



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PREDIKSI RISIKO STUNTING PADA BALITA MENGUNAKAN ALGORITMA *MACHINE LEARNING*

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
untuk Memperoleh Gelar Sarjana Komputer pada
Program Studi Sistem Informasi



Oleh:

EKATRI YULISARA

12250325372



UIN SUSKA RIAU

UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2026

LEMBAR PERSETUJUAN
PREDIKSI RISIKO STUNTING PADA BALITA
MENGGUNAKAN ALGORITMA *MACHINE LEARNING*

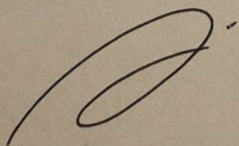
TUGAS AKHIR

Oleh:

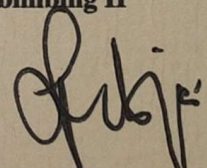
EKATRI YULISARA
12250325372

Telah diperiksa dan disetujui sebagai laporan tugas akhir
di Pekanbaru, pada tanggal 19 Januari 2026

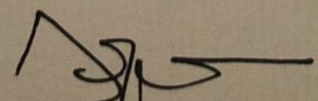
Pembimbing I


Inggi Permana, ST., M.Kom.
NIP. 198812102015031006

Pembimbing II


Febi Nur Salisah, S.Kom., M.Kom.
NIP. 199002222023212038

Ketua Program Studi


Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

LEMBAR PENGESAHAN
PREDIKSI RISIKO STUNTING PADA BALITA
MENGGUNAKAN ALGORITMA MACHINE LEARNING

TUGAS AKHIR

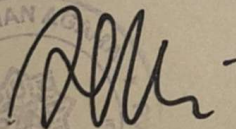
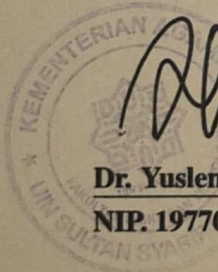
Oleh:

EKATRI YULISARA
12250325372

Telah dipertahankan di depan sidang dewan penguji
sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
di Pekanbaru, pada tanggal 07 Januari 2026

Pekanbaru, 07 Januari 2026
Mengesahkan,

Dekan

Dr. Yuslenita Muda, S.Si., M.Sc.
NIP. 197701032007102001

Ketua Program Studi


Angraini, S.Kom., M.Eng., Ph.D.
NIP. 198408212009012008

DEWAN PENGUJI:

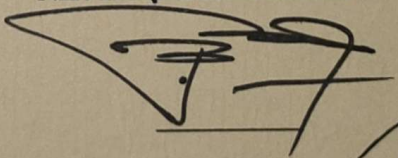
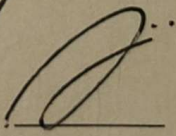
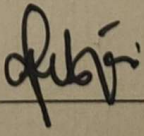
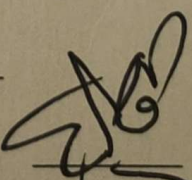
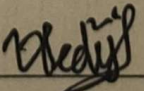
Ketua : Eki Saputra, S.Kom., M.Kom.

Sekretaris : Inggih Permana, ST., M.Kom.

Anggota 1 : Febi Nur Salisah, S.Kom., M.Kom.

Anggota 2 : M. Afdal, ST., M.Kom.

Anggota 3 : Medyantiwi Rahmawita M, ST., M.Kom.

Lampiran Surat:

Nomor : Nomor 25/2021

Tanggal : 10 September 2021

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Ekatri Yulisara

NIM : 12250325372

Tempat/Tgl. Lahir : Dumai/ 13 Juli 2003

Fakultas/Pascasarjana : Sains dan Teknologi

Prodi : Sistem Informasi

Judul Skripsi :

PREDIKSI RISIKO STUNTING PADA BALITA MENGGUNAKAN ALGORITMA
MACHINE LEARNING

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan Skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikianlah Surat Pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 19 Januari 2026

Yang membuat pernyataan



Ekatri Yulisara

NIM: 12250325372



LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum, dengan ketentuan bahwa hak cipta ada pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan atas izin penulis dan harus dilakukan mengikuti kaedah dan kebiasaan ilmiah serta menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin tertulis dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan dapat meminjamkan Tugas Akhir ini untuk anggotanya dengan mengisi nama, tanda peminjaman dan tanggal pinjam pada *form* peminjaman.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis terdapat dalam naskah ini dan disebutkan di dalam daftar pustaka.

Pekanbaru, 19 Januari 2026

Yang membuat pernyataan,

EKATRI YULISARA

NIM. 12250325372



UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillah Rabbil 'Alamin, segala puji bagi Allah Subhanahu Wa Ta'ala sebagai bentuk rasa syukur atas segala nikmat yang telah diberikan tanpa ada kekurangan sedikitpun. Shalawat beserta salam tak lupa pula kita ucapkan kepada Nabi Muhammad Shallallahu 'Alaihi Wa Sallam dengan mengucapkan *Allahumma Sholli'ala Sayyidina Muhammad Wa'ala Ali Sayyidina Muhammad*. Semoga kita semua selalu senantiasa mendapat syafaat-Nya di dunia maupun di akhirat. *Aamiin Ya Rabbal'alamiin*. Peneliti persembahkan karya kecil ini sebagai salah satu hadiah istimewa bentuk bakti, rasa terimakasih, dan hormat Peneliti kepada orang tua tercinta.

Ayahanda Nurhasan dan Ibunda Agustriyeni tersayang, terimakasih atas setiap perjuangan, doa, bimbingan, serta dukungan yang diberikan kepada Peneliti. Terimakasih atas segala kebaikan dan selalu ada saat keadaan tersulit sekalipun. Terimakasih untuk segala pengorbanan yang kalian lakukan. Sampai kapanpun tiada rasa dan cara yang dapat membalas semuanya. Peneliti akan selalu mendoakan yang terbaik agar bahagia dunia dan akhirat.

Terimakasih juga Peneliti ucapkan kepada abang Peneliti Zulkamar Rizki, Kakak Peneliti Dwi Sapta Yulisari dan adik – adik Peneliti Aidil OktoPiandra dan Perdiansyah Abrarruz, terimakasih atas doa, dukungan, serta semangat yang selalu diberikan kepada Peneliti. Kemudian Peneliti ucapkan terimakasih kepada sahabat tercinta Peneliti Nayla Husna Ryanda dan Irma Fitriani, terimakasih untuk segala waktu berharga yang telah dilalui bersama, doa, dukungan, dan perhatian yang diberikan, baik dalam suka maupun duka. Kemudian Peneliti ucapkan terimakasih kepada Bapak dan Ibu Dosen Program Studi Sistem Informasi yang telah mewariskan ilmu yang bermanfaat dan arahan kepada Peneliti untuk menyelesaikan studi di Program Studi Sistem Informasi ini. Semoga kita semua selalu diberikan kemudahan, rahmat, serta karunia-Nya. Aamin.

Wassalamu'alaikum Warahmatullahi Wabarakatuh



KATA PENGANTAR

Assalammu'alaikum wa rahmatullahi wa barakatuh

Alhamdulillah Rabbil 'Alamin, Puji Syukur Peneliti ucapkan kepada Allah *Subhanahu Wa Ta'ala* yang telah memberikan rahmat dan hidayah-Nya sehingga Peneliti bisa menyelesaikan Tugas Akhir ini Inshaallah dengan hasil yang baik. *Shalawat* serta salam juga senantiasa dihadiahkan kepada Nabi Muhammad *Shalallahu Wa'ala Alihi Sayyidina Muhammad*.

Laporan ini disusun sebagai syarat untuk memperoleh gelar sarjana dan sebagai pembelajaran akademis maupun spiritual, Peneliti mengucapkan terimakasih kepada semua pihak yang membantu dalam segala proses penelitian yang telah Peneliti lakukan baik berupa materi maupun motivasi dan doa, untuk itu pada kesempatan ini Peneliti mengucapkan terima kasih yang sangat mendalam kepada:

1. Ibu Prof. Dr. Hj. Leny Nofianti, MS, SE., M.Si., Ak., CA., sebagai Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Ibu Dr. Yuslenita Muda, S.Si., M.Sc., sebagai Dekan Fakultas Sains dan Teknologi.
3. Ibu Angraini, S.Kom., M.Eng., Ph.D., sebagai Ketua Program Studi Sistem Informasi, Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Dr. Rice Novita, S.Kom., M.Kom sebagai Sekretaris Program Studi Sistem Informasi.
5. Bapak Anofrizen, S.Kom., M.Kom selaku Dosen Pembimbing Akademik yang selalu memberikan arahan selama masa perkuliahan.
6. Bapak Inggi Permana, ST., M.Kom., sebagai Dosen Pembimbing I Tugas Akhir Peneliti yang selalu memberikan arahan serta ilmu pengalaman dan waktu yang dimiliki untuk membantu Peneliti menyelesaikan Tugas Akhir ini.
7. Ibu Febi Nur Salisah, S.Kom., M.Kom., sebagai Dosen Pembimbing II Tugas Akhir Peneliti yang selalu memberikan arahan serta ilmu pengalaman dan waktu yang dimiliki untuk membantu Peneliti menyelesaikan Tugas Akhir ini.
8. Bapak Eki Saputra, S.Kom., M.Kom, selaku Ketua Sidang yang telah meluangkan waktunya untuk melaksanakan sidang tugas akhir Peneliti dan memberikan arahan tambahan terkait penelitian Tugas Akhir ini.
9. Bapak M.Afdal, ST., M.Kom sebagai Koordinator Tugas Akhir Program



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Studi Sistem Informasi sekaligus Dosen Penguji I yang telah banyak memberikan arahan dan masukan dalam penulisan Laporan Tugas Akhir ini.

10. Ibu Medyantiwi Rahmawita M, ST., M.Kom sebagai Koordinator Kerja Praktek Program Studi Sistem Informasi sekaligus Dosen Penguji II Tugas Akhir yang memberikan arahan, kritik dan saran yang bermanfaat dalam penyelesaian Tugas Akhir ini.
11. Seluruh Bapak Ibu Dosen Program Studi Sistem Informasi yang telah banyak memberikan ilmunya kepada Peneliti selama masa perkuliahan.
12. Kedua orang tua Peneliti, Ayahanda Nurhasan, Ibunda Agustriyeni, Abang Zulkamar Rizki, Kakak Dwi Sapta Yulisari, serta adik-adik Peneliti Aidil Okto Piandra dan Perdiansyah Abrarruz yang telah memberikan doa, dukungan, dan semangat kepada Peneliti selama proses penyusunan skripsi.
13. Sahabat Peneliti saudari Irma Fitriani dan saudari Nayla Husna Ryanda yang telah memberi Peneliti semangat serta menemani peneliti selama masa perkuliahan berlangsung dalam suka maupun duka sehingga Peneliti dapat menyelesaikan Tugas Akhir ini.
14. Teman-teman seperjuangan Peneliti, yaitu ciwi pikmi yang terdiri atas Anggy, Dina, Rifsyia, Emil, Zhara, Mutiara, Alawiyah, dan Aliya, serta *Girl Squad* yang terdiri atas Devi dan Farah. Kebersamaan,dukungan, selama masa perkuliahan telah memberikan motivasi, kekuatan,dan pengalaman berharga bagi Peneliti dalam menyelesaikan studi dan Tugas Akhir ini.
15. Keluarga besar ClassSic yang selalu saling mendukung, menasehati, dan memberi semangat hingga Peneliti menyelesaikan Tugas Akhir ini.

Terimakasih sangat mendalam, semoga segala kebaikan yang telah diberikan menjadi ladang pahala serta mendapatkan balasan dari Allah *Subhanahu Wa Ta'ala* semoga kita semua selalu mendapatkan kebahagiaan dan kesehatan, *Amin Ya Rabbal'amin*.

Wassalamu'alaikum Warahmatullahi Wabarakaatuh

Pekanbaru, 19 Januari 2026

Penulis,

EKATRI YULISARA

NIM. 12250325372



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PREDIKSI RISIKO STUNTING PADA BALITA MENGUNAKAN ALGORITMA *MACHINE LEARNING*

EKATRI YULISARA
NIM: 12250325372

Tanggal Sidang: 07 Januari 2026
Periode Wisuda:

Program Studi Sistem Informasi
Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau
Jl. Soebrantas, No. 155, Pekanbaru

ABSTRAK

Stunting merupakan masalah gizi kronis pada balita yang berdampak jangka panjang terhadap pertumbuhan fisik, perkembangan kognitif, serta kualitas sumber daya manusia. Tingginya prevalensi stunting di Indonesia menunjukkan pentingnya upaya deteksi dini yang akurat dan berbasis data. Penelitian ini bertujuan untuk memprediksi risiko stunting pada balita menggunakan pendekatan machine learning serta membandingkan kinerja beberapa algoritma klasifikasi. Dataset yang digunakan berasal dari Stunting Balita Detection dengan jumlah 121.000 data yang diperoleh dari platform Kaggle, dengan atribut umur, jenis kelamin, dan tinggi badan sebagai variabel prediktor serta status gizi sebagai variabel target. Lima algoritma machine learning yang diterapkan dalam penelitian ini meliputi Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, dan Decision Tree C4.5. Tahapan penelitian mencakup preprocessing data, pembagian data latih dan data uji dengan rasio 80:20, pelatihan model, serta evaluasi kinerja menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) dan Random Forest memiliki kinerja paling optimal dengan nilai evaluasi yang tinggi dan relatif seimbang dibandingkan algoritma lainnya. Model yang dihasilkan diharapkan dapat membantu proses deteksi dini risiko stunting pada balita serta mendukung pengembangan sistem pendukung keputusan di bidang kesehatan dalam upaya penurunan angka stunting di Indonesia.

Kata Kunci: Stunting, Balita, Machine Learning, Klasifikasi, Prediksi Risiko



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

PREDICTING THE RISK OF STUNTING IN TODDLERS USING MACHINE LEARNING ALGORITHMS

EKATRI YULISARA
NIM: 12250325372

Date of Final Exam: January 07th 2026
Graduation Period:

Department of Information System
Faculty of Science and Technology
State Islamic University of Sultan Syarif Kasim Riau
Soebrantas Street, No. 155, Pekanbaru

ABSTRACT

Stunting is a chronic nutritional problem in toddlers that has long-term effects on physical growth, cognitive development, and the quality of human resources. The high prevalence of stunting in Indonesia highlights the importance of accurate, data-driven early detection efforts. This study aims to predict the risk of stunting in toddlers using a machine learning approach and to compare the performance of several classification algorithms. The dataset used is from Stunting Balita Detection, consisting of 121,000 data points obtained from the Kaggle platform, with age, gender, and height as predictor variables and nutritional status as the target variable. The five machine learning algorithms applied in this study include Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naïve Bayes, and Decision Tree C4.5. The research stages included data preprocessing, division of training and test data with a ratio of 80:20, model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results showed that the K-Nearest Neighbor (KNN) and Random Forest algorithms had the most optimal performance with high and relatively balanced evaluation scores compared to other algorithms. The resulting model is expected to assist in the early detection of stunting risk in toddlers and support the development of decision support systems in the health sector in an effort to reduce stunting rates in Indonesia.

Keywords: Stunting, Toddlers, Machine Learning, Classification, Risk Prediction



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR ISI

| | |
|--|-------------|
| LEMBAR PERSETUJUAN | ii |
| LEMBAR PENGESAHAN | iii |
| LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL | iv |
| LEMBAR PERNYATAAN | v |
| LEMBAR PERSEMBAHAN | vi |
| KATA PENGANTAR | vii |
| ABSTRAK | ix |
| ABSTRACT | x |
| DAFTAR ISI | xi |
| DAFTAR GAMBAR | xiv |
| DAFTAR TABEL | xv |
| DAFTAR SINGKATAN | xvii |
| 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Perumusan Masalah | 4 |
| 1.3 Batasan Masalah | 4 |
| 1.4 Tujuan | 5 |
| 1.5 Manfaat | 5 |
| 1.6 Sistematika Penulisan | 5 |
| 2 LANDASAN TEORI | 8 |
| 2.1 Stunting | 8 |
| 2.1.1 Klasifikasi dan Pengukuran Stunting | 9 |
| 2.1.2 Faktor Penyebab Stunting | 11 |
| 2.1.3 Dampak Stunting | 12 |
| 2.1.4 Upaya Pencegahan dan Penanganan Stunting | 13 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|----------|--|-----------|
| 2.1.5 | Data Prevalensi Stunting | 14 |
| 2.2 | <i>Data Mining</i> | 15 |
| 2.3 | Balita | 18 |
| 2.4 | Prediksi Risiko | 19 |
| 2.5 | <i>Machine Learning</i> | 20 |
| 2.5.1 | <i>Random Forest</i> | 21 |
| 2.5.2 | <i>Support Vector Machine (SVM)</i> | 22 |
| 2.5.3 | <i>K-Nearest Neighbor (KNN)</i> | 26 |
| 2.5.4 | <i>Naive Bayes</i> | 28 |
| 2.5.5 | <i>Decision Tree C4.5</i> | 29 |
| 2.6 | Matriks Evaluasi | 32 |
| 2.7 | <i>Google Colab</i> | 33 |
| 2.7.1 | Keunggulan <i>Python</i> dalam Lingkungan <i>Google Colab</i> | 35 |
| 2.7.2 | Pustaka <i>Python</i> Utama untuk Analisis dan <i>Machine Learning</i> | 36 |
| 2.7.3 | Struktur <i>Notebook Colab</i> dan Implementasi Kode | 37 |
| 2.8 | Penelitian Terkait | 37 |
| 3 | METODOLOGI PENELITIAN | 44 |
| 3.1 | Tahap Perencanaan | 44 |
| 3.2 | Pengumpulan Data | 46 |
| 3.3 | Pemodelan Klasifikasi | 46 |
| 3.3.1 | Dataset | 47 |
| 3.3.2 | Pemeriksaan Data Awal | 47 |
| 3.3.3 | Analisis Distribusi Status Gizi | 48 |
| 3.3.4 | <i>Encoding</i> Data Kategorikal | 48 |
| 3.3.5 | Pemisahan Data (<i>Data Splitting</i>) | 48 |
| 3.3.6 | Data latih (<i>Training Data</i>) | 48 |
| 3.3.7 | Data Uji (<i>Testing Data</i>) | 49 |
| 3.3.8 | Normalisasi Data (<i>Data Normalization</i>) | 49 |
| 3.3.9 | Pembuatan Model Klasifikasi | 49 |
| 3.3.10 | Evaluasi | 53 |
| 3.4 | Pemilihan Model Terbaik | 54 |
| 4 | HASIL DAN PEMBAHASAN | 55 |
| 4.1 | Proses Pengumpulan Data | 55 |
| 4.2 | <i>Preprocessing Data</i> | 58 |
| 4.2.1 | Pemeriksaan Data Awal | 58 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|-----------------------|---|--------------|
| 4.2.2 | Analisis Distribusi Kelas Status Gizi | 59 |
| 4.2.3 | <i>Encoding</i> Data Kategorikal | 59 |
| 4.2.4 | Pembagian Data Latih dan Data Uji | 60 |
| 4.2.5 | Normalisasi Data | 61 |
| 4.3 | Hasil Pembuatan Model Klasifikasi | 62 |
| 4.3.1 | <i>K – Neirest Neighbor</i> | 62 |
| 4.3.2 | <i>Random Forest</i> | 64 |
| 4.3.3 | <i>Naïve Bayes Classifier</i> | 66 |
| 4.3.4 | <i>Decision Tree C4.5</i> | 68 |
| 4.3.5 | <i>Support Machine Vector (SVM)</i> | 70 |
| 4.4 | Evaluasi Model | 72 |
| 4.5 | Tabel Matriks Evaluasi Model | 79 |
| 4.6 | Kesimpulan Hasil Terbaik | 80 |
| PENUTUP | | 81 |
| 5.1 | Kesimpulan | 81 |
| 5.2 | Saran | 81 |
| DAFTAR PUSTAKA | | |
| LAMPIRAN A | SURAT IZIN PENELITIAN | A - 2 |
| LAMPIRAN B | TRANSKIP WAWANCARA | B - 1 |
| LAMPIRAN C | BERITA ACARA WAWANCARA | C - 1 |
| LAMPIRAN D | DOKUMENTASI | D - 1 |



DAFTAR GAMBAR

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|------|--|-------|
| 2.1 | Visualisasi Balita Normal dan Balita Stunting | 8 |
| 2.2 | Visualisasi Klasifikasi Balita Berdasarkan Nilai Z-Score | 11 |
| 2.3 | Pravalensi Stunting di Indonesia | 15 |
| 2.4 | Model atau Pengetahuan Merupakan Output Data | 17 |
| 2.5 | Visualisasi <i>Support Vector Machine</i> | 22 |
| 2.6 | Visualisasi <i>K-Nearest Neighbor</i> (KNN) | 27 |
| 2.7 | Visualisasi Naïve Bayes | 29 |
| 2.8 | Visualisasi <i>Decision Tree C4.5</i> dengan K5 | 30 |
| 2.9 | Tampilan <i>Notebook</i> Baru di <i>Google Colab</i> | 34 |
| 2.10 | Tampilan Logo <i>Google Colab</i> | 34 |
| 3.1 | Alur Metodologi Penelitian | 44 |
| 3.2 | Pemodelan Prosedur Data | 47 |
| 4.1 | Hasil Evaluasi Akurasi dari 5 algoritma Machine Learning | 72 |
| 4.2 | Hasil Evaluasi F1- Score dari 5 algoritma Machine Learning | 73 |
| 4.3 | Hasil Evaluasi <i>Recall</i> dari 5 algoritma Machine Learning | 74 |
| 4.4 | Hasil Evaluasi Precision dari 5 algoritma Machine Learning | 75 |
| 4.5 | Hasil Evaluasi Perbandingan 4 Matriks Evaluasi | 76 |
| 4.6 | Pohon Keputusan | 77 |
| A.1 | Surat Izin Penelitian | A - 2 |
| C.1 | Berita Acara Wawancara | C - 1 |
| D.1 | Dokumentasi Wawancara | D - 1 |

