

**PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*, *EXTREME GRADIENT BOOSTING*, DAN *CATEGORICAL BOOSTING* DALAM  
KLASIFIKASI *STUNTING* DI KABUPATEN LAMPUNG BARAT**

**(SKRIPSI)**

**Oleh**

**Deta Sari Aslina  
NPM 2217051117**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG**

**2026**

**PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*, *EXTREME GRADIENT BOOSTING*, DAN *CATEGORICAL BOOSTING* DALAM KLASIFIKASI *STUNTING* DI KABUPATEN LAMPUNG BARAT**

**Oleh**

**Deta Sari Aslina**

**Skripsi**

Sebagai Salah Satu Syarat untuk Mencapai Gelar

**SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer**

**Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS LAMPUNG**

**BANDAR LAMPUNG**

**2026**

## ABSTRAK

### **PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE*, *EXTREME GRADIENT BOOSTING*, DAN *CATEGORICAL BOOSTING* DALAM KLASIFIKASI *STUNTING* DI KABUPATEN LAMPUNG BARAT**

Oleh

**DETA SARI ASLINA**

*Stunting* merupakan salah satu permasalahan gizi kronis pada balita yang masih menjadi perhatian dalam kesehatan masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan status *stunting* pada balita di Kabupaten Lampung Barat menggunakan pendekatan *machine learning* serta membandingkan kinerja algoritma *Support Vector Machine* (SVM), *Extreme Gradient Boosting* (XGBoost), dan *Categorical Boosting* (CatBoost). Penelitian ini menggunakan data sekunder berupa data antropometri balita usia 0–60 bulan yang diperoleh dari Dinas Kesehatan Kabupaten Lampung Barat periode Juni 2024–Juni 2025. Status *stunting* diklasifikasikan ke dalam empat kelas berdasarkan indikator tinggi badan menurut usia (TB/U). Data diproses melalui tahap pembersihan, normalisasi, serta penyeimbangan kelas menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE). Model dibangun dengan proses tuning *hyperparameter* dan dievaluasi menggunakan *confusion matrix* dengan metrik akurasi, precision, recall, F1-score, serta waktu komputasi. Hasil penelitian menunjukkan bahwa seluruh algoritma mampu mengklasifikasikan status *stunting* dengan akurasi di atas 98%. XGBoost menunjukkan kinerja terbaik dengan nilai akurasi, precision, recall, dan F1-score tertinggi, sementara SVM memiliki waktu komputasi yang lebih lama. CatBoost memberikan performa yang stabil dengan waktu komputasi yang relatif efisien. Hasil penelitian ini menunjukkan bahwa pendekatan *machine learning* efektif digunakan dalam klasifikasi status *stunting* berbasis data antropometri.

Kata Kunci: *Stunting*, Antropometri, *Machine Learning*, Sistem Pendukung Keputusan, Kesehatan Anak.

## **ABSTRACT**

### **APPLICATION OF SUPPORT VECTOR MACHINE, EXTREME GRADIENT BOOSTING, AND CATEGORICAL BOOSTING ALGORITHMS IN STUNTING CLASSIFICATION IN WEST LAMPUNG REGENCY**

**By**

**DETA SARI ASLINA**

Stunting is one of the chronic nutritional problems among children under five that remains a major concern in public health. This study aims to classify stunting status among children under five in West Lampung Regency using a machine learning approach and to compare the performance of Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) algorithms. This study used secondary data in the form of anthropometric data of children aged 0–60 months obtained from the West Lampung District Health Office for the period of June 2024 to June 2025. Stunting status was classified into four categories based on the height-for-age (H/A) indicator. The data were processed through data cleaning, normalization, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE). The models were developed through a hyperparameter tuning process and evaluated using a confusion matrix with accuracy, precision, recall, F1-score, and computational time as performance metrics. The results show that all algorithms were able to classify stunting status with accuracy above 98%. XGBoost demonstrated the best performance with the highest accuracy, precision, recall, and F1-score values, while SVM required longer computational time. CatBoost showed stable performance with relatively efficient computational time. These findings indicate that machine learning approaches are effective for classifying stunting status based on anthropometric data.

**Keywords:** Child Nutrition Disorders; Anthropometry; Machine Learning; Clinical Decision Support Systems; Child Health.

Judul Skripsi : **PENERAPAN ALGORITMA *SUPPORT VECTOR MACHINE, EXTREME GRADIENT BOOSTING, DAN CATEGORICAL BOOSTING* DALAM KLASIFIKASI *STUNTING* DI KABUPATEN LAMPUNG BARAT**

Nama Mahasiswa : **Deta Sari Aslina**

Nomor Pokok Mahasiswa : **2217051117**

Program Studi : **S-1 Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



**Dr. Aristoteles, S.Si., M.Si**  
NIP. 19810521 200604 1 002

**Muhaqiqin, S.Kom., M.T.I.**  
NIP. 19930525 202203 1 009

2. Ketua Jurusan Ilmu Komputer

3. Ketua Program Studi Ilmu Komputer

**Dwi Sakethi, S.Si., M.Kom.**  
NIP. 196806111998021001

**Tristiyanto, S.Kom., M.I.S., Ph.D**  
NIP. 198104142005011001

**MENGESAHKAN**

**1. Tim Penguji**

**Ketua : Dr. Aristoteles, S.Si., M.Si**

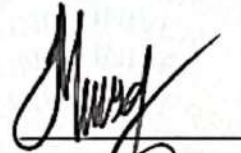
**Penguji**

**Sekretaris : Muhaqiqin, S.Kom., M.T.I.**

**Penguji**

**Penguji : Prof. Admi Syarif, Ph.D**

**Utama**

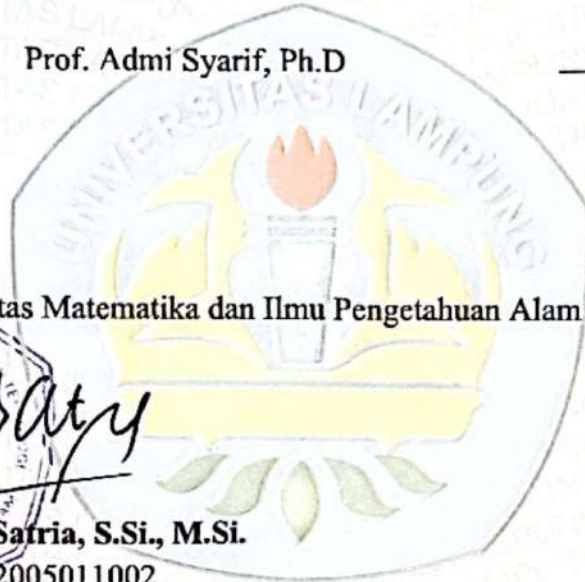


**2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam**



**Dr. Eng. Heri Sarria, S.Si., M.Si.**

**NIP. 197110012005011002**



**Tanggal Lulus Ujian Skripsi : 7 April 2026**

## PERNYATAAN

Saya bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul **“Penerapan Algoritma *Support Vector Machine*, *Extreme Gradient Boosting*, Dan *Categorical Boosting* Dalam Klasifikasi *Stunting* Di Kabupaten Lampung Barat”** merupakan hasil karya saya sendiri dan bukan karya orang lain. Seluruh penulisan yang terdapat dalam skripsi ini telah mematuhi aturan penulisan karya akademik dari Universitas Lampung. Apabila dikemudian hari ditemukan bahwa skripsi saya adalah hasil jiplakan atau dibuat orang lain, maka saya siap menerima segala konsekuensi berupa pencabutan gelar yang saya terima.

Bandar Lampung, 16 April 2026



**Deta Sari Aslina**

NPM. 2217051117

## RIWAYAT HIDUP



Penulis bernama Deta Sari Aslina, lahir di Turgak pada tanggal 9 Februari 2005. Penulis berdomisili di Pekon Turgak Kecamatan Belalau Kabupaten Lampung Barat. Penulis menempuh pendidikan formal di SDN Turgak dan lulus pada Tahun 2016 kemudian melanjutkan pendidikan di SMPN 1 Batu Brak dan lulus pada Tahun 2019 serta melanjutkan pendidikan di SMAN 1 Belalau dan lulus pada Tahun 2022. Pada Tahun 2022 penulis melanjutkan pendidikan tinggi pada Program Studi S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Selama masa perkuliahan, penulis aktif dalam kegiatan akademik dan organisasi dengan pengalaman sebagai berikut:

1. Asisten Dosen Sistem Operasi Program Studi D3 Manajemen Informatika (2024).
2. Asisten Dosen Sistem Interaksi Program Studi S1 Ilmu Komputer (2025).
3. Asisten Dosen Cloud Computing Program Studi S1 Ilmu Komputer (2025).
4. Studi Independen Bersertifikat (MSIB) Angkatan 7 Bangkit Academy – Cohort Cloud Computing (2024).
5. Kerja Praktik (KP) di Badan Pusat Statistik (BPS) Kota Bandar Lampung (2025).
6. Anggota Muda Ilmu Komputer (ADAPTER) Himpunan Mahasiswa Jurusan Ilmu Komputer (2022).
7. Kepala Biro Kesekretariatan HIMAKOM (2024).
8. Anggota Biro Kesekretariatan HIMAKOM (2023).
9. Tim Minat dan Bakat Fakultas MIPA (2023).
10. Ketua Pelaksana Pelatihan SPIA & SPIK HIMAKOM (2023).

11. Anggota Divisi Acara PRJxHT (2023).
12. Koordinator Divisi Kesekretariatan CSS (2024).
13. Panitia Seminar Nasional Bersama FMIPA (2024).
14. Kuliah Kerja Nyata (KKN) di Desa Karang Tanjung Kecamatan Padang Ratu Kabupaten Lampung Tengah (2025).

## **MOTTO**

“Janganlah kamu bersikap lemah, dan jangan pula bersedih hati, padahal kamulah yang paling tinggi derajatnya jika kamu beriman.”

*(QS. Ali Imran: 139)*

“Jika Anda tidak bisa melakukannya dengan baik, lakukanlah dengan cinta.”

*(Mother Teresa)*

“Menjadi lembut di dunia yang keras, seperti air yang tidak melawan batu, namun tetap menemukan jalan dengan caranya sendiri.”

## **PERSEMBAHAN**

*Alhamdulillahirobbilalamin*

Puji syukur kehadirat Allah Subhanahu Wa Ta'ala atas segala limpahan rahmat dan karunia-Nya, sehingga skripsi ini berhasil diselesaikan dengan baik. Shalawat serta salam selalu tercurahkan kepada Nabi Muhamad Shallahu 'Alaihi Wasallam sebagai suri teladan bagi seluruh umat manusia.

Kupersembahkan karya ini kepada:

### **Kedua Orang Tua Tercinta**

Untuk Ayah dan Ibu tercinta yang tidak pernah lelah memanjatkan doa, memberikan kasih sayang, dukungan serta pengorbanan yang tak terhingga. Terima kasih atas setiap usaha, nasihat dan kepercayaan yang selalu menjadi kekuatan bagi penulis untuk terus berjuang hingga sampai pada titik ini.

### **Seluruh Keluarga Besar Ilmu Komputer 2022**

Untuk teman-teman seperjuangan yang telah menjadi bagian dari perjalanan ini.

### **Almamater Tercinta Jurusan Ilmu Komputer FMIPA Universitas Lampung**

Tempat penulis belajar, bertumbuh dan menemukan berbagai pengetahuan serta pengalaman berharga yang menjadi bekal untuk menapaki masa depan.

## SANWACANA

Puji syukur kehadirat Allah Subhanahu Wa Ta'ala atas limpahan nikmat, rahmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul: “Penerapan Algoritma *Support Vector Machine*, *Extreme Gradient Boosting*, dan *Categorical Boosting* dalam Klasifikasi Stunting di Kabupaten Lampung Barat” dengan baik dan lancar. Dalam penyusunan skripsi ini, penulis menyadari bahwa terselesaikannya skripsi tidak terlepas dari bimbingan dan doa dari berbagai pihak. Pada kesempatan ini penulis menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas segala limpahan rahmat, pertolongan dan kemudahan yang selalu menyertai sehingga skripsi dapat terselesaikan dengan baik.
2. Mak, Bak, Ibu dan Ayah, yang senantiasa memberikan doa, motivasi dan dukungan. Terima kasih telah menjadi kekuatan terbesar bagi penulis meraih gelar Sarjana.
3. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
4. Bapak Dwi Sakethi, S.Si., M.Kom. selaku Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
5. Ibu Yunda Heningtyas, M.Kom. selaku Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung.
6. Bapak Tristiyanto, S.Kom., M.I.S., Ph.D. selaku Kepala Program Studi S1 Ilmu Komputer.
7. Bapak Dr. Aristoteles, S.Si., M.Si. selaku Pembimbing Utama yang telah memberikan bimbingan, arahan, kritik, serta saran yang membangun sehingga skripsi ini dapat terselesaikan dengan baik.

8. Bapak Muhaqiqin, S.Kom., M.T.I. selaku Pembimbing Kedua yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis menyelesaikan skripsi ini.
9. Bapak Prof. Admi Syarif, Ph.D. selaku Pembahas yang telah memberikan masukan, kritik, serta saran yang bermanfaat demi perbaikan skripsi ini.
10. Seluruh dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung, yang telah memberikan banyak ilmu pengetahuan selama perkuliahan.
11. Seluruh Staf dan Karyawan Jurusan Ilmu Komputer, khususnya Ibu Ade Nora Maela, Bang Zainuddin, Pak Dahud dan Mas Syam yang telah membantu segala urusan administrasi penulis di Jurusan Ilmu Komputer.
12. Seluruh keluarga besar penulis, Among Ajjong serta adik penulis Dian, Rizkya, Belia, Zaqi, Fadhil dan Fathan, yang selalu mendoakan dan memberi semangat.
13. Zuli Redo Putra, S.Tr.Kes. yang selalu memberikan dukungan dan menjadi tempat berbagi cerita selama proses penyusunan skripsi ini.
14. Sahabat dan Teman – teman penulis, Hawa Santuy, Lantai 3 Pride, Zalfa, Ayu, Oca, Zahra, Zainab, Afina, Nadya, Pimpinan HIMAKOM 2024 serta teman dekat lainnya yang tidak dapat disebutkan satu per satu.
15. Kebersamaan teman-teman Jurusan Ilmu Komputer angkatan 2022.
16. Diri penulis sendiri terima kasih karena tidak pernah berhenti berusaha.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Semoga skripsi ini dapat memberikan manfaat dan dapat digunakan sebagaimana mestinya.

