan tujuan penelitian, keterwakilan indikator determinan, serta potensi redundansi informasi antar tabel. Variabel yang bersifat terlalu rinci, deskriptif kategori, atau tidak lagi informatif pada tingkat agregasi provinsi dieliminasi dan, apabila memungkinkan, direpresentasikan melalui variabel status atau skor risiko komposit.

Pendekatan seleksi ini diterapkan untuk menyesuaikan karakteristik data dengan tingkat analisis penelitian, yaitu pada level provinsi, sehingga kompleksitas yang tidak diperlukan dapat dikurangi tanpa menghilangkan makna substantif indikator. Berdasarkan proses tersebut, sebanyak 71 tabel dipertahankan dan digunakan sebagai dasar pembentukan fitur penelitian. Hasil seleksi ini menghasilkan *dataset* yang lebih ringkas, informatif, dan relevan dalam mengidentifikasi faktor-faktor determinan *stunting* antarprovinsi.

3.2.3.3 Transformasi dan Agregasi Data

Setelah melalui proses seleksi tabel, langkah selanjutnya adalah transformasi data untuk memastikan setiap variabel yang terbentuk memiliki representasi informasi yang kuat terhadap risiko *stunting*. Pada tahap ini, tidak semua informasi dari tabel mentah digunakan secara langsung; sebaliknya, dilakukan peninjauan ulang untuk menetapkan satu indikator yang paling representatif dari setiap tabel. Untuk tabel yang hanya memuat indikator tunggal, nilai persentase diadopsi langsung sebagai fitur numerik. Namun, pada tabel yang memiliki kategori majemuk, dilakukan strategi transformasi khusus dengan memilih indikator yang mencerminkan risiko tertinggi atau melakukan penggabungan kategori agar menghasilkan variabel baru yang lebih ringkas dan substantif.

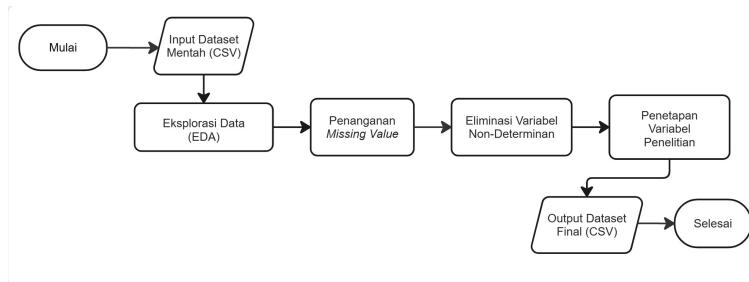
Selanjutnya, proses agregasi dilakukan untuk menyederhanakan struktur data sehingga seluruh variabel memiliki skala yang seragam dan mudah dibandingkan antarprovinsi. Seluruh variabel direpresentasikan dalam bentuk persentase guna menjaga konsistensi satuan data dan memudahkan proses pemodelan. Hasil dari tahap ini adalah *dataset* terstruktur berbasis provinsi yang siap digunakan sebagai input pada tahapan analisis selanjutnya.

3.2.4 Pra-pemrosesan Data

Setelah proses penyiapan dan pembentukan *dataset* terstruktur selesai dilakukan, tahap selanjutnya adalah pra-pemrosesan data. Tahapan ini bertujuan untuk memastikan bahwa data yang digunakan sebagai input model memiliki kualitas yang baik, konsisten, dan layak secara statistik sebelum masuk ke proses pemodelan *machine learning*. Mengingat *dataset* disusun dari hasil ekstraksi dan transformasi data sekunder, terdapat potensi munculnya permasalahan seperti nilai kosong atau ketidakkonsistenan format.