UIN SUSKA RIAU

DAFTAR TABEL

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

| | | |
|------|---|----|
| 2.1 | Standar WHO: Height-for-Age (TB/U atau PB/U) – Z-score untuk anak Laki-laki | 9 |
| 2.2 | Standar WHO: Height-for-Age (TB/U atau PB/U) – Z-score untuk anak Perempuan | 9 |
| 2.3 | Kernel-kernel pada algoritma <i>Support Vector Machine</i> (SVM) | 25 |
| 2.4 | Tabel <i>Confusion Matrix</i> | 33 |
| 2.5 | Pustaka <i>Python</i> Utama | 36 |
| 2.6 | Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting | 38 |
| 3.1 | Parameter <i>Decision Tree C4.5</i> | 50 |
| 3.2 | Parameter <i>Support Vector Machine</i> (SVM) | 51 |
| 3.3 | Parameter <i>K-Nearest Neighbor</i> (KNN) | 51 |
| 3.4 | Parameter <i>Random Forest</i> | 52 |
| 4.1 | Daftar Atribut <i>Dataset</i> | 55 |
| 4.2 | Distribusi Variabel Umur (bulan) | 56 |
| 4.3 | Distribusi Variabel Jenis Kelamin | 57 |
| 4.4 | Distribusi Variabel Tinggi Badan (Kategori) | 57 |
| 4.5 | Distribusi Variabel Status Gizi | 58 |
| 4.6 | Sampel data awal <i>dataset</i> | 58 |
| 4.7 | Distribusi Kelas Status Gizi | 59 |
| 4.8 | Hasil <i>Encoding</i> Data Kategorikal | 60 |
| 4.9 | Data Fitur Setelah Pembagian <i>Train-Test</i> | 61 |
| 4.10 | Hasil Normalisasi Data Latih | 61 |
| 4.11 | Hasil Normalisasi Data Uji | 62 |
| 4.12 | Hasil evaluasi kinerja dari algoritma <i>K-Nearest Neighbor</i> (KNN) | 63 |
| 4.13 | <i>Confusion Matrix</i> <i>K-Nearest Neighbor</i> (KNN) | 63 |
| 4.14 | Hasil evaluasi kinerja dari algoritma <i>Random Forest</i> | 65 |
| 4.15 | <i>Confusion Matrix</i> <i>Random Forest</i> | 66 |
| 4.16 | Hasil evaluasi kinerja dari algoritma <i>Naïve Bayes Classifier</i> | 67 |
| 4.17 | <i>Confusion Matrix</i> <i>Naïve Bayes Classifier</i> | 67 |
| 4.18 | Hasil evaluasi kinerja dari algoritma <i>Decision Tree C4.5</i> | 68 |
| 4.19 | <i>Confusion Matrix</i> <i>Decision Tree C4.5</i> | 69 |
| 4.20 | Hasil Evaluasi Kinerja Dari Algoritma <i>Support Vector Machine</i> (SVM) | 70 |

| | | |
|------|--|----|
| 4.21 | Confusion Matrix Support Vector Machine (SVM) | 71 |
| 4.22 | Tabel Matriks Evaluasi Model <i>Machine Learning</i> | 79 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR SINGKATAN

| | |
|------|--|
| AI | : <i>Artificial Intelligence</i> |
| AUC | : <i>Area Under the Curve</i> |
| C4.5 | : <i>Decision Tree C4.5</i> |
| EDA | : <i>Exploratory Data Analysis</i> |
| FN | : <i>False Negative</i> |
| FP | : <i>False Positive</i> |
| KNN | : <i>K-Nearest Neighbor</i> |
| ML | : <i>Machine Learning</i> |
| NB | : <i>Naïve Bayes</i> |
| RF | : <i>Random Forest</i> |
| ROC | : <i>Receiver Operating Characteristic</i> |
| SVM | : <i>Support Vector Machine</i> |
| TP | : <i>True Positive</i> |
| TN | : <i>True Negative</i> |



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Stunting pada anak adalah masalah kesehatan serius yang berdampak pada kualitas hidup manusia dalam jangka panjang. Kekurangan gizi jangka panjang, terutama selama 1.000 hari pertama kehidupan, menyebabkan stunting, yang seringkali tidak diketahui hingga anak berusia dua tahun. Kondisi ini menghambat pertumbuhan fisik anak dan mengganggu perkembangan organ penting seperti otak, jantung, dan ginjal (P. M. Putri, Shafira, dan Mahardhika, 2024). Dalam jangka panjang, itu dapat mengurangi daya tahan tubuh, meningkatkan risiko penyakit kronis, dan menurunkan produktivitas ekonomi. Dampak jangka pendek meliputi gangguan neurosensori, motorik, dan penurunan IQ. Menurut data *UNICEF* tahun 2023, sekitar 149 juta anak di bawah usia lima tahun di seluruh dunia mengalami stunting (M. Fazlur Rahman Assauqi dan Zaehol Fatah, 2024). Di Indonesia, prevalensi stunting masih berada pada angka 24,4% pada tahun 2022, berdasarkan data dari Kementerian Kesehatan, yang masih jauh dari target nasional sebesar 14% pada tahun 2024 (Sumanti, 2024).

Sekitar 6,3 juta balita di Indonesia terkena dampak stunting, menempatkan negara ini di peringkat ke-27 dari 154 negara di seluruh dunia dan di peringkat ke-5 di Asia. Untuk mencapai tujuan ini, pemerintah Indonesia harus menurunkan angka stunting menjadi 14% pada tahun 2024. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI) pada tahun 2024 yang dirilis pada Mei 2025, prevalensi stunting nasional turun menjadi 19,8%, dari 21,5% pada tahun 2023 (Sahamony dan Rianto, 2024). Namun, dikarenakan Prevalensi angka stunting masih kurang dari target, Pemerintah Indonesia masih menghadapi berbagai tantangan, terutama dalam hal pendataan yang terbatas, standar pengukuran yang tidak seragam, metode pengumpulan data yang tidak seragam, dan kekurangan sumber daya manusia dan teknologi. Akibatnya, untuk menangani stunting secara tepat sasaran, upaya pencegahan berbasis data dan teknologi sangat diperlukan (P. M. Putri dkk., 2024). Penelitian yang dilakukan di berbagai negara berkembang menunjukkan bahwa stunting berdampak buruk pada anak dan masa depan mereka. Anak-anak yang stunting cenderung memiliki pencapaian pendidikan yang lebih rendah, penghasilan yang lebih rendah, dan risiko yang lebih tinggi untuk terjebak dalam kemiskinan, yang semua berdampak pada kesehatan mereka di masa depan (Rais dan Warjiyono, 2024). Stunting juga meningkatkan risiko obesitas dan penyakit kronis saat dewasa karena



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

postur tubuh yang lebih pendek. Stunting pada perempuan juga dapat meningkatkan risiko melahirkan bayi dengan berat badan lahir rendah dan komplikasi persalinan lainnya. Penurunan tinggi badan anak sebesar 1% dan penurunan produktivitas ekonomi sebesar 1,4% di negara-negara berkembang di Asia dan Afrika dikaitkan, menurut Bank Dunia. Secara nasional kurang gizi pada balita masih menjadi masalah di Indonesia (Sukarene, Harjono, Sutaryono, dan Maslami, 2023).

Sampai saat ini, pemerintah Indonesia masih berusaha mengurangi dua pendekatan utama untuk mengurangi stunting adalah intervensi sensitif dan intervensi spesifik. Yang pertama bertujuan untuk mengatasi dan mencegah stunting secara langsung pada balita dan ibu hamil melalui pemberian imunisasi, suplemen makanan, zat besi, dan suplementasi mikronutrien seperti vitamin, zat besi, dan seng. Di sisi lain, intervensi sensitif adalah pendekatan lintas sektor yang menangani faktor sosial ekonomi yang meningkatkan risiko stunting. Dalam Rapat Kerja Nasional BKKBN, Kementerian Kesehatan mengumumkan hasil Survei Status Gizi Indonesia (SSGI) bahwa tingkat stunting di Indonesia menurun dari 24,4% pada tahun 2021 menjadi 21,6% pada tahun 2022 (Kementerian Kesehatan Republik Indonesia, 2023). Oleh karena itu, diperlukan tindakan segera di tingkat nasional, kabupaten, kecamatan, dan desa untuk menentukan wilayah mana yang paling memerlukan intervensi (Afriansyah, Saputra, Ardhana, dan Sa'adati, 2024).

Untuk menangani masalah kesehatan ini dengan lebih efisien, metode yang inovatif dan canggih harus digunakan. Salah satu solusi yang menjanjikan adalah penggunaan teknologi pembelajaran mesin, yang memungkinkan untuk memprediksi jumlah penderita stunting berdasarkan faktor-faktor yang relevan. Diharapkan prediksi stunting di Indonesia akan lebih akurat dengan melakukan analisis data yang menyeluruh dan mengembangkan model pembelajaran mesin yang berkualitas. Prediksi yang lebih akurat ini akan membantu dalam menciptakan intervensi yang lebih tepat dan efisien untuk mengatasi stunting, yang akan meningkatkan kualitas hidup anak-anak yang terdampak (Abdusyukur, 2023). Sebagai contoh, salah satu bentuk intervensi yang dapat dilakukan adalah melalui pemberian imunisasi untuk meningkatkan kekebalan tubuh anak. Imunisasi dilakukan dengan cara mengenalkan vaksin ke dalam tubuh, sehingga tubuh mampu membentuk antibodi yang berfungsi melindungi dari berbagai penyakit tertentu. Di sisi lain, vaksin merupakan zat yang digunakan untuk mendorong produksi antibodi dalam tubuh dengan cara diberikan melalui suntikan (seperti vaksin Campak) atau melalui konsumsi oral (seperti vaksin polio) (Hastuti dan Budiman, 2024).

Penelitian Fadellia Azzahra, Suarna, dan Arie Wijaya (2024) yang mem-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

prediksi status gizi anak menggunakan Random Forest hanya mencapai akurasi 78,33% dengan presisi dan *recall* yang rendah, serta terbatas pada satu algoritma, satu variabel (BB/TB), dan satu sumber data (Fadellia Azzahra dkk., 2024). Penelitian Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, dan Fitri Nurapriani (2023) terbatas pada perbandingan tiga algoritma (SVM tertinggi 90%) menggunakan data tunggal dari Puskesmas Jatilawang (Syahril Dwi Prasetyo dkk., 2023). A. I. Putri dkk. (2024) hanya membandingkan tiga algoritma (*Random Forest* dengan akurasi tertinggi 87,75%) pada 10.000 data, meski KNN lebih sensitif mendeteksi stunting (A. I. Putri dkk., 2024). Penelitian Jafar dan Aisyah (2022) menggunakan SVM dengan banyak variabel dari Puskesmas Anggadita dan hanya mencapai akurasi 80% untuk "Tidak Stunting", dengan kinerja rendah pada prediksi "Stunting" karena ketidakseimbangan data (Jafar & Aisyah, 2022). (?) terbatas pada penggunaan *Naïve Bayes Classifier* pada satu sumber data (Dinas P2KB Kota Dumai), di mana akurasi tertinggi 94,3% dicapai tanpa seleksi fitur (?, ?).

Penelitian Wahyudin dkk. (2023) menggunakan *Naïve Bayes* dengan hanya 300 data dari Puskesmas Pandanaran dan menghasilkan akurasi 85,33% menggunakan beberapa variabel (Wahyudin dkk., 2023). Penelitian Wardita dkk. (2022) merupakan *studi cross-sectional* pada 92 ibu balita yang mengidentifikasi faktor risiko stunting (infeksi, pola asuh), namun masih menggunakan pendekatan manual dan statistik deskriptif (Yulia Wardita, Eka Meiri Kurniyati, Cory Nelia Damayanti, Mujib Hannan, 2022). Penelitian Ratnasari dkk. (2024) menggunakan empat algoritma dengan 220 data balita dari Puskesmas Wates (*Random Forest* mencapai akurasi 92,70%) untuk deteksi dini (Wahidin & Andika, 2024). Penelitian Simbolon, Pangaribuan, dan ... (2022) berfokus pada peningkatan kapasitas kader melalui pelatihan alat ukur, namun pendekatannya masih manual dan observasional (Simbolon dkk., 2022). Penelitian Gunawan dkk. (2023) membangun sistem informasi Posyandu berbasis web menggunakan *Naïve Bayes*, namun akurasi prediksinya tergolong rendah, hanya 60%, meski usability-nya baik (Gunawan & Andika, 2023).

Meskipun pendekatan *machine learning* sudah umum digunakan untuk memprediksi risiko stunting pada balita, sebagian besar penelitian sebelumnya masih terbatas. Keterbatasan ini terlihat dari minimnya jumlah algoritma yang dibandingkan, seringkali hanya satu hingga tiga (seperti *Random Forest*, SVM, atau *Naïve Bayes*), serta penggunaan data yang relatif kecil dan bersumber dari wilayah terbatas. Akibatnya, hasil yang diperoleh sulit untuk digeneralisasi. Selain itu, evaluasi kinerja model terdahulu seringkali tidak komprehensif karena hanya

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber;

- menggunakan satu atau dua metrik, tanpa analisis mendalam tentang presisi, sensitivitas, atau keseimbangan antarkelas. Untuk mengatasi keterbatasan ini, penelitian ini bertujuan memperluas cakupan analisis dengan membandingkan lima algoritma klasifikasi *Random Forest*, *SVM*, *K-Nearest Neighbor* (KNN), *Naïve Bayes*, dan *Decision Tree C4.5*. Penelitian ini menggunakan dataset yang lebih besar dan bervariasi, yaitu "*Stunting Balita Detection (121K Rows)*" yang bersumber dari platform *Kaggle*.

Diharapkan bahwa penelitian ini akan menghasilkan sebuah model prediksi yang akurat dan dapat diandalkan untuk menentukan risiko stunting pada balita. Tujuan dari model ini adalah untuk mendukung proses deteksi dini risiko stunting sehingga tenaga kesehatan dan pihak terkait dapat melakukan intervensi yang tepat dan cepat berdasarkan data. Penelitian ini juga bertujuan untuk mengevaluasi dan membandingkan performa lima algoritma *machine learning*, yaitu *Random Forest*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), *Naive Bayes*, dan *Decision Tree* (Hastuti & Budiman, 2024). Dengan membandingkan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score* dari masing-masing algoritma, diharapkan dapat diidentifikasi model terbaik yang paling efektif untuk diterapkan dalam konteks prediksi stunting di Indonesia. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan di bidang kesehatan masyarakat serta membantu upaya penurunan angka stunting secara nasional (Wahyuni & Kusumodestoni, 2024).

Berdasarkan latar belakang dan identifikasi permasalahan diatas, maka rumusan permasalahan dari penelitian ini adalah bagaimana mendapatkan model prediksi yang akurat dan dapat diandalkan untuk menentukan risiko stunting pada balita menggunakan algoritma *machine learning*.

Masalah yang akan dibahas dalam penelitian Tugas Akhir ini dibatasi pada beberapa hal yaitu:

- 4



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dom Forest, SVM, KNN, Naïve Bayes, dan Decision Tree C4.5.

4. Penilaian kinerja model hanya menggunakan empat metrik utama, yaitu akurasi (*accuracy*), presisi (*precision*), recall, dan F1-score.
5. Hasil model hanya digunakan untuk prediksi risiko stunting secara umum.
6. Penelitian ini menggunakan metode *hold-out validation* dengan rasio 80% data latih (*training data*) dan 20% data uji (*testing data*).
7. Implementasi dan pelatihan dilakukan menggunakan bahasa pemrograman *Python* dengan *Google Colab*.

1.4 Tujuan

Adapun tujuan yang ingin dicapai melalui pelaksanaan penelitian ini adalah sebagai berikut:

1. Menerapkan lima algoritma *machine learning*, yaitu *Random Forest*, *SVM*, *KNN*, *Naive Bayes*, dan *Decision Tree C4.5*, dalam memprediksi status stunting balita berdasarkan data usia, jenis kelamin, dan tinggi badan.
2. Membandingkan performa masing-masing algoritma berdasarkan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Penelitian ini diharapkan dapat menambah wawasan dalam penerapan algoritma *machine learning* untuk prediksi masalah kesehatan, khususnya dalam memprediksi risiko stunting pada balita.
2. Dengan penerapan berbagai algoritma *machine learning* seperti *Random Forest*, *SVM*, *KNN*, *Naive Bayes*, dan *Decision Tree C4.5*, penelitian ini dapat menjadi landasan dalam menentukan model prediksi yang paling akurat dan efisien.
3. Memberikan referensi atau rekomendasi kepada praktisi kesehatan dan pengambil kebijakan terkait pemanfaatan model prediksi berbasis data untuk mendukung deteksi dini dan intervensi stunting yang lebih tepat sasaran.
4. Memberikan kontribusi dalam pengembangan sistem pendukung keputusan berbasis data di bidang kesehatan masyarakat, khususnya untuk mendukung program penurunan angka stunting di Indonesia.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

BAB 1. PENDAHULUAN



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Bab ini berisi gambaran umum mengenai penelitian yang dilakukan. Pada bagian awal, dijelaskan mengenai latar belakang masalah yang melandasi pentingnya penelitian ini serta fenomena stunting yang masih menjadi permasalahan nasional. Selanjutnya dijabarkan rumusan masalah yang dirumuskan berdasarkan kesenjangan pengetahuan maupun kebutuhan praktis di lapangan. Kemudian disampaikan pula batasan masalah agar penelitian lebih terarah dan fokus pada aspek tertentu. Bagian berikutnya memuat tujuan penelitian, baik tujuan umum maupun tujuan khusus, yang ingin dicapai melalui penelitian ini. Selain itu, juga dipaparkan manfaat penelitian yang terbagi menjadi manfaat teoritis dan praktis, baik bagi akademisi, praktisi kesehatan, maupun masyarakat. Terakhir, bab ini ditutup dengan sistematika penulisan yang menjelaskan struktur penulisan skripsi secara keseluruhan.

BAB 2. LANDASAN TEORI

Bab ini memuat teori-teori yang menjadi dasar dalam penelitian. Bagian pertama membahas tentang pengertian dan konsep stunting, faktor penyebab, dampak, serta upaya pencegahan yang telah dilakukan berdasarkan kajian literatur. Selanjutnya dipaparkan mengenai konsep *data mining* dan *machine learning*, termasuk definisi, prinsip kerja, dan peranannya dalam bidang kesehatan. Bab ini juga menguraikan teori terkait algoritma klasifikasi yang digunakan dalam penelitian, seperti *Decision Tree*, *Naïve Bayes*, *Random Forest*, *SVM*, dan *KNN*. Selain itu, juga disajikan penelitian terdahulu yang relevan untuk memperkuat landasan teoritis, membandingkan hasil penelitian sebelumnya, serta menunjukkan posisi kebaruan penelitian ini.

BAB 3. METODOLOGI PENELITIAN

Bab ini menjelaskan secara rinci mengenai langkah-langkah penelitian yang dilakukan. Bagian pertama memaparkan desain penelitian yang digunakan serta alasan pemilihannya. Selanjutnya dijelaskan mengenai sumber data dan teknik pengumpulan data, baik dari dataset terbuka, instansi kesehatan, maupun data sekunder lain yang relevan. Setelah itu, dijabarkan mengenai tahapan penelitian, mulai dari perencanaan, *preprocessing data*, pembagian data latih dan data uji, hingga implementasi algoritma. Bab ini juga menguraikan alur penelitian dalam bentuk flowchart untuk memberikan gambaran visual. Bagian berikutnya menjelaskan pengelolaan dan analisis data, termasuk tahapan evaluasi performa algoritma menggunakan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*. Terakhir, disampaikan pula alat dan perangkat yang digunakan, seperti bahasa pemrograman, software pendukung, dan spesifikasi perangkat keras.



BAB 4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang telah dilakukan, yang meliputi hasil preprocessing data, implementasi algoritma machine learning, serta perbandingan performa antar model. Hasil pengujian algoritma ditampilkan dalam bentuk tabel, grafik, maupun *confusion matrix* untuk memperjelas performanya. Selanjutnya, dilakukan analisis dan pembahasan yang menginterpretasikan hasil penelitian, menjelaskan kelebihan dan kelemahan masing-masing algoritma, serta membandingkannya dengan penelitian terdahulu. Bab ini juga membahas faktor-faktor yang memengaruhi hasil penelitian, tantangan dalam penerapan model, serta potensi pengembangan sistem di masa depan.

BAB 5. PENUTUP

Bab terakhir ini berisi kesimpulan dari penelitian yang telah dilakukan. Kesimpulan disusun berdasarkan hasil penelitian dan menjawab rumusan masalah yang diajukan pada bab pertama. Selain itu, disajikan pula saran-saran yang ditujukan bagi peneliti selanjutnya, praktisi kesehatan, dan pembuat kebijakan, agar hasil penelitian ini dapat dikembangkan lebih lanjut dan diimplementasikan secara nyata. Bab ini juga memuat rekomendasi terkait penggunaan algoritma *machine learning* terbaik untuk prediksi risiko stunting serta arah penelitian di masa mendatang.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 2

LANDASAN TEORI

2.1 Stunting

Stunting merupakan kondisi gagal tumbuh yang dialami anak balita akibat kekurangan gizi kronis yang terjadi dalam jangka waktu panjang, terutama pada periode seribu hari pertama kehidupan yang dimulai sejak masa kehamilan hingga anak berusia dua tahun (Simbolon dkk., 2022). Masalah ini sering kali dipicu oleh rendahnya asupan nutrisi, infeksi berulang, serta pola asuh yang kurang memadai (Hartati dan Wahyuningsih, 2021). Anak yang mengalami stunting umumnya memiliki tinggi badan yang lebih rendah dari standar usianya dan berisiko mengalami keterlambatan perkembangan kognitif, motorik, serta gangguan kemampuan belajar di kemudian hari (Fitriani dan Darmawi, 2022). Selain itu, stunting juga berkaitan erat dengan kondisi sosial ekonomi keluarga, tingkat pendidikan orang tua, pengetahuan gizi ibu, dan akses terhadap pelayanan kesehatan. Di Indonesia sendiri, stunting menjadi salah satu isu kesehatan prioritas nasional mengingat prevalensinya masih tergolong tinggi (M. Fazlur Rahman Assauqi dan Zaehol Fatah, 2024). Oleh karena itu, upaya pencegahan yang menyeluruh sangat diperlukan, termasuk intervensi gizi sejak masa kehamilan, pemberian ASI eksklusif, pemantauan tumbuh kembang anak secara rutin, serta peningkatan kesadaran masyarakat mengenai pentingnya pemenuhan gizi seimbang bagi anak sejak dini (S. Handayani, 2023). Visualisasi balita normal dan stunting dapat dilihat pada Gambar 2.1



Gambar 2.1. Visualisasi Balita Normal dan Balita Stunting

Sumber: Buku Pedoman Pencegahan dan Penanggulangan Stunting (BKKBN, 2021).

2.1.1 Klasifikasi dan Pengukuran Stunting

Stunting diukur dengan membandingkan tinggi atau panjang badan anak dengan standar pertumbuhan anak sebayanya. Indeks yang digunakan adalah Panjang Badan menurut Usia (PB/U) atau Tinggi Badan menurut Usia (TB/U). Data pengukuran ini kemudian diubah menjadi nilai Z-score untuk menentukan status gizi anak (Akhir, 2023). Nilai Z-score adalah alat statistik yang menunjukkan seberapa jauh pengukuran seorang anak dari nilai rata-rata populasi rujukan WHO yang sehat, dalam satuan Standar Deviasi (SD). Dengan menggunakan Z-score, klasifikasi stunting menjadi lebih objektif dan seragam.