Bandar Lampung, 16 April 2026  
Penulis



Deta Sari Aslina  
NPM. 2217051117

## DAFTAR ISI

	Halaman
DAFTAR GAMBAR .....	iv
DAFTAR TABEL .....	v
I. PENDAHULUAN .....	10
1.1 Latar Belakang.....	10
1.2 Rumusan Masalah .....	13
1.3 Batasan Masalah.....	13
1.4 Tujuan.....	14
1.5 Manfaat.....	14
II. TINJAUAN PUSTAKA.....	15
2.1 Penelitian Terdahulu.....	15
2.2 Landasan Teori .....	21
2.2.1 <i>Stunting</i> .....	21
2.2.2 Data Antropometri .....	23
2.2.3 <i>Machine learning</i> .....	30
2.2.4 Klasifikasi .....	33
2.2.5 <i>Support Vector Machine (SVM)</i> .....	34
2.2.6 <i>Xtreme Gradient Boosting (XGBoost)</i> .....	38
2.2.7 <i>Categorical Boosting</i> .....	41
2.2.8 <i>Encoding dan Non encoding</i> .....	41
2.2.9 <i>Feature Engineering</i> .....	44
2.2.10 Teknik Validasi Model.....	46

2.2.11	Evaluasi Model .....	46
III	METODOLOGI PENELITIAN.....	52
3.1	Tempat Dan Waktu.....	52
3.1.1	Tempat .....	52
3.1.2	Waktu .....	52
3.2	Alat pendukung .....	54
3.2.1	Perangkat Keras .....	54
3.2.2	Perangkat Lunak.....	54
3.3	Data .....	55
3.4.1	Kajian Pustaka.....	60
3.4.4	Pembagian Dataset .....	63
3.4.5	Implementasi Model SVM, XGBoost dan Catboost.....	64
3.4.6	Pengujian.....	65
3.4.8	Perbandingan Hasil Prediksi .....	65
IV	PEMBAHASAN .....	67
4.1	Pengumpulan Data.....	67
4.2	<i>Pre-processing</i> Data .....	67
4.2.1	<i>Data Cleaning</i> .....	68
4.2.2	<i>Feature Engineering</i> .....	73
4.2.3	Normalisasi Data.....	78
4.3	Pembagian Data.....	80
4.4	Balancing Data .....	80
4.5	Pemodelan ( <i>Modeling</i> ) .....	82
4.5.1	Model Klasifikasi <i>Support Vector Machine</i> (SVM) .....	82
4.5.2	Model Klasifikasi <i>Extreme Gradient Boosting</i> (XGBoost) .....	91

4.5.3	Model Klasifikasi <i>Categorical Boosting (CatBoost)</i> .....	105
4.6	Perbandingan Algoritma SVM, XGBoost, dan CatBoost .....	119
V	PENUTUP.....	123
5.1	Kesimpulan.....	123
5.2	Saran.....	124
	DAFTAR PUSTAKA .....	125

**DAFTAR GAMBAR**

<b>Gambar</b>	<b>Halaman</b>
Gambar 1. Penurunan Angka Stunting Nasional. ....	22
Gambar 2. Klasifikasi metode Machine learning.....	31
Gambar 3. Diagram Supervised Learning (Tuays Almuqati et al., 2024). ....	32
Gambar 4. Ilustrasi Pemilihan Hyperplane Optimal oleh SVM. ....	35
Gambar 5. Arsitektur Model XGBoost dalam Proses Boosting. ....	39
Gambar 6. Diagram Alur CatBoost.....	42
Gambar 7. Proses K-Fold Cross Validation.....	48
Gambar 8. FlowChart Tahapan Penelitian.....	59
Gambar 9. Ilustrasi fitur tanggal setelah feature engineering. ....	74
Gambar 10. Fitur Z-Score TB/U terbaru.....	75
Gambar 11. Ilustrasi perubahan fitur kategorikal ke numerik. ....	77
Gambar 12. Ilustrasi data awal sebelum Normalisasi. ....	79
Gambar 13. Ilustrasi data awal setelah normalisasi. ....	79
Gambar 14. Grafik Perbandingan hasil tuning XGBoost.....	104
Gambar 15. Grafik perbandingan hasil tuning CatBoost. ....	119

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
Tabel 1. Penelitian terdahulu .....	15
Tabel 2. Rekap Data Antropometri Indonesia.....	23
Tabel 3. Kategori Status Gizi Anak .....	26
Tabel 4. Standar TB/U Anak Perempuan Umur 0-60 Bulan .....	26
Tabel 5. Standar TB/U Anak Laki-Laki Usia 0–60 Bulan.....	28
Tabel 6. Parameter pada XGboost .....	40
Tabel 7. Matriks Konfusi .....	49
Tabel 8. Waktu Penelitian .....	53
Tabel 9. Perangkat Lunak .....	54
Tabel 10. Atribut Data Balita .....	56
Tabel 11. Jumlah Data Stunting Per Kelas .....	56
Tabel 12. Atribut Data Penelitian.....	57
Tabel 13. Contoh Data Penelitian .....	57
Tabel 14. Makna simbol flowchart .....	66
Tabel 15. Ilustrasi Nilai kosong .....	69
Tabel 16. Penanganan Nilai Kosong.....	69
Tabel 17. Ilustrasi data setelah cleaning data kosong .....	70
Tabel 18. Ilustrasi data dengan nilai ekstream .....	70
Tabel 19. Ilustrasi Data setelah Cleaning nilai eskream .....	71
Tabel 20. Ilustrasi Data dengan Tanggal tidak valid.....	71
Tabel 21. Rekap data cleaning .....	72
Tabel 22. Perbedaan jumlah distribusi tiap kelas.....	72
Tabel 23. Ilustrasi fitur tanggal sebelum feature engineering.....	73
Tabel 24. Distribusi data awal dan data baru setelah penentuan label.....	76

Tabel 25. Rekap jumlah perubahan TB/U awal dan TB/U baru .....	76
Tabel 26. Nilai minimum dan maksimum dari masing-masing data .....	78
Tabel 27. Distribusi data training awal .....	81
Tabel 28. Distribusi data training setelah dilakukan SMOTE .....	82
Tabel 29. Nilai Hyperparameter SVM .....	83
Tabel 30. Percobaan metode SVM tanpa parameter .....	83
Tabel 31. Confussion matrix tanpa parameter .....	84
Tabel 32. Hasil akurasi model SVM dengan menggunakan kernel .....	84
Tabel 33. Confussion matrix kernel RBF .....	85
Tabel 34. hasil percobaan penggunaan beberapa cost .....	85
Tabel 35. Confussion matrix Cost 100.....	86
Tabel 36. hasil percobaan penggunaan beberapa gamma .....	86
Tabel 37. Confussion matrix Gamma Scale.....	87
Tabel 38. hasil percobaan penggunaan max iter .....	87
Tabel 39. Confussion matrix Max Iter 10000 .....	88
Tabel 40. Akurasi terbaik dari masing parameter SVM .....	88
Tabel 41. Hasil akurasi model SVM dengan Gabungan 4 parameter .....	88
Tabel 42. Confussion matrix SVM 4 parameter gabungan.....	89
Tabel 43. Hasil Perbandingan Accuracy Parameter SVM.....	89
Tabel 44. Hasil Perbandingan Precision Parameter SVM .....	90
Tabel 45. Hasil Perbandingan Recall Parameter SVM .....	90
Tabel 46. Hasil Perbandingan F1-Score Parameter SVM.....	90
Tabel 47. Nilai Hyperparameter XGBoost.....	92
Tabel 48. Percobaan metode XGBoost tanpa parameter .....	92
Tabel 49. Confussion matrix XGBoost tanpa parameter .....	92
Tabel 50. Hasil Akurasi XGBoost menggunakan Learning rate.....	93
Tabel 51. Confussion matrix Learning Rate 0.5 .....	94
Tabel 52. Hasil Akurasi XGBoost menggunakan Max depth.....	94
Tabel 53. Confussion Max Depth .....	95
Tabel 54. Hasil Akurasi XGBoost menggunakan N Estimator.....	95
Tabel 55. Confussion matrix N Estimator 300.....	96

Tabel 56. Hasil Akurasi XGBoost menggunakan Subsample .....	96
Tabel 57. Confussion matrix Subsample 1.0 .....	97
Tabel 58. Hasil Akurasi XGBoost menggunakan Colsample Bytree .....	97
Tabel 59. Confussion matrix Colsample Bytree 1.0 .....	98
Tabel 60. Hasil Akurasi XGBoost menggunakan Gamma .....	98
Tabel 61. Confussion matrix Colsample Bytree 0.0 .....	99
Tabel 62. Hasil Akurasi XGBoost menggunakan Lambda & Alpha .....	99
Tabel 63. Confussion matrix Lambda 0 & Alpha 0.0 .....	100
Tabel 64. Akurasi terbaik dari masing parameter XGBoost .....	100
Tabel 65. Hasil akurasi XGBoost dengan Gabungan 7 parameter.....	101
Tabel 66. Confussion matrix XGBoost 7 parameter gabungan .....	101
Tabel 67. Hasil Perbandingan Accuracy Parameter XGBoost.....	102
Tabel 68. Hasil Perbandingan Precision Parameter XGBoost .....	102
Tabel 69. Hasil Perbandingan Recall Parameter XGBoost.....	103
Tabel 70. Hasil Perbandingan F1-Score Parameter XGBoost .....	103
Tabel 71. Nilai Hyperparameter CatBoost.....	105
Tabel 72. Percobaan metode CatBoost tanpa parameter.....	106
Tabel 73. Confussion matrix CatBoost tanpa parameter .....	106
Tabel 74. Hasil Akurasi CatBoost Learning rate data encoding .....	107
Tabel 75. Hasil Akurasi CatBoost Learning rate data non-encoding.....	107
Tabel 76. Confussion matrix Learning Rate 0.7 .....	108
Tabel 77. Hasil Akurasi CatBoost menggunakan depth encoding.....	108
Tabel 78. Hasil Akurasi CatBoost depth non-encoding.....	109
Tabel 79. Confussion Max Depth 12 .....	109
Tabel 80. Hasil Akurasi CatBoost menggunakan iteration encoding .....	110
Tabel 81. Hasil Akurasi CatBoost iteration non-encoding .....	110
Tabel 82. Confussion matrix iterations 1000 .....	111
Tabel 83. Hasil Akurasi CatBoost Regularization encoding .....	111
Tabel 84. Hasil Akurasi CatBoost Regularization non-encoding .....	112
Tabel 85. Confussion matrix Regularization 0.5 dan 0.8.....	113
Tabel 86. Hasil Akurasi CatBoost Random Strength encoding.....	113

Tabel 87. Hasil Akurasi CatBoost Random Strength non-encoding.....	113
Tabel 88. Confussion matrix Random Strength 1 .....	114
Tabel 89. Akurasi terbaik dari masing parameter CatBoost .....	115
Tabel 90. Hasil akurasi CatBoost dengan Gabungan 5 parameter.....	115
Tabel 91. Confussion matrix SVM 5 parameter gabungan.....	116
Tabel 92. Hasil Perbandingan Accuracy Parameter CatBoost.....	116
Tabel 93. Hasil Perbandingan Precision Parameter CatBoost .....	117
Tabel 94. Hasil Perbandingan Recall Parameter CatBoost .....	117
Tabel 95. Hasil Perbandingan F1-Score Parameter SVM.....	118
Tabel 96. Perbandingan Algoritma SVM, XGBoost, dan CatBoost.....	121
Tabel 97. Perbandingan Waktu Algoritma SVM, XGBoost, dan CatBoost .....	121

**DAFTAR KODE**

<b>Kode</b>	<b>Halaman</b>
Kode 1. Menghitung Usia dalam Bulan.....	74
Kode 2. Membersihkan Kolom yang Tidak Relevan.....	78
Kode 3. Kode Normalisai dengan MinMaxScaler .....	79
Kode 4.Kode Pembagian Data Training dan testing .....	80
Kode 5. Proses Balancing Data Menggunakan SMOTE .....	81
Kode 6. Proses Balancing Data Menggunakan SMOTENC .....	81

# I. PENDAHULUAN

## 1.1 Latar Belakang

Masalah *Stunting* adalah salah satu penanda utama kekurangan gizi pada balita, yang menjadi perhatian utama pemerintah Indonesia. Anak-anak yang menderita *stunting* mengalami kegagalan pertumbuhan akibat kekurangan gizi dan penyakit yang berulang, yang mengakibatkan tinggi badan di bawah rata-rata untuk usianya. Kondisi ini sering disebut sebagai fase 1.000 Hari Pertama Kehidupan (HPK), biasanya berlangsung sejak janin masih dalam kandungan hingga anak berusia dua tahun. (Kementerian Kesehatan RI, 2022).

*Stunting* berdampak pada penampilan fisik seseorang serta perkembangan mental anak. Hal ini dapat memengaruhi pembelajaran dan perkembangan kognitif, menurunkan produktivitas individu saat dewasa, dan meningkatkan risiko penyakit metabolik termasuk diabetes dan hipertensi. *Stunting* juga berpengaruh terhadap pembangunan nasional, dan kualitas sumber daya manusia (World Health Organization, 2020).

Survei Status Gizi Indonesia (SGI) 2024 menemukan bahwa 19,8% penduduk Indonesia mengalami *stunting*. Angka ini sudah lebih rendah dari kriteria 20% yang ditetapkan oleh Organisasi Kesehatan Dunia (WHO) sebagai indikasi kedaruratan kesehatan masyarakat, meskipun terjadi penurunan dari tahun-tahun sebelumnya. Prevalensi *Stunting* sebesar 15,2% di Provinsi Lampung dan 16,4% di Kabupaten Lampung Barat yang ditargetkan akan turun menjadi 14% melalui regulasi daerah (Bappeda Provinsi Lampung, 2025). Data ini menunjukkan bahwa Kabupaten Lampung Barat memerlukan strategi yang lebih efektif dalam

deteksi dini dan klasifikasi *stunting* untuk mencapai target tersebut (Kementerian Kesehatan Republik Indonesia, 2024b).

Saat ini, pencatatan dan pemantauan status gizi balita di Indonesia telah mengalami transformasi digital. Sistem memungkinkan tenaga kesehatan mencatat data antropometri balita (berat badan, tinggi badan, usia, dan identitas) secara real-time serta melakukan integrasi dan sinkronisasi data secara terpusat (Kementerian Kesehatan RI, 2023). Digitalisasi ini meningkatkan efisiensi pencatatan dan mempermudah akses riwayat kesehatan balita. Namun demikian, meskipun sistem pencatatan telah terdigitalisasi, proses klasifikasi status *stunting* masih bergantung pada perhitungan berbasis standar z-score WHO dan belum sepenuhnya memanfaatkan pendekatan analitik prediktif berbasis *machine learning*. Proses validasi data tetap memerlukan ketelitian tinggi, karena ketidaksesuaian data seperti NIK, tanggal lahir, dan jenis kelamin dapat menyebabkan data tidak terintegrasi secara optimal dalam sistem (Kementerian Kesehatan RI, 2022). Selain itu, tingginya beban kerja tenaga kesehatan di layanan primer dapat memengaruhi kualitas input data dan akurasi interpretasi hasil (Nababan, 2021). Penelitian menunjukkan bahwa sistem otomatis berbasis komputasi memiliki tingkat deviasi kesalahan yang lebih rendah dibandingkan pendekatan manual, dengan deviasi z-score absolut median sebesar 0,01 dibandingkan 0,17 pada perhitungan manual (Cohen-Sela et al., 2025). Hal ini menunjukkan bahwa penguatan sistem melalui pendekatan *machine learning* berpotensi meningkatkan akurasi, konsistensi, serta kecepatan klasifikasi status gizi. Oleh karena itu, diperlukan evaluasi dan pengembangan model klasifikasi berbasis *machine learning* yang tidak hanya terintegrasi dengan sistem digital yang telah ada, tetapi juga mampu memberikan hasil prediksi yang lebih optimal dan stabil.

Oleh karena itu, dibutuhkan metode yang tidak hanya sekadar mendigitalisasi pencatatan data, tetapi juga dapat melakukan klasifikasi

secara otomatis. *Machine learning* dapat menjadi solusi karena mampu memproses data antropometri dengan cepat dan mengurangi kemungkinan kesalahan dalam perhitungan, sehingga dapat mendukung proses pengambilan keputusan (Pudji Andjar et al., 2024). Penelitian sebelumnya oleh Cahyani (2024) di Kabupaten Lampung Barat membandingkan algoritma *Support Vector Machine* (SVM) dan *Extreme Gradient Boosting* (XGBoost) untuk klasifikasi stunting pada balita usia 0–60 bulan. Hasil penelitian menunjukkan bahwa XGBoost mencapai akurasi tertinggi sebesar 99,79% dengan waktu komputasi yang lebih efisien dibandingkan SVM (Cahyani, 2024). Namun, penelitian tersebut hanya membandingkan dua algoritma dan belum mengeksplorasi model *boosting* modern yang memiliki keunggulan khusus dalam menangani data kategorikal, seperti *Categorical Boosting* (CatBoost).

CatBoost, algoritma *boosting* berbasis gradien yang dikembangkan oleh Yandex, memiliki keunggulan spesifik dalam menangani data kategorikal tanpa memerlukan proses *encoding* manual yang rumit serta mengurangi risiko *overfitting* melalui pendekatan *ordered boosting* (Dorogush et al., 2018). Algoritma ini menggunakan mekanisme yang lebih stabil dalam mengolah fitur kategorikal dan menghasilkan model yang lebih robust terhadap *noise* dalam data. Penelitian terbaru menunjukkan bahwa CatBoost mengungguli algoritma *boosting* lain dalam klasifikasi data tabular, seperti dalam prediksi tingkat kemiskinan di Filipina yang mencapai akurasi 91%, lebih tinggi dibandingkan XGBoost (89%) dan Gradient Boosting Machine/GBM (88%) (Salvador, 2024). Meskipun CatBoost telah terbukti efektif dalam berbagai domain klasifikasi, penerapannya dalam klasifikasi *stunting* khususnya pada data balita Indonesia masih sangat terbatas dan belum dibandingkan secara komprehensif dengan SVM dan XGBoost dalam konteks deteksi *stunting*.

Dengan membandingkan ketiga algoritma SVM, XGBoost, dan CatBoost penelitian ini bertujuan untuk mengevaluasi performa klasifikasi

*stunting* dari sisi akurasi, waktu komputasi, dan stabilitas prediksi. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem klasifikasi status gizi berbasis data antropometri. Sistem ini dapat digunakan sebagai alat bantu dalam program intervensi *stunting* oleh Dinas Kesehatan Kabupaten Lampung Barat. Selain itu, hasil penelitian ini juga diharapkan menjadi dasar pengembangan model prediksi otomatis berbasis *machine learning* yang mampu membantu petugas dalam mendeteksi dini risiko *stunting* secara lebih cepat, efisien, dan akurat tanpa perlu melakukan perhitungan manual.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan diatas, rumusan masalah pada penelitian ini yaitu:

1. Bagaimana klasifikasi status *stunting* balita di Kabupaten Lampung Barat dengan menggunakan algoritma *Support Vector Machine*, *Extreme Gradient Boosting*, dan *Categorical Boosting*?
2. Bagaimana kinerja ketiga algoritma *Support Vector Machine*, *Extreme Gradient Boosting*, dan *Categorical Boosting* dalam kategorisasi status *Stunting* dibandingkan satu sama lain?