Alur pra-pemrosesan data ditunjukkan pada Gambar 3.3. Proses ini diawali dengan pemuatan *dataset* ke dalam lingkungan komputasi serta pemeriksaan

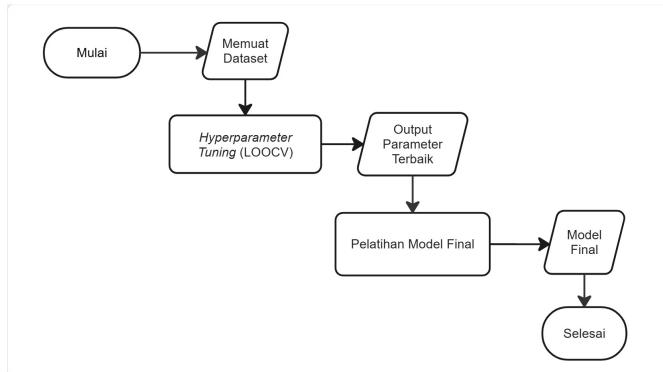
struktur dan konsistensi data. Selanjutnya dilakukan *Exploratory Data Analysis* (EDA) untuk memperoleh gambaran umum karakteristik data. Tahap akhir pra-pemrosesan mencakup eliminasi variabel non-determinan serta penetapan prevalensi *stunting* sebagai variabel target (*y*) dan indikator determinan sebagai variabel prediktor (*X*).



Gambar 3.3 Alur Pra-pemrosesan Data

3.2.5 Pembangunan Model *Random Forest*

Setelah melalui tahapan penyiapan dan pra-pemrosesan data, *dataset* yang telah bersih dan terstruktur selanjutnya digunakan pada tahap pemodelan. Pada penelitian ini, pendekatan yang digunakan adalah *Random Forest Regression* untuk memodelkan hubungan antara variabel determinan *stunting* dan nilai prevalensi *stunting* pada tingkat provinsi. Pemilihan metode ini didasarkan pada kemampuannya dalam menangani data dengan jumlah fitur yang relatif banyak serta hubungan non-linier antarvariabel tanpa asumsi distribusi data yang ketat **referensi 'rf'**. Alur proses pemodelan disajikan secara ringkas pada Gambar 3.4.



Gambar 3.4 Alur Pembangunan Model *Random Forest Regression*

3.2.5.1 Validasi Silang dan Optimasi Hiperparameter

Tahap awal pembangunan model difokuskan pada pencarian konfigurasi model yang paling optimal melalui strategi validasi yang ketat. Mengingat keterbatasan jumlah sampel data yang hanya tersedia pada tingkat provinsi ($n=36$), penelitian ini menerapkan metode *Leave-One-Out Cross Validation* (LOOCV) sebagai pengganti pembagian data latih-uji konvensional **referensi** 'loocv'. Dalam skema ini, proses pelatihan dilakukan secara iteratif sebanyak jumlah sampel, di mana pada setiap iterasi, satu provinsi digunakan sebagai data uji sementara provinsi sisanya menjadi data latih. Pendekatan ini dipilih untuk memaksimalkan pemanfaatan informasi dari data yang terbatas serta menghasilkan estimasi performa model yang tidak bias terhadap pemilihan acak data uji tertentu.

Bersamaan dengan proses validasi LOOCV, dilakukan optimasi kinerja model melalui mekanisme *Hyperparameter Tuning* menggunakan metode *Grid Search*. Tahap ini bertujuan untuk mencari kombinasi parameter terbaik meliputi jumlah pohon (*n_estimators*), kedalaman maksimal (*max_depth*), dan jumlah sampel minimum percabangan yang menghasilkan tingkat kesalahan prediksi terendah. Eksplorasi ruang parameter ini dinilai krusial untuk mencegah terjadinya *overfitting*, kondisi di mana model terlalu "menghafal" pola data latih

namun gagal melakukan generalisasi pada data baru. Kombinasi parameter yang menghasilkan nilai rata-rata *Root Mean Squared Error* (RMSE) terendah selama proses validasi silang kemudian ditetapkan sebagai konfigurasi model optimal.

3.2.5.2 Pelatihan Model Final

Setelah parameter optimal diperoleh dari tahap validasi, langkah selanjutnya adalah pembangunan model final (*final model fitting*). Pada tahap ini, model *Random Forest* dilatih ulang (*retraining*) menggunakan keseluruhan *dataset* provinsi yang tersedia dengan menerapkan parameter terbaik hasil optimasi sebelumnya. Tujuannya adalah untuk menangkap pola data secara utuh tanpa menyisakan data untuk pengujian, karena evaluasi performa telah diselesaikan pada tahap validasi silang. Model terbaik ini kemudian digunakan untuk memetakan hubungan non-linier yang kompleks antara seluruh variabel determinan dengan prevalensi *stunting*.