Untuk memperoleh nilai Z-score, hasil pengukuran tinggi atau panjang badan anak dibandingkan dengan nilai median dan standar deviasi yang ditetapkan dalam standar pertumbuhan WHO sesuai dengan usia dan jenis kelamin anak. Secara konseptual, perhitungan Z-score dapat dinyatakan dengan rumus sebagai berikut:

$$Z = \frac{\left(\frac{TB}{PB \text{ anak}} - \text{Median WHO} \right)}{\text{Standar Deviasi WHO}} \quad (2.1)$$

Untuk memberikan gambaran yang lebih jelas mengenai batasan tinggi badan balita dalam menentukan status stunting, Tabel 2.1 dan 2.2 menyajikan contoh standar tinggi badan menurut umur (TB/U atau PB/U) berdasarkan nilai Z-score WHO yang disesuaikan dengan usia dan jenis kelamin anak.

Tabel 2.1. Standar WHO: Height-for-Age (TB/U atau PB/U) – Z-score untuk anak Laki-laki

| Umur (bulan) | -3 SD (cm) | -2 SD (cm) | Median (0 SD) (cm) | +2 SD (cm) | +3 SD (cm) |
|-----------------|---------------|---------------|-----------------------|---------------|---------------|
| 24 | 78.0 | 81.0 | 87.1 | 93.2 | 96.3 |
| 36 | 85.0 | 88.7 | 96.1 | 103.5 | 107.2 |
| 48 | 90.7 | 94.9 | 103.3 | 111.7 | 115.9 |
| 60 | 96.1 | 100.7 | 110.0 | 119.2 | 123.9 |

Tabel 2.2. Standar WHO: Height-for-Age (TB/U atau PB/U) – Z-score untuk anak Perempuan

| Umur (bulan) | -3 SD (cm) | -2 SD (cm) | Median (0 SD) (cm) | +2 SD (cm) | +3 SD (cm) |
|-----------------|---------------|---------------|-----------------------|---------------|---------------|
| 24 | 76.0 | 79.3 | 85.7 | 92.2 | 95.4 |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

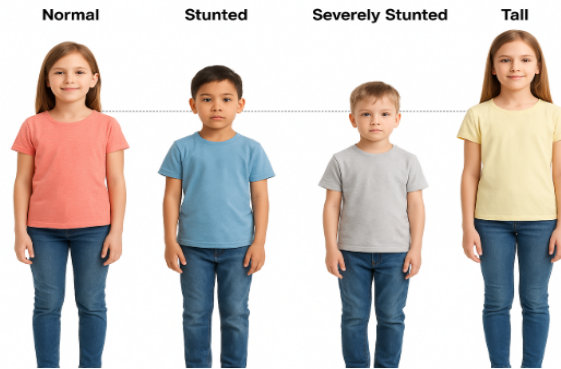
Tabel 2.2 Standar WHO Anak Perempuan

| Umur (bulan) | -3 SD (cm) | -2 SD (cm) | Median (0 SD) (cm) | +2 SD (cm) | +3 SD (cm) |
|-----------------|---------------|---------------|-----------------------|---------------|---------------|
| 36 | 83.6 | 87.4 | 95.1 | 102.7 | 106.5 |
| 48 | 89.8 | 94.1 | 102.7 | 111.7 | 115.7 |
| 60 | 95.2 | 99.9 | 109.4 | 109.4 | 123.7 |

Tabel 2.1 dan Tabel 2.2 menyajikan standar tinggi badan menurut umur (*Height-for-Age*) berdasarkan nilai Z-score yang ditetapkan oleh *World Health Organization* (WHO) untuk anak laki-laki dan anak perempuan usia 24 hingga 60 bulan. Kedua tabel tersebut menampilkan batasan tinggi badan pada setiap kelompok usia yang dikelompokkan ke dalam kategori -3 standar deviasi (SD), -2 SD, median (0 SD), $+2$ SD, dan $+3$ SD sebagai acuan dalam penentuan status stunting. Balita dengan nilai tinggi badan menurut umur berada di bawah batas -2 SD diklasifikasikan sebagai *stunted*, sedangkan balita dengan nilai di bawah -3 SD diklasifikasikan sebagai *severely stunted*. Perbedaan nilai standar antara anak laki-laki dan perempuan pada setiap kelompok usia menunjukkan bahwa penilaian status stunting perlu mempertimbangkan usia, jenis kelamin, dan tinggi badan secara simultan agar klasifikasi yang dihasilkan bersifat objektif serta sesuai dengan standar pertumbuhan yang ditetapkan oleh WHO.

Klasifikasi Stunting Berdasarkan Z-score WHO, Berdasarkan nilai Z-score, status gizi anak diklasifikasikan sebagai berikut (Sahamony dan Rianto, 2024):

1. Normal: Nilai Z-score berada di antara -2 SD hingga $+3$ SD. Anak dengan Z-score di atas -2 SD dianggap memiliki tinggi badan yang sesuai untuk usianya.
2. Pendek (Stunted): Nilai Z-score ≤ -2 SD. Kondisi ini menunjukkan anak mengalami stunting.
3. Sangat Pendek (*Severely Stunted*): Nilai Z-score ≤ -3 SD. Kondisi ini menunjukkan stunting yang parah dan memerlukan penanganan segera.
4. Tinggi (Tall): Nilai Z-score $\geq +3$ SD. Kondisi ini menunjukkan bahwa tinggi badan anak berada jauh di atas rata-rata usianya sesuai standar WHO. Pada Gambar 2.2 dapat dilihat gambar klasifikasi balita berdasarkan nilai Z-Score.



Gambar 2.2. Visualisasi Klasifikasi Balita Berdasarkan Nilai Z-Score

Sumber: Buku Peraturan BPK - Permenkes Nomor 2 Tahun 2020 (Permenkes, 2020).

Klasifikasi ini penting untuk membedakan anak yang memiliki tinggi badan di bawah rata-rata karena faktor genetik dari anak yang gagal tumbuh akibat kekurangan gizi kronis (Adzhima, Budianita, Nazir, dan Syafria, 2023).

2.1.2 Faktor Penyebab Stunting

Banyak faktor yang menyebabkan stunting, termasuk gizi buruk yang dialami oleh responden dan anak baik selama hamil maupun selama masa balita; kurangnya layanan kesehatan, terutama *antenatal care* (ANC), yang memberikan layanan khusus kepada responden hamil untuk memantau kesehatan janin dan responden; dan kurangnya layanan kesehatan postpartum yang memberikan perhatian khusus pada responden hamil untuk memantau kesehatan mereka sendiri dan bayi (Sahyati dan Islami, 2022).

Selain itu, banyak ibu-ibu tidak memahami nutrisi yang diperlukan oleh seseorang anak selama pertumbuhannya. Mereka hanya merekam jumlah makanan yang diberikan tetapi tidak memperhatikan nutrisi yang terkandung dalam makanan tersebut, akibatnya banyak ibu yang memberikan makanan yang tidak sehat kepada anak-anak mereka. Paramita, Helmyati, Basrowi, dan Dilantika (2023) melakukan penelitian yang mendukung teori bahwa pengetahuan responden tentang pemberian nutrisi juga akan berkorelasi dengan angka kejadian stunting. Selain itu, sikap responden juga ternyata mendukung terjadinya stunting pada anak-anak (Paramita dkk., 2023).

Hasil penelitian Beal dkk. (2018), yang menyelidiki faktor penyebab stunting di Indonesia, menemukan beberapa penyebab utama stunting, termasuk pekerjaan kepala keluarga, pemberian ASI eksklusif pada usia 6 bulan pertama, status



sosial ekonomi yang rendah, kelahiran prematur, bayi yang lahir pendek, dan responden berperawakan pendek. Studi Bustami Ampera (2020) menemukan bahwa pekerjaan kepala keluarga, pemberian ASI eksklusif pada usia 6 bulan pertama, usia penyapihan, pemberian (Suryani, Rini, Hardika, dan Widiastari, 2023).

2.1.3 Dampak Stunting

Menurut jurnal yang berjudul Bagaimana konsistennya hubungan antara stunting dan perkembangan anak, penurunan kecerdasan atau IQ Miller, Ann C., Megan B. Murray, Dana R. Thomson, dan Mary Catherine Arbour meneliti bukti dari meta-analisis hubungan antara stunting dan perkembangan anak berbagai dimensi di lima belas negara dengan tingkat pendapatan menengah. Penerbitan *Public Health Nutrition*, 2016 Analisis PICO, P: anak-anak berusia antara 36 dan 59 bulan. I menunjukkan bahwa tidak ada intervensi dan C menunjukkan bahwa tidak ada perbandingan. O: mengurangi prevalensi stunting untuk menghindari dampak pada penurunan kecerdasan (Anwar, Winarti, dan Sunardi, 2022).

Dalam jurnal dengan judul *Impact of stunting on early childhood cognitive development in Benin: evidence from Demographic and Health Survey*, Michael Ekholuenetale¹, Amadou Barrow, Charity Ehimwenma Ekholuenetale, dan Godson Tudeme menemukan bahwa anak stunting akan mengalami penurunan 7% dalam perkembangan kognitif yang optimal dibandingkan dengan anak tidak stunting. *Gazette of the Egyptian Pediatric Association*, 2020 Analisis PICO, P: anak-anak berusia antara 36 dan 59 bulan. I menunjukkan bahwa tidak ada intervensi dan C menunjukkan bahwa tidak ada perbandingan. O: mengurangi prevalensi stunting sehingga tidak mempengaruhi penurunan kecerdasan (Khalid, Patimah, dan Asrina, 2022).

Dampak yang ditimbulkan stunting dapat dibagi menjadi dampak jangka pendek dan jangka panjang (KHOTIMAH, 2022):

1. Dampak Jangka Pendek

- (a) Peningkatan insiden kesakitan dan kematian
- (b) Perkembangan motorik, kognitif, dan verbal anak yang buruk
- (c) Peningkatan biaya kesehatan

2. Dampak Jangka Panjang

- (a) Postur tubuh yang tidak ideal saat dewasa (lebih pendek dibandingkan orang lain)
- (b) Peningkatan kemungkinan terkena obesitas dan penyakit lainnya
- (c) Menurunnya kesehatan reproduksi

- (d) Kurang kemampuan untuk belajar dan berprestasi di sekolah

2.1.4 Upaya Pencegahan dan Penanganan Stunting

Pencegahan stunting harus dimulai sejak masa pra-kehamilan, selama kehamilan, dan selama 1.000 hari pertama kehidupan (HPK). Karena pertumbuhan fisik dan otak anak pada saat ini sangat cepat, periode ini disebut sebagai *window of opportunity*. Pemenuhan nutrisi ibu hamil, suplementasi asam folat dan zat besi, dan pelatihan kesehatan reproduksi adalah beberapa tindakan yang diberikan. Sebuah penelitian yang dilakukan oleh Rusyda dkk. (2024) menemukan bahwa intervensi gizi esensial yang dikombinasikan dengan layanan kesehatan ibu dan anak memiliki dampak yang signifikan terhadap penurunan tingkat stunting di Indonesia (Bandur dan Madur, 2025).

Selain itu, dalam upaya pencegahan, intervensi gizi sensitif dan gizi spesifik digunakan. Intervensi gizi spesifik biasanya menyasar kelompok berisiko seperti ibu hamil, ibu menyusui, dan balita. Sementara itu, program gizi sensitif berfokus pada faktor tidak langsung seperti akses air bersih yang lebih baik, sanitasi yang lebih baik, pendidikan kesehatan, dan kemandirian ekonomi keluarga. Program ini mencakup pemberian makanan tambahan, imunisasi, dan pengobatan infeksi, serta mendorong pemberian ASI eksklusif selama enam bulan. Intervensi berbasis komunitas dan keluarga dapat memperkuat pelaksanaan kedua jenis intervensi tersebut untuk mencapai hasil yang lebih baik (Grathima, Yunitasari, dan Indarwati, 2024).

Selain itu, terbukti bahwa intervensi berupa pemberian makanan tambahan, juga dikenal sebagai pemberian makanan tambahan, berguna untuk membantu balita yang memiliki status gizi kurang atau stunting. Bandur dan Madur (2025) dalam tinjauan pustakanya, sebagian besar penelitian menemukan bahwa makanan tambahan, baik yang berbasis produk instan maupun bahan pangan lokal, meningkatkan berat badan dan tinggi badan balita. Hasil ini sejalan dengan Grathima dkk. (2024), yang menunjukkan bahwa pemberian makanan pendamping ASI dengan kualitas dan kuantitas yang baik sangat penting untuk mencegah stunting pada masa awal kehidupan anak (Hadna, Ardiansyah, dan Habib, 2023).

Pengetahuan ibu dan perilakunya juga penting. Sebuah penelitian yang dilakukan oleh P. M. Putri dkk. (2024) menemukan bahwa ada korelasi signifikan antara pengetahuan ibu tentang gizi dan status gizi anak mereka. Pendidikan gizi dapat membantu ibu hamil dan menyusui memahami pola makan seimbang, pemberian MP-ASI, dan perawatan kesehatan dasar. Puspitasari, Indarwati, Wahyuni, dan Suraya (2025) juga menunjukkan bahwa edukasi gizi pada ibu hamil dan

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak Cipta Dilindungi Undang-Undang



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

menyusui dapat meningkatkan pengetahuan mereka tentang nutrisi dan diharapkan akan berdampak pada praktik yang lebih baik dalam pemberian makanan bergizi kepada anak-anak (Puspitasari dkk., 2025).

Penanggulangan stunting juga membutuhkan partisipasi masyarakat dan kolaborasi lintas sektor. Karena ada dukungan sosial dan pengawasan langsung dari masyarakat, intervensi yang melibatkan kelompok keluarga, kader posyandu, dan tokoh masyarakat terbukti lebih berhasil. Puspitasari dkk. (2025) menekankan bahwa keterlibatan komunitas tidak hanya membuat program gizi lebih efektif, tetapi juga membuat intervensi lebih berkelanjutan karena masyarakat merasa memiliki tanggung jawab atas apa yang mereka lakukan. Oleh karena itu, pendekatan pencegahan dan penanggulangan stunting harus diterapkan secara menyeluruh, berkesinambungan, dan terintegrasi pada berbagai tingkat intervensi (P. M. Putri dkk., 2024).

2.1.5 Data Prevalensi Stunting

Pada rentang waktu 2019 hingga 2024, prevalensi stunting di Indonesia menunjukkan tren penurunan yang relatif stabil meskipun pada beberapa periode laju penurunannya melambat. Berdasarkan data resmi Survey Status Gizi Indonesia (SSGI), angka stunting nasional tercatat sebesar $\pm 27,7\%$ pada tahun 2019, kemudian menurun menjadi sekitar $24,4\%$ pada tahun 2021. Selanjutnya, pada tahun 2022 angka prevalensi terus menurun ke kisaran $21,6\%$, dan pada 2023 hampir stabil di angka $21,5\%$. Pada tahun 2024 tercapai angka $19,8\%$, yang menjadi titik penting karena untuk pertama kalinya prevalensi stunting Indonesia turun di bawah 20% (Kuala, 2022).

Penurunan ini merupakan hasil dari berbagai upaya, termasuk intervensi gizi esensial dan peningkatan layanan kesehatan ibu dan anak. Sebagai contoh, studi Rusyda Baliwati (2025) menemukan bahwa integrasi antara layanan kesehatan ibu-anak dengan intervensi gizi esensial memiliki efek langsung dan tidak langsung yang signifikan dalam menurunkan stunting di tingkat provinsi. Meskipun demikian, penurunan angka stunting antara tahun-tahun 2021 dan 2022 relatif kecil apabila dibandingkan dengan periode sebelumnya, yang menunjukkan adanya tantangan dalam mempertahankan laju penurunan yang cepat (Rusyda dkk., 2024).

Tren ini dipengaruhi oleh kualitas intervensi gizi, sanitasi dan lingkungan, ketersediaan dan akses terhadap layanan kesehatan ibu dan anak, dan pendidikan ibu. Selain itu, diduga bahwa pandemi COVID-19 memengaruhi implementasi berbagai inisiatif intervensi dan pengukuran antropometri. Akibatnya, perkiraan un-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tuk tahun 2020 lebih sensitif terhadap perubahan dalam metode dan cakupan (Belay, 2022).

Meskipun terjadi kemajuan, tujuan nasional seperti RPJMN untuk menurunkan prevalensi stunting ke angka 14% masih jauh dari tercapai. Selain itu, ini menunjukkan betapa strategi percepatan, intervensi lintas sektor, dan mitigasi hambatan lokal sangat penting untuk mempercepat penurunan (Aulifia Riski Putra Wahyu, Fayi Farozzi, Putra Mahendra, dan Kembang Hapsari, 2023). Pembaca dapat melihat grafik tren stunting dari tahun 2019 sampai dengan 2024 pada Gambar 2.3.



Gambar 2.3. Pravalensi Stunting di Indonesia

Sumber: (Garcia et al., 2025).

Grafik tren prevalensi stunting di Indonesia menunjukkan adanya penurunan yang signifikan dan konsisten dari tahun 2019 hingga 2024. Pada tahun 2019, prevalensi berada di angka tinggi 27,67%, yang berarti hampir tiga dari sepuluh balita mengalami stunting. Angka ini mulai turun menjadi 24,4% pada 2021, dan penurunan signifikan berlanjut hingga 21,6% pada 2022, mencerminkan keberhasilan upaya percepatan penurunan stunting oleh pemerintah. Meskipun penurunannya lebih kecil, tren positif berlanjut pada tahun 2023 dengan angka 21,5%. Puncak pencapaian terjadi pada tahun 2024, di mana prevalensi stunting berhasil ditekan hingga 19,8%, menandai pertama kalinya angka nasional berada di bawah batas 20%. Secara keseluruhan, tren ini membuktikan kemajuan nyata dalam program nasional; namun, upaya berkelanjutan masih esensial untuk mencapai target nasional dan standar yang ditetapkan oleh WHO.

2.2 Data Mining

Data mining merupakan sebuah prosedur sistematis yang bertujuan untuk mengeksplorasi serta mengidentifikasi pola-pola signifikan dan informasi bernilai dari kumpulan data berskala besar. Proses ini mencakup serangkaian tahapan teknis



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

yang komprehensif, mulai dari pengumpulan dan ekstraksi hingga analisis statistik mendalam. Dalam literatur akademik, istilah ini sering kali direferensikan sebagai *knowledge discovery* atau ekstraksi pengetahuan, yang berfungsi sebagai instrumen logis dalam mengungkap wawasan tersembunyi. Secara fungsional, penerapan data mining berperan krusial sebagai sistem pendukung keputusan strategis guna mengoptimalkan pengembangan bisnis melalui pemanfaatan temuan informasi dan tren data yang akurat (Dewi, Nurwati, dan Rahayu, 2022).

Selain berfungsi sebagai alat deskriptif, *data mining* berperan sebagai instrumen prediktif yang memungkinkan identifikasi kecenderungan masa depan guna mendukung pengambilan keputusan yang lebih efektif dan responsif. Teknik ini menjamin bahwa pengetahuan yang diekstraksi dari basis data memiliki akurasi yang sesuai dengan ekspektasi riset. Contoh nyata urgensi teknologi ini terlihat pada manajemen data di instansi kesehatan (seperti rumah sakit), di mana data rekam medis dan radiologi cenderung terfragmentasi dan sulit diintegrasikan akibat minimnya protokol standardisasi. Integrasi *data mining* menawarkan solusi sistematis untuk mengolah data tersebut secara akurat. Output yang dihasilkan berupa pengetahuan strategis yang sangat esensial bagi pengambil kebijakan di level pemerintahan guna mengoptimalkan layanan kesehatan masyarakat secara menyeluruh (Sidik dkk., 2020).

Data mining didefinisikan sebagai serangkaian prosedur logis yang bertujuan untuk mengungkap pola tersembunyi guna ditransformasikan menjadi informasi yang bernilai guna. Melalui penyimpanan data yang tersentralisasi dalam *data warehouse*, algoritma data mining berperan dalam mengekstraksi pengetahuan yang krusial bagi strategi pengembangan bisnis. Implementasi analisis matematika dalam proses ini memungkinkan penemuan kecenderungan data yang tidak mampu diidentifikasi oleh metode statistik tradisional. Kompleksitas hubungan antardata serta besarnya skala *dataset* (*Big Data*) sering kali menjadi hambatan dalam eksplorasi data biasa; oleh karena itu, penggunaan data mining menjadi solusi fundamental untuk memetakan hubungan-hubungan rumit tersebut secara akurat dan efisien (Dewi dkk., 2022).

Secara fungsional, *data mining* berorientasi pada penemuan pola-pola mutakhir yang esensial dalam penyelesaian berbagai masalah di berbagai sektor. Teknologi ini merepresentasikan kekuatan baru dalam dunia digital yang memungkinkan organisasi untuk memprioritaskan analisis pada variabel-variabel paling relevan mengenai dinamika perilaku pelanggan. Dengan memanfaatkan keberagaman algoritma dalam *data mining*, pengolahan big data dapat dilakukan dengan tingkat

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

akurasi yang lebih tinggi. Hal ini memungkinkan perusahaan untuk mengekstraksi informasi berharga secara lebih tepat guna, meskipun dihadapkan pada volume data yang sangat masif, yang sulit dijangkau oleh metode analisis konvensional (Sidik dkk., 2020). Secara sederhana, *data mining* dapat di lihat pada Gambar 2.4.



Gambar 2.4. Model atau Pengetahuan Merupakan Output Data

Sumber: (Arhami et al,2020).

Menurut Pan Ning Tan et al (2019:25), terdapat dua jenis proses dalam penemuan pengetahuan dalam basis data (KDD), yaitu:

1. Pra-pemrosesan data bertujuan untuk mengubah data mentah menjadi format yang sesuai untuk analisis. Tahapan pra-pemrosesan mencakup penggabungan data dari berbagai sumber, pembersihan data untuk menghapus observasi duplikat dan noise, serta pemilihan catatan dan fitur yang relevan dalam proses penambangan data.
2. Data *postprocessing* bertujuan untuk menjamin bahwa hasil yang valid dan bermanfaat akan diintegrasikan ke dalam pengambilan keputusan melalui pola penyaringan, visualisasi, dan interpretasi pola (Wasito, 2024).

Beberapa metode pengolahan data mining, menurut Pan Ning Tan et al. (2019:29), termasuk (Wasito, 2024):

1. *Predictive Tasks*
Pemodelan ini bertujuan memproyeksikan variabel target (*dependen*) berdasarkan kontribusi variabel eksplanatori (bebas). Identifikasi yang tepat terhadap kedua kategori variabel ini merupakan tahapan krusial untuk menjamin validitas serta konsistensi hasil prediksi dalam analisis data.
2. Tugas Deskriptif
Tujuan utama analisis ini adalah mengeksplorasi hubungan antar-data melalui penemuan pola-pola strategis. Melalui tahap *post-processing*, hasil ekstraksi data divalidasi lebih lanjut untuk menghasilkan eksplanasi yang presisi serta mendalam mengenai perilaku data yang diteliti (Wasito, 2024).

Beberapa metode pengolahan data mining, menurut Pan Ning Tan et al. (2019:29), termasuk (Wasito, 2024):



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Prediksi (*Predictive*).

Teknik prediktif diimplementasikan pada data dengan karakteristik yang heterogen melalui penggunaan algoritma seperti *Linear Regression* dan *Neural Network*.