## 1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan sesuai dengan tujuan yang ingin dicapai, maka ruang lingkup penelitian dibatasi pada hal-hal berikut:

1. Penelitian ini hanya menggunakan tiga algoritma *Machine learning*, yaitu: *Support Vector Machine* (SVM), *Extreme Gradient boosting* (XGBoost), dan *Categorical Boosting* (CatBoost).
2. Data yang digunakan merupakan data sekunder yang berasal dari Dinas Kesehatan Kabupaten Lampung Barat dengan rentang waktu satu tahun dari bulan Juni 2024 sampai dengan juni 2025, berupa data antropometri balita usia 0–60 bulan yang terindikasi *stunting*.

## 1.4 Tujuan

Berdasarkan latar belakang yang telah dijabarkan diatas, tujuan pada penelitian ini sebagai berikut.

1. Mengembangkan dan menerapkan model prediksi status gizi balita menggunakan *algoritma Support Vector Machine (SVM)*, *Extreme Gradient Boosting (XGBoost)*, dan *Categorical Boosting (CatBoost)* berdasarkan data antropometri balita di Kabupaten Lampung Barat.
2. Mengevaluasi serta membandingkan kinerja ketiga algoritma melalui pengukuran akurasi, precision, recall, F1-score, dan waktu komputasi untuk menentukan model yang paling stabil, efektif, dan akurat dalam klasifikasi status *stunting*.
3. Menentukan model *machine learning* yang paling optimal berdasarkan hasil evaluasi dan perbandingan kinerja ketiga algoritma tersebut.

## 1.5 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Mempermudah dan mempercepat proses penilaian status gizi balita dengan menyediakan metode prediksi otomatis berbasis *machine learning*, sehingga mengurangi ketergantungan pada perhitungan manual yang memakan waktu.
2. Mengurangi potensi kesalahan dalam penentuan *stunting*, karena model *machine learning* dapat memberikan hasil yang konsisten dan akurat.
3. Memberikan dasar ilmiah bagi instansi kesehatan untuk mengembangkan sistem pendukung keputusan dalam deteksi dini *stunting*.

## II. TINJAUAN PUSTAKA

### 2.1 Penelitian Terdahulu

Penelitian terdahulu digunakan sebagai referensi untuk mendukung penelitian yang dilakukan. Ringkasan penelitian terdahulu dapat dilihat pada Tabel 1 berikut.

Tabel 1. Penelitian terdahulu

No	Peneliti	Judul	Metode	Objek	Hasil
1	(Cahyani, 2024)	Klasifikasi <i>Stunting</i> Di Kabupaten Lampung Barat Berdasarkan Data Antropometri Menggunakan Analisis Algoritma <i>Support Vector Machine</i> Dan <i>Xtreme Gradient Boosting</i>	SVM, Xgboost	Data <i>Stunting</i> Di Kabupaten Lampung Barat	Xgboost Mencapai Akurasi 99,79%, Lebih Baik Dari SVM dengan Akurasi 99.35%.
2	(Anku & Duah, 2024)	Predicting and Identifying Factors Associated with Undernutrition among Children	XGBoost, SVM, Random Forest, LASSO, Logistic	Data MICS 2017 Ghana (5.168 anak balita)	Untuk klasifikasi <i>stunting</i> : SVM dan XGBoost mencapai akurasi tertinggi 98% dengan AUC 100%. XGBoost

No	Peneliti	Judul	Metode	Objek	Hasil
		Under Five Years in Ghana Using Machine Learning Algorithms	Regression, LDA, Ridge		menunjukkan keseimbangan optimal antara <i>sensitivity</i> (93%) dan <i>specificity</i> (99%).
3	(Sugihartono et al., 2025a)	<i>Comparative Analysis of Machine Learning Algorithms for Classification about Stunting Genesis</i>	<i>Random Forest, K-Nearest Neighbors, XGBoost</i>	Data stunting dari Kaggle dengan 10.000 sampel balita	<i>Random Forest</i> mencapai akurasi tertinggi 99,95%, <i>precision</i> 99,89%, <i>recall</i> 99,94%, dan <i>F1-score</i> 99,91%. KNN mencapai akurasi 99,93%, sedangkan XGBoost mencapai akurasi 99,36%.
4	(Mgomezulu et al., 2025a)	<i>Advancing Predictive Analytics in Child Malnutrition: Machine, Ensemble and Deep Learning Models with Balanced Class Distribution for Early Detection of Stunting and Wasting</i>	<i>Random Forest, XGBoost, Deep Neural Networks (DNN), SVM, Logistic Regression, KNN, Gradient Boosting</i>	Kanker Data LSMS World Bank dari Malawi dengan fitur antropometri, sosial-ekonomi, dan diet.	Hasil Terbaik Didapatkan Dengan Metode <i>Random Forest</i> mencapai 100% <i>Accuracy</i> untuk <i>wasting</i> dan 99,98% <i>Accuracy</i> untuk <i>stunting</i> . XGBoost menunjukkan performa <i>excellent</i> dengan 99,49% <i>Accuracy</i> untuk <i>wasting</i> dan

No	Peneliti	Judul	Metode	Objek	Hasil
					95,52% Accuracy untuk <i>stunting</i> .
5	(Abdu-Aljabar et al., 2025)	<i>A Comparative Study of Breast Cancer Detection and Recurrence Prediction Using CatBoost Classifier</i>	CatBoost, XGBoost, Random Forest, SVM, KNN, Naive Bayes	Data <i>breast cancer</i>	SVM Akurasi CatBoost mencapai AUC di atas 98% pada dataset <i>Wisconsin Breast Cancer</i> , mengungguli XGBoost, Random Forest, SVM, KNN, dan Naive Bayes. CatBoost menunjukkan performa konsisten dengan AUC >90% pada dataset WBCD dan >83% pada dataset <i>Coimbra</i> . Model ini juga menghasilkan well-calibrated predictions karena mengoptimasi log-loss. (91,30%), Dibanding MLP (78,26%) Dan Xgboost (73,91%)
6	(Salvador, 2024)	Penggunaan Algoritma Boosting dalam	CatBoost, XGBoost, GBM,	Data sosial-ekonomi rumah	CatBoost mencapai akurasi tertinggi 90,93%, presisi

No	Peneliti	Judul	Metode	Objek	Hasil
		Pengukuran Kemiskinan Tingkat Rumah Tangga: Pendekatan Pembelajaran Mesin untuk Memprediksi dan Mengklasifikasikan Kuintil Kekayaan Rumah Tangga di Filipina	LightGBM, AdaBoost	tangga dari DHS 2022 Filipina	90,92%, recall 90,93%, dan F1-score 90,92%. Mengungguli XGBoost (89%) dan GBM (88%).
7	Penelitian Yang Akan Dilakukan	Penerapan Algoritma Support Vector Machine, Extreme Gradient Boosting, Dan Categorical Boosting Dalam Klasifikasi <i>Stunting</i> Di Kabupaten Lampung Barat	SVM, Xgboost, Dan Catboost	<i>Stunting</i> di Kabupaten Lampung Barat	Penelitian Ini Akan Membahas Tentang Klasifikasi Status <i>Stunting</i> Pada Anak Balita Di Kabupaten Lampung Barat Dengan SVM, Xgboost, Dan Catboost.

a. (Cahyani, 2024)

Dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Extreme Gradient Boosting* (XGBoost), penelitian yang dilakukan oleh Nesa Dwi Cahyani bertujuan untuk mengklasifikasikan status *stunting* anak balita di Kabupaten Lampung Barat. Data yang digunakan adalah data antropometri balita usia 0 hingga 60 bulan, yang terdiri dari informasi seperti usia, berat badan, tinggi badan, dan jenis

kelamin. Penelitian ini bertujuan untuk mengetahui model klasifikasi terbaik antara dua algoritma yang digunakan. Berdasarkan hasil evaluasi menggunakan *confusion matrix*, diketahui bahwa algoritma XGBoost menghasilkan akurasi tertinggi sebesar 99,79%, dengan efisiensi waktu komputasi yang lebih baik dibandingkan SVM.

b. (Anku & Duah, 2024)

Penelitian yang dilakukan oleh Anku dan Duah (2024) menggunakan tujuh algoritma *machine learning* untuk memprediksi status gizi kurang pada anak usia di bawah lima tahun di Ghana, yaitu: *Linear Discriminant Analysis* (LDA), *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest* (RF), *LASSO*, *Ridge Regression*, dan *XGBoost*. Dataset yang digunakan berasal dari *Multiple Indicator Cluster Survey* (MICS) 2017 dengan total 5.168 anak balita yang mencakup fitur antropometri dan sosiodemografi. Hasil penelitian menunjukkan bahwa untuk klasifikasi stunting, SVM dan XGBoost mencapai akurasi tertinggi sebesar 98% dengan nilai AUC 100%, yang mengindikasikan kemampuan diskriminasi sempurna antara kelas *stunting* dan normal. XGBoost menunjukkan keseimbangan optimal antara sensitivity (93%) dan specificity (99%), menjadikannya model terbaik untuk prediksi *stunting*.

c. (Sugihartono et al., 2025)

Sugihartono (2024) melakukan analisis komparatif terhadap tiga algoritma *machine learning* untuk klasifikasi kejadian *stunting* pada balita, yaitu *Random Forest*, *K-Nearest Neighbors* (KNN), dan *Extreme Gradient Boosting* (XGBoost). Dataset yang digunakan bersumber dari *Kaggle* dengan total 10.000 sampel balita yang mencakup data antropometri lengkap. Hasil evaluasi menunjukkan bahwa *Random Forest* mencapai performa tertinggi dengan akurasi 99,95%, precision 99,89%, recall 99,94%, dan F1-score 99,91%. *K-Nearest Neighbors* mencapai akurasi 99,93%, precision 99,87%, recall

99,88%, dan F1-score 99,88%. Sementara itu, *Extreme Gradient Boosting* mencapai akurasi 99,36%, precision 98,86%, recall 98,95%, dan F1-score 98,90%. Penelitian ini menunjukkan bahwa dengan penerapan *preprocessing* data yang optimal, ketiga algoritma mampu menghasilkan performa klasifikasi yang sangat tinggi (>99%) dalam mendeteksi status *stunting* pada balita.

d. (Mgomezulu et al., 2025)

Penelitian yang dilakukan oleh Mgomezulu et al., (2025) menggunakan pendekatan *machine learning*, *ensemble learning*, dan *deep learning* untuk deteksi dini *stunting* dan *wasting* pada anak-anak di Malawi. Tujuh algoritma dievaluasi, meliputi *Random Forest*, *XGBoost*, *Deep Neural Networks (DNN)*, *Support Vector Machine (SVM)*, *Logistic Regression*, *K-Nearest Neighbors (KNN)*, dan *Gradient Boosting*. Dataset yang digunakan berasal dari Living Standards Measurement Surveys (LSMS) World Bank yang mencakup fitur antropometri, sosial-ekonomi, dan pola diet. Hasil penelitian menunjukkan performa yang luar biasa tinggi, dengan *Random Forest* mencapai akurasi 100% untuk prediksi *wasting* dan 99,98% untuk prediksi *stunting* (dengan *precision*, *recall*, *F1-score*, dan *AUC-ROC* yang juga sempurna). *XGBoost* mendemonstrasikan performa excellent dengan akurasi 99,49% untuk *wasting* dan 95,52% untuk *stunting*.

e. (Abdu-Aljabar et al., 2025)

Penelitian yang dilakukan oleh Abdu-Aljabar et al. (2025) melakukan studi komparatif terhadap enam algoritma *machine learning* untuk deteksi dan prediksi rekurensi kanker payudara, yaitu: *CatBoost*, *XGBoost*, *Random Forest*, *Support Vector Machine (SVM)*, *K-Nearest Neighbors (KNN)*, dan *Naive Bayes*. Dataset yang digunakan berasal dari tiga sumber berbeda, yaitu *Wisconsin Breast Cancer Dataset*, *WBCD (Wisconsin Breast Cancer Database)*, dan *Coimbra Breast Cancer Dataset*, untuk menguji konsistensi performa

model pada data dengan karakteristik berbeda. Hasil penelitian menunjukkan bahwa *CatBoost* secara konsisten mengungguli semua algoritma lain dalam hal metrik evaluasi. Pada *dataset Wisconsin Breast Cancer*, *CatBoost* mencapai nilai AUC (Area Under Curve) di atas 98%, yang merupakan performa tertinggi dibandingkan XGBoost, Random Forest, SVM, KNN, dan Naive Bayes. Model ini juga menunjukkan performa yang stabil pada dataset WBCD dengan AUC di atas 90% dan dataset Coimbra dengan AUC di atas 83%. Selain itu, *CatBoost* menghasilkan *well-calibrated predictions* karena algoritma ini mengoptimasi fungsi *log-loss*, sehingga prediksi yang dihasilkan lebih dapat diandalkan untuk aplikasi klinis.

f. (Salvador, 2024)

Penelitian oleh Salvador (2024) berfokus pada penerapan algoritma *boosting* untuk mengklasifikasikan tingkat kemiskinan rumah tangga di Filipina dengan memanfaatkan data *Demographic and Health Survey (DHS) 2022*. Algoritma yang diuji adalah *CatBoost*, XGBoost, GBM, LightGBM, dan AdaBoost. Hasil penelitian menunjukkan bahwa *CatBoost* unggul dibanding XGBoost (89%) dan GBM (88%). Dengan akurasi sebesar 90,93%, presisi 90,92%, recall 90,93%, dan F1-score 90,92%.

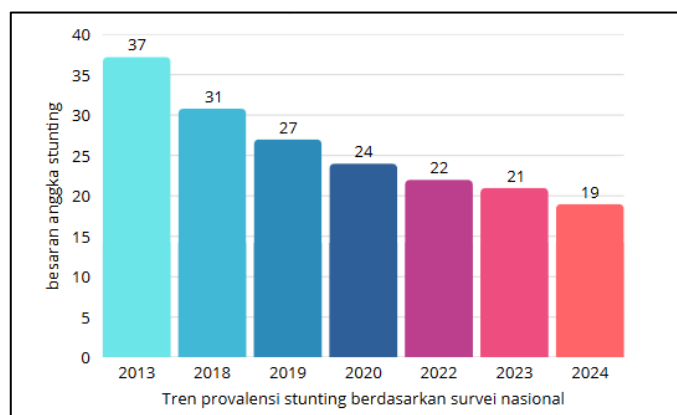
## 2.2 Landasan Teori

### 2.2.1 *Stunting*

*Stunting* adalah kondisi pertumbuhan dan perkembangan anak sebagai akibat dari kekurangan gizi kronis dan infeksi berulang, terutama pada anak usia 1.000 Hari Pertama Kehidupan (HPK) dari masa kanak-kanak hingga dewasa. Menurut pertumbuhan normal Organisasi Kesehatan Dunia (WHO), *Stunting* didefinisikan sebagai ketika Z skor tinggi badan terhadap usia anak kurang dari -2 standar deviasi. *Stunting* tidak terbatas pada gangguan fisik, hal itu juga memengaruhi perkembangan kognitif,

kapasitas belajar, daya tahan tubuh, dan peningkatan risiko penyakit nonspesifik pada populasi umum (World Health Organization, 2020).

Survei Status Gizi Indonesia (SSGI) 2024 menunjukkan bahwa Indonesia sangat berhasil menurunkan jumlah anak yang mengalami *Stunting* di Indonesia. Pada tahun 2013, 37,2% penduduk negara Indonesia mengalami *stunting*. Pada tahun 2024, angka tersebut turun menjadi 19,8%, turun 17,4 poin persentase dalam kurun waktu sebelas tahun dapat dilihat pada diagram penurunan *stunting* nasional pada Gambar 1 (Tim Penyusun SSGI, 2024).



Gambar 1. Penurunan Angka Stunting Nasional.

Secara keseluruhan, penurunan *Stunting* dari tahun ke tahun merupakan hasil dari upaya bersama dan komitmen pemerintah untuk mempercepat pencegahan *stunting*, dengan fokus pada intervensi sejak dini, edukasi gizi, dan peningkatan kualitas data kesehatan.

### 2.2.2 Data Antropometri

Antropometri adalah ilmu yang mempelajari pengukuran dimensi tubuh manusia, termasuk ukuran seperti berat, tinggi, volume, dan ruang gerak, serta karakteristik fisik lainnya. Antropometri dapat dibedakan menjadi dua jenis utama, yaitu antropometri statis yang mengukur tubuh dalam keadaan diam, dan antropometri dinamis yang mengukur perubahan dimensi tubuh saat bergerak atau beraktivitas. Pendekatan ini membantu menghasilkan produk dan lingkungan yang lebih sesuai dengan kebutuhan fisik manusia (Petra et al., 2023).