Selain menghasilkan prediksi angka prevalensi, model final ini memiliki fungsi krusial untuk mengekstraksi nilai kepentingan fitur (*feature importance*). Nilai ini mengkuantifikasi kontribusi relatif setiap indikator determinan terhadap hasil prediksi model secara global. Informasi mengenai bobot kontribusi fitur inilah yang nantinya menjadi landasan utama dalam analisis prioritas intervensi penanganan *stunting* pada tahap interpretasi hasil menggunakan metode SHAP.

3.2.6 Evaluasi Performa Model

Evaluasi performa model dilakukan untuk mengukur sejauh mana model *Random Forest Regression* mampu memprediksi prevalensi *stunting* secara akurat pada tingkat provinsi. Mengingat penelitian ini bersifat regresi, digunakan beberapa metrik evaluasi berbasis kesalahan prediksi dan kekuatan penjelasan model. Penggunaan lebih dari satu metrik bertujuan untuk memberikan gambaran performa model secara komprehensif, baik dari sisi rata-rata kesalahan absolut, sensitivitas terhadap kesalahan besar, maupun kemampuan model dalam

menjelaskan variasi data target. Seluruh metrik dievaluasi pada skema validasi silang yang telah ditetapkan pada tahap pemodelan.

1. Mean Absolute Error (MAE)

MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai prevalensi *stunting* hasil prediksi model dengan nilai aktual pada masing-masing provinsi. Metrik ini memberikan gambaran kesalahan prediksi secara langsung dalam satuan persentase, sehingga memudahkan interpretasi seberapa jauh hasil prediksi model menyimpang dari data SSGI 2024 yang sebenarnya.

2. Mean Squared Error (MSE)

MSE digunakan untuk mengevaluasi kesalahan prediksi dengan memberikan penalti lebih besar pada selisih prediksi yang tinggi antarprovinsi. Penggunaan metrik ini penting untuk memastikan bahwa model tidak menghasilkan kesalahan ekstrem pada provinsi tertentu yang dapat mengaburkan pola determinan *stunting* secara nasional.

3. Root Mean Squared Error (RMSE)

RMSE digunakan sebagai indikator utama akurasi model karena mengembalikan nilai kesalahan ke dalam skala yang sama dengan prevalensi *stunting*. Dalam konteks penelitian ini, RMSE membantu menilai seberapa besar deviasi prediksi model terhadap nilai aktual *stunting* antarprovinsi secara keseluruhan.

4. R-squared (R^2)

Koefisien determinasi R^2 digunakan untuk mengukur sejauh mana variasi prevalensi *stunting* antarprovinsi dapat dijelaskan oleh variabel determinan yang dimasukkan ke dalam model. Metrik ini memberikan gambaran kemampuan model dalam menangkap hubungan kompleks antara faktor kesehatan ibu, balita, dan lingkungan dengan tingkat *stunting*.

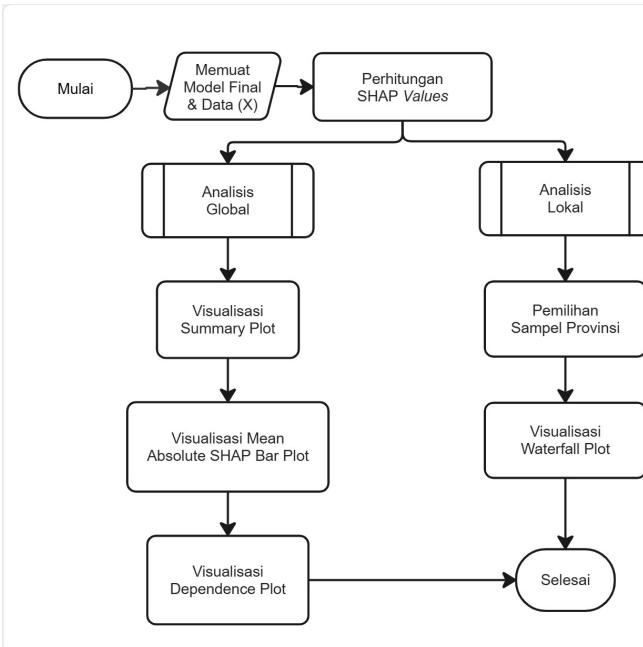
5. Mean Absolute Percentage Error (MAPE)