2. Asosiasi (*Association*)

Metode asosiasi diterapkan untuk menganalisis interaksi antar-data, di mana algoritma Apriori menjadi salah satu instrumen utama yang sering diimplementasikan.

3. Klastering (*Clustering*)

Metodologi clustering berfungsi untuk mengorganisasikan data ke dalam klaster tertentu berdasarkan kemiripan karakteristik, dengan menggunakan algoritma seperti *K-Means*, *K-Medoids*, *Self-Organizing Map (SOM)*, serta *Fuzzy C-Means*.

4. Klasifikasi (*Classification*)

Teknik klasifikasi mengandalkan variabel target dalam proses pengelompokan datanya. Pendekatan ini merupakan bentuk supervised learning yang diwakili oleh algoritma seperti ID3 serta K-Nearest Neighbor.

2.3 Balita

Balita merupakan anak dengan rentang usia 0 hingga 59 bulan yang berada pada fase pertumbuhan dan perkembangan yang sangat pesat (Hariani, 2024). Masa ini sering disebut sebagai masa keemasan atau golden age karena menjadi periode kritis dalam menentukan kualitas hidup anak di masa mendatang (Felicia Watratan, Puspita, dan Moeis, 2020). Pada fase ini, anak mengalami perkembangan fisik, kognitif, motorik, serta emosional yang berlangsung sangat cepat dan membutuhkan dukungan gizi yang cukup, baik dari segi jumlah maupun kualitasnya (Laila, Hardiansyah, dan Susilowati, 2023).

Secara umum, balita terbagi menjadi dua kelompok, yaitu batita (anak usia 1-3 tahun) yang masih sangat bergantung pada orang tua dalam hal perawatan dan pemberian makanan, serta anak prasekolah (usia 3-5 tahun) yang mulai memiliki kemandirian dalam memilih makanan dan berinteraksi dengan lingkungan sekitar (Yulia, Sari, Khoiriyah, Ade, dan Nuarita, 2021). Karena tingkat kerentanannya yang tinggi terhadap gangguan gizi dan infeksi, perhatian khusus perlu diberikan terhadap pola makan, perawatan, serta stimulasi tumbuh kembang pada periode ini. Pemenuhan gizi yang optimal, lingkungan yang sehat, serta peran orang tua yang responsif merupakan faktor penting dalam mendukung pertumbuhan balita

agar mencapai potensi maksimalnya (Laloan, Masithah, Rahmaniar MB, Wahyuni, dan Selvia, 2024).

2.4 Prediksi Risiko

Prediksi risiko adalah metode analitis yang bertujuan untuk menemukan, mengukur, dan memperkirakan kemungkinan suatu peristiwa yang dapat menyebabkan kerugian di masa depan. Prediksi risiko digunakan dalam sistem informasi dan organisasi untuk mencegah efek buruk dari ketidakpastian. Banyak kali, proses ini didukung oleh penggunaan metode statistik, pembelajaran mesin, dan model matematis yang dapat memproyeksikan tingkat risiko berdasarkan pola dan data historis tertentu (Rais dan Warjiyono, 2024).

Proses prediksi risiko tidak hanya mengandalkan intuisi, tetapi juga didukung oleh metode ilmiah, seperti:

1. Analisis statistik: Menggunakan data historis untuk menemukan pola dan tren.
2. Algoritma pembelajaran mesin (*machine learning*): Menggunakan model cerdas untuk memproyeksikan tingkat risiko berdasarkan data yang tersedia.
3. Model matematis: Menciptakan rumus untuk memproyeksikan risiko secara kuantitatif.

Prediksi risiko biasanya dilakukan dalam beberapa langkah: pengumpulan data, pemilihan variabel yang relevan, penerapan model prediksi, dan evaluasi hasil. Hasil dari proses ini dapat digunakan sebagai dasar untuk pengambilan keputusan strategis dalam bidang seperti pengelolaan proyek, keamanan data, kesehatan, dan keuangan. Oleh karena itu, prediksi risiko tidak hanya membantu dalam mengantisipasi masalah yang mungkin terjadi, tetapi juga menawarkan perspektif yang berbeda untuk membangun metode pengurangan yang lebih efisien (Yusuf, Razi, Aman, Terisia, dan Nurjayanti, 2025).

Prediksi risiko juga membantu organisasi dalam merancang strategi mitigasi yang efektif. Dengan mengetahui potensi masalah, pengambil keputusan dapat menyiapkan langkah antisipatif, seperti penyediaan sumber daya, penyusunan rencana darurat, atau penguatan sistem pengendalian internal. Dengan demikian, prediksi risiko tidak hanya berfungsi untuk mengantisipasi masalah, tetapi juga menjadi landasan untuk membangun strategi pencegahan dan pengurangan risiko (M. Fazlur Rahman Assauqi dan Zaehol Fatah, 2024).



2.5 Machine Learning

Machine learning didefinisikan sebagai sub-bidang kecerdasan buatan yang membekali sistem dengan kemampuan belajar secara mandiri dari data tanpa instruksi pemrograman manual (Ananto dkk., 2023). Melalui pemanfaatan algoritma matematis, teknologi ini mengidentifikasi tren dalam data masa lalu untuk menghasilkan estimasi atau kategorisasi pada informasi baru (Roihan, Sunarya, dan Rafika, 2020). Klasifikasi utamanya terdiri dari tiga tipe: *supervised*, *unsupervised*, dan *reinforcement learning*, yang masing-masing memiliki mekanisme unik, mulai dari penggunaan data berlabel hingga optimalisasi melalui sistem penghargaan (*reward*) (Gevindo dan Hendrik, 2025). Saat ini, implementasinya telah mencakup berbagai sektor krusial seperti keamanan siber, diagnostik medis, hingga sistem rekomendasi, menjadikannya pilar utama inovasi teknologi di masa sekarang (Pratama dkk., 2025).

Tujuan utama dari *machine learning* adalah menciptakan sistem yang memiliki kapasitas untuk belajar secara mandiri serta menetapkan keputusan tanpa instruksi manual yang berulang. Kapasitas ini memungkinkan perangkat untuk beradaptasi dan mengambil tindakan berdasarkan perubahan situasi secara dinamis. Operasional teknologi ini sangat bergantung pada ketersediaan data, di mana algoritma akan memproses dataset berskala besar (*big data*) guna mengidentifikasi pola-pola spesifik. Dalam proses ini, data berfungsi sebagai elemen fundamental pembelajaran agar mesin mampu memberikan *output* analisis yang akurat. Secara teknis, data dikelompokkan menjadi dua bagian: *data training* yang digunakan untuk membangun model algoritma, dan *data testing* yang berfungsi untuk mengevaluasi performa model saat dihadapkan pada informasi baru yang belum pernah dipelajari sebelumnya (Retnoningsih dan Pramudita, 2020).

Salah satu peran utama *machine learning* adalah meningkatkan kecerdasan sistem komputer melalui kemampuan generalisasi secara otomatis. Dengan memanfaatkan pengalaman dari data yang ada, sistem dapat membangun model pengklasifikasi yang akurat untuk mengestimasi nilai-nilai yang belum teridentifikasi. Secara umum, model klasifikasi dikategorikan ke dalam dua kelompok utama, yaitu:

1. Melakukan pelatihan data guna mengelompokkannya ke dalam kategori yang sesuai;
2. Menilai kemampuan model dalam mengenali objek yang tidak teridentifikasi sebelumnya.

Implementasi metode pembelajaran tanpa pengawasan menuntut persiapan data yang optimal serta dukungan bahasa pemrograman yang mumpuni untuk men-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber.

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



jalankan fungsi-fungsi cerdas pada mesin (Badri, Alvionita, Ramdani, Safitri, dan ...: 2025).

2.5.1 Random Forest

Random Forest (RF) merupakan algoritma yang beroperasi melalui mekanisme pemisahan biner rekursif untuk mencapai simpul akhir (*node*) dalam struktur pohon, baik untuk tugas klasifikasi maupun regresi. Algoritma ini dikenal unggul karena mampu meminimalkan tingkat kesalahan, memiliki performa klasifikasi yang stabil, serta efisien dalam memproses dataset pelatihan berskala besar. Selain itu, RF efektif dalam melakukan estimasi terhadap data yang hilang (*missing values*). Secara teknis, metode ini membangun kumpulan pohon keputusan secara mandiri, di mana setiap pohon disusun menggunakan teknik bootstrap pada subset data serta pemilihan variabel secara acak pada setiap node-nya (Pamuji dan Ramadhan, 2021).

Menurut Polamuri (2017), *Random Forest* merupakan algoritma yang sangat populer untuk mengolah dataset berskala besar karena akurasi prediksinya yang unggul serta basis strukturnya yang terdiri dari banyak pohon keputusan (Lukhia Britanthia Christina Tanujaya, Bambang Susanto, dan Asido Saragih, 2020). Pembentukan pohon-pohon ini dilakukan secara acak untuk kemudian diintegrasikan menjadi satu kesatuan model (Bimawan, Astuti, dan Arsi, 2024). Adapun metodologi dalam membangun struktur "hutan acak" ini didasarkan pada prinsip yang dikembangkan oleh Breiman (2001) serta Breiman Cutler (2003), yakni:

1. Mengaplikasikan teknik bootstrap melalui pengambilan sampel acak berukuran n dengan prinsip pengembalian.
2. Membangun setiap pohon secara maksimal tanpa pemangkasan, di mana pemilihan variabel penjelas dilakukan secara acak pada setiap *node* atau simpul.
3. Melakukan iterasi pada langkah-langkah sebelumnya hingga terbentuk sekumpulan pohon acak yang menyusun *Random Forest*.

Struktur ansambel pohon keputusan pada *Random Forest* menyebabkan algoritma ini tidak dapat mengidentifikasi signifikansi individual variabel, melainkan sebatas menunjukkan derajat kepentingan fiturnya (Fadellia Azzahra dkk., 2024). Prosedur penghitungan untuk menentukan tingkat kepentingan variabel ini telah diformulasikan oleh Sandri Zuccolotto (2008). Sebagai ilustrasi, jika terdapat variabel penjelas dengan

Untuk $h = 1, 2, \dots, q$, maka *Mean Decrease in Gini* (MDG) mengukur



tingkat kepentingan variabel penjelas X_h sebagai berikut:

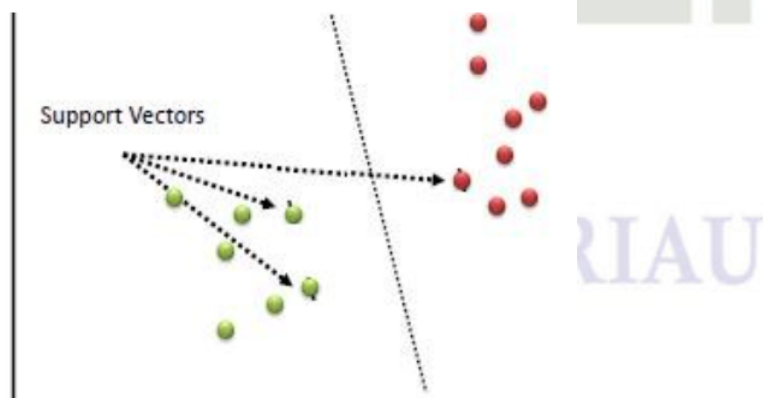
$$\text{MDG}_h = \frac{1}{k} \sum_t [d(h, t) \cdot I(h, t)] \quad (2.2)$$

Dengan:

- k = banyaknya pohon dalam *random forest*
- $d(h, t)$ = besar penurunan indeks Gini untuk variabel penjelas X_h pada simpul t
- $I(h, t)$ = indikator, bernilai 1 jika X_h memilih simpul t , dan 0 selainnya

2.5.2 Support Vector Machine (SVM)

SVM merupakan serangkaian metode *supervised learning* yang sering diimplementasikan dalam bidang medis untuk kebutuhan klasifikasi maupun regresi. Keunggulan SVM terletak pada kemampuannya untuk meminimalkan galat klasifikasi empiris sekaligus memaksimalkan margin geometris secara simultan, sehingga algoritma ini kerap disebut sebagai *Maximum Margin Classifier*. Secara teoretis, SVM berpijak pada prinsip *Structural Risk Minimization* (SRM) yang menjamin batasan risiko berdasarkan teori pembelajaran statistik. Algoritma ini bekerja dengan mencari *hyperplane optimal* yang mampu memisahkan kelompok data secara maksimal. Secara geometris, hyperplane ini merepresentasikan fungsi pemisah yang berbentuk garis pada ruang dua dimensi (2-D), bidang pada ruang tiga dimensi (3-D), dan menjadi *hyperplane* pada dimensi yang lebih tinggi (Abdusyukur, 2023). Visualisasi SVM dapat dilihat pada Gambar 2.5.



Gambar 2.5. Visualisasi *Support Vector Machine*

Sumber: (Kowalczyk, 2023)



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Namun ketika hasil akurasi yang didapatkan masih belum cukup baik dan tidak dapat digeneralisasikan dengan baik, maka diperlukan peningkatan akurasi ke dimensi yang lebih tinggi. Hal ini disebut sebagai fungsi kernel. Fungsi kernel adalah suatu fungsi k yang untuk semua vektor input x, z akan memenuhi kondisi:

$$k(x, z) = \phi(x)^T \phi(z) \quad (2.3)$$

di mana $\phi(\cdot)$ adalah fungsi pemetaan dari ruang input ke ruang fitur. Dengan kata lain, fungsi kernel merupakan fungsi perkalian dalam pada ruang fitur (F Handayani, 2021).

Teknik ini memungkinkan model beroperasi pada ruang fitur yang lebih luas tanpa perlu mendefinisikan transformasi fungsi pemetaan secara manual dari ruang asal. Hal ini memberikan keleluasaan bagi *hyperplane* untuk bertindak sebagai sekat pemisah antar kategori data. Dalam praktiknya, kernel *Gaussian* atau RBF menjadi pilihan populer karena karakteristiknya yang berbasis pada jarak antar titik data. Pemahaman mendalam mengenai SVM mensyaratkan analisis terhadap persamaan pembentuk *hyperplane* serta kendala optimasi yang menyertainya (Hovi, Id Hadiana, dan Rakhmat Umbara, 2023).

1. Persamaan *Hyperplane*

Untuk klasifikasi linear, *hyperplane* (bidang pemisah) didefinisikan oleh persamaan berikut:

$$W^T X + b = 0 \quad (2.4)$$

Dengan:

- W adalah vektor bobot (weight vector), tegak lurus terhadap *hyperplane*.
- X adalah vektor fitur input.
- b adalah bias (*intercept*).

2. Fungsi Keputusan (*Decision Function*)

Setelah *hyperplane* ditemukan, klasifikasi untuk titik data baru X_{new} dilakukan menggunakan fungsi keputusan:

$$h(x_{\text{new}}) = \text{sign}(W^T x_{\text{new}} + b) \quad (2.5)$$

Dengan:



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Jika $h(x_{\text{new}}) \geq 0$, maka data diklasifikasikan ke Kelas 1 (misalnya, $y_i = +1$).
- Jika $h(x_{\text{new}}) < 0$, maka data diklasifikasikan ke Kelas 2 (misalnya, $y_i = -1$).

3. Masalah Optimasi Inti (*Margin Maksimum*)

Tujuan utama SVM adalah untuk memaksimalkan margin, yaitu jarak antara hyperplane dan titik-titik data terdekat (*support vectors*) dari kedua kelas. Secara matematis, hal ini setara dengan meminimalkan norma dari vektor bobot, w (Damar Pratama, 2024).

Masalah Optimasi (Bentuk Primal untuk Separasi Linear Keras - *Hard Margin*):

$$\text{Minimize } \frac{1}{2} \|w\|^2 \quad (2.6)$$

$$\text{subject to } y_i(w^T x_i + b) \geq 1, \quad \text{untuk semua } i \quad (2.7)$$

Dengan:

- $\frac{1}{2} \|w\|^2$ adalah fungsi objektif yang harus diminimalkan (memaksimalkan margin $\frac{2}{\|w\|}$).
- y_i adalah label kelas untuk titik data ke- i (yaitu $+1$ atau -1).
- $y_i(w^T x_i + b) \geq 1$ adalah kendala yang memastikan setiap titik data diklasifikasikan dengan benar dan berada di luar margin.

4. *Soft Margin dan Kernel Trick*

Untuk data yang tidak terpisah secara linear atau memiliki outlier, SVM menggunakan:

- (a) *Soft Margin*: Memperkenalkan variabel slack ϵ_i dan parameter regulasi C pada masalah optimasi untuk memungkinkan beberapa misklasifikasi. Hal ini dikenal sebagai masalah *Hinge Loss* yang diminimalkan.

$$\min \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \max(0, 1 - y_i(w^T x_i + b)) \right) \quad (2.8)$$

- (b) *Kernel Trick*: Untuk data yang tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel (seperti *Radial Basis Function (RBF)* atau *Polynomial*) untuk memetakan data ke ruang dimensi

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumpukan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang lebih tinggi sehingga data menjadi terpisah secara linear. Kernel trick diimplementasikan dengan mengganti perkalian titik $\mathbf{X}_i^T \mathbf{X}_j$ dalam masalah optimasi bentuk *Dual* dengan fungsi kernel $K(\mathbf{X}_i, \mathbf{X}_j)$ (Simbolon dkk., 2022).

Beberapa kernel yang umum digunakan dalam SVM dapat dilihat pada Tabel 2.3.

Tabel 2.3. Kernel-kernel pada algoritma *Support Vector Machine* (SVM)

| No | Nama Kernel | Rumus | Penjelasan | Penggunaan Terbaik |
|----|--------------------------|---|--|--|
| 1 | Kernel Linear | $K(x_i, x_j) = x_i^T x_j$ | Bentuk paling sederhana, melakukan klasifikasi linear tanpa <i>kernel trick</i> . | Cocok untuk data yang terpisah secara linear atau dataset dengan fitur sangat banyak. Cepat dan efisien. |
| 2 | Kernel <i>Polinomial</i> | $K(x_i, x_j) = (x_i^T x_j + c)^d$ | Memungkinkan batas keputusan melengkung, mempertimbangkan interaksi fitur hingga derajat d . | Cocok untuk data dengan struktur polinomial, seperti pengenalan gambar. |
| 3 | Kernel RBF | $K(x_i, x_j) = e^{-\gamma \ x_i - x_j\ ^2}$ | Juga dikenal sebagai <i>Gaussian Kernel</i> . Memetakan data ke ruang fitur berdimensi tak hingga. | Cocok untuk klasifikasi non-linear tanpa pengetahuan awal tentang struktur data. |

Tabel 2.3 (lanjutan)

| No | Nama Kernel | Rumus | Penjelasan | Penggunaan Terbaik |
|----|----------------|---|--|--|
| 4 | Kernel Sigmoid | $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + c)$ | Berasal dari fungsi aktivasi JST. Kadang digunakan untuk membandingkan SVM dengan JST. | Cocok untuk eksperimen perbandingan SVM dan ANN dalam pemodelan hubungan non-linear. |

2.5.3 K-Nearest Neighbor (KNN)

KNN merupakan algoritma pembelajaran terawasi yang melakukan klasifikasi data dengan mengandalkan informasi dari *training record* terdekat. Dalam mekanisme ini, parameter k menentukan jumlah tetangga terdekat yang menjadi acuan dalam penentuan kelas. Prinsip operasionalnya berfokus pada penghitungan derajat perbedaan atau jarak antara data uji dengan seluruh data latih di dalam sistem, di mana klasifikasi dilakukan dengan mengidentifikasi nilai kemiripan tertinggi atau jarak terkecil (Supriyanto, Alita, dan Isnain, 2023). Penentuan kategori objek didasarkan pada mayoritas kelas yang muncul pada sejumlah K tetangga tersebut. Sebagai contoh, apabila nilai k=1, maka objek secara otomatis akan mengikuti kelas tetangga tunggal yang paling dekat dengannya. Nilai k sendiri merupakan konstanta yang didefinisikan oleh pengguna pada tahap awal klasifikasi guna memetakan data tes ke dalam kategori yang paling dominan di antara sampel referensi (Akbar, 2024). Visualisasi KNN dapat dilihat pada Gambar 2.6.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

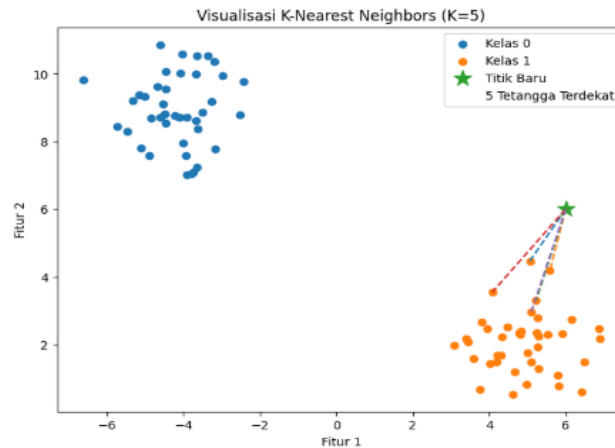
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.6. Visualisasi *K-Nearest Neighbor* (KNN)

Sumber: (Permana et al., 2023)

Jarak geometris, atau jarak antara dua titik, dihitung dengan rumus Pythagoras. Ini adalah panjang garis miring yang menghubungkan titik a dan b di ruang. Garis miring juga terbentang di antara sumbu x dan sumbu y dengan koordinat yang diberikan untuk titik a dan b. Rumus perhitungan untuk menentukan jarak antara dua titik, yaitu titik pelatihan (x) dan titik pengujian (y), ditunjukkan pada gambar berikut (Yusuf dkk., 2024). Untuk melakukan ini, digunakan rumus geometri, seperti yang ditunjukkan dalam persamaan berikut.

$$D(x,y) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2} \quad (2.9)$$

Dengan D adalah jarak antara titik pada *data training* x dan titik *data testing* yang akan diklasifikasi, dimana $x=x_1, x_2, \dots, x_i$ dan y_1, y_2, \dots, y_i dan i mempresentasikan nilai atribut serta n merupakan dimensi atribut.

Langkah-langkah untuk menghitung metode Algoritma KNN Rahman, Rahmah, dan Saribulan (2023):

1. Menentukan parameter k yang akan digunakan dalam proses pemungutan suara (*voting*).
2. Melakukan perhitungan jarak Euclidean untuk mengukur kemiripan antara objek input dengan seluruh data referensi.
3. Melakukan pemeringkatan objek dari jarak yang paling minimum hingga maksimum.
4. Mengelompokkan label kategori dari sekumpulan tetangga terdekat yang telah teridentifikasi.



5. Mengambil keputusan prediksi terhadap *query instance* dengan merujuk pada kelas mayoritas dalam himpunan tetangga tersebut.