Adapun parameter rekap data antropometri Indonesia secara umum dapat dilihat pada Tabel 2, yang memuat jenis pengukuran antropometri yang digunakan secara nasional (Kementerian Kesehatan Republik Indonesia, 2024a). Dimensi-dimensi tersebut mencakup 36 parameter pengukuran yang meliputi tinggi tubuh, tinggi mata, tinggi bahu, hingga dimensi tangan dan kaki.

Tabel 2. Rekap Data Antropometri Indonesia

<b>Dimensi</b>	<b>Keterangan</b>
D1	Tinggi tubuh
D2	Tinggi mata
D3	Tinggi bahu
D4	Tinggi siku
D5	Tinggi pinggul
D6	Tinggi tulang ruas
D7	Tinggi ujung jari
D8	Tinggi dalam posisi duduk
D9	Tinggi mata dalam posisi duduk
D10	Tinggi bahu dalam posisi duduk
D11	Tinggi siku dalam posisi duduk
D12	Tebal paha

<b>Dimensi</b>	<b>Keterangan</b>
D13	Panjang lutut
D14	Panjang popliteal
D15	Tinggi lutut
D16	Tinggi popliteal
D17	Lebar sisi bahu
D18	Lebar bahu bagian atas
D19	Lebar pinggul
D20	Tebal dada
D21	Tebal perut
D22	Panjang lengan atas
D23	Panjang lengan bawah
D24	Panjang rentang tangan ke depan
D25	Panjang bahu-genggaman tangan ke depan
D26	Panjang kepala
D27	Lebar kepala
D28	Panjang tangan
D29	Lebar tangan
D30	Panjang kaki
D31	Lebar kaki
D32	Panjang rentangan tangan ke samping
D33	Panjang rentangan siku
D34	Tinggi genggaman tangan ke atas dalam posisi berdiri
D35	Tinggi genggaman ke atas dalam posisi duduk
D36	Panjang genggaman tangan ke depan

Dari berbagai parameter antropometri tersebut, petugas kesehatan di posyandu dan puskesmas umumnya hanya mengumpulkan data dasar yang paling representatif untuk menilai status gizi balita, yaitu berat badan, tinggi badan, usia, dan jenis kelamin. Data tersebut kemudian digunakan untuk menghitung Z-score tinggi badan menurut umur (TB/U) dan menentukan kategori status gizi anak.

Dalam praktik kesehatan anak, antropometri statis meliputi pengukuran berat badan (BB), tinggi atau panjang badan (TB/PB), usia, dan jenis kelamin, yang selanjutnya dibandingkan dengan kurva pertumbuhan standar WHO (Dewi Marfuah et al., 2024).

Pemerintah Indonesia melalui Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 menetapkan bahwa status gizi anak usia 0–60 bulan dinilai berdasarkan empat indeks utama, yaitu:

- a. Berat badan menurut umur (BB/U)
- b. Tinggi atau panjang badan menurut umur (TB/U atau PB/U)
- c. Berat badan menurut tinggi/panjang badan (BB/TB atau BB/PB)
- d. Indeks massa tubuh menurut umur (IMT/U)

Indeks TB/U atau PB/U merupakan acuan utama dalam menentukan status *stunting*. Seorang anak digolongkan *stunting* apabila nilai Z-score lebih rendah dari  $-2$  standar deviasi (SD) dari median kurva pertumbuhan WHO (World Health Organization, 2006). Definisi ini juga selaras dengan literatur global terbaru yang menegaskan bahwa batasan *height-for-age*  $< -2$  SD merupakan indikator utama dalam mengidentifikasi *stunting* di negara berpendapatan rendah dan menengah (Ssentongo et al., 2021).

Selain itu, penelitian lain menunjukkan bahwa *stunting* tidak hanya berdampak pada pertumbuhan fisik, tetapi juga berhubungan erat dengan perkembangan kognitif serta risiko penyakit metabolik di masa dewasa (Prendergast & Humphrey, 2014). Berikut adalah kategori status gizi anak menurut nilai Z-score dapat dilihat pada Tabel 3.

Tabel 3. Kategori Status Gizi Anak

<b>Kategori Status Gizi</b>	<b>Rentang Z-Score (SD)</b>	<b>Interpretasi</b>
Sangat Pendek c)	$Z < -3 \text{ SD}$	Anak mengalami gangguan pertumbuhan sangat parah
Pendek (Stunted)	$-3 \text{ SD} \leq Z < -2 \text{ SD}$	Anak mengalami gangguan pertumbuhan kronis
Normal	$-2 \text{ SD} \leq Z \leq +3 \text{ SD}$	Anak memiliki pertumbuhan normal
Tinggi (Tall)	$Z > +3 \text{ SD}$	Anak memiliki pertumbuhan linier di atas standar WHO

Secara khusus, Standar Antropometri Anak yang tercantum dalam Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2020 membedakan antara anak laki-laki dan perempuan karena pola pertumbuhan biologis keduanya tidak identik. Berdasarkan ketentuan tersebut, Tabel 4 dan Tabel 5 berikut menyajikan Standar Panjang Badan menurut Umur (PB/U) untuk Anak Perempuan dan Laki-Laki usia 0–60 bulan.

Tabel 4. Standar TB/U Anak Perempuan Umur 0-60 Bulan

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
0	43.6	45.4	47.3	49.1	51.0	52.9	54.7
1	47.8	49.8	51.7	53.7	55.6	57.6	59.5
2	51.0	53.0	55.0	57.1	59.1	61.1	63.2
3	53.5	55.6	57.7	59.8	61.9	64.0	66.1
4	55.6	57.8	59.9	62.1	64.3	66.4	68.6
5	57.4	59.6	61.8	64.0	66.2	68.5	70.7
6	58.9	61.2	63.5	65.7	68.0	70.3	72.5
7	60.3	62.7	65.0	67.3	69.6	71.9	74.2

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
8	61.7	64.0	66.4	68.7	71.1	73.5	75.8
9	62.9	65.3	67.7	70.1	72.6	75.0	77.4
10	64.1	66.5	69.0	71.5	73.9	76.4	78.9
11	65.2	67.7	70.3	72.8	75.3	77.8	80.3
12	66.3	68.9	71.4	74.0	76.6	79.2	81.7
13	67.3	70.0	72.6	75.2	77.8	80.5	83.1
14	68.3	71.0	73.7	76.4	79.1	81.7	84.4
15	69.3	72.0	74.8	77.5	80.2	83.0	85.7
16	70.2	73.0	75.8	78.6	81.4	84.2	87.0
17	71.1	74.0	76.8	79.7	82.5	85.4	88.2
18	72.0	74.9	77.8	80.7	83.6	86.5	89.4
19	72.8	75.8	78.8	81.7	84.7	87.6	90.6
20	73.7	76.7	79.7	82.7	85.7	88.7	91.7
21	74.5	77.5	80.6	83.7	86.7	89.8	92.9
22	75.2	78.4	81.5	84.6	87.7	90.8	94.0
23	76.0	79.2	82.3	85.5	88.7	91.9	95.0
24	76.0	79.3	82.5	85.7	88.9	92.2	95.4
25	76.8	80.0	83.3	86.6	89.9	93.1	96.4
26	77.5	80.8	84.1	87.4	90.8	94.1	97.4
27	78.1	81.5	84.9	88.3	91.7	95.0	98.4
28	78.8	82.2	85.7	89.1	92.5	96.0	99.4
29	79.5	82.9	86.4	89.9	93.4	96.9	100.3
30	80.1	83.6	87.1	90.7	94.2	97.7	101.3
31	80.7	84.3	87.9	91.4	95.0	98.6	102.2
32	81.3	84.9	88.6	92.2	95.8	99.4	103.1
33	81.9	85.6	89.3	92.9	96.6	100.3	103.9
34	82.5	86.2	89.9	93.6	97.4	101.1	104.8
35	83.1	86.8	90.6	94.4	98.1	101.9	105.6
36	83.6	87.4	91.2	95.1	98.9	102.7	106.5
37	84.2	88.0	91.9	95.7	99.6	103.4	107.3
38	84.7	88.6	92.5	96.4	100.3	104.2	108.1

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
39	85.3	89.2	93.1	97.1	101.0	105.0	108.9
40	85.8	89.8	93.8	97.7	101.7	105.7	109.7
41	86.3	90.4	94.4	98.4	102.4	106.4	110.5
42	86.8	90.9	95.0	99.0	103.1	107.2	111.2
43	87.4	91.5	95.6	99.7	103.8	107.9	112.0
44	87.9	92.0	96.2	100.3	104.5	108.6	112.7
45	88.4	92.5	96.7	100.9	105.1	109.3	113.5
46	88.9	93.1	97.3	101.5	105.8	110.0	114.2
47	89.3	93.6	97.9	102.1	106.4	110.7	114.9
48	89.8	94.1	98.4	102.7	107.0	111.3	115.7
49	90.3	94.6	99.0	103.3	107.7	112.0	116.4
50	90.7	95.1	99.5	103.9	108.3	112.7	117.1
51	91.2	95.6	100.1	104.5	108.9	113.3	117.7
52	91.7	96.1	100.6	105.0	109.5	114.0	118.4
53	92.1	96.6	101.1	105.6	110.1	114.6	119.1
54	92.6	97.1	101.6	106.2	110.7	115.2	119.8
55	93.0	97.6	102.2	106.7	111.3	115.9	120.4
56	93.4	98.1	102.7	107.3	111.9	116.5	121.1
57	93.9	98.5	103.2	107.8	112.5	117.1	121.8
58	94.3	99.0	103.7	108.4	113.0	117.7	122.4
59	94.7	99.5	104.2	108.9	113.6	118.3	123.1
60	95.2	99.9	104.7	109.4	114.2	118.9	123.7

Tabel 5. Standar TB/U Anak Laki-Laki Usia 0–60 Bulan

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
0	44.2	46.1	48.0	49.9	51.8	53.7	55.6
1	48.9	50.8	52.8	54.7	56.7	58.6	60.6
2	52.4	54.4	56.4	58.4	60.4	62.4	64.4
3	55.3	57.3	59.4	61.4	63.5	65.5	67.6
4	57.6	59.7	61.8	63.9	66.0	68.0	70.1

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
5	59.6	61.7	63.8	65.9	68.0	70.1	72.2
6	61.2	63.3	65.5	67.6	69.8	71.9	74.0
7	62.7	64.8	67.0	69.2	71.3	73.5	75.7
8	64.0	66.2	68.4	70.6	72.8	75.0	77.2
9	65.2	67.5	69.7	72.0	74.2	76.5	78.7
10	66.4	68.7	71.0	73.3	75.6	77.9	80.1
11	67.6	69.9	72.2	74.5	76.9	79.2	81.5
12	68.6	71.0	73.4	75.7	78.1	80.5	82.9
13	69.6	72.1	74.5	76.9	79.3	81.8	84.2
14	70.6	73.1	75.6	78.0	80.5	83.0	85.5
15	71.6	74.1	76.6	79.1	81.7	84.2	86.7
16	72.5	75.0	77.6	80.2	82.8	85.4	88.0
17	73.3	76.0	78.6	81.2	83.9	86.5	89.2
18	74.2	76.9	79.6	82.3	85.0	87.7	90.4
19	75.0	77.7	80.5	83.2	86.0	88.8	91.5
20	75.8	78.6	81.4	84.2	87.0	89.8	92.6
21	76.5	79.4	82.3	85.1	88.0	90.9	93.8
22	77.2	80.2	83.1	86.0	89.0	91.9	94.9
23	78.0	81.0	83.9	86.9	89.9	92.9	95.9
24	78.0	81.0	84.1	87.1	90.2	93.2	96.3
25	78.6	81.7	84.9	88.0	91.1	94.2	97.3
26	79.3	82.5	85.6	88.8	92.0	95.2	98.3
27	79.9	83.1	86.4	89.6	92.9	96.1	99.3
28	80.5	83.8	87.1	90.4	93.7	97.0	100.3
29	81.1	84.5	87.8	91.2	94.5	97.9	101.2
30	81.7	85.1	88.5	91.9	95.3	98.7	102.1
31	82.3	85.7	89.2	92.7	96.1	99.6	103.0
32	82.8	86.4	89.9	93.4	96.9	100.4	103.9
33	83.4	86.9	90.5	94.1	97.6	101.2	104.8
34	83.9	87.5	91.1	94.8	98.4	102.0	105.6
35	84.4	88.1	91.8	95.4	99.1	102.7	106.4
36	85.0	88.7	92.4	96.1	99.8	103.5	107.2

<b>Umur</b>	<b>-3SD</b>	<b>-2SD</b>	<b>-1SD</b>	<b>Median</b>	<b>+1SD</b>	<b>+2SD</b>	<b>+3SD</b>
37	85.5	89.2	93.0	96.7	100.5	104.2	108.0
38	86.0	89.8	93.6	97.4	101.2	105.0	108.8
39	86.5	90.3	94.2	98.0	101.8	105.7	109.5
40	87.0	90.9	94.7	98.6	102.5	106.4	110.3
41	87.5	91.4	95.3	99.2	103.2	107.1	111.0
42	88.0	91.9	95.9	99.9	103.8	107.8	111.7
43	88.4	92.4	96.4	100.4	104.5	108.5	112.5
44	88.9	93.0	97.0	101.0	105.1	109.1	113.2
45	89.4	93.5	97.5	101.6	105.7	109.8	113.9
46	89.8	94.0	98.1	102.2	106.3	110.4	114.6
47	90.3	94.4	98.6	102.8	106.9	111.1	115.2
48	90.7	94.9	99.1	103.3	107.5	111.7	115.9
49	91.2	95.4	99.7	103.9	108.1	112.4	116.6
50	91.6	95.9	100.2	104.4	108.7	113.0	117.3
51	92.1	96.4	100.7	105.0	109.3	113.6	117.9
52	92.5	96.9	101.2	105.6	109.9	114.2	118.6
53	93.0	97.4	101.7	106.1	110.5	114.9	119.2
54	93.4	97.8	102.3	106.7	111.1	115.5	119.9
55	93.9	98.3	102.8	107.2	111.7	116.1	120.6
56	94.3	98.8	103.3	107.8	112.3	116.7	121.2
57	94.7	99.3	103.8	108.3	112.8	117.4	121.9
58	95.2	99.7	104.3	108.9	113.4	118.0	122.6
59	95.6	100.2	104.8	109.4	114.0	118.6	123.2
60	96.1	100.7	105.3	110.0	114.6	119.2	123.9

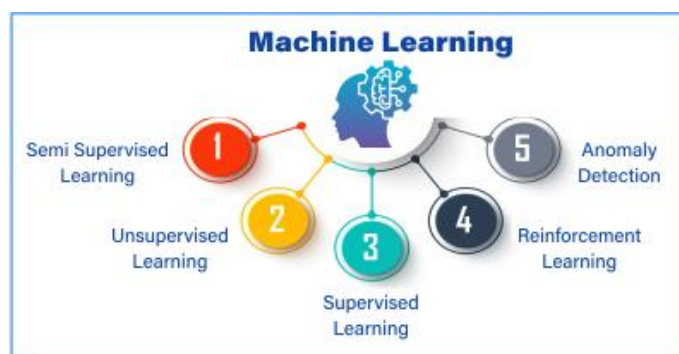
Sumber : (Kementerian Kesehatan RI, 2020)

### 2.2.3 Machine learning

*Machine learning* (ML) atau pembelajaran mesin merupakan salah satu cabang dari kecerdasan buatan (*Artificial Intelligence*) yang berfokus pada bagaimana sistem komputer dapat belajar secara otomatis dari data yang tersedia, mengidentifikasi pola, dan membuat prediksi atau

keputusan tanpa diprogram secara eksplisit (Alnuaimi & Albaldawi, 2024).

Secara umum, *machine learning* dapat dikategorikan ke dalam beberapa pendekatan utama berdasarkan cara model memperoleh pengetahuan dari data (Band et al., 2022; Drogkoula et al., 2023). Klasifikasi metode *machine learning* dapat dilihat pada Gambar 2 dibawah ini.



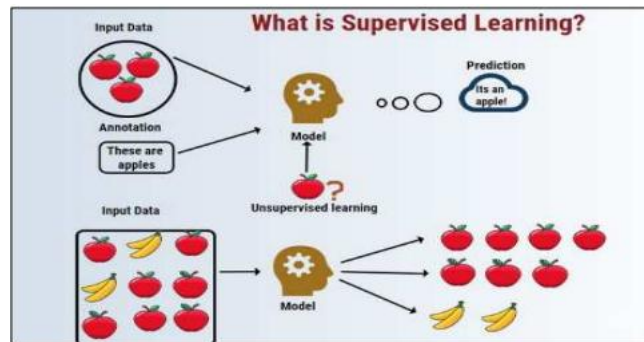
Gambar 2. Klasifikasi metode *Machine learning*.