MAPE digunakan untuk mengukur kesalahan prediksi dalam bentuk

persentase relatif terhadap nilai aktual prevalensi *stunting*. Metrik ini membantu melihat konsistensi performa model antarprovinsi, terutama ketika terdapat perbedaan tingkat *stunting* yang cukup besar antara satu wilayah dan wilayah lainnya.

3.3 Interpretasi Hasil dengan SHAP

Setelah model final terbentuk, tahap selanjutnya adalah interpretasi hasil menggunakan metode *SHAP* (*SHapley Additive exPlanations*) untuk membuka karakteristik *black-box* dari algoritma *Random Forest*. Pendekatan ini digunakan untuk mengkuantifikasi kontribusi setiap variabel determinan terhadap prediksi prevalensi *stunting*, baik secara agregat nasional maupun spesifik per provinsi. Proses interpretasi ini dilakukan melalui perhitungan nilai *Shapley* yang mengukur dampak marginal marjinal setiap fitur terhadap *output* model dibandingkan dengan rata-rata prediksi. Alur kerja analisis interpretabilitas model dengan SHAP dalam penelitian ini diilustrasikan pada Gambar 3.5.



Gambar 3.5 Alur Interpretasi Model dengan SHAP

3.3.1 Interpretasi Global

Interpretasi global bertujuan untuk memberikan gambaran menyeluruh mengenai faktor-faktor yang paling dominan mempengaruhi prevalensi *stunting* di tingkat nasional. Analisis ini menggabungkan nilai SHAP mutlak dari seluruh sampel data untuk mengidentifikasi hierarki kepentingan fitur serta pola hubungan umum antara determinan dan kejadian *stunting*. Dalam penelitian ini, interpretasi global diuraikan melalui dua komponen visualisasi utama:

1. *Feature Importance Plot*

Visualisasi berbentuk diagram batang yang mengurutkan variabel determinan berdasarkan besaran rata-rata nilai SHAP absolut ($\text{meanOf}|\text{SHAP}|$). Grafik ini berfungsi untuk menentukan fitur mana yang memiliki dampak paling signifikan terhadap model secara keseluruhan tanpa memandang arah pengaruhnya (positif atau negatif).

2. *Summary Plot (Beeswarm)*

Visualisasi distribusi nilai SHAP yang menyajikan arah pengaruh setiap variabel terhadap prediksi. Pada grafik ini, titik berwarna merah merepresentasikan nilai fitur yang tinggi, sedangkan biru merepresentasikan nilai rendah. Sebaran titik tersebut menunjukkan bagaimana nilai fitur tertentu dapat menaikkan (nilai SHAP positif) atau menurunkan (nilai SHAP negatif) risiko *stunting*.

3.3.2 Interpretasi Lokal

Interpretasi lokal dilakukan untuk memahami mekanisme prediksi model *Random Forest* pada tingkat wilayah spesifik, yaitu provinsi. Berbeda dengan interpretasi global yang menyoroti pola umum secara nasional, pendekatan ini berfokus pada kontribusi variabel determinan terhadap prediksi prevalensi *stunting* pada masing-masing provinsi. Analisis ini penting karena faktor dominan penyebab *stunting* dapat bervariasi antarwilayah meskipun memiliki tingkat prevalensi yang relatif serupa. Alur interpretasi lokal menggunakan metode SHAP dalam penelitian ini ditunjukkan pada Gambar 3.5.