2.5.4 Naive Bayes

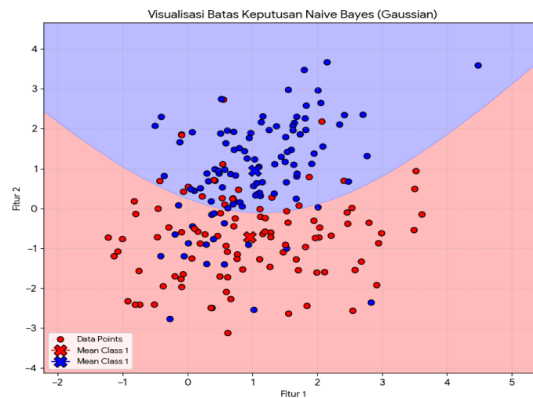
Naive Bayes Classifier merupakan algoritma yang mengimplementasikan Teorema Bayes melalui pendekatan probabilitas dan statistika guna memproyeksikan peluang di masa mendatang berdasarkan data historis. Karakteristik fundamental dari metode ini terletak pada asumsi "naif" yang sangat kuat, di mana setiap variabel atau kondisi dianggap bersifat independen atau berdiri sendiri (Wijaya dan Dwiasnati, 2020). Merujuk pada Olson Delen (2008), algoritma ini mengkalkulasi probabilitas setiap kelas keputusan dengan anggapan bahwa kelas tersebut valid berdasarkan vektor informasi yang tersedia, sambil mengasumsikan tidak adanya keterkaitan antar atribut objek. Proses estimasi akhir dilakukan dengan menghitung frekuensi dari tabel keputusan utama sebagai dasar probabilitas. Dalam berbagai pengujian, *Naive Bayes Classifier* terbukti menunjukkan performa yang sangat kompetitif dan unggul dibandingkan dengan model klasifikasi lainnya (Felicia Wattratan dkk., 2020).

Dalam jurnalnya yang berjudul "*NaiveBayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*", Xhemali Hinde Stone menunjukkan bahwa "*NaiveBayes Classifier* memiliki tingkat akurasi yang lebih baik dibanding model classifier lainnya (Syahril Dwi Prasetyo dkk., 2023). Keuntungan utama metode ini terletak pada kemampuannya mengestimasi parameter secara akurat dengan jumlah data pelatihan yang minimum. Karena setiap variabel dianggap saling bebas (*independen*), klasifikasi dapat dilakukan cukup dengan menghitung varians per variabel tanpa melibatkan matriks kovarians total (Afriansyah dkk., 2024). Visualisasi *Naive Bayes* dapat dilihat pada Gambar 2.7.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 2.7. Visualisasi Naïve Bayes

Sumber: (Mulyoto, 2024)

Persamaan Metoda *Naïve Bayes*

Persamaan dari teorema Bayes adalah:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.10)$$

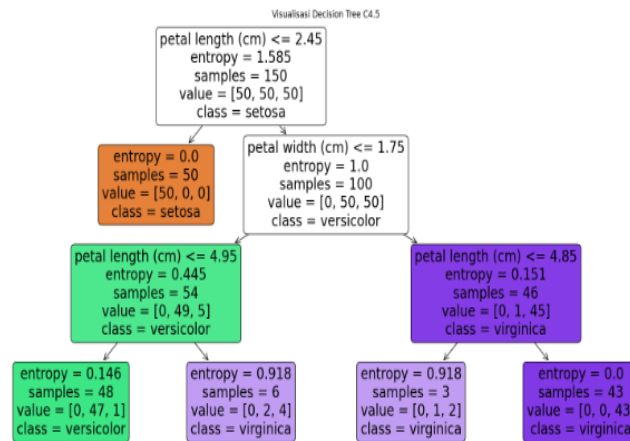
Dimana:

- X : Data dengan kelas yang belum diketahui.
- H : Hipotesis bahwa data merupakan suatu kelas spesifik.
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (posterior probabilitas).
- $P(H)$: Probabilitas hipotesis H (prior probabilitas).
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H .
- $P(X)$: Probabilitas X .

2.5.5 *Decision Tree C4.5*

Dalam disiplin ilmu data, *Decision Tree* didefinisikan sebagai metodologi pemodelan *prediktif* yang mengonversi *dataset* kompleks menjadi struktur keputusan yang menyerupai pohon. Keunggulan algoritma ini terletak pada kemampuannya untuk memecah masalah besar menjadi sub-masalah yang lebih sederhana melalui kriteria pemisahan (*splitting criteria*) pada setiap node. Struktur ini dimulai dari *root node* sebagai titik awal, kemudian diikuti oleh berbagai cabang yang

masing-masing merepresentasikan aturan atau syarat atribut yang harus dipenuhi sebelum mencapai keputusan *final* di bagian daun. Pada Penelitian Hana (2020) menekankan bahwa operasional algoritma ini sangat bergantung pada pengorganisasian data dalam bentuk tabel yang memuat atribut kategorikal maupun numerik (Hana, 2020). Dengan mengandalkan pemisahan rekaman data berdasarkan kemiripan pola, *Decision Tree* mampu memberikan transparansi logika dalam setiap klasifikasi yang dihasilkan. Visualisasi *Decision Tree* dapat dilihat pada Gambar 2.8.



Gambar 2.8. Visualisasi *Decision Tree* C4.5 dengan K5

Sumber: (Raharjo, 2022)

C4.5 merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) yang diperkenalkan oleh J. Ross Quinlan pada tahun 1993. Algoritma ini dirancang untuk mengatasi beberapa keterbatasan yang dimiliki oleh ID3. Secara umum, C4.5 dikenal karena kemampuannya dalam menangani berbagai jenis data, menjadikannya salah satu algoritma *Decision Tree* yang paling populer dan efisien (Saleh, Informatika, dan Komputer, 2020).

Fitur utama dan keunggulan algoritma C4.5 meliputi:

1. Menangani Data Kontinu: C4.5 dapat memproses fitur dengan nilai diskrit maupun kontinu (*numerik*). Untuk data kontinu, algoritma ini secara dinamis menentukan batas ambang (*threshold*) terbaik untuk membagi data menjadi dua kelompok biner.
2. Menangani Data Hilang (*Missing Values*): C4.5 memiliki mekanisme untuk menangani nilai fitur yang hilang (*missing value*) dengan memberikan bobot (*weight*) pada observasi berdasarkan probabilitas nilai tersebut.
3. Pruning Pohon: C4.5 menggunakan teknik pruning (pemangkasan) post-hoc untuk menghilangkan cabang-cabang yang tidak signifikan. Tujuan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

pruning adalah mencegah *overfitting* (terlalu cocok dengan data latih) dan meningkatkan generalisasi model.

- Menangani Fitur dengan Biaya yang Berbeda: C4.5 dapat memperhitungkan biaya yang berbeda terkait dengan kesalahan klasifikasi.

Algoritma C4.5 terletak di tengah pilihan karakteristik utama untuk pembagi (pembagi) setiap saat di tengah penggunaan metrik Rasio Perolehan (Rasio Perolehan). C4.5 terletak di tengah pemilihan karakteristik utama untuk pembagi (pembagi) setiap titik di tengah penggunaan metrik Rasio Perolehan (Rasio Perolehan) (Amananti, 2024).

- Entropi (*Entropy*)

Entropi adalah ukuran ketidakmurnian (*impurity*) atau ketidakpastian dalam sekumpulan data S

$$\text{Entropy}(S) = - \sum_{i=1}^C p_i \log_2(p_i) \quad (2.11)$$

Dengan:

- S adalah *dataset* atau *subset data* yang akan dihitung entropinya.
- C adalah jumlah kelas yang berbeda dalam S .
- p_i adalah proporsi sampel yang termasuk dalam kelas ke- i di dalam himpunan S .

- Information Gain* (Perolehan Informasi)

Information Gain mengukur seberapa banyak pengurangan *entropi* (peningkatan kemurnian) yang dihasilkan jika data dibagi berdasarkan fitur S (Amananti, 2024a)

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v) \quad (2.12)$$

Dengan:

- A adalah fitur yang diuji sebagai pemisah.
- $\text{values}(A)$ adalah semua nilai unik dari fitur A .
- $|S_v|$ adalah jumlah sampel di subset S yang memiliki nilai v untuk fitur A .
- $|S|$ adalah total jumlah sampel dalam himpunan S .

- Split Information* (Informasi Pemisahan)



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Split Information digunakan untuk menormalkan *Information Gain*. Ini mengukur keragaman distribusi nilai fitur A dan dirancang untuk menghukum fitur yang membagi data menjadi banyak subset kecil (sehingga bias *Information Gain* teratasi) (Darmawan, 2022).

$$Split\ Information(S,A) = - \sum_{v \in values(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right) \quad (2.13)$$

4. *Gain Ratio* (Rasio Perolehan)

Gain Ratio adalah metrik utama yang digunakan oleh C4.5. Ia menyeimbangkan *Information Gain* dengan *Split Information*. Fitur dengan nilai *Gain Ratio* tertinggi yang akan dipilih sebagai pemisah terbaik untuk *node* tersebut (Elisabeth, 2025).

$$Gain\ Ratio(S,A) = \frac{Gain(S,A)}{Split\ Information(S,A)} \quad (2.14)$$

Perhitungan ini akan diulang untuk semua fitur yang tersisa pada *node* saat ini menggunakan algoritma C4.5. Fitur yang menghasilkan nilai persentase peningkatan tertinggi akan dipilih untuk membagi data pada *node* tersebut, dan proses akan diulang secara rekursif hingga pohon selesai terbentuk (Amananti, 2024b).

2.6 Matriks Evaluasi

Confusion Matrix atau matriks evaluasi diterapkan sebagai instrumen untuk mengukur tingkat akurasi model. Dalam mengevaluasi performa hasil klasifikasi, metode ini menggunakan empat parameter utama sebagai indikator penilaiannya (Retnoningsih dan Pramudita, 2020).

Adapun keempat istilah tersebut yaitu:

1. *False Positive* (FP), yaitu data negatif tapi terprediksi sebagai data positif.
2. *False Negative* (FN), yaitu data positif yang terprediksi sebagai data negatif.
3. *True Positive* (TP), yaitu data positif yang terprediksi benar.
4. *True Negative* (TN), yaitu data negatif yang terprediksi dengan benar.

Untuk melihat Tabel *confusion matrix* dapat dilihat pada Tabel 2.4.

Tabel 2.4. Tabel *Confusion Matrix*

| Kondisi Aktual (Sebenarnya) | Prediksi Positif | Prediksi Negatif |
|--------------------------------|----------------------------|----------------------------|
| Positif | <i>True Positive</i> (TP) | <i>False Negative</i> (FN) |
| Negatif | <i>False Positive</i> (FP) | <i>True Negative</i> (TN) |

Untuk menghitung akurasi digunakan rumus sebagai berikut:

$$\text{Akurasi} = \frac{TP + FN}{TP + FN + FP + TN} \times 100\% \quad (2.15)$$

Metrik *Recall* atau *sensitivitas* digunakan untuk mengevaluasi proporsi hasil positif sebenarnya yang berhasil dideteksi oleh sistem. Rasio ini membandingkan jumlah prediksi positif yang benar terhadap seluruh populasi data yang memiliki label positif asli. Relevansi sensitivitas terletak pada kemampuan model dalam mengenali target positif dari total kasus yang memang seharusnya bernilai positif (Paryatno, Hari Linda Lukitowati, dan M. Arna Ramadhan, 2022). Guna melakukan kalkulasi pada metrik ini, diperlukan

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \quad (2.16)$$

Sementara itu, *Precision* didefinisikan sebagai perbandingan antara hasil prediksi positif yang benar dengan seluruh data yang diprediksi sebagai positif oleh sistem. Metrik ini merepresentasikan efektivitas model dalam menyajikan informasi yang relevan dan akurat. Dalam konteks *data mining*, nilai *precision* diperoleh dari pembagian antara jumlah data *true positive* dengan total data yang diidentifikasi sebagai kelas positif. Untuk menghitung *precision* menggunakan persamaan dibawah ini:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.17)$$

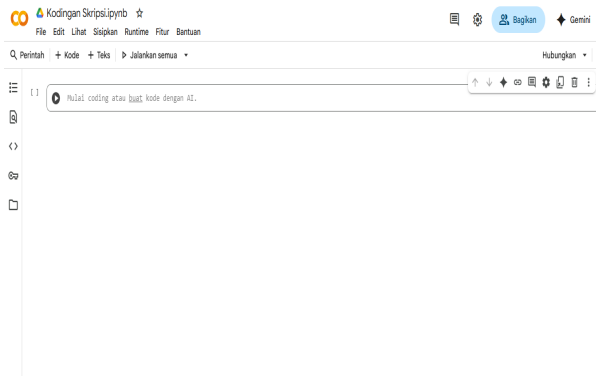
2.7 Google Colab

Google Colaboratory (*Colab*) adalah lingkungan notebook berbasis cloud yang disediakan oleh *Google*, memungkinkan pengguna untuk menulis dan menjalankan kode *Python* langsung melalui *browser*. *Colab* pada dasarnya merupakan implementasi dari *Jupyter Notebook* yang di-hosting dan diperkuat oleh infrastruktur *Google*, menjadikannya standar ideal untuk proyek-proyek *data science* dan *machine learning* (ML) (Fajrizal, Handayani, dan Taslim, 2024). Gambar 2.9

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

merupakan tampilan *notebook* baru di *Google Colab* dan Gambar 2.10 adalah logo *Google Colab*.



Gambar 2.9. Tampilan *Notebook* Baru di *Google Colab*



Gambar 2.10. Tampilan Logo *Google Colab*

Google Collaboratory, juga disebut *Google Colab*, adalah alat penelitian gratis yang berbasis cloud. Ini mirip dengan *Jupyter Notebook* karena dibuat dengan lingkungan *Jupyter* dan mendukung hampir semua buku bacaan yang diperlukan dalam lingkungan pengembangan AI (Handika, 2024). *Google Colab* memiliki beberapa keuntungan untuk digunakan (Sukarene dkk., 2023).

1. *Free Access* (Penggunaan Gratis)

Google Colab ditujukan untuk para peneliti yang sedang mengembangkan penelitian dan membutuhkan komputer yang canggih. *Google Colab* membutuhkan koneksi internet

2. *Good Specification* (Spesifikasi yang Baik)

Ketika kita menginstall *Google Colab* untuk pertama kalinya, kita akan menerima akses ke cloud komputer dengan spesifikasi berikut (diupdate 2019):



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- (a) GPU Nvidia K80s, T4s, P4s dan P100s.
- (b) RAM 13 GB
- (c) Disk 130 GB

3. Zero Configuration (Tanpa Konfigurasi)

Tidak ada konfigurasi yang diperlukan saat menggunakan *Google Colab*; namun, ketika kita ingin menambahkan *library* baru, kita harus menginstal paket *library*. Dengan *Easy Sharing* (Pengiriman Mudah), kita dapat mengintegrasikan *google drive* milik kita dan menyimpan *script* ke dalam proyek github (Wilyani, Nuryan Arif, dan Aslimar, 2024).

Google Internal Research telah meluncurkan alat baru yang disebut *Google Colaboratory*, atau *Colab*. Alat ini bertujuan untuk membantu peneliti melakukan eksperimen dengan pengolahan data, terutama dalam bidang pembelajaran mesin. Penggunaan *Google Colab* mirip dengan *Jupyter Notebook* karena dibangun di atas lingkungan *Jupyter*, sehingga tidak perlu mengatur apa pun sebelum digunakan. Selain itu, dengan menggunakan media penyimpanan *Google Drive*, *Google Colab* berjalan sepenuhnya di *Cloud* (Handika, 2024).

Alat *Colab* yang menarik ini memungkinkan penggunanya menggunakan layanan GPU sebagai backend komputasi secara gratis. Selain itu, memungkinkan penggunaan bersama selama 12 jam. Kita tidak perlu mengeluarkan banyak uang atau membeli perangkat komputer dengan GPU tambahan jika kita ingin belajar mesin pembelajaran. Proses pelatihan mesin pembelajaran cukup sulit, terutama jika sumber daya tidak memadai, dan mungkin gagal jika kita hanya mencobanya (Handika, 2024).

Google Colab memungkinkan untuk membuat aplikasi berbasis *Deep Learning* dengan menggunakan pustaka terkenal seperti *Pandas*, *TensorFlow*, *PyTorch*, dan *OpenCV*. Alat-alat ini hampir mirip dengan *Jupyter Notebook*, tetapi *Colab* beroperasi di *cloud Google* dan menyimpan file ke *Google Drive*. Salah satu perbedaan utama dari *Jupyter Notebook* adalah kita tidak dapat menggunakan *syntax Python* dan *Markdown*, tetapi kita dapat menjalankan perintah komando langsung pada *cell notebook* dengan tanda `"""` (Yukandri et al., 2024).

2.7.1 Keunggulan Python dalam Lingkungan Google Colab

Colab berfungsi sebagai jembatan yang menyederhanakan penggunaan Python dengan menghilangkan hambatan konfigurasi lingkungan lokal. Fitur-fitur utamanya meliputi:

1. item Akses Gratis ke Sumber Daya Komputasi: *Colab* menyediakan ak-



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

ses gratis ke Unit Pemrosesan Grafis (GPU) dan Unit Pemrosesan Tensor (TPU). Sumber daya ini sangat penting bagi *Python*, terutama saat menjalankan proses komputasi intensif seperti pelatihan model *Deep Learning* menggunakan *framework* seperti *TensorFlow* atau *PyTorch*.

- Integrasi Ekosistem *Google: Colab* terintegrasi langsung dengan layanan *cloud Google*, seperti *Google Drive*. Hal ini memungkinkan pengguna *Python* untuk memuat dataset besar, menyimpan model yang telah dilatih, dan berbagi *notebook* dengan mudah, menyederhanakan alur kerja kolaboratif.
- Lingkungan *Python* Siap Pakai: *Colab* hadir dengan sebagian besar Pustaka (*Library*) *Python* utama yang sudah terinstal, termasuk *NumPy* (untuk operasi numerik), *Pandas* (untuk manipulasi data), *Matplotlib* dan *Seaborn* (untuk visualisasi), serta pustaka-pustaka ML kritis seperti *Scikit-learn*. Ini memastikan pengguna dapat langsung mengimpor dan menjalankan kode tanpa melalui proses instalasi yang rumit (Yuliyannah Sain, Aynun Andriani, dan Nahla Nurhidayah, 2023).

2.7.2 Pustaka *Python* Utama untuk Analisis dan *Machine Learning*

Dalam konteks skripsi ini, peran *Python* di *Colab* sangat vital karena mendukung seluruh siklus proyek *machine learning* melalui beberapa pustaka utama pada Tabel 2.5 (Tritularsih dan Prasetyo, 2025).

Tabel 2.5. Pustaka *Python* Utama

| Pustaka (<i>Library</i>) | Peran Kunci dalam <i>Colab</i> |
|----------------------------|--|
| <i>Pandas</i> | Digunakan untuk mengimpor data (<i>dataset</i>), memuatnya ke dalam struktur data <i>DataFrame</i> , serta melakukan pembersihan, agregasi, dan transformasi data. |
| <i>NumPy</i> | Pustaka dasar untuk operasi matematika tingkat tinggi pada data, termasuk manipulasi <i>array</i> dan matriks, yang menjadi fondasi bagi semua perhitungan numerik dalam ML. |
| <i>Scikit-learn</i> | Pustaka standar untuk algoritma ML klasik, seperti SVM, KNN, dan regresi. Digunakan untuk melatih model klasifikasi dan regresi. |



Tabel 2.5 Pustaka *Python* Utama

| Pustaka (Library) | Peran Kunci dalam <i>Colab</i> |
|----------------------------------|--|
| <i>Matplotlib/Seaborn</i> | Pustaka visualisasi data untuk menghasilkan grafik, plot, dan diagram, yang penting untuk Eksplorasi Data Analisis (EDA) dan presentasi hasil model. |

2.7.3 Struktur *Notebook Colab* dan Implementasi Kode

Colab membagi kode *Python* ke dalam sel (*cell*) yang dapat dijalankan secara independen atau berurutan. Pendekatan ini mendukung pengembangan kode yang bersifat *interaktif* dan *eksploratif*, sangat cocok untuk proses prototyping dan dokumentasi ilmiah (Soen, Marlina, dan Renny, 2022).

Dalam implementasinya, setiap langkah proyek (misalnya, pemuatan data, pra-pemrosesan, pelatihan model, dan evaluasi) dapat diwakili oleh sel terpisah, memastikan bahwa:

1. **Reproduksibilitas:** Seluruh langkah analisis terdokumentasi dengan jelas dan dapat direplikasi oleh pihak lain.
2. ***Debugging* yang Mudah:** Pengembang dapat menguji dan memperbaiki bagian kode secara terisolasi tanpa harus menjalankan ulang seluruh skrip.

Secara keseluruhan, *Colab* menyajikan *Python* bukan hanya sebagai bahasa pemrograman, tetapi sebagai lingkungan komputasi *end-to-end* yang efisien, kolaboratif, dan kuat untuk melaksanakan penelitian *machine learning* yang akurat dan dapat direplikasi (Soen dkk., 2022).

2.8 Penelitian Terkait

Penelitian terkait disusun untuk memberikan gambaran mengenai penelitian-penelitian sebelumnya yang relevan dengan topik yang dibahas dalam skripsi ini. Tinjauan ini bertujuan untuk mengetahui metode, objek, serta hasil penelitian terdahulu sehingga dapat menjadi dasar perbandingan, rujukan ilmiah, dan pendukung dalam menentukan kebaruan serta posisi penelitian yang dilakukan. Ringkasan penelitian terkait tersebut disajikan pada Tabel 2.6 berikut.