Sumber : (Plutus Education, n.d.)

a. *Supervised Learning*

*Supervised learning* adalah metode pembelajaran mesin di mana algoritma dilatih menggunakan data berlabel, yaitu data input yang disertai dengan output yang benar. Tujuannya adalah agar model dapat mengenali pola dari data tersebut untuk mengklasifikasikan atau memprediksi data baru yang belum pernah dilihat sebelumnya. Pendekatan ini sangat penting dalam berbagai aplikasi yang membutuhkan prediksi atau klasifikasi berdasarkan data berlabel. (Tuays Almuqati et al., 2024).

Secara ringkas *supervised learning* dapat dilihat pada Diagram di Gambar 3.



Gambar 3. Diagram Supervised Learning (Tuays Almuqati et al., 2024).

b. *Unsupervised Learning*

Berbeda dengan *supervised learning*, pendekatan ini digunakan ketika data tidak memiliki label. Tujuan utama *unsupervised learning* adalah untuk menemukan struktur tersembunyi atau pola alami dalam data. Teknik ini sering digunakan untuk segmentasi pelanggan, pengelompokan data pasien, atau analisis asosiasi (Naeem et al., 2023).

c. *Semi-Supervised Learning*

*Semi-supervised learning* adalah metode yang menggabungkan sebagian data berlabel dan sebagian data tidak berlabel. Pendekatan ini sangat berguna ketika pelabelan data memerlukan biaya atau waktu yang besar, seperti pada data kesehatan yang memerlukan validasi ahli. Dengan memanfaatkan data tidak berlabel, model dapat memperkaya pembelajaran dari jumlah data terbatas (Zhang et al., 2024).

d. *Reinforcement Learning*

*Reinforcement learning* adalah pendekatan pembelajaran berbasis *trial-and-error*, di mana agen (sistem) berinteraksi dengan lingkungan dan menerima *feedback* dalam bentuk *reward* atau *punishment*. Tujuannya adalah mempelajari strategi terbaik untuk memaksimalkan *reward* kumulatif seiring waktu (Kommey et al., 2024).

e. *Anomaly Detection*

*Anomaly detection* adalah pendekatan yang digunakan untuk mengidentifikasi data yang menyimpang dari pola normal atau mayoritas. Pendekatan ini sangat berguna dalam mendeteksi *fraud*, kesalahan sistem, atau kasus kesehatan yang ekstrem. Dalam konteks *Machine learning*, *anomaly detection* dapat digunakan sebagai metode pengawasan kualitas data sebelum proses klasifikasi dilakukan (Kumari et al., 2024).

#### 2.2.4 **Klasifikasi**

Klasifikasi merupakan salah satu teknik dalam *supervised learning* yang digunakan untuk mengelompokkan data berdasarkan fitur yang dimilikinya. Menurut (Alnuaimi & Albaldawi, 2024) klasifikasi berperan penting dalam berbagai bidang, seperti biologi hingga kesehatan Masyarakat.

(IBM, 2024) mendefinisikan klasifikasi sebagai proses dalam *machine learning* yang membagi data ke dalam kelas yang telah ditentukan sebelumnya. Model dibangun dengan mempelajari karakteristik setiap kelas pada data *training*, kemudian digunakan untuk memprediksi kelas dari data baru. Dengan demikian, klasifikasi tidak hanya sekadar mengelompokkan data, tetapi juga membuat prediksi kategorikal terhadap data yang belum diketahui kelasnya.

Secara umum, menurut (Aragão et al., 2025) klasifikasi dibedakan menjadi tiga bentuk utama:

a. Klasifikasi Biner (Binary Classification)

Pengelompokan data ke dalam dua kelas saja, misalnya "ya" dan "tidak" atau "positif" dan "negatif". Contohnya adalah deteksi spam email yang hanya mengklasifikasikan email sebagai spam atau bukan spam.

b. Klasifikasi Multikelas (Multiclass Classification)

Pengelompokan data ke dalam lebih dari dua kelas yang saling eksklusif. Contohnya adalah klasifikasi status gizi seperti "normal", "pendek", "sangat pendek", dan "tinggi".

c. Klasifikasi Multilabel (Multilabel Classification)

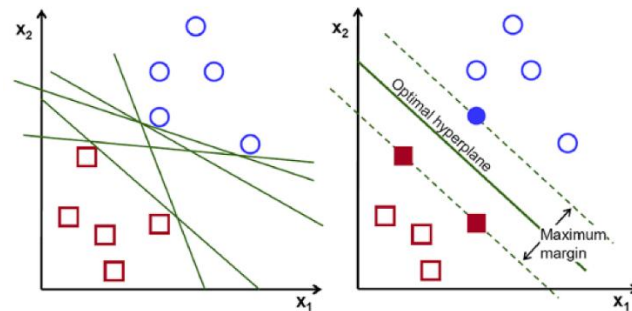
Pengelompokan di mana setiap sampel dapat memiliki lebih dari satu label secara bersamaan. Misalnya, dalam pengenalan gambar, satu gambar bisa berisi label "mobil", "jalan", dan "orang" sekaligus.

Ketiga jenis klasifikasi ini memiliki tingkat kompleksitas output yang berbeda dan sering menggunakan algoritma yang berbeda pula untuk menyelesaikan tugasnya.

### 2.2.5 *Support Vector Machine (SVM)*

*Support Vector Machine (SVM)* merupakan algoritma supervised *learning* yang sangat efektif digunakan dalam tugas klasifikasi maupun regresi. SVM bekerja dengan mencari *Hyperplane* optimal yang mampu memisahkan data ke dalam dua kelas dengan margin terbesar (Guido et al., 2024). Konsep ini pertama kali diperkenalkan oleh Vapnik dan Lerner

pada tahun 1963, lalu diperluas menjadi *soft-margin SVM* oleh Cortes dan Vapnik (Cortes et al., 1995).



Gambar 4. Ilustrasi Pemilihan *Hyperplane* Optimal oleh SVM.

Sumber : (Goldschmidt & Redline, 2023)

Gambar 4 memperlihatkan proses pemilihan *hyperplane* optimal pada SVM. Titik biru dan merah mewakili dua kelas data, sementara garis lurus di tengah adalah *hyperplane* yang memisahkan keduanya. Garis putus-putus di kedua sisi menunjukkan margin, yaitu jarak ke data terdekat dari masing-masing kelas. Titik pada margin disebut *support vectors* yang menentukan posisi *hyperplane*. Semakin besar margin, semakin baik model dalam membedakan kelas dan menggeneralisasi data baru.

*Support Vector Machine* (SVM) bertujuan untuk membangun *Hyperplane* optimal yang dapat memisahkan data dua kelas dengan margin terbesar. Margin didefinisikan sebagai jarak antara *hyperplane* dengan titik data terdekat dari masing-masing kelas. Titik-titik data yang paling dekat dengan *Hyperplane* disebut *support vectors*.

Secara matematis, fungsi *Hyperplane* pada SVM dituliskan sebagai:

$$f(x) = W^T x + b$$

di mana :

$w$  merupakan vektor bobot;

$x$  adalah vektor fitur input; dan

$b$  merupakan bias yang menentukan posisi *hyperplane*.

Pada SVM, data diklasifikasikan sedemikian rupa sehingga setiap data berada pada sisi *hyperplane* yang sesuai dengan kelasnya. Apabila data tidak bisa dipisahkan secara linear, SVM menggunakan pendekatan *kernel trick* untuk memetakan data ke ruang berdimensi lebih tinggi, sehingga memungkinkan terbentuknya batas keputusan yang bersifat non-linear tanpa melakukan transformasi fitur secara eksplisit. Menurut Liu, Shen & Wang (2014) terdapat empat kernel pada SVM, sebagaimana persamaan-persamaan berikut:

#### *Linear Kernel*

Kernel *linear* digunakan ketika data dapat dipisahkan secara *linear*, dan didefinisikan sebagai:

$$K(x_i, x_j) = x_i^T x_j$$

#### *Polynomial Kernel*

Kernel *polynomial* memungkinkan pembentukan batas keputusan *non-linear* dengan tingkat kompleksitas tertentu, yang dirumuskan sebagai:

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$$

### *Radial Basis Function (RBF)*

Kernel *Radial Basis Function* (RBF) merupakan kernel yang paling umum digunakan karena kemampuannya dalam menangani data yang tidak terpisahkan secara *linear*. Kernel ini didefinisikan sebagai:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

di mana  $\gamma$  merupakan parameter kernel yang mengontrol pengaruh jarak antar data.

### *Sigmoid Kernel*

Kernel sigmoid menyerupai fungsi aktivasi pada jaringan saraf dan dirumuskan sebagai:

$$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$$

Selain kernel, SVM memiliki beberapa parameter penting untuk mendapatkan akurasi terbaik, di antaranya

a. *Cost*

*Cost* pada SVM berfungsi untuk mengatur tingkat penalti terhadap kesalahan klasifikasi. Nilai *cost* yang besar menyebabkan model lebih sensitif terhadap kesalahan, sedangkan nilai *cost* yang kecil membuat model lebih toleran terhadap kesalahan (Guido et al., 2023).

b. *Gamma*

*Gamma* digunakan untuk mengatur sejauh mana pengaruh jarak antar data dalam pembentukan model SVM (Tharwat, 2020). Nilai *gamma* yang terlalu besar dapat menyebabkan model menjadi terlalu sensitif terhadap data tertentu, sedangkan nilai *gamma* yang terlalu kecil membuat model kurang mampu menangkap pola data.

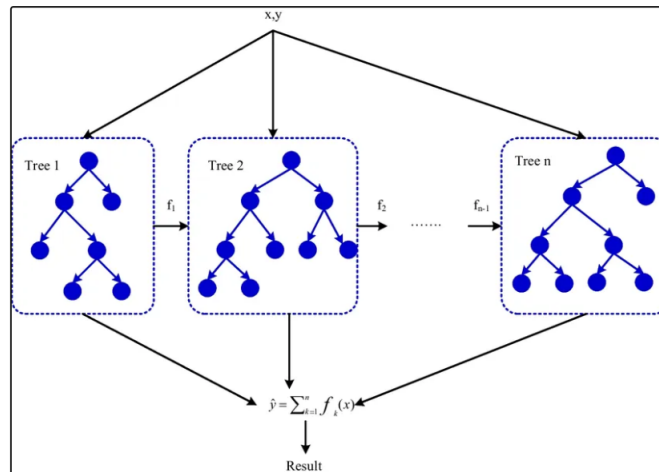
c. *Maximum Iterations*

*Maximum iterations* digunakan untuk membatasi jumlah iterasi dalam proses pelatihan SVM (James et al., 2021). Semakin besar nilai *maximum iterations*, semakin lama waktu komputasi yang dibutuhkan.

### 2.2.6 *Xtreme Gradient Boosting (XGBoost)*

*Extreme Gradient Boosting* atau XGBoost adalah algoritma pembelajaran mesin berbasis *ensemble* yang menggunakan metode *gradient boosting* pada pohon keputusan. Algoritma ini dikembangkan oleh Tianqi Chen dan Carlos Guestrin pada tahun 2016 sebagai peningkatan dari metode *gradient boosting* konvensional dengan fokus pada efisiensi, fleksibilitas, dan performa tinggi. XGBoost mengoptimalkan proses pelatihan melalui paralelisasi, pemangkasan pohon, penanganan nilai hilang, dan regularisasi untuk menghindari *overfitting*. Algoritma ini banyak digunakan dalam berbagai aplikasi seperti prediksi pasar keuangan, deteksi penipuan, klasifikasi medis, dan prediksi status kesehatan masyarakat termasuk stunting karena kemampuannya dalam menangani dataset terstruktur dengan cepat dan akurat (Arif Ali et al., 2023).

XGBoost bekerja dengan membentuk sejumlah model pohon keputusan secara bertahap. Setiap pohon baru bertugas untuk mengoreksi kesalahan prediksi dari pohon sebelumnya. Proses ini disebut sebagai *boosting*, dan hasil akhir merupakan gabungan dari seluruh pohon yang dibentuk (Brownlee, 2019). Keunggulan utamanya terletak pada efisiensi komputasi tinggi, akurasi prediksi yang unggul, serta kemampuan menangani *missing values* dan regularisasi otomatis (Wang & Dong, 2020).



Gambar 5. Arsitektur Model XGBoost dalam Proses *Boosting*.

Sumber : (DQLab, 2023)

Pada Gambar 5 ditunjukkan bahwa dataset  $x, y$  pertama kali diproses oleh pohon keputusan pertama. Setiap pohon berikutnya menghasilkan nilai koreksi terhadap prediksi sebelumnya, dan seluruh hasil pohon dijumlahkan untuk memperoleh prediksi akhir  $y$ . Proses ini membuat model bersifat adaptif terhadap kesalahan, sehingga menghasilkan prediksi yang lebih akurat dan stabil.

Fungsi objektif dalam XGBoost terdiri atas dua komponen, yaitu fungsi *loss* dan fungsi regularisasi (Chen & Guestrin, 2016; Zhang et al., 2024). Serta sumber lain DQLab, 2023; MySkill.id, 2023, beberapa *hyperparameter* penting dalam XGBoost ditunjukkan pada Tabel 6.

Tabel 6. Parameter pada XGboost

Parameter	Fungsi
<i>learning_rate (eta)</i>	Menentukan kontribusi tiap pohon. Nilai kecil = <i>training</i> lebih stabil.
<i>max_depth</i>	Kedalaman <i>maksimum</i> pohon, terlalu dalam = <i>overfitting</i> .
<i>n_estimators</i>	Jumlah total pohon dalam <i>boosting</i> .
<i>subsample</i>	Proporsi data yang digunakan di setiap iterasi.
<i>colsample_bytree</i>	Proporsi fitur yang dipakai untuk membangun setiap pohon.
<i>gamma</i>	<i>Minimum loss reduction</i> untuk membagi <i>leaf</i> .
<i>lambda, alpha</i>	Regularisasi L2 dan L1.

d. *learning\_rate (eta)*

Parameter ini menentukan seberapa besar kontribusi setiap pohon terhadap hasil akhir. Semakin kecil nilainya, semakin hati-hati model belajar dari data, sehingga mengurangi risiko *overfitting* namun memerlukan lebih banyak pohon (MySkill.id, 2023).

e. *Max depth*

Menentukan kedalaman *maksimum* setiap pohon keputusan. Nilai yang lebih besar memungkinkan model menangkap lebih banyak pola kompleks, tetapi dapat menyebabkan *overfitting* jika tidak dikontrol (DQLab, 2023).

f. *N\_estimator*

Menunjukkan jumlah total pohon yang akan dibentuk selama proses *boosting*. Nilai ini harus dioptimalkan bersama *learning\_rate*, karena terlalu banyak estimator bisa meningkatkan akurasi tetapi memperpanjang waktu komputasi (Brownlee, 2019).

g. *Subsample*

Mengatur jumlah data pelatihan yang digunakan dalam setiap iterasi. Dengan memilih nilai kurang dari 1, model menjadi lebih bervariasi dan cenderung tidak *overfit* (Chen & Guestrin, 2016).

h. *colsample\_bytree*

Mengatur persentase fitur yang dipilih secara acak untuk membangun masing-masing pohon. Parameter ini membantu mengurangi korelasi antar pohon dan meningkatkan performa generalisasi model (DQLab, 2023).

i. *gamma*

Merupakan ambang batas minimum pengurangan *loss* yang dibutuhkan agar pembelahan node dilakukan. Nilai yang lebih besar akan membuat model lebih selektif dan dapat menghindari *overfitting* pada data kecil (Zhang et al., 2024).

j. *lambda* dan *alpha*

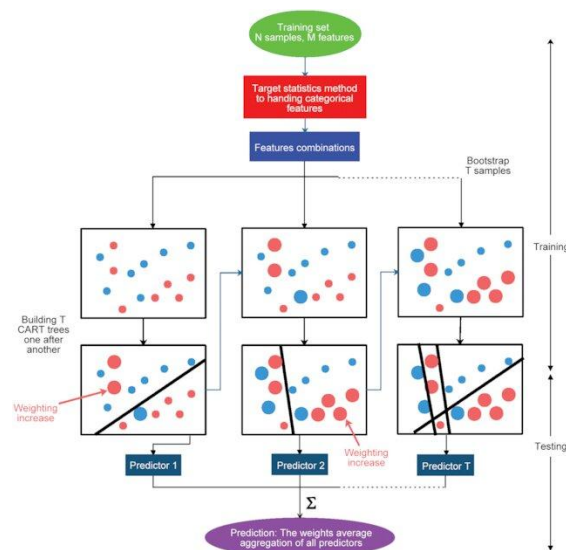
Lambda ( $\lambda$ ) dan alpha ( $\alpha$ ) adalah parameter regularisasi dalam XGBoost yang mengontrol kompleksitas model untuk mencegah *overfitting*. Lambda mengatur regularisasi L2 yang menghaluskan bobot model, sedangkan alpha mengatur regularisasi L1 yang mendorong *sparsity* dengan membuat beberapa bobot menjadi nol. Keduanya membantu model menjadi lebih sederhana dan lebih baik dalam generalisasi.(Chen & Guestrin, 2016b).