3.3.2.1 Pemilihan Sampel Provinsi

Mengingat interpretasi lokal bersifat mendalam dan kontekstual, tidak seluruh provinsi dianalisis secara individual. Oleh karena itu, dilakukan pemilihan sampel provinsi yang bertujuan untuk merepresentasikan variasi kondisi *stunting* di Indonesia. Pemilihan sampel didasarkan pada distribusi nilai prevalensi *stunting* hasil prediksi model, sehingga mencakup provinsi dengan tingkat prevalensi relatif tinggi, rendah, serta kondisi menengah. Pendekatan ini memungkinkan analisis yang lebih terfokus tanpa mengurangi representativitas karakteristik wilayah secara nasional.

3.3.2.2 Visualisasi dan Analisis *Waterfall Plot*

Setelah sampel provinsi ditetapkan, dilakukan interpretasi lokal menggunakan visualisasi *Waterfall Plot*. Visualisasi ini digunakan untuk menguraikan proses pembentukan nilai prediksi prevalensi *stunting* oleh model pada satu provinsi tertentu. Proses dimulai dari nilai rata-rata prediksi model (*base value*), kemudian setiap variabel determinan ditampilkan sebagai kontribusi aditif yang dapat meningkatkan atau menurunkan nilai prediksi hingga mencapai nilai akhir.

Melalui *Waterfall Plot*, kontribusi masing-masing variabel dapat diamati secara eksplisit baik dari segi arah maupun besar pengaruhnya. Variabel dengan kontribusi positif merepresentasikan faktor yang mendorong peningkatan risiko *stunting*, sedangkan kontribusi negatif menunjukkan faktor yang berperan dalam menurunkan risiko tersebut. Pendekatan ini memberikan transparansi terhadap keputusan model serta menjadi dasar analisis perbedaan determinan *stunting* antarprovinsi pada tahap pembahasan hasil.

3.4 Alat dan Bahan

3.4.1 Alat Penelitian

Perangkat yang digunakan dalam penelitian ini meliputi perangkat keras dan perangkat lunak yang mendukung proses pengolahan data, pemodelan, serta interpretasi hasil. Adapun rincian alat penelitian yang digunakan adalah sebagai berikut:

1. **Perangkat Keras:** Laptop dengan spesifikasi prosesor Intel® Core™ i7-11600H (2.9 GHz, 6 core), RAM 16 GB, penyimpanan 256 GB SSD, dan kartu grafis NVIDIA® GeForce RTX™ 3050. Perangkat ini digunakan untuk menjalankan seluruh proses komputasi, mulai dari pra-pemrosesan data hingga pelatihan dan interpretasi model *Random Forest*.
2. **Sistem Operasi:** Windows 11 sebagai lingkungan kerja utama dalam menjalankan perangkat lunak dan bahasa pemrograman yang digunakan

pada penelitian ini.

3. **Bahasa Pemrograman:** Python versi 3.10 yang digunakan sebagai bahasa utama dalam pengolahan data, pembangunan model *machine learning*, serta analisis interpretabilitas model.
4. **Library Python:** Pandas dan NumPy digunakan untuk manipulasi dan pengolahan data numerik, Scikit-learn untuk implementasi model *Random Forest* dan validasi silang, Matplotlib dan Seaborn untuk visualisasi data dan hasil analisis, serta SHAP untuk interpretasi kontribusi variabel determinan terhadap prediksi *stunting*.
5. **Perangkat Lunak Pendukung:** Visual Studio Code digunakan sebagai editor kode utama untuk pengembangan skrip program, sedangkan Jupyter Notebook dimanfaatkan untuk eksplorasi data, dokumentasi proses analisis, dan visualisasi hasil secara interaktif.