Tabel 2.6. Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|---|---|---|
| 1 | 2024 | Wulan Sri Lestari, Yuni Marlina Saragih, Caroline | Multiclass Classification for Stunting Prediction Using Deep Neural Networks | Penelitian menggunakan <i>Deep Neural Networks</i> untuk klasifikasi multikelas status stunting. Model diuji dengan beberapa learning rate dan skema pembagian data, dengan akurasi terbaik mencapai 93,83%, sehingga DNN terbukti efektif untuk prediksi stunting. |
| 2 | 2024 | Fadellia Azzahra, Nana Suarna, Yudhistira Arie Wijaya | Penerapan Algoritma <i>Random Forest</i> dan <i>Cross Validation</i> untuk Prediksi Data Stunting | <i>Random Forest</i> digunakan untuk memprediksi status gizi balita dengan akurasi 77,55% pada 40 pohon dan meningkat menjadi 78,33% menggunakan 10-fold cross validation, sehingga meningkatkan ketepatan identifikasi stunting. |

Hak Cipta Dilindungi Undang-Undang

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.6 Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|--|---|--|
| 4 | 2024 | Ramadya Wahyu Dwiananto, Arif Setia Sandi A, Rian Ardianto | Klasifikasi Berisiko Stunting Pada Balita: Perbandingan KNN, <i>Naïve Bayes</i> , dan SVM | Hasil menunjukkan SVM memiliki performa terbaik dengan akurasi 90%, diikuti KNN 89% dan <i>Naïve Bayes</i> 85%. Model terbaik diimplementasikan dalam website berbasis <i>Flask</i> untuk mendukung deteksi dini stunting. |
| | 2024 | Indah Pratiwi Putri, Terttiaavini, Nur Arminarahmah | Analisis Perbandingan Algoritma <i>Machine Learning</i> untuk Prediksi Stunting | <i>Random Forest</i> menjadi model terbaik dengan akurasi 87,75% dan F1-score 0,922, sedangkan KNN unggul pada nilai recall, menunjukkan keunggulan masing-masing algoritma. |
| | 2025 | Romlah, Sutan Faisal, Rahmat, Jamaludin Indra | Prediksi Risiko Angka Stunting Menggunakan SVM | Model SVM mencapai akurasi 80% pada kategori tidak stunting, namun performa pada kategori stunting masih rendah akibat ketidakseimbangan data sehingga diperlukan <i>SMOTE</i> dan <i>tuning parameter</i> . |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.6 Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|----------------------------|---|---|
| 6 | 2024 | Umairah Rizky Gurning dkk. | Prediksi Risiko Stunting Menggunakan <i>Naïve Bayes</i> dan <i>Chi-Square</i> | <i>Naïve Bayes</i> mencapai akurasi 94,3%. Seleksi fitur <i>Chi-Square</i> menghasilkan lima atribut utama yang mempercepat pemodelan tanpa menurunkan akurasi. |
| | 2023 | Widya Cholid Wahyudin dkk. | Prediksi Stunting Menggunakan <i>Naïve Bayes</i> | Menggunakan 300 data balita dan menghasilkan akurasi 85,33%, menunjukkan efektivitas <i>Naïve Bayes</i> dalam deteksi dini stunting. |
| | 2022 | Yulia Wardita dkk. | Model Prediksi Kejadian Stunting Berdasarkan Faktor Ibu dan Pola Asuh | Infeksi menjadi faktor utama penyebab stunting, sedangkan pengetahuan dan pola asuh ibu berperan tidak langsung melalui imunisasi dan pencegahan infeksi. |
| | 2024 | Ratnasari dkk. | Deteksi Dini Stunting Berdasarkan Indikator Antropometri | <i>Random Forest</i> menghasilkan akurasi tertinggi sebesar 92,7%. Faktor utama prediksi adalah tinggi badan dan usia balita. |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.6 Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|----------------------------------|--|---|
| 0 | 2021 | Demsa Simbolon dkk. | Peningkatan Kemampuan Kader Kesehatan dalam Deteksi Stunting | Pelatihan tiga hari terbukti meningkatkan pengetahuan dan keterampilan kader kesehatan dalam mendeteksi stunting pada anak usia 6–24 bulan. |
| 1 | 2023 | Dedi Gunawan, Verania Nur Andika | Implementasi Teorema Bayes pada Sistem Informasi Posyandu | Sistem berbasis web menggunakan <i>Naïve Bayes</i> dengan akurasi 60%. Pengujian usability menghasilkan skor SUS 76,5 (kategori baik). |
| 12 | 2025 | Taufik Hidayat dkk. | Prediksi Prevalensi Stunting dengan SVM dan SMOTE | SVM yang dipadukan dengan SMOTE menghasilkan akurasi 94%, precision 95%, recall 94%, dan F1-score 94%. |
| 3 | 2024 | Nurul Qisthi dkk. | Prediksi Stunting Menggunakan Algoritma C4.5 | Model C4.5 mencapai akurasi 98,88% dan ROC-AUC 98,98%, serta mudah diinterpretasikan untuk tenaga kesehatan. |
| 4 | 2024 | Natalia Situmeang dkk. | Penggunaan Algoritma KNN untuk Deteksi Stunting | KNN mencapai akurasi 97% dengan nilai precision, recall, dan F1-score yang seimbang setelah proses normalisasi data. |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tabel 2.6 Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|----------------------------|---|---|
| 15 | 2024 | Yudi Mulyanto dkk. | Klasifikasi Stunting Menggunakan <i>Naïve Bayes</i> | Menggunakan 5.829 data balita dan menghasilkan akurasi 95,08%, sehingga metode ini cepat dan efektif untuk deteksi dini stunting. |
| 16 | 2025 | Azza M. Nisfi dkk. | Perbandingan Algoritma ML dalam Pendeteksian Stunting | Random Forest memberikan akurasi tertinggi 81,78%, namun <i>Decision Tree</i> memiliki keseimbangan <i>precision</i> dan <i>recall</i> yang lebih stabil. |
| 17 | 2025 | Novita Ranti Muntiari dkk. | Deteksi Stunting dengan Perbandingan Algoritma ML | <i>Decision Tree</i> , KNN, dan <i>Random Forest</i> mencapai akurasi 99%, sedangkan <i>Naïve Bayes</i> hanya 48%. |
| 18 | 2025 | Diana Yusuf dkk. | Prediksi Risiko Stunting dengan <i>Logistic Regression</i> dan <i>Decision Tree</i> | <i>Logistic Regression</i> menunjukkan performa stabil, sedangkan <i>Decision Tree</i> mengalami <i>overfitting</i> pada data uji. |
| 19 | 2025 | Muhammad Verdyan-syah dkk. | Analisis Faktor Risiko Stunting dengan <i>Random Forest</i> | <i>Random Forest</i> menghasilkan akurasi 85% dan AUC 0,91. Faktor dominan adalah sanitasi, gizi buruk, dan usia ibu. |

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Tabel 2.6 Ringkasan Penelitian Terkait Prediksi dan Deteksi Stunting

| No | Tahun | Peneliti | Judul | Keterangan |
|----|-------|--------------------------------------|---|---|
| 20 | 2024 | Nazhifatul Muthohharoh, Zaehol Fatah | Perbandingan <i>Decision Tree</i> dan <i>Deep Learning</i> untuk Klasifikasi Stunting | <i>Deep Learning</i> memiliki akurasi lebih tinggi dibanding C4.5 dan direkomendasikan untuk sistem deteksi stunting berbasis data. |

Penelitian terdahulu mengenai prediksi risiko stunting pada balita seringkali terbatas, baik dari sisi jumlah algoritma *machine learning* yang dibandingkan, kebanyakan hanya satu hingga tiga seperti *Random Forest*, SVM, dan *Naïve Bayes*, maupun dari sisi variasi data yang hanya berasal dari satu wilayah. Akibatnya, generalisasi model menjadi sulit, dan beberapa sistem berbasis web yang dikembangkan pun masih menunjukkan akurasi rendah serta evaluasi yang kurang komprehensif. Sementara itu, penelitian yang menggunakan algoritma SVM dengan akurasi yang baik masih bermasalah dengan ketidakseimbangan data dan tidak membandingkan dengan algoritma lain yang potensial.

Oleh karena itu, penelitian yang dilakukan saat ini menawarkan kebaruan signifikan dengan membandingkan lima algoritma sekaligus (*Random Forest*, SVM, KNN, *Naïve Bayes*, dan *Decision Tree C4.5*) dan menggunakan sumber data dari platform Kaggle yaitu *stunting balita detection 121k-rows* untuk mendapatkan data yang lebih *representatif*. Fokus pada variabel utama seperti usia, jenis kelamin, dan tinggi badan bertujuan untuk menyederhanakan prediksi namun tetap menjaga relevansi klinis. Selain itu, evaluasi model dilakukan secara menyeluruh dengan empat metrik utama (akurasi, *precision*, *recall*, dan *F1-score*), yang diharapkan akan menghasilkan model prediksi risiko stunting yang lebih akurat, seimbang, dan praktis untuk deteksi dini.

Hak Cipta Dilindungi Undang-Undang

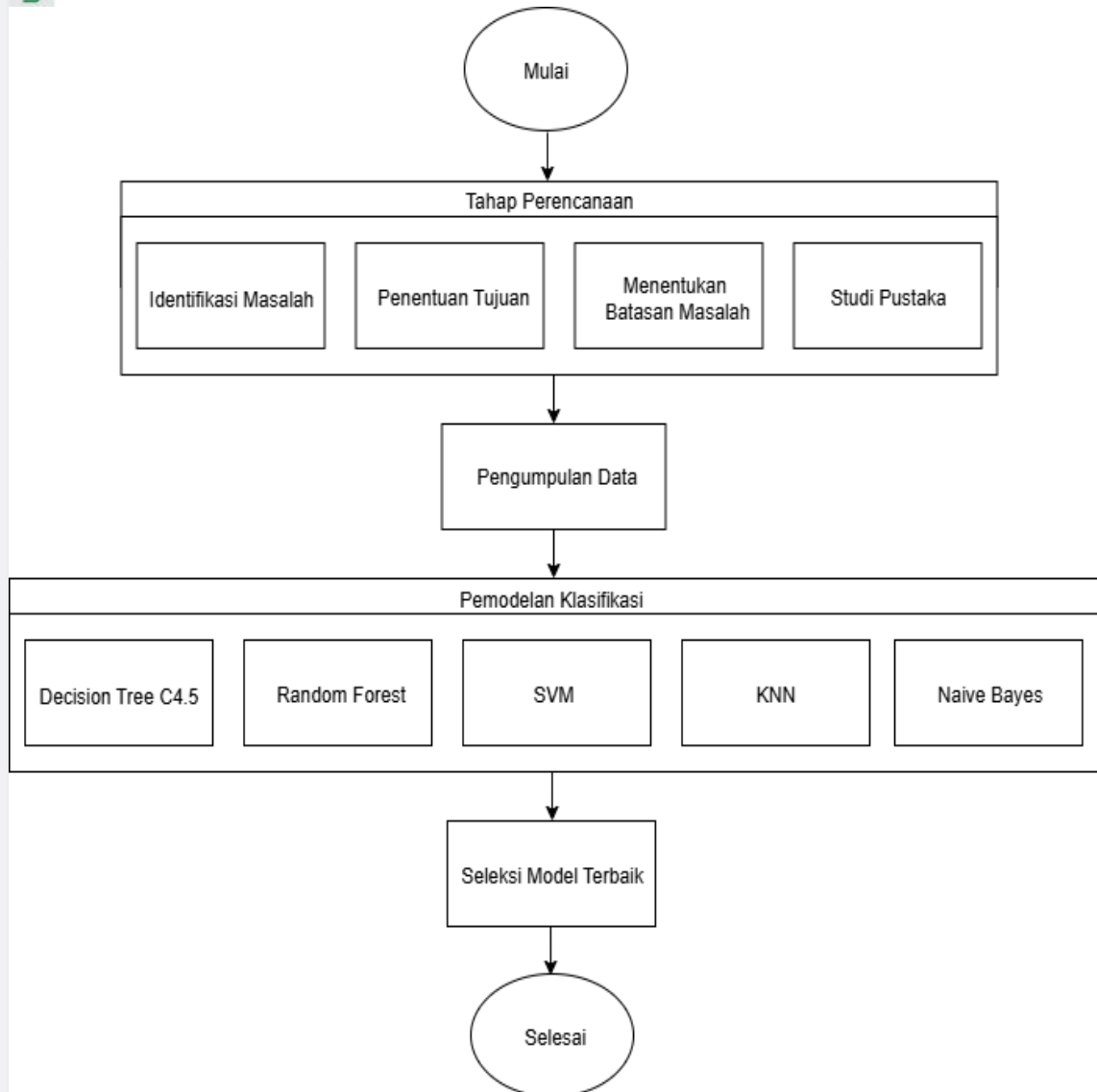
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

BAB 3 METODOLOGI PENELITIAN

Adapun proses penelitian adalah Tahap Perencanaan, Pengumpulan Data, *Preprocessing Data*, Pemodelan *Machine Learning*, serta Evaluasi dan Seleksi Model. Alur metodologi untuk penelitian kali ini dapat dilihat pada Gambar 3.1.



Gambar 3.1. Alur Metodologi Penelitian

3.1 Tahap Perencanaan

Tahap perencanaan merupakan langkah awal yang perlu dilakukan agar tujuan penelitian lebih jelas dan teratur.

1. Identifikasi Masalah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pemilihan metode dalam penelitian ini didasarkan pada permasalahan tingginya angka stunting di Indonesia serta keterbatasan sistem deteksi dini yang masih bersifat manual dan kurang akurat. Berdasarkan hasil kajian terhadap penelitian terdahulu, sebagian besar studi masih terbatas pada penggunaan satu hingga dua algoritma *machine learning* dengan jumlah data yang kecil dan kurang bervariasi, sehingga hasil prediksi belum optimal. Untuk mengatasi hal tersebut, penelitian ini membandingkan lima algoritma klasifikasi, yaitu *Random Forest*, SVM, KNN, *Naïve Bayes*, dan *Decision Tree C4.5*, menggunakan dataset “*Stunting Balita Detection (121K Rows)*” dari platform Kaggle. Pendekatan ini diharapkan dapat menghasilkan model klasifikasi risiko stunting yang lebih akurat dan dapat digunakan sebagai dasar pengembangan sistem deteksi dini berbasis data

2. Penentuan Tujuan

Tujuan utama penelitian ini, yang didorong oleh tingginya angka stunting dan keterbatasan deteksi dini, adalah mengembangkan model prediksi risiko stunting yang akurat dan efisien dengan cara menerapkan dan membandingkan lima algoritma *machine learning* (*Random Forest*, SVM, KNN, *Naïve Bayes*, dan *Decision Tree C4.5*). Selain itu, penelitian ini juga bertujuan menentukan algoritma terbaik berdasarkan evaluasi komprehensif menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*, sebagai dasar untuk membangun sistem pendukung keputusan deteksi dini stunting.

3. Menentukan Batasan Masalah

Batasan masalah dalam penelitian ini ditetapkan agar fokus dan terarah, yaitu hanya menggunakan data sekunder dari platform Kaggle, variabel yang dianalisis dibatasi pada usia, jenis kelamin, dan tinggi badan, hanya melibatkan penerapan dan evaluasi lima algoritma *machine learning* (*Random Forest*, SVM, KNN, *Naïve Bayes*, dan *Decision Tree C4.5*) tanpa optimasi atau seleksi fitur lanjutan; dan performa model diukur hanya dengan empat metrik utama (akurasi, *precision*, *recall*, dan *F1-score*). Dengan batasan ini, penelitian difokuskan pada pengujian dan perbandingan efektivitas kelima algoritma tersebut dalam memprediksi risiko stunting.

4. Studi Pustaka

Studi pustaka dilakukan dengan menelaah referensi (jurnal, buku, laporan resmi) untuk memperoleh pemahaman mendalam tentang konsep, faktor penyebab, dampak, dan upaya pencegahan stunting. Fokus lain dari studi ini adalah mengkaji penelitian terdahulu yang menggunakan algoritma *machine*



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

learning untuk prediksi stunting, yang mengungkapkan bahwa kebanyakan studi masih terbatas pada satu atau dua algoritma dan dataset berskala kecil, sehingga akurasi kurang maksimal. Temuan ini krusial karena menjadi dasar untuk menentukan arah penelitian saat ini, yaitu dengan mengembangkan pendekatan yang lebih komprehensif dengan membandingkan lima algoritma menggunakan data yang lebih besar dan beragam, sekaligus menegaskan kebaruan (*novelty*) dari penelitian yang dilakukan.

3.2 Pengumpulan Data

Penelitian ini memanfaatkan satu sumber utama data, yaitu *dataset* publik dari *platform Kaggle*, yang berjudul “*Stunting Balita Detection (121K Rows)*”. *Dataset* ini dipilih karena memiliki jumlah data yang besar, bervariasi, serta memuat informasi penting terkait faktor-faktor yang berpengaruh terhadap stunting. Variabel utama yang digunakan dalam penelitian ini meliputi usia balita, jenis kelamin, dan tinggi badan, yang merupakan indikator kunci dalam memprediksi risiko stunting dan variabel target status gizi yang meliputi normal, *stunted*, *severely stunted* dan tinggi. *Dataset* tersebut kemudian diproses melalui tahap data preprocessing agar siap digunakan dalam proses analisis dan pelatihan model *machine learning* yang dikembangkan.

3.3 Pemodelan Klasifikasi

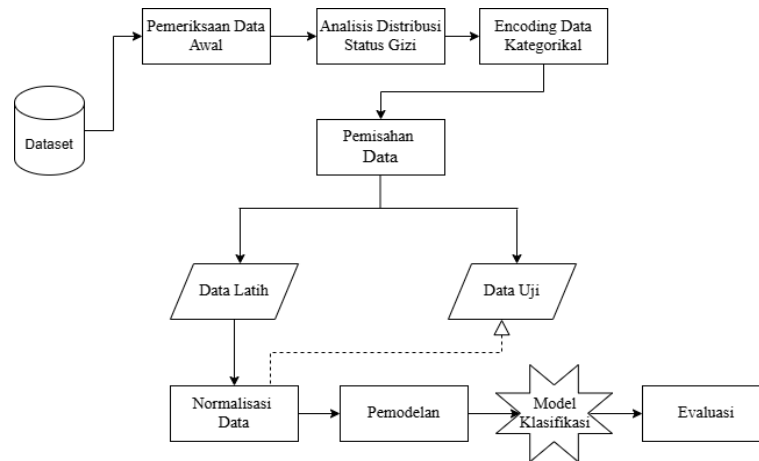
Tahap pemodelan klasifikasi merupakan proses penting dalam penelitian ini karena bertujuan untuk membangun model *machine learning* yang mampu memprediksi risiko stunting pada balita secara akurat. Proses pemodelan dilakukan melalui beberapa langkah sistematis yang dimulai dari pengolahan dataset hingga evaluasi hasil klasifikasi model dapat dilihat pada Gambar 3.2.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar 3.2. Pemodelan Prosedur Data

3.3.1 Dataset

Data yang digunakan dalam penelitian ini berasal dari dataset publik yang diperoleh melalui *platform Kaggle*. *Dataset* ini mengandung informasi tentang karakteristik balita, termasuk data tentang status gizi dan faktor antropometri, dataset berisi informasi karakteristik balita dengan empat label yaitu Normal, Tinggi, *Stunted*, dan *Severely Stunted*. *Dataset* ini digunakan sebagai sumber utama untuk melatih dan menguji model pembelajaran mesin untuk memprediksi risiko stunting.

Dataset ini sebelumnya juga telah digunakan dalam penelitian Lestari, Saragih, dan Caroline (2024) yang menerapkan metode *Deep Neural Networks* (DNN) untuk melakukan klasifikasi multikelas status gizi balita. Penelitian tersebut membuktikan bahwa DNN mampu memberikan performa akurasi tinggi hingga 93,83%, sehingga menunjukkan bahwa *dataset* ini memiliki kualitas yang baik dan relevan untuk pengembangan model prediksi stunting. Dengan demikian, pemanfaatan *dataset* yang sama dalam penelitian ini diharapkan dapat menghasilkan model yang akurat serta memungkinkan dilakukan perbandingan dengan penelitian terdahulu.

3.3.2 Pemeriksaan Data Awal

Tahap pemeriksaan data awal dilakukan untuk memastikan bahwa *dataset* yang digunakan berada dalam kondisi layak sebelum memasuki proses pengolahan lebih lanjut. Pada tahap ini dilakukan peninjauan terhadap kelengkapan data, konsistensi nilai pada setiap atribut, serta identifikasi potensi permasalahan seperti data hilang, duplikasi, atau nilai yang tidak wajar. Proses ini juga mencakup pengecekan tipe data setiap fitur untuk memastikan kesesuaiannya dengan kebutuhan analisis



dan pemodelan. Dengan dilakukannya pemeriksaan data awal, kualitas *dataset* dapat terjamin sehingga proses transformasi data dan pembangunan model dapat dilakukan secara lebih akurat dan terstruktur.

3.3.3 Analisis Distribusi Status Gizi

Analisis distribusi kelas status gizi bertujuan untuk mengetahui proporsi masing-masing kategori target, yaitu Normal, Tinggi, *Stunted*, dan *Severely Stunted*, dalam dataset. Tahap ini penting dilakukan untuk menilai apakah terdapat ketidakseimbangan jumlah data pada setiap kelas yang dapat memengaruhi proses pelatihan model. Ketidakseimbangan kelas berpotensi menurunkan kemampuan model dalam mengenali kelas dengan jumlah data yang lebih sedikit. Dengan memahami pola distribusi ini, peneliti dapat memperoleh gambaran awal mengenai karakteristik dataset dan menilai apakah diperlukan penanganan khusus untuk memastikan bahwa model yang dibangun mampu belajar secara optimal pada seluruh kelas.

3.3.4 Encoding Data Kategorikal

Encoding data kategorikal merupakan proses mengubah variabel dalam bentuk kategori menjadi nilai numerik agar dapat diproses oleh algoritma pembelajaran mesin. Hal ini dilakukan karena sebagian besar algoritma tidak dapat bekerja secara langsung dengan data dalam bentuk teks atau label kategorikal. Pada penelitian ini, fitur yang bersifat kategorikal dikonversi menggunakan metode *encoding* yang tepat untuk menjaga makna dan informasi pada setiap kategori. Proses ini memastikan bahwa seluruh fitur dalam dataset berada dalam bentuk numerik yang seragam sehingga dapat digunakan dalam tahap normalisasi dan pemodelan secara optimal.

3.3.5 Pemisahan Data (*Data Splitting*)

Tahap Berikutnya adalah pemisahan data menjadi dua bagian, yaitu data latih dan data uji. Proses ini bertujuan agar model machine learning dapat dibangun menggunakan data latih dan kemudian di evaluasi dengan data uji yang belum pernah digunakan dalam pelatihan. Pemisahan data dilakukan dengan metode *hold-out*, dengan perbandingan 80% data latih dan 20% data uji. Dengan adanya pembagian ini, tingkat generalisasi model dapat dinilai secara objektif.

3.3.6 Data latih (*Training Data*)

Data latih merupakan komponen dataset yang difungsikan untuk mengonstruksi model prediksi. Dalam tahapan ini, algoritma pembelajaran mesin melakukan ekstraksi terhadap pola, keterkaitan, serta karakteristik dari berbagai variabel yang tersedia. Penelitian ini mengalokasikan sebesar 80% dari total dataset

sebagai data latih, dengan asumsi bahwa ketersediaan data latih yang masif akan meningkatkan kapabilitas model dalam mengidentifikasi pola secara lebih akurat dan mendalam.

3.3.7 Data Uji (*Testing Data*)

Data uji adalah bagian dari dataset yang dipisahkan dan tidak digunakan dalam proses pelatihan model. Data ini digunakan khusus untuk menguji kinerja model yang telah di bangun. Data uji yang digunakan sebesar 20% pada penelitian ini, melalui data uji ini, dapat diketahui seberapa baik model mampu melakukan prediksi pada data baru yang sebelumnya tidak dikenal.

3.3.8 Normalisasi Data (*Data Normalization*)

Sebelum digunakan dalam proses pelatihan, data latih dilakukan normalisasi untuk menyamakan skala antar fitur. Normalisasi penting dilakukan karena perbedaan skala antar variabel dapat memengaruhi kinerja algoritma machine learning. Dalam penelitian ini digunakan metode Min-Max Normalization, yaitu metode normalisasi yang mengubah nilai data ke dalam rentang tertentu, biasanya [0,1]. Rumus Min-Max Normalization adalah sebagai berikut:

$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (3.1)$$

Dengan:

- $x_{\text{normalized}}$ adalah nilai data yang telah dinormalisasi.
- x adalah nilai data asli.
- x_{\min} adalah nilai minimum dari fitur (kolom data) tersebut.
- x_{\max} adalah nilai maksimum dari fitur (kolom data) tersebut.