### 2.2.7 *Categorical Boosting*

*Categorical Boosting* (CatBoost) adalah algoritma *machine learning* berbasis *gradient boosting* yang dikembangkan oleh Yandex

untuk menyelesaikan berbagai tugas regresi dan klasifikasi. Ini memiliki keunggulan utama dalam menangani data kategorikal, dan, seperti algoritma penguatan lainnya, secara bertahap membangun sejumlah model pohon keputusan. Namun, CatBoost menambahkan strategi regularisasi baru dan pengolahan fitur kategorikal untuk mencegah *overfitting* (Zhao, 2022).

Setiap pohon CatBoost bertanggung jawab untuk membagi ruang fitur dan membuat perkiraan target yang lebih baik. CatBoost mengurangi bias dan kesalahan target dengan menggunakan metode statistik seperti mean target *encoding* dan *permutation-based encoding*. Ini membedakannya dari metode tradisional yang bergantung pada *encoding* eksplisit, seperti *one-hot* (GeeksforGeeks, 2023).



Gambar 6. Diagram Alur CatBoost.

Sumber : (Yao et al., 2022)

Gambar 6 memperlihatkan alur proses algoritma *Categorical Boosting (CatBoost)*. Pada tahap awal, dataset dengan jumlah  $N$  samples dan  $M$  features diproses, termasuk variabel kategorikal yang ditangani menggunakan metode *target statistics* untuk menghasilkan representasi

numerik yang lebih stabil. Fitur-fitur tersebut kemudian dikombinasikan dan dilakukan proses *bootstrap sampling* untuk menghasilkan data *training* yang berbeda. Selanjutnya, pohon keputusan (*CART trees*) dibangun secara berurutan (Predictor 1, Predictor 2, hingga Predictor T). Setiap pohon baru dirancang untuk memperbaiki kesalahan prediksi dari pohon sebelumnya dengan cara memberikan bobot (*weighting increase*) yang lebih besar pada data yang salah diprediksi. Pada tahap akhir, seluruh prediksi dari pohon keputusan digabungkan melalui proses agregasi berbobot (*weighted average aggregation*) untuk menghasilkan prediksi akhir yang lebih akurat dan stabil (Yao et al., 2022).

Dalam penelitian ini, CatBoost digunakan untuk membangun model klasifikasi status *stunting* berdasarkan data antropometri anak usia 0–60 bulan. Berdasarkan studi terbaru oleh Wahyudi (2023), algoritma CatBoost menunjukkan performa unggul dalam klasifikasi data kesehatan, dengan akurasi tinggi serta efisiensi waktu yang baik dibandingkan XGBoost maupun LightGBM.

### **2.2.8 Encoding dan Non encoding**

Encoding merupakan proses transformasi data kategorikal ke dalam bentuk representasi numerik agar dapat diproses oleh algoritma pembelajaran mesin (Kedar Potdar et al., 2017)). Tahapan ini menjadi bagian penting dalam proses *preprocessing* data karena sebagian besar algoritma *machine learning*, seperti Support Vector Machine (SVM), Neural Networks, dan Logistic Regression, hanya mampu melakukan komputasi matematis pada data numerik ((Hancock & Khoshgoftaar, 2020).

Data kategorikal yang dimaksud meliputi variabel dengan nilai diskrit atau berbentuk kategori, seperti jenis kelamin (Laki-laki, Perempuan), tingkat pendidikan (SD, SMP, SMA, Sarjana), maupun

kategori produk (Elektronik, Pakaian, Makanan). Tanpa dilakukan proses encoding, algoritma tersebut tidak dapat mengolah data kategorikal karena ketidakmampuannya dalam menginterpretasikan nilai non-numerik pada operasi matematis, seperti perkalian matriks, perhitungan jarak Euclidean, serta proses optimasi seperti *gradient descent* (Kedar Potdar et al., 2017).

### 2.2.9 *Data Processing*

*Data preprocessing* merupakan tahap awal yang sangat penting dalam *machine learning* karena efektivitas aplikasi *artificial intelligence* sangat bergantung pada kuantitas dan kualitas data yang digunakan. Data mentah biasanya masih mengandung ketidaksempurnaan, seperti nilai hilang, skala fitur yang berbeda, atau distribusi kelas yang tidak seimbang. Tanpa *preprocessing* yang tepat, teknik preprocessing dapat mempengaruhi secara signifikan performa algoritma machine learning (Amato & Di Lecce, 2023), sehingga model berisiko menghasilkan prediksi yang bias atau kurang akurat.

*Preprocessing* dilakukan untuk memastikan data dalam kondisi optimal sebelum masuk ke tahap pemodelan. Adapun langkah yang umum dilakukan antara lain:

#### 1. *Cleaning Data*

Pembersihan data merupakan tahap penting dalam pengolahan data untuk mengatasi *missing values* dan *outliers* yang dapat menurunkan kualitas analisis. Nilai yang hilang sering muncul akibat kesalahan pencatatan, kegagalan sistem, atau inkonsistensi sumber data. Untuk mengatasinya, digunakan metode imputasi sesuai jenis data, pada data numerik dengan median karena lebih tahan terhadap *outliers*, dan pada data kategorikal dengan modus agar distribusi tetap stabil (Santoso & Priyadi, 2024).

## 2. *Transforming Data*

Transformasi data merupakan proses penting dalam preprocessing yang bertujuan untuk mengubah skala atau bentuk distribusi atribut agar memenuhi persyaratan analisis. Variabel dalam dataset sering memiliki skala yang berbeda, seperti tinggi badan yang diukur dalam centimeter dan berat badan dalam kilogram. Perbedaan skala ini dapat menyebabkan beberapa fitur mendominasi algoritma yang sensitif terhadap jarak atau besaran nilai. Oleh karena itu, teknik normalisasi dan standarisasi diterapkan untuk mengubah skala atribut ke dalam rentang yang sebanding, sehingga algoritma machine learning dapat bekerja secara optimal dan menghasilkan model yang lebih akurat (Agung et al., 2023).

## 3. Normalisasi Data

Normalisasi data adalah proses menskalakan ulang nilai fitur numerik agar berada dalam rentang tertentu, biasanya antara 0 dan 1, untuk mengatasi perbedaan skala antar fitur dalam dataset. Hal ini penting agar setiap fitur memiliki kontribusi yang seimbang dalam analisis dan algoritma pembelajaran mesin, terutama yang sensitif terhadap skala. Dengan normalisasi, stabilitas perhitungan meningkat dan konvergensi algoritma dapat dipercepat, sehingga hasil model menjadi lebih akurat dan tidak bias terhadap fitur dengan nilai lebih besar (Palinggik Allorerung et al., 2024).

## 4. *Balancing Data*

Ketidakseimbangan kelas menyebabkan model lebih mudah mengenali kelas mayoritas dibandingkan kelas minoritas, sehingga diperlukan teknik *balancing data* seperti *oversampling*, *undersampling*, dan kombinasi keduanya. *Oversampling* seperti SMOTE membuat data sintesis untuk kelas minoritas agar mengurangi

*overfitting*, sedangkan *undersampling* mengurangi data kelas mayoritas tapi bisa kehilangan informasi penting. Metode k-means-SMOTE menggabungkan *clustering* dengan SMOTE untuk menghasilkan data sintetis yang lebih baik dan meningkatkan akurasi klasifikasi pada dataset (Hairani et al., 2020).

### **2.2.10 Feature Engineering**

*Feature engineering* adalah proses mengubah data mentah menjadi fitur yang lebih bermakna dan relevan agar model *machine learning* dapat bekerja lebih optimal. Tahapan ini mencakup pembuatan, transformasi, serta pemilihan fitur dengan tujuan meningkatkan akurasi, mempercepat komputasi, dan mengurangi risiko *overfitting* (Huang et al., 2023).

Menurut Mumuni & Mumuni, (2024), *feature engineering* berperan penting karena kualitas fitur yang digunakan sering kali lebih menentukan kinerja model dibandingkan kompleksitas algoritma. Dengan fitur yang tepat, model dapat mengenali pola dengan lebih baik dan menghasilkan prediksi yang akurat.

### **2.2.11 Teknik Validasi Model**

Validasi model merupakan proses untuk menguji sejauh mana model *machine learning* mampu melakukan generalisasi terhadap data baru yang tidak pernah dilihat sebelumnya. Tahap ini penting untuk mencegah *overfitting*, yaitu kondisi ketika model hanya bekerja baik pada data *training* tetapi buruk pada data uji (Ghojogh & Crowley, 2023).

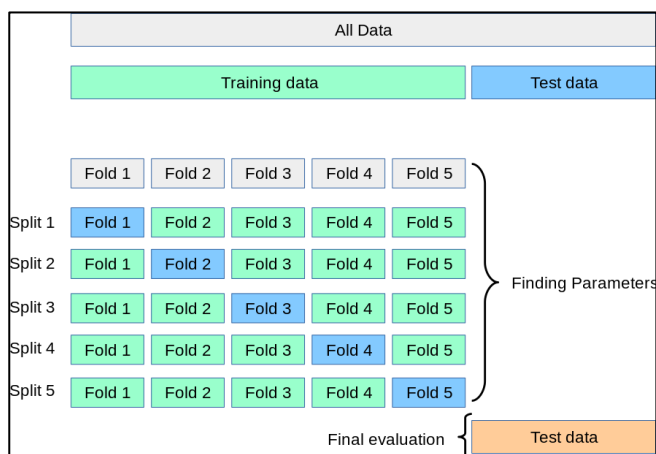
Beberapa teknik validasi yang umum digunakan dalam penelitian klasifikasi antara lain:

1. *Train-test split*

*Train-test split* adalah metode sederhana dengan cara membagi dataset menjadi dua bagian yaitu data *training*. dan data *test*. Data *training* digunakan untuk membangun model, sementara data *testing* digunakan untuk menilai kinerjanya. Proporsi yang sering digunakan adalah 80:20 atau 70:30 (Sivakumar et al., 2024). Kelebihan metode ini adalah sederhana dan cepat, tetapi hasil evaluasi bisa sangat dipengaruhi oleh cara data dibagi.

2. *K-Fold Cross Validation*

*K-Fold Cross Validation* adalah metode validasi model yang digunakan untuk meningkatkan akurasi algoritma *machine learning* dan menghindari bias yang disebabkan oleh pembagian data yang tidak merata. Metode ini membagi dataset menjadi  $k$  bagian (*folds*) yang sama besar. Proses pelatihan dan pengujian dilakukan sebanyak  $k$  kali, di mana pada setiap iterasi, satu *fold* digunakan sebagai data *testing* dan sisanya sebagai data *training*. Untuk menghasilkan perkiraan performa model yang lebih stabil, nilai metrik performa seperti skor F1 dan akurasi dirata-rata (Lumumba et al., 2024).



Gambar 7. Proses *K-Fold Cross Validation*.

Sumber : ((scikit-learn, 2024)

Gambar 7. memperlihatkan alur kerja *K-Fold Cross Validation* dengan 5 *fold*. Setiap baris mewakili satu iterasi, di mana satu *fold* (warna biru) dijadikan data *testing* dan empat lainnya (warna hijau) sebagai data *training*. Proses ini diulang hingga semua *fold* pernah digunakan sebagai data *testing* tepat satu kali. Evaluasi seperti akurasi, precision, recall, dan F1-score dihitung pada setiap iterasi, lalu dirata-ratakan untuk mengukur performa model secara keseluruhan.

### 2.2.12 Evaluasi Model

Evaluasi model dalam *machine learning* penting untuk mengetahui seberapa baik model tersebut dalam membuat prediksi dan apakah model tersebut dapat diandalkan untuk digunakan. Untuk melakukan ini, kita menggunakan berbagai metrik yang sesuai dengan jenis tugas, seperti klasifikasi atau regresi. Selain itu, performa model harus diukur secara adil dengan menggunakan data yang belum pernah dilihat oleh model, seperti data pengujian, agar hasilnya tidak bias dan mencerminkan kemampuan sebenarnya dari model tersebut (Varoquaux et al., 2023).

Metode evaluasi yang umum digunakan pada penelitian klasifikasi antara lain:

### 1. *Confusion Matrix*

*Confusion Matrix* adalah salah satu metode evaluasi performa model klasifikasi yang sangat populer karena mampu memberikan gambaran menyeluruh mengenai prediksi model terhadap data aktual dalam bentuk matriks. Matriks ini terdiri dari kombinasi antara prediksi benar dan salah yang diklasifikasikan dalam empat komponen utama, yakni *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) (Sokolova & Lapalme, 2009; Tharwat, 2018).

Bentuk umum *confusion matrix* pada kasus klasifikasi biner dapat dilihat pada Tabel 7 berikut:

Tabel 7. Matriks Konfusi

	<b>Prediksi Positif</b>	<b>Prediksi Negatif</b>
<b>Aktual Positif</b>	<i>True Positive</i> (TP)	<i>False Negative</i> (FN)
<b>Aktual Negatif</b>	<i>False Positive</i> (FP)	<i>True Negative</i> (TN)

Keterangan:

- a. *True Positive* (TP): Prediksi positif yang benar (misalnya model memprediksi anak *stunting* dan memang *stunting*).
- b. *False Positive* (FP): Prediksi positif yang salah (model memprediksi *stunting*, padahal tidak).
- c. *False Negative* (FN): Prediksi negatif yang salah (model tidak memprediksi *stunting*, padahal seharusnya).
- d. *True Negative* (TN): Prediksi negatif yang benar (model tidak memprediksi *stunting* dan memang tidak).

Setelah mengetahui nilai TP, FP, dan FN untuk dataset secara keseluruhan, langkah berikutnya adalah menemukan nilai *matrix confusion*. Nilai-nilai berikut dapat diperoleh dari matriks yang akan digunakan untuk menghitung kinerja model klasifikasi:

a. *Accuracy* (Akurasi)

Akurasi menunjukkan proporsi prediksi yang benar dari total keseluruhan data. Meskipun populer, metrik ini bisa menyesatkan jika data tidak seimbang, karena model dapat tampak "akurat" hanya dengan memprediksi kelas mayoritas, sehingga persamaannya adalah sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

(Tharwat, 2018).

b. *Precision* (Presisi)

Presisi menggambarkan seberapa tepat model dalam memprediksi kelas positif. Semakin tinggi presisi, semakin sedikit kesalahan dalam memprediksi positif. Ini penting ketika konsekuensi dari *false positive* tinggi, seperti memberikan perlakuan pada anak yang tidak *stunting*. Adapun persamaannya adalah sebagai berikut:

$$Precision = \frac{TP}{TP + FP}$$

(Grandini et al., 2020).

c. *Recall* (Sensitivitas)

*Recall* mengukur kemampuan model untuk menangkap semua kasus aktual positif. Ini penting dalam konteks *stunting* karena kita tidak

ingin anak *stunting* tidak terdeteksi. *Recall* yang rendah berarti banyak kasus *stunting* tidak dikenali oleh model. Adapun persamaannya adalah sebagai berikut:

$$Reccal = \frac{TP}{(TP + FN)}$$

(Powers & Ailab, 2020).

d. *F1-Score*

*F1-score* adalah rata-rata harmonis dari *precision* dan *recall*, yang memberikan gambaran menyeluruh terhadap kinerja model, terutama saat data tidak seimbang. Nilai F1 yang tinggi menunjukkan bahwa model memiliki keseimbangan yang baik antara akurasi dalam prediksi positif dan kemampuannya dalam mengenali semua data positif. Adapun persamaannya adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times recall}{Precision + Recall}$$

(Chicco & Jurman, 2020).

### **III. METODOLOGI PENELITIAN**

#### **3.1 Tempat Dan Waktu**

Adapun tempat dan waktu pelaksanaan penelitian ini adalah sebagai berikut. Penjelasan ini disampaikan untuk memberikan gambaran mengenai lokasi penelitian serta waktu pelaksanaan secara lebih jelas.

##### **3.1.1 Tempat**

###### **a. Tempat Pengumpulan Data**

Data pada penelitian ini diperoleh dari Dinas Kesehatan Kabupaten Lampung Barat.

###### **b. Tempat Penelitian**

Penelitian ini dilakukan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang terletak di Jalan Soemantri Brojonegoro No. 1 Gedung Meneng, Bandar Lampung.

##### **3.1.2 Waktu**

Penelitian ini dilaksanakan mulai bulan Agustus 2025 sampai dengan bulan Februari 2026. Rincian alokasi waktu penelitian dapat dilihat pada Tabel 8.