3.4.2 Bahan Penelitian

Dataset yang digunakan dalam penelitian ini bersumber dari Laporan Publikasi Survei Status Gizi Indonesia (SSGI) tahun 2024 yang diterbitkan oleh Kementerian Kesehatan Republik Indonesia [3]. Data tersebut merupakan data sekunder berskala nasional yang menyajikan berbagai indikator status gizi, kesehatan ibu dan anak, serta kondisi lingkungan dan sosial ekonomi pada tingkat provinsi. Setelah melalui tahap persiapan dan seleksi data, diperoleh satu dataset terstruktur yang terdiri dari 56 variabel independen dan satu variabel dependen berupa prevalensi *stunting*. Untuk memudahkan pemahaman struktur data dan menjaga konsistensi analisis determinan, seluruh variabel independen kemudian dikelompokkan ke dalam beberapa kategori tematik berdasarkan karakteristik indikator yang diwakilinya.

Tabel 3.1 Distribusi Variabel Determinan Stunting berdasarkan Kategori

No	Kategori Variabel	Jumlah Fitur
1	Riwayat Kehamilan dan Kelahiran	15
2	Riwayat Pemberian ASI dan MP-ASI	17
3	Akses dan Pemanfaatan Pelayanan Kesehatan	9
4	Imunisasi dan Morbiditas Balita	8
5	Kepemilikan dan Pemanfaatan Jaminan Kesehatan	2
6	Pendampingan Keluarga dan Pengetahuan Stunting	4
7	Kontrasepsi Pasca Bersalin	1
8	Bantuan Sosial	1
9	Akses Air Minum dan Sanitasi	4
Total		56

Tabel 3.1 menunjukkan bahwa variabel determinan yang digunakan dalam penelitian ini didominasi oleh indikator terkait riwayat kehamilan, pola pemberian makan bayi dan anak, serta akses pelayanan kesehatan. Komposisi ini mencerminkan pendekatan multidimensi dalam menganalisis *stunting*, di mana faktor biologis, perilaku, layanan kesehatan, serta kondisi lingkungan dianalisis secara simultan. Pengelompokan ini juga menjadi dasar dalam interpretasi model, khususnya pada tahap analisis kontribusi fitur menggunakan metode SHAP.

3.5 Metode Pengembangan

Penelitian ini menerapkan kerangka kerja pengembangan sistem berbasis data (*data-driven*) yang mengacu pada tahapan standar *Knowledge Discovery in Databases* (KDD). Kerangka kerja ini digunakan untuk memastikan bahwa proses transformasi data sekunder dari laporan SSGI 2024 menjadi informasi analitis dilakukan secara sistematis dan terstruktur. Sebagai metode utama dalam tahap pemodelan, penelitian ini menggunakan algoritma *Random Forest Regression*, yang secara luas dikenal memiliki kemampuan dalam memodelkan hubungan non-linier dan menangani interaksi kompleks antarvariabel determinan. Pemilihan algoritma ini didukung oleh temuan penelitian sebelumnya yang menunjukkan kinerja *Random Forest* yang kompetitif dalam konteks pemodelan

prevalensi *stunting* di Indonesia [10].

Tantangan metodologis utama dalam penelitian ini adalah keterbatasan jumlah observasi data pada tingkat provinsi ($n = 36$), yang berpotensi menimbulkan risiko *overfitting* apabila digunakan skema pembagian data latih dan uji konvensional. Oleh karena itu, diterapkan strategi validasi *Leave-One-Out Cross Validation* (LOOCV), di mana setiap provinsi secara bergantian digunakan sebagai data uji sementara provinsi lainnya berperan sebagai data latih. Pendekatan ini memungkinkan pemanfaatan data secara maksimal dan memberikan estimasi performa model yang lebih stabil pada *dataset* berukuran kecil, sebagaimana direkomendasikan dalam penelitian sebelumnya [11].

Selain aspek akurasi prediksi, penelitian ini juga menekankan pada interpretabilitas model melalui integrasi metode *SHapley Additive exPlanations* (SHAP). Metode ini digunakan untuk mengkuantifikasi kontribusi relatif setiap variabel determinan terhadap hasil prediksi model, baik secara global maupun pada tingkat provinsi. Pendekatan SHAP dipilih karena kemampuannya dalam menjelaskan model berbasis *ensemble* secara transparan, sehingga hasil pemodelan tidak hanya menghasilkan estimasi angka prevalensi *stunting*, tetapi juga menyediakan dasar analitis yang dapat mendukung perumusan kebijakan intervensi gizi yang lebih tepat sasaran [12].