3.3.9 Pembuatan Model Klasifikasi

Penelitian ini memulai tahap pemodelan dengan menggunakan lima algoritma pembelajaran mesin untuk membangun model klasifikasi prediksi risiko stunting. Untuk memberikan gambaran komparatif tentang kinerja masing-masing model, algoritma yang dipilih adalah *Decision Tree C4.5*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), *Random Forest* dan *Naive Bayes*. Kelima algoritma ini dipilih berdasarkan perbedaan karakteristik dan pendekatan klasifikasi.

1. *Decision Tree C4.5*

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Pembuatan model klasifikasi C4.5 dalam penelitian ini diimplementasikan menggunakan *Python* di *Google Colab*, memanfaatkan pustaka *Scikit-learn* dengan metode *Decision Tree Classifier*. Model pohon keputusan ini dibangun untuk memprediksi risiko stunting. Setelah menetapkan parameter, model dilatih menggunakan data latih dan diuji dengan data uji untuk mengevaluasi performanya. Penilaian akhir dilakukan secara komprehensif menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk mengukur keakuratan dan efektivitas klasifikasi model pada Tabel 3.1.

Tabel 3.1. Parameter *Decision Tree C4.5*

| No | Parameter | Nilai |
|----|-------------------------|-----------------------------------|
| 1 | <i>Criterion</i> | <i>Entropy (Information Gain)</i> |
| 2 | <i>Min Samples Leaf</i> | 10, 20, 30, 40, 50 |
| 3 | <i>Random State</i> | 42 |

Parameter yang digunakan pada algoritma *Decision Tree C4.5* pada penelitian ini terdiri dari tiga komponen utama. Pertama, *criterion* menggunakan *entropy* sebagai dasar perhitungan *information gain* untuk menentukan atribut terbaik dalam proses pemecahan *node*. Kedua, *min samples leaf* diatur dengan beberapa variasi nilai, yaitu 10, 20, 30, 40, dan 50, guna melihat pengaruh jumlah minimum sampel pada node daun terhadap performa model serta mencegah terbentuknya *leaf* yang terlalu kecil. Ketiga, *random state* ditetapkan pada nilai 42 untuk memastikan proses pelatihan model berjalan secara konsisten dan dapat direplikasi. Konfigurasi parameter ini bertujuan untuk memperoleh model C4.5 yang optimal dan stabil pada data yang digunakan.

2. *Support Vector Machine (SVM)*

Pembuatan model klasifikasi *Support Vector Machine (SVM)* ini diimplementasikan menggunakan bahasa *Python* di lingkungan *Google Colab* dengan bantuan pustaka *Scikit-learn*. Algoritma SVM berfungsi untuk memprediksi risiko stunting dengan cara membentuk *hyperplane* terbaik yang memisahkan data ke dalam kelas stunting dan *non-stunting*. Model dilatih menggunakan data latih dan kemudian diuji menggunakan data uji untuk menilai seberapa akurat ia mampu mengenali pola data. Kinerja model SVM dievaluasi secara komprehensif menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk menentukan keefektifannya dalam mem-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

prediksi risiko stunting pada Tabel 3.2.

Tabel 3.2. *Parameter Support Vector Machine (SVM)*

| Parameter | Nilai |
|-----------|---|
| Kernel | <i>Linear, Polynomial, RBF (Radial Basis Function), Sigmoid</i> |

Model Support Vector Machine (SVM) menggunakan daftar jenis kernel yang umum digunakan, yaitu *Linear, Polynomial, Radial Basis Function (RBF)*, dan *Sigmoid*, sebagaimana ditampilkan pada tabel parameter. Kernel-kernel tersebut merupakan pilihan yang tersedia dalam konfigurasi SVM untuk menangani berbagai pola data, baik linear maupun non-linear. Pemilihan kernel yang tepat memungkinkan model menyesuaikan proses pemisahan kelas sesuai karakteristik data pada *dataset Stunting Balita Detection (121K Rows)* dari *platform Kaggle*.

3. *K-Nearest Neighbor (KNN)*

Pembuatan model klasifikasi *K-Nearest Neighbor (KNN)* dalam penelitian ini diimplementasikan menggunakan *Python* di *Google Colab* dengan bantuan pustaka *Scikit-learn*. Prinsip kerja KNN adalah mengklasifikasikan data baru berdasarkan mayoritas kelas dari sejumlah tetangga terdekat yang ditentukan oleh jarak. Untuk mengoptimalkan kinerja, model diuji dengan tiga variasi nilai K (3, 5, dan 7). Pemilihan variasi K ini penting karena nilai K yang tidak tepat dapat meningkatkan sensitivitas terhadap noise (jika terlalu kecil) atau mengaburkan batas kelas (jika terlalu besar). Kinerja model dinilai secara komprehensif menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk memprediksi risiko stunting pada Tabel 3.3.

Tabel 3.3. *Parameter K-Nearest Neighbor (KNN)*

| No | Parameter | Nilai |
|----|----------------------------|----------------------------|
| 1 | Perhitungan jarak terdekat | <i>Euclidean Distance</i> |
| 2 | Nilai K | K = 1, K = 3, K = 5, K = 7 |

Model *K-Nearest Neighbor (KNN)* diuji dengan memvariasikan nilai K (1, 3, 5, dan 7). Tujuan dari pengujian ini adalah untuk membandingkan kinerja model pada setiap konfigurasi dan menentukan nilai K yang paling optimal. Konfigurasi terbaik tersebut kemudian diterapkan pada dataset *Stunting Balita Detection (121K Rows)* yang diperoleh dari *platform Kaggle*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

4. *Random Forest*

Pengembangan model klasifikasi berbasis *Random Forest* pada studi ini dilakukan dengan menggunakan bahasa pemrograman *Python* dalam ekosistem *Google Colab* serta dukungan pustaka *Scikit-learn*. Algoritma ini beroperasi melalui konstruksi sekumpulan pohon keputusan secara acak, yang kemudian diintegrasikan menggunakan mekanisme *voting* mayoritas untuk menetapkan prediksi *final*. Pemilihan metode ini didasari oleh kapabilitasnya dalam mengoptimalkan akurasi serta meminimalisir potensi *overfitting* yang sering terjadi pada penggunaan pohon keputusan tunggal. Model tersebut diproses melalui tahap pelatihan dan pengujian guna memvalidasi efektivitasnya dalam mengategorikan risiko stunting secara akurat pada Tabel 3.4.

Tabel 3.4. Parameter *Random Forest*

| No | Parameter | Nilai |
|----|---------------------|-----------------------|
| 1 | <i>n_estimators</i> | 10, 50, 100, 200, 500 |
| 2 | <i>Criterion</i> | <i>Gini</i> |
| 3 | <i>Max Depth</i> | <i>None</i> |
| 4 | <i>Random State</i> | 42 |

Model *Random Forest* pada penelitian ini dikonfigurasi menggunakan variasi jumlah pohon keputusan melalui parameter *n_estimators* yang diatur pada nilai 10, 50, 100, 200, dan 500. Parameter ini digunakan untuk mengevaluasi pengaruh banyaknya pohon terhadap kualitas prediksi model. Selanjutnya, *criterion* = ‘*gini*’ dipilih sebagai metode pengukuran kualitas pemisahan antar kelas. Selain itu, *max_depth* = *None* diterapkan agar setiap pohon dapat tumbuh hingga seluruh data terklasifikasi, dan *random_state* = 42 ditetapkan untuk memastikan konsistensi serta reproduksibilitas hasil pelatihan. Kombinasi parameter tersebut disesuaikan untuk mencapai performa klasifikasi yang optimal pada dataset *Stunting Balita Detection (121K Rows)* yang diperoleh dari platform *Kaggle*.

5. *Naive Bayes*

Pembuatan model klasifikasi *Naive Bayes* dilakukan di *Google Colab* menggunakan *Python* dan pustaka *Scikit-learn*. Penelitian ini mengadopsi model *Gaussian Naive Bayes* karena cocok untuk data numerik dan kontinu seperti usia, tinggi badan, dan jenis kelamin balita. Untuk meningkatkan stabili-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

tas dan akurasi perhitungan, digunakan parameter $\text{var_smoothing} = 1e-9$. Fungsi dari var_smoothing ini adalah menambahkan nilai yang sangat kecil pada varians agar terhindar dari probabilitas nol. Setelah dilatih, model diuji menggunakan data uji untuk mengevaluasi performa klasifikasi risiko stunting.

Dengan menggunakan kelima algoritma ini, penelitian dapat melakukan perbandingan performa untuk mengetahui model mana yang paling optimal dalam memprediksi risiko stunting pada balita.

3.3.10 Evaluasi

Tahap terakhir adalah evaluasi, yaitu mengukur kinerja model klasifikasi dengan menggunakan *confusion matrix* dan metrik evaluasi yang umum dipakai dalam penelitian machine learning. Beberapa metrik yang digunakan antara lain:

1. Akurasi (*Accuracy*)

Mengukur persentase prediksi yang benar dari keseluruhan data uji.

$$\text{Akurasi} = \left(\frac{TP + TN}{TP + FP + TN + FN} \right) \times 100\% \quad (3.2)$$

2. Presisi (*Precision*)

Mengukur tingkat ketepatan model dalam memprediksi kelas positif dibandingkan dengan keseluruhan prediksi positif yang dihasilkan.

$$\text{Precision} = \left(\frac{TP}{TP + FP} \right) \times 100\% \quad (3.3)$$

3. Recall (*Sensitivitas*)

Mengukur kemampuan model dalam menemukan seluruh data yang benar-benar termasuk kelas positif.

$$\text{Recall} = \left(\frac{TP}{TP + FN} \right) \times 100\% \quad (3.4)$$

4. F1-Score

Merupakan rata-rata harmonik dari presisi dan *recall*. Metrik ini penting digunakan apabila terdapat ketidakseimbangan kelas pada dataset.

$$\text{F1-score} = \left(\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \times 100\% \quad (3.5)$$

keterangan:

- (a) TP (*True Positive*): jumlah data positif yang berhasil diklasifikasikan



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dengan benar.

- (b) TN (*True Negative*): jumlah data negatif yang berhasil diklasifikasikan dengan benar.
- (c) FP (*False Positive*): jumlah data negatif yang salah diklasifikasikan sebagai positif.
- (d) FN (*False Negative*): jumlah data positif yang salah diklasifikasikan sebagai negatif.

3.4 Pemilihan Model Terbaik

Penentuan model terbaik dari hasil evaluasi menjadi langkah penutup dalam siklus pengembangan algoritma ini. Proses seleksi dilakukan dengan membandingkan nilai akurasi, presisi, *recall*, serta *F1-score* demi mendapatkan hasil klasifikasi stunting yang stabil pada semua kelas data. Selain performa metrik, efisiensi waktu komputasi dan kemudahan interpretasi model juga menjadi faktor penentu utama. Hasil dari pemilihan ini diproyeksikan sebagai solusi cerdas yang tangguh untuk memetakan risiko stunting, sehingga dapat menjadi referensi penting bagi pengambilan keputusan strategis atau penelitian di masa mendatang.

BAB 5

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan mengenai prediksi risiko stunting pada balita menggunakan lima algoritma *Machine Learning*, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Penerapan Algoritma *Machine Learning*: Penelitian ini berhasil menerapkan lima algoritma klasifikasi yaitu KNN, *Random Forest*, SVM, *Naïve Bayes*, dan *Decision Tree C4.5* pada *dataset Stunting Balita Detection (121K Rows)*. Seluruh algoritma dapat membangun model klasifikasi untuk memprediksi status gizi balita berdasarkan usia, jenis kelamin, dan tinggi badan setelah melalui tahap preprocessing data dan normalisasi.
2. Hasil Akurasi dan Evaluasi Model: Hasil evaluasi menggunakan empat metrik utama yaitu akurasi, *precision*, *recall*, dan *F1-score*, menunjukkan bahwa KNN dan *Random Forest* merupakan algoritma dengan performa paling unggul. KNN memperoleh akurasi 99,94% dan *Random Forest* 99,93%, yang mencerminkan kemampuan klasifikasi yang sangat stabil, konsisten, serta akurat dalam mengenali seluruh kelas. Di sisi lain, SVM menunjukkan performa kategori menengah, sedangkan *Naïve Bayes* dan *Decision Tree* menghasilkan nilai evaluasi yang jauh lebih rendah sehingga kurang efektif dalam memprediksi pola data stunting.
3. Model Terbaik untuk Prediksi Stunting: Berdasarkan hasil evaluasi menyeluruh, KNN dan *Random Forest* dinyatakan sebagai model paling optimal untuk digunakan dalam mendeteksi risiko stunting. Kedua model ini memiliki nilai *precision*, *recall*, dan *F1-Score* yang hampir sempurna, sehingga dapat diandalkan untuk aplikasi deteksi dini karena minim kesalahan klasifikasi baik pada kelas positif maupun keseluruhan kelas.

5.2 Saran

Agar hasil penelitian semakin representatif dan dapat digeneralisasikan, penelitian selanjutnya disarankan menggunakan cakupan data yang lebih luas serta melibatkan variasi karakteristik balita dari wilayah yang berbeda. Penggunaan data yang lebih beragam diharapkan mampu meningkatkan kemampuan model dalam mengenali pola stunting pada populasi yang lebih heterogen. Selain itu, penelitian berikutnya dapat mempertimbangkan penambahan variabel atau fitur pendukung, seperti faktor lingkungan, gizi keluarga, serta kondisi sosial ekonomi, yang



berpotensi memberi pengaruh signifikan terhadap risiko stunting namun belum digunakan dalam penelitian ini. Untuk meningkatkan keakuratan model, metode seleksi fitur juga dapat dikembangkan lebih lanjut dengan menerapkan teknik seperti *Mutual Information*, *Recursive Feature Elimination* (RFE), atau *Recursive Feature Elimination with Cross-Validation* (RFECV) guna memastikan bahwa fitur yang digunakan benar-benar relevan terhadap hasil prediksi. Temuan penelitian ini diharapkan dapat menjadi dasar atau bahan acuan bagi instansi kesehatan maupun pihak terkait dalam melakukan pengembangan lanjutan, baik dalam pengolahan data maupun pengembangan sistem pendukung keputusan secara mandiri, dengan merujuk pada hasil analisis dan evaluasi model yang telah dilakukan dalam penelitian ini.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU



DAFTAR PUSTAKA

- Abdusyukur, F. (2023). KOMPUTA : Jurnal Ilmiah Komputer dan Informatika PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK KOMPUTA : Jurnal Ilmiah Komputer dan Informatika. *Komputa : Jurnal Ilmiah Komputer dan Informatika*, 12(1), 73–82.
- Adzhima, F., Budianita, E., Nazir, A., dan Syafria, F. (2023). Klasifikasi Status Stunting Balita Dengan Metode Support Vector Machine Berbasis Web. *INOVTEK Polbeng - Seri Informatika*, 8(2), 381. doi: 10.35314/isi.v8i2.3641
- Arriansyah, M., Saputra, J., Ardhana, V. Y. P., dan Sa'adati, Y. (2024). Algoritma naive bayes yang efisien untuk klasifikasi buah pisang raja berdasarkan fitur warna. *Journal of Information Systems Management and Digital Business*, 1(2), 236–248.
- Akbar, F. M. N. (2024). Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air. *Jurnal Tekno Kompak*, 18(1), 28. doi: 10.33365/jtk.v18i1.3241
- Ananto, D. T., Duta Mahardewantoro, D., Mustafa, F., Ardianto, M. G., Rafi, M. M., Zein, R. A., ... Adharani, Y. (2023). Edukasi dan pelatihan pengenalan machine learning dan computer vision untuk mengeksplorasi potensi visual. *Prosiding Seminar Nasional LPPM UMJ*. Retrieved from <http://jurnal.umj.ac.id/index.php/semnaskat>
- Anwar, S., Winarti, E., dan Sunardi, S. (2022). Systematic Review Faktor Risiko, Penyebab Dan Dampak Stunting Pada Anak. *Jurnal Ilmu Kesehatan*, 11(1), 88. doi: 10.32831/jik.v11i1.445
- Alifia Riski Putra Wahyu, B., Fayi Farozzi, A., Putra Mahendra, C., dan Kembang Hapsari, R. (2023). Klasifikasi Penderita Penyakit Diabetes Berdasarkan Decision Tree Menggunakan Algoritma C4. *INTEGER: Journal of Information Technology*, 5(Widyasari 2017), 80.
- Badri, R. M., Alvionita, M., Ramdani, A. L., Safitri, I., dan ... (2025). Pelatihan Algoritma Machine Learning dalam Robotik di MAN 1 Bandar Lampung. *RENATA: Jurnal ...*, 3(1), 125–131. Retrieved from <https://www.jurnalpkm.id/index.php/renata/article/view/135%0Ahttps://www.jurnalpkm.id/index.php/renata/article/download/135/92>
- Bandur, P. M. Y., dan Madur, J. P. (2025). Effectiveness of Providing Supplementary Food for Stunted Toddlers in Indonesia: A Literature Review.



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Dunia Keperawatan: Jurnal Keperawatan dan Kesehatan, 13(1), 75–89. doi: 10.20527/jdk.v13i1.783

Bimawan, Z. I., Astuti, T., dan Arsi, P. (2024). Komparasi Algoritma Random Forest , K-Nearest Neighbor ,. *Jurnal Teknik Informatika (Jutif)*, 5(6), 1599–1607.

Cahyati, N., dan Islami, C. C. (2022). Pemahaman Ibu Mengenai Stunting dan Dampaknya. , 2(2), 175–191.

Damar Pratama, A. (2024). JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika) Journal homepage: <https://jurnal.stkipppgritulungagung.ac.id/index.php/jipi> ANALISA SENTIMEN MASYARAKAT TERHADAP PENGGUNAAN CHATGPT MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM). , 9(1), 327–338. Retrieved from <https://doi.org/10.29100/jipi.v9i1.4285>

Darmawan, A. C. (2022). Pengembang Aplikasi Berbasis Web dengan Python Flask untuk Klasifikasi Data Menggunakan Metode Decision Tree C4.5. *Jurnal Pendiidikan Konseling*, 4(5), 5351 – 5362.

Dewi, S. P., Nurwati, N., dan Rahayu, E. (2022). Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor. *Building of Informatics, Technology and Science (BITS)*, 3(4), 639–648. doi: 10.47065/bits.v3i4.1408

Elisabeth, K. (2025). Implementasi Algoritma C4.5 Untuk Prediksi Stunting Pada Data Kasus Gizi Di Posyandu Kelurahan Nangalimang. *Jurnal In Create (Inovasi dan Kreasi dalam Teknologi Informasi)*, 11(1), 38–44.

Fadellia Azzahra, Suarna, N., dan Arie Wijaya, Y. (2024). Penerapan Algoritma Random Forest Dan Cross Validation Untuk Prediksi Data Stunting. *Kopertip : Jurnal Ilmiah Manajemen Informatika dan Komputer*, 8(1), 1–6. doi: 10.32485/kopertip.v8i1.238

Fajrizal, Handayani, S., dan Taslim. (2024). Pelatihan Pemrograman Python Dengan Google Colab Bagi Siswa Smk Inovasi Riau. *J-COSCIS : Journal of Computer Science Community Service*, 4(2), 133–141. doi: 10.31849/jcscis.v4i2.19106

Felicia Watratan, A., Puspita, A. B., dan Moeis, D. (2020). Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia. *Journal of Applied Computer Science and Technology*, 1(1), 7–14. Retrieved from <http://journal.isas.or.id/index.php/JACOST>

Fitriani, dan Darmawi. (2022). *Jurnal Biology Education* Volume. 10 Nomor 1



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Edisi Khusus 2022. *Jurnal Biology Education*, 10(1), 23–33.

Gevindo, A., dan Hendrik, B. (2025). Penerapan Machine Learning Untuk Mendeteksi Serangan Anomali Dalam Jaringan Komputer : Systematic Literature Review. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4659–4664. doi: 10.36040/jati.v9i3.13746

Grathima, E. F., Yunitasari, E., dan Indarwati, R. (2024). the Impact of Complementary Food Interventions on Adolescence Stunting Prevention: a Systematic Review. *Indonesian Journal of Community Health Nursing*, 9(2), 62–66. doi: 10.20473/ijchn.v9i2.47246

Hadna, A. H., Ardiansyah, J., dan Habib, F. (2023). Stand-Alone Conditional Cash Transfer in Regards to Reducing Stunting in Indonesia : Evidence From a Randomized Controlled Trial Keywords. , 435–454.

Handayani, F. (2021). Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 7(3), 329. doi: 10.26418/jp.v7i3.48053

Handayani, S. (2023). Save the Nation's Generation From the Dangers of Stunting. *Journal of Midwifery Science and Women's Health*, 3(2), 87–92. doi: 10.36082/jmswh.v3i2.1082

Handika. (2024). Edumatnesia: Prosiding Seminar Nasional Matematika dan Pendidikan Matematika Pemanfaatan Python dan Google Colab Dalam Pembelajaran Statistika Deskriptif. , 379–389.

Hariani, A. L. (2024). HUBUNGAN BERAT BADAN LAHIR RENDAH DENGAN KEJADIAN STUNTING PADA ANAK USIA 1-5 TAHUN DI WILAYAH KERJA The Relationship Between Low Birth Weight and the Incidence of Stunting in Children Aged 1-5 Years in the Jorongon Health Center Work Area PENDAHULUAN Bal. *Assifa Jurnal Ilmu Kesehatan*, 1(1), 133–142.

Hartati, L., dan Wahyuningsih, A. (2021). Stunting adalah kondisi gagal tumbuh pada anak balita (bayi di bawah lima tahun) yang disebabkan oleh kekurangan gizi kronis, sehingga tinggi badan anak tidak sesuai dengan usianya. Kekurangan gizi ini dapat terjadi sejak bayi dalam kandungan dan pada mas. *INVOLUSI: Jurnal Ilmu Kebidanan*, 11(1), 28–34.