## 3.2 Alat pendukung

Alat pendukung dalam penelitian ini digunakan untuk membantu proses pengolahan data sampai pembangunan model. Alat pendukung yang digunakan terdiri dari perangkat keras dan perangkat lunak.

### 3.2.1 Perangkat Keras

- a. Processor : Intel(R) Core(TM) i5-8350U CPU @ 1.70GHz
- b. Memori : 8,00 GB
- c. System Type : 64-bit Operating System
- d. Tipe Sistem : x64-based processor

### 3.2.2 Perangkat Lunak

Adapaun Perangkat Lunak yang digunakan pada penelitian ini diantaranya dapat dilihat pada Tabel 9 berikut.

Tabel 9. Perangkat Lunak

No	Perangkat Lunak	Keterangan
1.	Sistem Operasi	Windows 11 Pro Version 24H2, Build 26100.4349
2.	Google Colaboratory (Jupyter Notebook Online)	Lingkungan pengembangan berbasis cloud dari Google yang digunakan untuk eksekusi kode Python.
3.	Bahasa Pemrograman	Python versi 3.11.13
4.	pandas (v2.2.2)	Digunakan untuk manipulasi dan analisis data.
5.	numpy (v2.0.2)	Digunakan untuk komputasi numerik.
6.	scikit-learn (v1.6.1)	Digunakan untuk preprocessing data dan evaluasi model machine learning.
7.	matplotlib (v3.10.1)	Digunakan untuk visualisasi data.
8.	imbalanced-learn (v0.13.0)	Digunakan untuk menangani dataset yang tidak seimbang.
9.	xgboost (v2.1.4)	Digunakan untuk implementasi algoritma XGBoost.

10. scikit-optimize (v0.10.2)	Digunakan untuk tuning hyperparameter model.
11. catboost (v1.2.8)	Digunakan untuk implementasi algoritma CatBoost.
12. seaborn (v0.13.2)	Digunakan untuk visualisasi tambahan dan analisis korelasi antar fitur.
13. time (built-in Python)	Digunakan untuk mengukur waktu proses eksekusi model.
14. Microsoft Excel 2019	Digunakan untuk pengolahan dan pemeriksaan awal data.
15. Microsoft Word 2019	Digunakan untuk penulisan laporan penelitian.

### 3.3 Data

Penelitian ini menggunakan data sekunder yang bersumber dari Dinas Kesehatan Kabupaten Lampung Barat. Data ini diperoleh dari kegiatan pemantauan pertumbuhan anak balita yang dilaksanakan oleh petugas kesehatan di puskesmas dan posyandu pada periode Juni 2024 hingga September 2025.

Seperti dijelaskan pada Bab 2, data antropometri mencakup berbagai parameter, namun tidak semuanya dikumpulkan secara rutin dalam pemantauan di lapangan sesuai pedoman Standar Antropometri Anak (Permenkes RI No. 2 Tahun 2020). Dalam praktiknya, atribut yang dikumpulkan pada kegiatan pemantauan pertumbuhan balita ditampilkan pada Tabel 10 berikut.

Tabel 10. Atribut Data Balita

No	Atribut
1.	NIK
2.	Nama
3.	Jenis Kelamin (JK)
4.	Tanggal Lahir
5.	Berat Badan Lahir (BB Lahir)
6.	Tinggi Badan Lahir (TB Lahir)
7.	Nama Orang Tua
8.	Alamat
9.	Usia Saat Ukur
10.	Tanggal Pengukuran
11.	KIA

Dalam penelitian ini, penggunaan atribut dibatasi pada variabel yang tersedia secara konsisten dan memiliki relevansi langsung dengan indikator tinggi badan menurut umur (TB/U). Dataset yang digunakan terdiri dari 53.119 *record*, di mana setiap *record* merepresentasikan satu individu balita dengan hasil pengukuran antropometri yang dilakukan langsung oleh petugas kesehatan di lapangan. Adapun detail record data yang didapatkan dapat dilihat pada tabel 11 berikut:

Tabel 11. Jumlah Data *Stunting* Per Kelas

No	Kelas	Jumlah Record Data	Jumlah Record Data perempuan	Jumlah Record Data Laki-Laki
1.	Normal	22664	9856	12808
2.	Pendek	17061	7452	9609
3.	Sangat Pendek	12325	5403	6922
4.	Tinggi	1069	434	635

Data yang didapatkan terdiri dari 7 atribut sebagaimana Atribut-atribut yang disajikan pada Tabel 12 berikut:

Tabel 12. Atribut Data Penelitian

No	Atribut Data Penelitian
1	Jenis Kelamin
2	Tanggal Lahir
3	Berat Badan Lahir
4	Tinggi Badan Lahir
5	Berat Badan (Saat Pengukuran)
6	Tinggi Badan (Saat Pengukuran)
7	Tanggal Ukur
8	TB/U

Data ini digunakan sebagai variabel input (fitur) dalam proses pelatihan model *machine learning*, sedangkan hasil klasifikasi TB/U digunakan sebagai variabel target (label).

Contoh data penelitian adalah seperti pada tabel 13 berikut:

Tabel 13. Contoh Data Penelitian

JK	Tanggal Lahir	BB Lahir	TB Lahir	Berat	Tinggi	Tanggal Pengukuran	TB/U
P	2019-11-27		48,62	14,02	92,32	2024-06-03	Pendek
L	2020-05-25	2,83	48,78	6,14	60	2024-06-03	Pendek
P	2019-11-27	2,86	48,47	13,68	88,66	2024-06-03	Normal
P	2023-03-11	3,2	49,64	9,68	67,48	2024-06-03	Normal
L	2021-06-19		48,57	8,62	78,33	2024-06-03	Sangat Pendek
P	2020-06-24	3,13	52,19	11,2	93,53	2024-06-03	Pendek
L	2019-12-27	2,77	48,17	10,32	76,51	2024-06-03	Normal
P	2023-11-06	2,76	45,06	5	60	2024-06-03	Normal

Dalam penelitian ini, label dikategorikan menjadi empat kelas, yaitu:

1. Pendek ( $Z\text{-score } -3 \text{ SD} \leq Z < -2 \text{ SD}$ )
2. Sangat Pendek ( $Z\text{-score} < -3 \text{ SD}$ )
3. Normal ( $Z\text{-score } -2 \text{ SD} \leq Z \leq +3 \text{ SD}$ )
4. Tinggi ( $Z\text{-score} > +3 \text{ SD}$ )

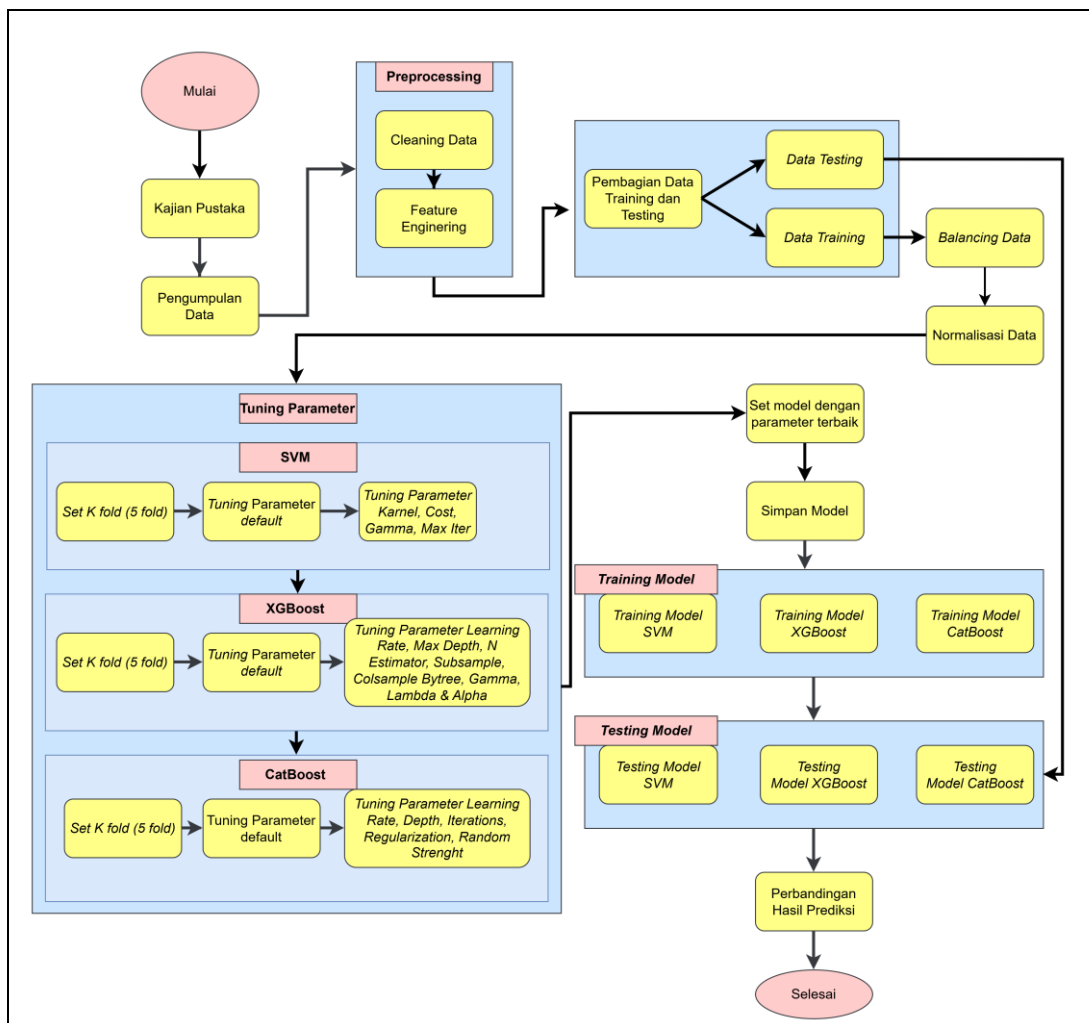
### 3.4 Tahap Penelitian

Tahapan penelitian ini terdiri dari beberapa langkah utama yang dilakukan secara sistematis untuk memperoleh model klasifikasi *stunting* yang optimal. Proses diawali dengan tahap kajian pustaka untuk mengumpulkan teori, konsep, dan hasil penelitian terdahulu yang relevan dengan topik *stunting* dan *machine learning*. Selanjutnya dilakukan pengumpulan data yang bersumber dari Dinas Kesehatan Kabupaten Lampung Barat. Tahap berikutnya adalah *data processing*, yang mencakup tiga langkah utama:

- a. *Cleaning data*, yaitu menghapus data duplikat dan nilai kosong;
- b. *Feature engineering*, yaitu menambahkan atribut baru seperti usia, serta Z-score TB/U;
- c. Setelah seluruh tahapan *data processing* selesai, dataset kemudian dibagi menjadi *data training* dan *data testing*.

Pada tahap *data training*, diterapkan metode *K-Fold Cross Validation* untuk memperoleh hasil evaluasi yang lebih stabil, namun sebelum itu akan diterapkan terlebih dahulu Normalisasi data, yaitu menyeragamkan skala nilai pada fitur numerik seperti berat dan tinggi badan agar memiliki rentang yang seimbang. *Balancing data* kemudian juga diterapkan agar jumlah data antar kelas seimbang. *Data training* tersebut selanjutnya digunakan untuk melatih tiga algoritma *machine learning*, yaitu *Support Vector Machine* (SVM), *Extreme Gradient Boosting* (XGBoost), dan *Categorical Boosting* (CatBoost).

Setelah proses *training* selesai, masing-masing model dilakukan pengujian menggunakan *data testing*. Hasil prediksi dari setiap algoritma kemudian dievaluasi menggunakan beberapa metrik kinerja, yaitu akurasi, presisi, recall, dan F1-score, guna menilai kemampuan model dalam mengklasifikasikan status *stunting*. Tahap akhir dari penelitian ini adalah melakukan perbandingan performa ketiga model berdasarkan hasil evaluasi tersebut untuk menentukan algoritma dengan kinerja terbaik dalam klasifikasi status *stunting* berdasarkan data antropometri balita. Alur tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 8.



Gambar 8. *FlowChart* Tahapan Penelitian.

### 3.4.1 Kajian Pustaka

Tahap kajian pustaka dilakukan untuk memperoleh dasar teori dan acuan metodologis yang relevan dengan penelitian mengenai klasifikasi *stunting* menggunakan algoritma *machine learning*. Proses pencarian literatur dilakukan melalui beberapa sumber, di antaranya *Google Scholar* dengan menggunakan keyword seperti “klasifikasi *stunting*”, “Support Vector Machine”, “XGBoost”, “CatBoost”. Dari hasil pencarian awal yang menghasilkan sekitar 20 jurnal, dilakukan proses seleksi berdasarkan kesesuaian topik dan tahun publikasi, sehingga diperoleh 5 jurnal yang relevan sebagai referensi utama.

Selain itu, referensi tambahan berupa skripsi terdahulu diperoleh dari perpustakaan jurusan Ilmu Komputer Universitas Lampung, khususnya skripsi milik mahasiswa sebelumnya yang memiliki topik serupa. Sumber literatur juga dilengkapi dengan buku teks dan *e-book* ilmiah yang diakses melalui platform *ResearchGate*, seperti buku *Data Mining: Concepts and Techniques* oleh Han, Pei, dan Tong (2022), serta beberapa *e-book* terkait teori *machine learning* dan pengolahan data kesehatan.

### 3.4.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data antropometri yang diperoleh dari Dinas Kesehatan Kabupaten Lampung Barat. Data tersebut berfokus pada informasi terkait balita usia 0–60 bulan, yang diukur menggunakan indikator antropometri dengan standar dan Peraturan Menteri Kesehatan RI Nomor 2 Tahun 2020.

### 3.4.3 Pre-Processing Data

Tahapan *pre-processing* merupakan langkah penting dalam menyiapkan data agar siap digunakan oleh model *machine learning*.

Proses ini bertujuan untuk meningkatkan kualitas data dengan mengurangi kesalahan, ketidakkonsistenan, serta memastikan bahwa seluruh variabel berada dalam format yang sesuai untuk analisis lebih lanjut.

Pada penelitian ini, tahapan *pre-processing* mencakup beberapa langkah utama sebagai berikut :

a. *Data Cleaning*

Tahap *data cleaning* dilakukan untuk memastikan *dataset* berada dalam kondisi bersih, konsisten, dan siap digunakan dalam proses analisis. Langkah ini mencakup pemeriksaan serta penanganan terhadap *missing values*, data duplikat, nilai ekstrem, serta penyamaan format dan satuan pengukuran.

Untuk menjaga kelengkapan data tanpa mengurangi jumlah sampel yang signifikan, *missing values* akan ditangani dengan dilakukan penghapusan karena jumlahnya sangat kecil dan tidak berpengaruh signifikan terhadap keseluruhan data.

Selain itu, dilakukan pemeriksaan terhadap nilai ekstrem yang tidak logis, seperti berat badan lahir yang belum dikonversi ke kilogram, serta tinggi badan lahir 0 cm atau sangat kecil. Format tanggal pada kolom tanggal lahir dan tanggal pengukuran juga diseragamkan ke dalam format *datetime* agar konsisten dan dapat digunakan untuk perhitungan usia.

b. *Feature Engineering*

Tahap *feature engineering* dilakukan untuk menyiapkan fitur-fitur baru yang lebih bermakna bagi proses klasifikasi *stunting*. Pada langkah ini, dilakukan penambahan dua atribut baru, yaitu:

1. Usia (dalam bulan) yang dihitung dari selisih antara tanggal lahir dan tanggal ukur.
2. Nilai Z-score TB/U yang dihitung berdasarkan standar antropometri dari Permenkes RI Nomor 2 Tahun 2020. Nilai Z-score tersebut kemudian digunakan untuk menentukan label kategori status gizi.
3. Variabel kategorikal seperti jenis kelamin dikonversi menjadi numerik pada data yang akan di *encoding* menggunakan label *encoding* (misalnya 0 untuk laki-laki dan 1 untuk perempuan).
4. Atribut yang sudah tidak relevan seperti tanggal lahir dan tanggal ukur dihapus karena informasi tersebut telah diubah menjadi variabel usia.

c. *Normalization Data*

Tahap *normalization* dilakukan untuk menyamakan skala antar fitur numerik seperti berat badan dan tinggi badan sebelum model dilatih. Normalisasi diperlukan karena setiap fitur memiliki rentang nilai yang berbeda. Jika tidak dinormalisasi, fitur dengan nilai lebih besar dapat lebih dominan dalam proses pembelajaran model dan dapat memengaruhi hasil prediksi.