3.6 Ilustrasi Perhitungan Evaluasi

Evaluasi model regresi dihitung menggunakan persamaan berikut:

1. **Mean Absolute Error (MAE):** $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ (Rumus 3.1)

2. **Root Mean Squared Error (RMSE):** $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ (Rumus 3.2)

3. **Koefisien Determinasi (R^2):** $R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$ (Rumus 3.3)

Dimana y_i adalah nilai aktual, \hat{y}_i adalah nilai prediksi, dan \bar{y} adalah rata-rata nilai aktual.

3.7 Rancangan Pengujian

Pengujian dilakukan untuk membuktikan kinerja model dalam memprediksi stunting dan kemampuan SHAP dalam mengisolasi faktor determinan. Validitas dijamin melalui skema LOOCV dimana setiap provinsi diuji secara independen sebagai *unseen data* untuk memastikan objektivitas hasil.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Berisi hasil penelitian berdasarkan rancangan yang sudah dijelaskan pada Bab III, terutama dari Subbab ???. Bagi yang membuat alat, jelaskan alat yang jadi dalam bentuk apa. Bagi yang membuat aplikasi, jelaskan aplikasi yang jadi dalam bentuk seperti apa. Jabarkan dalam bentuk pseudocode dan dijelaskan per bagian kodennya. Gunakan gambar dan tabel sebagai alat bantu menjelaskan hasil.

Contoh implementasi kode dapat dituliskan menggunakan `\begin{lstlisting}`. Contoh kode dapat dilihat pada Kode 4.1.

Kode 4.1 Akuisisi Gambar

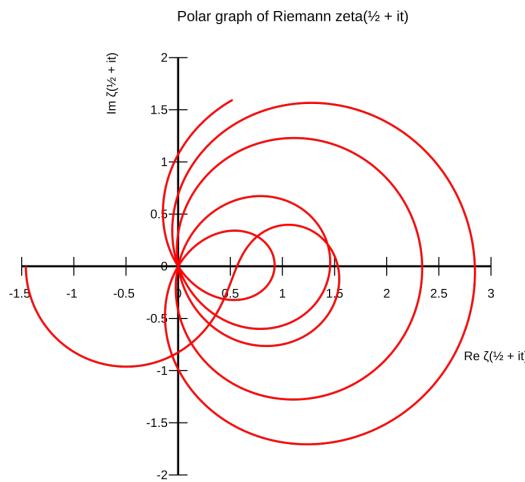
```
1 def process_dataset(dataset_path):
2     image_files = glob(os.path.join(dataset_path, '*.png'))
3     image_files.sort()
4     for image_file in image_files:
5         frame = cv2.imread(image_file)
6         if frame is None:
7             continue
8         frame_rgb = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
9         cv2.imshow('Frame', frame)
10        if cv2.waitKey(1) & 0xFF == ord('q'):
11            break
12    cv2.destroyAllWindows()
13 def main():
14     datasets = get_all_dataset_folders(DATASET_ROOT)
15     for dataset in datasets:
16         process_dataset(dataset)
17         print("print string")
```

4.2 Hasil Pengujian

Berikan hasil pengujian berdasarkan rancangan & skenario yang sudah direncanakan sebelumnya pada Subbab ??.

Tabel 4.1 Data *dummy* Pengujian

Subjek	Hasil Prediksi (BPM)							GT
	F	NA	NO	RC	LC	M	C	
1	68	69	68	70	68	71	69	68
2	69	69	68	70	68	71	69	69
3	70	70	69	71	68	73	69	70
4	71	70	70	72	69	73	70	71
5	72	72	70	72	70	74	70	72



Gambar 4.1 Contoh Graf Pengujian

4.3 Analisis Hasil Penelitian

Berikan analisis hasil penelitian & pengujian, berupa data yang didapatkan dari penelitian & pengujian Tugas Akhir yang sudah anda kerjakan. Gunakan gambar dan tabel sebagai alat bantu menjelaskan analisis hasil. Data luaran penelitian yang dapat dianalisis berupa:

1. Hasil pengujian

2. Hasil kuesioner
3. Aplikasi yang dikembangkan

Analisis dapat membandingkan dengan hasil penelitian sebelumnya yang memiliki kemiripan topik.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berisi kesimpulan dari hasil dan pembahasan terkait penelitian yang dilakukan, dapat juga berupa temuan yang Anda dapatkan setelah melakukan penelitian atau analisis terhadap tugas akhir Anda. Memberikan jawaban dari poin pada subbab Rumusan Masalah dan Tujuan Penelitian.