Hastuti, N. T., dan Budiman, F. (2024). Optimasi Klasifikasi Stunting Balita dengan Teknik Boosting pada Decision Tree. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 655–664. doi: 10.29408/edumatic.v8i2.27913



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

- Hovi, H. S. W., Id Hadiana, A., dan Rakhmat Umbara, F. (2023). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, 4(1), 40–45. doi: 10.36423/index.v4i1.895
- Jafar, M., dan Aisyah, D. (2022). Romlah1, Sutan Faisal2, Rahmat3, Jamaludin Indra4. , 2338, 13–34.
- Khalid, S. H., Patimah, S., dan Asrina, A. (2022). Persepsi Masyarakat mengenai Penyebab dan Dampak Stunting di Kabupaten Majene Sulawesi Barat Tahun 2020. *Journal of Muslim Community Health (JMCH)*, 3(4), 80–94. Retrieved from <https://pasca-umi.ac.id/index.php/jmch/article/view/1177/1366>
- KHOTIMAH, K. (2022). Dampak Stunting dalam Perekonomian di Indonesia. *JISP (Jurnal Inovasi Sektor Publik)*, 2(1), 113–132. doi: 10.38156/jisp.v2i1.124
- Laila, F. N., Hardiansyah, A., dan Susilowati, F. (2023). Pengetahuan Gizi Ibu, Pendapatan Orang Tua, Pemberian Susu Formula, dan Kaitannya dengan Status Gizi Balita di Posyandu Desa Welahan Kabupaten Jepara. *Jurnal Gizi dan Kuliner (Journal of Nutrition and Culinary)*, 3(1), 24. doi: 10.24114/jnc.v3i1.42426
- Laloan, N. N., Masithah, S., Rahmaniar MB, A., Wahyuni, F., dan Selvia, S. (2024). Faktor Risiko Status Gizi Kurang Pada Anak Usia 0-59 Bulan Di Wilayah Kerja Puskesmas Kimi Nabire Papua Tahun 2023. *Jurnal Kesehatan Tambusai*, 5(4), 11432–11440. doi: 10.31004/jkt.v5i4.35925
- Lukhia Britanthia Christina Tanujaya, Bambang Susanto, dan Asido Saragih. (2020). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Fitur Mode Audio Spotify. *Indonesian Journal of Data and Science (IJODAS)*, 1(3), 68–78.
- M. Fazlur Rahman Assauqi, dan Zaehol Fatah. (2024). Prediksi Risiko Demam Berdarah Menggunakan Decision Tree Berdasarkan Gejala Klinis Dan Data Laboratorium. *Jurnal Ilmiah Sains Teknologi Dan Informasi*, 2(4), 42–53. doi: 10.59024/jiti.v2i4.972
- Pamuji, F. Y., dan Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy. *Jurnal Teknologi dan Manajemen Informatika*, 7(1), 46–50. doi: 10.26905/jtmi.v7i1.5982
- Paramita, M., Helmyati, S., Basrowi, R. W., dan Dilantika, C. (2023). Risk Factors for Stunting among Infants on Java Island: A Cross-Sectional Study using



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

the 2021 Indonesian Nutritional Status Survey Data. *Amerta Nutrition*, 7(3), 20–29. doi: 10.20473/amnt.v7i3SP.2023.20-29

Paryatno, L., Hari Linda Lukitowati, dan M. Arna Ramadhan. (2022). Evaluasi Pembelajaran Berbasis Webinar Dengan Importance Performance Matrix Analysis (Ipma). *Jurnal Teknodik*, 26(April 2021), 89–100. doi: 10.32550/teknodik.vi.790

Pratama, R. A., Wabula, F., Imandry, H., Isabela, L. M., Raharjo, M., dan Sianipar, R. (2025). Literature Review the Impact of Machine Learning in Modern Industries. *Nian Tana Sikka: Jurnal ilmiah Mahasiswa*, 3(1), 177–182.

Puspitasari, Y. D., Indarwati, R., Wahyuni, S. D., dan Suraya, A. S. (2025). Community And Family-Based Intervention Strategies To Prevent Stunting: A Systematic Review. *Care : Jurnal Ilmiah Ilmu Kesehatan*, 13(2), 286–298. doi: 10.33366/jc.v13i2.6613

Putri, A. I., Husna, N. A., Cia, N. M., Arba, M. A., Aisyi, N. R., Pramesthi, C. H., dan Irdayusman, A. S. (2024). Implementation of K-Nearest Neighbors, Naïve Bayes Classifier, Support Vector Machine and Decision Tree Algorithms for Obesity Risk Prediction. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2(1), 26–33. doi: 10.57152/predatecs.v2i1.1110

Putri, P. M., Shafira, A. S., dan Mahardhika, G. S. (2024). Stunting Reduction Strategy in Indonesia: Maternal Knowledge Aspects. *Indonesian Journal of Public Health*, 19(2), 329–343. doi: 10.20473/ijph.v19i2.2024.329-343

Rahman, H., Rahmah, M., dan Saribulan, N. (2023). Stunting Treatment Efforts in Indonesia. *Journal of Science of the Voice of Khatul Event (JIPSK)*, VIII(01), 44–59.

Rais, A. N., dan Warjiyono, W. (2024). Sistem Informasi Prediksi Risiko Kredit Keuangan Berbasis Web Machine Learning. *Justika: Jurnal Sistem Informasi Akuntansi*, 4(2), 59–62. doi: 10.31294/justika.v4i2.7509

Retnoningsih, E., dan Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. doi: 10.51211/biict.v7i2.1422

Roihan, A., Sunarya, P. A., dan Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. doi: 10.31294/ijcit.v5i1.7951

Rusyda, A. L., Baliwati, Y. F., Gizi, S. S., Masyarakat, D. G., Manusia, F. E.,



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Masyarakat, D. G., dan Manusia, F. E. (2024). (Transforming Maternal and Child Health Systems and Nutritional. , 47(1), 43–52.

Sahamony, N. F., dan Rianto, H. (2024). Analysis of Performance Comparison of Machine Learning Models for Predicting Stunting Risk in Children ' s Growth Analisis Perbandingan Kinerja Model Machine Learning untuk Memprediksi Risiko Stunting pada Pertumbuhan Anak. *MALCOM : Indonesian Journal of Machine Learning and Computer Science*, 4(2), 413–422.

Saleh, H., Informatika, J. T., dan Komputer, F. I. (2020). (Saleh, 2020;13). , 3(1), 11–17.

Sidik, A. D. W. M., Himawan Kusumah, I., Suryana, A., Edwinanto, Artiyasa, M., dan Pradiftha Junfithrana, A. (2020). Gambaran Umum Metode Klasifikasi Data Mining. *FIDELITY : Jurnal Teknik Elektro*, 2(2), 34–38. doi: 10.52005/fidelity.v2i2.111

Simbolon, A. S., Pangaribuan, N. I., dan ... (2022). Analisis Sentimen Aplikasi E-Learning Selama Pandemi Covid-19 Dengan Menggunakan Metode Support Vector Machine Dan *Prosiding ...*, 9–17. Retrieved from <http://journal.universitasmulia.ac.id/index.php/seminastika/article/view/236%0Ahttp://journal.universitasmulia.ac.id/index.php/seminastika/article/download/236/199>

Soen, G. I. E., Marlina, M., dan Renny, R. (2022). Implementasi Cloud Computing dengan Google Colaboratory pada Aplikasi Pengolah Data Zoom Participants. *JITU : Journal Informatic Technology And Communication*, 6(1), 24–30. doi: 10.36596/jitu.v6i1.781

Skarene, Harjono, Sutaryono, Y., dan Maslami, V. (2023). PENDAMPINGAN PENGGUNAAN GOOGLE COLAB PADA PEMBELAJARAN PYTHON DAN MACHINE LEARNING BAGI DOSEN MATEMATIKA DI PALEM-BANG Sisca. *Jurnal Pepadu*, 4(4), 541–546.

Supriyanto, J., Alita, D., dan Isnain, A. R. (2023). Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring. *Jurnal Informatika dan Rekayasa Perangkat Lunak*, 4(1), 74–80. doi: 10.33365/jatika.v4i1.2468

Suryani, K., Rini, M. T., Hardika, B. D., dan Widiastari, N. K. (2023). Analisis Faktor Penyebab Kejadian Stunting. *Jurnal Keperawatan Florence Nightingale*, 6(1), 8–12. doi: 10.52774/jkfn.v6i1.112

Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, dan Fitri Nurapriani. (2023). Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

dan KNN. *Jurnal KomtekInfo*, 10, 1–7. doi: 10.35134/komtekinfo.v10i1.330

Titularsi, Y., dan Prasetyo, H. (2025). Penerapan Machine Learning untuk Pencarian Pelanggan Loyal Berpotensi Menggunakan Metode Python Pandas Seaborn. *Integrasi: Jurnal Ilmiah Teknik Industri*, 10(1), 12–24. doi: 10.32502/integrasi.v10i1.292

Wahyudin, W. C., Hana, F. M., Prihandono, A., Kudus, U. M., No, J. G., Email, I., ... Classifier, N. B. (2023). P Rediksi S Tunting P Ada B Alita D I R Umah S Akit K Ota. , 2019, 32–36.

Wijaya, H. D., dan Dwiasnati, S. (2020). Implementasi Data Mining dengan Algoritma Naïve Bayes pada Penjualan Obat. *Jurnal Informatika*, 7(1), 1–7. doi: 10.31311/ji.v7i1.6203

Wilyani, F., Nuryan Arif, Q., dan Aslimar, F. (2024). Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory Basic Introduction to Python Programming With Google Colaboratory. *Jurnal Pengabdian Pada Masyarakat Indonesia*, 8(2), 788–796. Retrieved from <https://doi.org/10.55606/jppmi.v3i1.1087>

Yulia, F., Sari, K., Khoiriyah, N., Ade, D., dan Nuarita, W. (2021). Pola Konsumsi Balita Selama Ppkm. *Jurnal Medika Indonesia*, 2(2), 13–20.

Yuliyannah Sain, Aynun Andriani, dan Nahla Nurhidayah. (2023). Pelatihan Dasar Menganalisis Data dengan Menggunakan Google Colab di SMA Muhammadiyah Kendari. *KREATIF: Jurnal Pengabdian Masyarakat Nusantara*, 3(1), 01–08. doi: 10.55606/kreatif.v3i1.1163

Yusuf, D., Rachmawati, I., Triwibowo, R. N., Fathimah, A., Zahra, A., dan Nuraeni, F. P. (2024). Sains Indonesiana: Jurnal Ilmiah Nusantara Pemanfaatan Metode K-Nearest Neighbors (KNN) untuk Memprediksi Strategi Promosi Efektif Berdasarkan Lokasi Asal Mahasiswa di Universitas Al-Irsyad Cilacap. , 2, 2024.

Yusuf, D., Razi, F., Arman, S. A., Terisia, V., dan Nurjayanti, R. (2025). Prediksi Risiko Stunting pada Balita menggunakan Algoritma Logistic Regression dan Decision Tree berbasis Data Terbuka. *Prosiding Semnastek*, 1–7.




Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

LAMPIRAN A

SURAT IZIN PENELITIAN



KEMENTERIAN AGAMA REPUBLIK INDONESIA
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU
FAKULTAS SAINS DAN TEKNOLOGI
كلية العلوم و التكنولوجيا
FACULTY OF SCIENCES AND TECHNOLOGY
 JL. HR Soebrantas KM 15 No. 155 Tuah Madani Kec. Tuah Madani- Pekanbaru 28298 PO Box 1004
 Fax (0761) 589 025 Web www.uin-suska.ac.id E-mail fasite@uin-suska.ac.id

Nomor : B- 6792 / F.V/PP.009/09/2025 Pekanbaru, 29 September 2025
 Sifat : Biasa
 Hal : Mohon Izin Penelitian dan Pengambilan Data Tugas Akhir/Skripsi


Kepada Yth.
 Kepala Dinas Kesehatan Kota Pekanbaru
 Komplek Perkantoran Walikota Pekanbaru, Gedung Belah Bubung, Lt. IV, Jalan Abdul Rahman Hamid, Kelurahan Tuah Negeri, Kecamatan Tenayan Raya.

Assalamualaikum Wr. Wb.
 Dengan hormat, Sehubungan telah dimulainya mata kuliah Tugas Akhir pada Fakultas Sains dan Teknologi UIN Suska Riau, kami bermaksud mengirimkan mahasiswa:


| | |
|-------------------|------------------------|
| Nama | : Ekatri Yulisara |
| NIM | : 12250325372 |
| Fakultas | : Sains dan Teknologi |
| Program Studi/Smt | : Sistem Informasi / 7 |
| No. Hp/E-mail | : 085247814251 |

untuk pengambilan data yang sangat dibutuhkan dalam Tugas akhir mahasiswa tersebut yang berjudul "PREDIKSI RISIKO STUNTING PADA BALITA MENGGUNAKAN ALGORITMA MACHINE LEARNING". Kami mohon kiranya saudara berkenan memberikan izin dan fasilitas demi kelancaran Tugas Akhir mahasiswa yang bersangkutan.
 Demikian surat ini kami sampaikan, atas bantuan dan kerjasama Saudara kami ucapkan terima kasih.

Wassalam
 Dekan



Heniita Muda, M.Sc
 NIP. 197701032007102001

 Dokumen ini telah ditanda tangani secara elektronik.
 Token : 183E2vhy

CS Dipindai dengan CamScanner

Gambar A.1. Surat Izin Penelitian



LAMPIRAN B

TRANSKIP WAWANCARA

Jenis Data : Wawancara
Perihal : Mengenai Stunting pada balita di Puskesmas Simpang Baru
Hari / Tanggal : Rabu, 19 Novemver 2025
Tempat : Puskesmas Simpang Baru

Peneliti : Izin kak, mau wawancara tentang Stunting pada balita untum validasi data kak, apakah boleh kak?

Narasumber : Iya, boleh langsung saja ya di mulai.

Peneliti : Baik kak.

Peneliti : Bagaimana gambaran kasus stunting di puskesmas simpang baru ini dalam satu tahun terakhir ini kak dan berapa orang yang terkena stunting di tahun 2025 ini kak?

Narasumber : Kalau di tahun 2025 ini terdapat 5 anak yang terkena Stunting, dari 5 anak yang diteliti terkena Stunting terdapat dari 3 kelurahan yaitu di Kelurahan di puskesmas ini Kelurahan Simpang baru, Kelurahan Binawidya dan Kelurahan Sungai Sibam. Kalau Stunting itu tinggi badan per umur ya.

Peneliti : Jadi kalau stunting itu tidak di lihat dari jenis kelamin, tinggi badan / per umur dan berat badan ya kak?

Narasumber : Kalau stunting ini dilihat dari tinggi badan per umur dan jenis kelamin kalau berat badan tidak, untuk jenis kelamin itu laki laki dan perempuan biasanya berbeda

Peneliti : Baik kak, Untuk faktor- faktor yang menyebabkan stunting itu apa saja ya kak?

Narasumber : Biasanya karena faktor ekonomi si, biasanya dari pola asuh keluarga juga

Peneliti : Kalau pemeriksaan status gizi di puskesmas Simpang baru ini bagaimana kak?

Narasumber : Kalau pemeriksaan status gizi pertama kita timbang dulu nih dari psyandu berat badan, tinggi badannya, lyla nya, lingkak kepala, nanti dilihat kalau stunting ini buku PMK Tahun 2010 Atropametri, dan selanjutnya anak – anak yang terkena stunting ini harus di rujuk dulu jika dikatakan stunting oleh dokter spesialis



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

anak

Peneliti : Untuk pemantauan stunting di posyandu itu di lakukan sekali dalam sebulan ya kak?

Narasumber : Iya, untuk pemantauan stunting itu dilakukan sekali dalam sebulan termasuk di puskesmas Simpang Baru ini, disini kita ada 20 Posyandu dalam 3 Kelurahan.

Peneliti : Kalau Program yang dilakukan untuk Stunting itu hanya Posyandu saja ya kak?

Narasumber : Kalau dari Posyandu nanti ada namanya makanan tambahan lokal, PMT lokal namanya dari Puskesmas jugatapi yang mengelola kadernya, itu selama 90 hari.

Peneliti : Kalau misalnya disini sudah ada yang terkena stunting, apa yang dilakukan oleh pihak puskesmas kak?

Narasumber : Pertama tu, anak ini harus di rujuk dulu ke Rumah Sakit Madani, nanti kalau sudah di Rumah Sakit Madani nanti akan dikasih susu, susu SGM warna ungu.

Peneliti : Itu bakal di cek secara rutin ga kak? Sampai dia ga stunting lagi kak?

Narasumber : Iya, setiap bulan, nanti setiap bulan kami juga kunjungan ke rumah anak yang terkena stunting itu.

Peneliti : Kalau edukasi kepada ibu terkait pencegahan stunting ini apa saja kak?

Narasumber : Kalau anak-anak ni kan yang terpenting itu gizi seimbang, pola asuh, makannya terkadang karena ekonomi nya jadi orang tua ga mampu memberikan makanan yang bergizi untuk anak-anaknya. Padahal untuk makanan yang bergizi itu ga perlu mahal bisa saja telur dan ikan teri pun bisa kan itu juga makanan bergizi.

Peneliti : Kendala apa yang sering di hadapi oleh petugas saat mendeteksi Stunting ini kak?

Narasumber : Terkadang kita susah menemui balita yang terkena Stunting itu, terkadang sudah datang ke rumahnya tetapi balita nya tidak ada, dan terkadang orang tua balita yang terkena Stunting itu tidak membawa anaknya ke Posyandu secara rutin

Peneliti : Atau disini sudah ada alat teknologi terbaru, untuk mendeteksi Stunting ini kak?

Narasumber : Belum ada, kalau mengukur stunting untuk tinggi badan itu masih



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

menggunakan yang dari kemenkes nya langsung

Peneliti : Oh, berarti untuk alat pendeteksi dari teknologi terbaru nya belum ada ya kak?

Narasumber : Belum, tapi kami menggunakan sesuai dengan buku PMK Tahu 2010 itu

Peneliti : Untuk pencatatan nya masih manual atau sudah digital kak?

Narasumber : Kalau ini pencatatan kami sudah digital berupa aplikasi yang bernama Sigizikesga namanya yang kita nama kan dulunya itu FPPGBM

Peneliti : Kalau untuk petugas Stunting ini apakah sudah melakukan pelatihan kak? Untuk tau mendeteksi Stunting ini kak?

Narasumber : Ada, kami paling tidak dalam setahun itu satu kali atau 2 kali dalam pelatihan ke dinas.

Peneliti : Kalau disini ada kesalahan dalam pengukuran Stunting ga kak?

Narasumber : Kalau itu, misalnya dari kader ni ada yang memberitahu kalau anak ini stunting, jadi kita harus turun dulu kita benar atau ga, kalau iya dia stunting rujuk lagi ke rumah sakit.

Peneliti : Kalau stunting itu kurus ya kak?

Narasumber : Kalau Stunting ini pendek, kalau kurus tu berat badan kurang namanya. Jadi Stunting itu Pendek makanya untuk mengetahui stunting itu di ukur tinggi badan/umur.

Peneliti : Kalau sistem evaluasi program pencegahan stunting yang ada saat ini disini apa ya kak?

Narasumber : Itulah yang PMT Lokal nanti di berikan susu dan dikasih roti

Peneliti : Ada inovasi yang diharapkan ga kak? Inovasi baru gitu misalnya?

Narasumber : Berharap semua anak di indonesia ini mendapatkan MBG dari pemerintah, agar gizi nya terpenuhi, agar mendapatkan susu, pemerintah harus lebih tanggap lagi untuk menanggapi masalah stunting ini. Jadi menangani stunting itu ga mudah kan dia masalah tinggi badan kan yang mudah itu menaikkan berat badan kalau tinggi badan ini susah, jadi kalau anak stunting ni kita kasih makan banyak banyak malah yang nambah berat badan bukan tinggi badan. Kalau tinggi badan ni paling setiap bulan ini cuman naik 1 cm atau 2 cm itu sudah bagus itu. Jadi bertahap emang lama. Beda sama anak gizi kurang dan gizi buruk, kalau anak yang gizi buruk rutin saja kita kasih makan saja pasti dia berat badannya akan bertambah. Kalau tinggi badan ini agak susah memang



Peneliti : Baik kak, segitu aja kak, terimakasih banyak ya kak..

Narasumber : Iya, sama – sama

Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



UIN SUSKA RIAU



LAMPIRAN C

BERITA ACARA WAWANCARA

BERITA WAWANCARA ATAU HASIL OBSERVASI

Pada hari ini,

Telah dilaksanakan wawancara yang berkaitan dengan penelitian yang akan dilakukan untuk penyusunan Tugas Akhir

Tempat : Puskesmas Simpang Baru
 Pukul : 09:00
 Nama Narasumber : Ade Purrama Sari, S.Gz
 Jabatan : Staff J121

Pihak pewawancara melakukan wawancara dengan pihak narasumber yang berkaitan dengan penelitian yang akan dilakukan di Puskesmas Simpang Baru, kemudian narasumber memberikan jawaban terkait pertanyaan yang diajukan oleh pewawancara.

Adapun pertanyaan yang diajukan serta hasil wawancara terlampir.

Mengetahui,

Puskesmas Simpang Baru,

2025

Peneliti

Narasumber

Ekatri Yulisara

NIM. 12250325372

Ade Purrama Sari, S.Gz

Gambar C.1. Berita Acara Wawancara

LAMPIRAN D

DOKUMENTASI

© Hak cipta

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



Gambar D.1. Dokumentasi Wawancara

Syarif Kasim Riau

DAFTAR RIWAYAT HIDUP



Ekatri Yulisara lahir di Dumai, 13 Juli 2003. Merupakan anak ke-tiga dari lima bersaudara dari Bapak Nurhasan dan Ibu Agustriyeni, beralamatkan di Jalan Nuri, Kelurahan Purnama, Kecamatan Dumai Barat, Kota Dumai, Provinsi Riau. Peneliti pertama kali menempuh pendidikan formal Peneliti di TK Ratu Sima II di Kota Dumai, dan menyelesaikan pendidikan pada tahun 2009. Selanjutnya, Peneliti melanjutkan Sekolah Dasar di SDN 007 Purnama di Kota Dumai, Peneliti menyelesaikan pendidikan Sekolah Dasar pada tahun 2015. Setelah menyelesaikan pendidikan dasar, Peneliti melanjutkan ke tingkat SLTP di SMPN 7 Dumai di Kota Dumai, dan menyelesaikan pendidikan pada tahun 2018. Selanjutnya Peneliti melanjutkan ke tingkat SLTA di SMKN 1 Dumai di Kota Dumai, dan menyelesaikan pendidikan pada tahun 2021. Saat ini Peneliti sedang menempuh pendidikan tingkat sarjana di Universitas Islam Negeri Sultan Syarif Kasim Riau Fakultas Sains dan Teknologi Program Studi Sistem Informasi. Selama masa kuliah, penulis telah menyelesaikan kerja praktek yang bertempat di Dinas Sosial Kota Dumai yang memberikan Peneliti pengalaman praktis dalam belajar memahami tugas dan fungsi Dinas Sosial dalam penyelenggaraan pelayanan kesejahteraan sosial kepada masyarakat. Selanjutnya, Peneliti telah menyelesaikan Kuliah Kerja Nyata (KKN) di Desa Simpang Tanah Lapang, Kecamatan Kauantan Hilir, Kabupaten Kuantan Singingi, dimana Peneliti diajarkan bagaimana menjaga kelestarian budaya di masyarakat. Sebagai bagian dari Tugas Akhir Peneliti di universitas. Peneliti telah menyelesaikan Tugas Akhir dengan mengambil judul “Prediksi Risiko Stunting Pada Balita Menggunakan Algoritma Machine Learning”, yang bertujuan untuk memprediksi status stunting pada balita menggunakan lima algoritma machine learning untuk memperoleh model dengan performa terbaik berdasarkan metrik evaluasi.

UIN SUSKA RIAU

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.