5.2 Saran

Berisi saran mengenai aspek tugas akhir atau temuan yang dapat dikembangkan dan diperkaya di tugas akhir selanjutnya. Saran dapat berkaitan erat pada subbab Analisis Hasil Penelitian.

DAFTAR PUSTAKA

- [1] D. H. Said and A. Rahmah, “Pencegahan stunting melalui edukasi gaya hidup sehat dan penguatan ekonomi,” *Jurnal Pembelajaran Pemberdayaan Masyarakat (JP2M)*, vol. 5, no. 2, pp. 349–358, 2024, 2721-5148.
- [2] R. Rahagia, N. Sriyanah, I. A. Tyarini, A. Lontaan, and M. Yunus, “Upaya Pencegahan Stunting Berbasis Edukasi dan Sosialisasi,” *Abdimas Polsaka: Jurnal Pengabdian Kepada Masyarakat*, vol. 2, no. 1, 2023, 2829-162X.
- [3] Kementerian Kesehatan RI, “Survei Status Gizi Indonesia (SSGI) 2024 Dalam Angka,” Kementerian Kesehatan RI, tech. rep., 2025.
- [4] Badan Kebijakan Pembangunan Kesehatan, “Survei Kesehatan Indonesia (SKI) 2023 Dalam Angka,” Kementerian Kesehatan RI, tech. rep., 2023.
- [5] Kementerian Sekretariat Negara RI, “Strategi Nasional Percepatan Pencegahan dan Penurunan Stunting 2025-2029,” Kementerian Sekretariat Negara RI, tech. rep., 2024.
- [6] H. Supriyanto, “Implementasi Kebijakan Penurunan Stunting di Kecamatan Salem Kabupaten Brebes,” *Jurnal Multidisiplin Indonesia*, vol. 2, no. 2, 2023, 2964-9048.
- [7] A. S. Saleh, T. Hasan, and U. K. S. Saleh, “Edukasi Penerapan Gizi Seimbang Masa Kehamilan Berbasis Pangan Lokal Sebagai Pencegahan Stunting,” *AHMAR METAKARYA: JURNAL PENGABDIAN MASYARAKAT*, vol. 2, no. 2, pp. 49–53, 2023, 2807-3576.

- [8] A. B. Zemariam et al., “Prediction of stunting and its socioeconomic determinants among adolescent girls in Ethiopia using machine learning algorithms,” *PLOS ONE*, vol. 20, no. 1, e0316452, 2025.
- [9] T. Tamanna, S. Mahmud, N. Salma, M. M. Hossain, and M. R. Karim, “Identifying determinants of malnutrition in under-five children in Bangladesh: insights from the BDHS-2022 cross-sectional study,” *Scientific Reports*, vol. 15, no. 14336, 2025.
- [10] M. A. E. Pratama, S. Hendra, H. R. Ngembra, R. Nur, R. Azhar, and R. Laila, “Comparison of Machine Learning Algorithms for Predicting Stunting Prevalence in Indonesia,” *SISFOKOM Journal (Information and Computer Systems)*, vol. 13, no. 2, pp. 200–209, 2024.
- [11] G.-W. Cha, H.-J. Moon, and Y.-C. Kim, “Comparison of Random Forest and Gradient Boosting Machine Models for Predicting Demolition Waste Based on Small Datasets and Categorical Variables,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 16, p. 8530, 2021.
- [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, p. 1638, 2023.

LAMPIRAN

- A Dataset**
- B Hasil Wawancara**
- C Rincian Kasus Uji**