Pada penelitian ini *normalization* dilakukan menggunakan metode *Min-Max Scaling*. Proses *normalization* dilakukan setelah pembagian data menjadi *data training* dan *data testing*. *MinMaxScaler* di-fit hanya pada *data training*, kemudian hasil scaler tersebut digunakan untuk melakukan transformasi pada *data training* dan *data testing*. Cara ini dilakukan untuk menghindari *data leakage* agar *data testing* tidak ikut mempengaruhi proses pembentukan model.

#### d. *Balancing Data*

Pada penelitian ini, tahap *balancing data* dilakukan untuk mengatasi ketidakseimbangan jumlah data antar kelas. Proses *balancing* hanya diterapkan pada data *training* untuk mencegah bias model terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali kelas minoritas. Untuk skema data *encoding*, metode SMOTE digunakan untuk menyeimbangkan distribusi kelas dengan cara menghasilkan data sintetis pada kelas minoritas. Sementara itu, pada skema data *non-encoding*, digunakan metode SMOTE-NC. Metode ini dipilih karena mampu menangani data dengan kombinasi fitur numerik dan kategorikal, sehingga fitur kategorikal tetap dipertahankan tanpa dilakukan proses *encoding*.

#### 3.4.4 **Pembagian Dataset**

Pada tahap ini, dataset dibagi menjadi dua bagian menggunakan metode *train test split* dengan proporsi 80% sebagai data *training* dan 20% sebagai data *testing*. Pembagian ini dilakukan dengan teknik *stratified sampling* untuk memastikan distribusi kelas TB/U pada data *training* dan data *testing* tetap seimbang dan merepresentasikan kondisi data asli. Data *training* digunakan dalam proses pelatihan dan validasi model, sedangkan data *testing* disimpan dan digunakan pada tahap akhir untuk mengevaluasi kinerja model secara objektif.

Selanjutnya, data *training* divalidasi menggunakan metode *K-Fold Cross Validation* dengan *5 Fold*, di mana pada setiap iterasi empat *Fold* digunakan sebagai data *training* dan satu *Fold* sebagai data validasi. Proses ini dilakukan secara bergantian hingga setiap *Fold* berperan sebagai data validasi satu kali. Nilai performa model diperoleh dari rata-rata hasil evaluasi seluruh iterasi.

### 3.4.5 Implementasi Model SVM, XGBoost dan Catboost

Pada tahap ini dilakukan proses pelatihan (*training*) dan pengujian (*testing*) model menggunakan tiga algoritma *machine learning*, yaitu *Support Vector Machine* (SVM), *Extreme Gradient Boosting* (XGBoost), dan *Categorical Boosting* (CatBoost). Dataset yang telah melalui tahap *preprocessing* meliputi *cleaning*, *feature engineering*, *normalization*, dan *balancing*.

Sebelum proses pelatihan, masing-masing algoritma dilakukan *tuning* parameter untuk memperoleh kombinasi parameter yang optimal. Pada model SVM, *tuning* parameter dilakukan terhadap *kernel*, *cost*, *gamma*, dan *max-iter*. Pada model XGBoost, parameter yang dituning meliputi *learning rate*, *maximum depth*, *jumlah estimator*, *subsample*, *colsample bytree*, *gamma*, serta regularisasi *lambda* dan *alpha*. Sementara itu, pada model CatBoost, *tuning* dilakukan pada *learning rate*, *depth*, *jumlah iterasi*, *parameter regularisasi*, dan *random strength*.

Model SVM digunakan untuk membangun *hyperplane* yang memisahkan data balita berdasarkan atribut antropometri seperti jenis kelamin, berat badan lahir, tinggi badan lahir, berat dan tinggi saat ukur, serta usia. Sementara itu, XGBoost dan CatBoost digunakan untuk membangun model berbasis pohon keputusan dengan pendekatan *boosting* yang bertujuan memperbaiki kesalahan prediksi dari iterasi sebelumnya.

Setelah seluruh model selesai dilatih, proses *testing* dilakukan untuk menguji kemampuan model dalam mengklasifikasikan status gizi balita ke dalam empat kategori, yaitu Sangat Pendek, Pendek, Normal, dan Tinggi, berdasarkan nilai Z-score TB/U.

### 3.4.6 Pengujian

Setelah proses pelatihan model dilakukan menggunakan data *training*, tahap berikutnya adalah melakukan pengujian (*testing*) untuk menilai sejauh mana model mampu mengenali pola baru yang belum pernah dipelajari sebelumnya. Tahap ini bertujuan untuk memastikan bahwa model tidak hanya menghafal data *training* (*overfitting*), tetapi juga mampu melakukan generalisasi terhadap data baru.

Evaluasi dilakukan dengan menggunakan *confusion matrix*, yang menampilkan perbandingan antara hasil prediksi model dengan label aktual dari data *testing*. Berdasarkan matriks ini, dihitung beberapa metrik utama, yaitu *Accuracy*, *precision*, *recall*, dan *F1-score*, untuk mengukur kinerja model dari berbagai aspek. Akurasi menunjukkan tingkat ketepatan keseluruhan prediksi, presisi menilai sejauh mana model tepat dalam memprediksi kelas positif, *recall* mengukur kemampuan model dalam mendeteksi seluruh data aktual positif, sedangkan *F1-score* memberikan keseimbangan antara presisi dan *recall*.


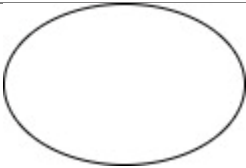

Melalui kombinasi keempat metrik tersebut, performa model SVM, XGBoost, dan CatBoost dapat dinilai secara menyeluruh dan objektif, sehingga diperoleh model dengan hasil klasifikasi terbaik dalam memprediksi status gizi balita berdasarkan data antropometri.

### 3.4.8 Perbandingan Hasil Prediksi

Tahap ini merupakan langkah akhir dari proses analisis, yaitu membandingkan hasil prediksi dari ketiga algoritma *Support Vector Machine* (SVM), *Extreme Gradient Boosting* (XGBoost), dan *Categorical Boosting* (CatBoost) berdasarkan nilai metrik evaluasi yang telah diperoleh sebelumnya. Perbandingan dilakukan terhadap empat parameter utama, yaitu *Accuracy*, *precision*, *recall*, dan *F1-score*, untuk menilai performa setiap model dari berbagai sudut pandang.

Melalui analisis perbandingan ini, dapat diketahui algoritma mana yang memiliki kinerja paling optimal dalam mengklasifikasikan status gizi balita ke dalam kategori Sangat Pendek, Pendek, Normal, dan Tinggi. Hasil perbandingan ini kemudian menjadi dasar dalam menentukan model terbaik yang dapat direkomendasikan untuk diterapkan dalam sistem klasifikasi stunting berbasis data antropometri di Kabupaten Lampung Barat.

Tabel 14. Makna simbol *flowchart*

<b>Simbol</b>	<b>Nama</b>	<b>Keterangan</b>
	<i>Flow</i>	Digunakan untuk menentukan arah dalam proses atau program
	<i>Terminator</i>	Digunakan untuk menyatakan awal atau akhir pada proses atau program
	<i>Process</i>	Digunakan untuk menyatakan proses yang dilakukan

## V. PENUTUP

### 5.1 Kesimpulan

Berdasarkan hasil pengolahan dan analisis data dalam penelitian ini, dapat disimpulkan beberapa hal sebagai berikut:

1. Penelitian ini berhasil mengklasifikasikan status gizi balita (*stunting*) menggunakan tiga algoritma *machine learning*, yaitu SVM, XGBoost, dan CatBoost, dengan seluruh model menghasilkan akurasi di atas 98%, sehingga pendekatan *machine learning* terbukti efektif untuk permasalahan klasifikasi *stunting* berbasis data antropometri.
2. Berdasarkan hasil evaluasi, XGBoost menjadi algoritma dengan performa terbaik, dengan nilai akurasi 99,29%, precision 98,44%, recall 99,01%, dan F1-score 98,72%. Nilai recall yang tinggi menunjukkan kemampuan XGBoost dalam mendeteksi sebagian besar kasus *stunting* dengan baik.
3. Algoritma SVM juga menunjukkan performa klasifikasi yang sangat kompetitif dengan akurasi sebesar 99,15%. Namun, berdasarkan hasil pengujian waktu komputasi, SVM membutuhkan waktu pemrosesan yang lebih lama dibandingkan XGBoost dan CatBoost. Hal ini menunjukkan bahwa meskipun SVM memiliki akurasi yang tinggi, efisiensi komputasinya relatif lebih rendah ketika diterapkan pada dataset dengan jumlah data yang besar.
4. Pada algoritma CatBoost, pengujian dilakukan menggunakan dua skema *preprocessing*, yaitu skema *encoding* dan *non-encoding*. Hasil pengujian menunjukkan bahwa CatBoost dengan skema *encoding* menghasilkan kinerja yang cukup efisien dengan waktu komputasi rata-rata sebesar 0,0476 detik, serta akurasi yang tetap tinggi, yaitu 98,70%. Sementara itu, CatBoost dengan skema *non-encoding*

menghasilkan akurasi sedikit lebih tinggi, yaitu 98,79%, dengan waktu komputasi yang jauh lebih cepat yaitu rata-rata sebesar 0,0390.

## 5.2 Saran

Berdasarkan kesimpulan penelitian, berikut adalah beberapa saran untuk penelitian selanjutnya diantaranya:

1. Penelitian selanjutnya disarankan untuk menambahkan variabel lain yang berkaitan dengan faktor penyebab *stunting*, seperti pola asuh, asupan gizi, atau kondisi lingkungan, agar model dapat memberikan hasil klasifikasi yang lebih akurat.
2. Model klasifikasi *stunting* yang dihasilkan dapat dikembangkan lebih lanjut dengan menggunakan algoritma lainnya, serta diuji pada data dari wilayah yang berbeda untuk mengetahui kemampuan generalisasi model terhadap karakteristik data yang lebih beragam.
3. Model terbaik pada penelitian ini disarankan untuk diimplementasikan dalam bentuk aplikasi berbasis *website* atau sistem pendukung keputusan, sehingga dapat dimanfaatkan oleh tenaga kesehatan sebagai alat bantu dalam melakukan deteksi dini *stunting* secara lebih cepat, objektif, dan efisien.

## DAFTAR PUSTAKA

- Abdu-Aljabar, R. D., Aljafaar, K. D., Ameen, Z. J. M., & Naman, H. A. (2025). A COMPARATIVE STUDY OF BREAST CANCER DETECTION AND RECURRENCE PREDICTION USING CATBOOST CLASSIFIER. *Acta Polytechnica*, 65(2), 136–142. <https://doi.org/10.14311/AP.2025.65.0136>
- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05.
- Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97. <https://doi.org/10.1051/bioconf/20249700133>
- Amato, A., & Di Lecce, V. (2023). Data preprocessing impact on machine learning algorithm performance. *Open Computer Science*, 13(1). <https://doi.org/10.1515/comp-2022-0278>
- Anku, E. K., & Duah, H. O. (2024). Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms. *PLoS ONE*, 19(2 February). <https://doi.org/10.1371/journal.pone.0296625>
- Aragão, M. V. C., Afonso, A. G., Ferraz, R. C., Ferreira, R. G., Leite, S. G., de Figueiredo, F. A. P., & Mafra, S. B. (2025). A practical evaluation of AutoML tools for binary, multiclass, and multilabel classification. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-02149-x>

- Arif Ali, Z., H. Abduljabbar, Z., A. Tahir, H., Bibo Sallow, A., & Almufti, S. M. (2023). eXtreme Gradient Boosting Algorithm with Machine Learning: a Review. *Academic Journal of Nawroz University*, 12(2), 320–334. <https://doi.org/10.25007/ajnu.v12n2a1612>
- Band, S. S., Ardabili, S., Yarahmadi, A., Pahlevanzadeh, B., Kiani, A. K., Beheshti, A., Alinejad-Rokny, H., Dehzangi, I., Chang, A., Mosavi, A., & Moslehpour, M. (2022). A Survey on Machine Learning and Internet of Medical Things-Based Approaches for Handling COVID-19: Meta-Analysis. In *Frontiers in Public Health* (Vol. 10). Frontiers Media S.A. <https://doi.org/10.3389/fpubh.2022.869238>
- Bappeda Provinsi Lampung. (2025, January 20). *Rapat Pembahasan Pencegahan dan Percepatan Penurunan Stunting di Provinsi Lampung Tahun 2025*. Bappeda.Lampungprov.Go.Id.
- Brownlee Jason. (2019). *XGBoost with Python- Gradient Boosted Trees with XGBoost and -- Jason Brownlee*.
- Cahyani, N. D. (2024). *Klasifikasi Stunting di Kabupaten Lampung Barat Berdasarkan Data Antropometri Menggunakan Analisis Algoritma Support Vector Machine dan XGBoost*. Universitas Lampung.
- Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. In *Journal of Artificial Intelligence Research* (Vol. 16).
- Chen, T., & Guestrin, C. (2016a). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Chen, T., & Guestrin, C. (2016b). *XGBoost: A Scalable Tree Boosting System*. <https://doi.org/10.1145/2939672.2939785>

- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1). <https://doi.org/10.1186/s12864-019-6413-7>
- Cohen-Sela, E., Lebenthal, Y., Brener, A., Regev, R., & Hagenäs, L. (2025). An AI-assisted tool for automated growth monitoring in pediatric achondroplasia. *European Journal of Pediatrics*, 184(8). <https://doi.org/10.1007/s00431-025-06321-3>
- Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Dewi Marfuah, Siti Sarifah, Siti Khusnul Khotimah, & Dhinda Kusuma Hatifah. (2024). Pengukuran Antropometri dan Penentuan Status Gizi Balita di Posyandu Balita Bina Sejahtera Kadipiro Banjarsari Surakarta. *ALKHIDMAH: Jurnal Pengabdian Dan Kemitraan Masyarakat*, 2(3), 138–149. <https://doi.org/10.59246/alkhidmah.v2i3.983>
- DQLab. (2023). *Tools Data Science dengan Algoritma XGBoost*. DQLab.
- Drogkoula, M., Kokkinos, K., & Samaras, N. (2023). A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management. In *Applied Sciences (Switzerland)* (Vol. 13, Number 22). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/app132212147>
- GeeksforGeeks. (2023). *CatBoost ML Algorithm*. 2023.
- Ghojogh, B., & Crowley, M. (2023). *The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial*. <http://arxiv.org/abs/1905.12787>

- Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: an Overview*. <http://arxiv.org/abs/2008.05756>
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information (Switzerland)*, 15(4). <https://doi.org/10.3390/info15040235>
- Guido, R., Groccia, M. C., & Conforti, D. (2023). A hyper-parameter tuning approach for cost-sensitive support vector machine classifiers. *Soft Computing*, 27(18), 12863–12881. <https://doi.org/10.1007/s00500-022-06768-8>
- Hairani, H., Saputro, K. E., & Fadli, S. (2020). K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes. *Jurnal Teknologi Dan Sistem Komputer*, 8(2), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00369-8>
- Huang, H. N., Chen, H. M., Lin, W. W., Huang, C. J., Chen, Y. C., Wang, Y. H., & Yang, C. T. (2023). Employing feature engineering strategies to improve the performance of machine learning algorithms on echocardiogram dataset. *Digital Health*, 9. <https://doi.org/10.1177/20552076231207589>
- IBM. (2024). *What is Classification in Machine Learning?* IBM Cloud Learn Hub.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning*.

- Kedar Potdar, Taher S. Pardawala, & Chinmay D. Pai. (2017). *A Comparative Study of Categorical Variable Encoding Techniques for Neural Network Classifiers*. Packt Publishing.
- Kementerian Kesehatan Republik Indonesia. (2024a). *Data Antropometri Indonesia*.
- Kementerian Kesehatan Republik Indonesia. (2024b, May 26). *SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%*. Sehat Negeriku.
- Kementerian Kesehatan RI. (2022).
- Komme, B., Isaac, O. J., Tamakloe, E., & Opoku, D. (2024). A Reinforcement Learning Review: Past Acts, Present Facts and Future Prospects. *Journal Research and Development (ITJRD)*, 8(2). <https://doi.org/10.25299/itjrd.2024.13474>
- Kumari, S., Prabha, C., Karim, A., Hassan, M. M., & Azam, S. (2024). A Comprehensive Investigation of Anomaly Detection Methods in Deep Learning and Machine Learning: 2019–2023. In *IET Information Security* (Vol. 2024, Number 1). John Wiley and Sons Ltd. <https://doi.org/10.1049/2024/8821891>
- Lumumba, V., Kiprotich, D., Mpaine, M., Makena, N., & Kavita, M. (2024). Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models. *American Journal of Theoretical and Applied Statistics*, 13(5), 127–137. <https://doi.org/10.11648/j.ajtas.20241305.13>
- Mgomezulu, W. R., Thangata, P., Mkandawire, B., & Amoah, N. (2025a). Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting. *Human Nutrition and Metabolism*, 42. <https://doi.org/10.1016/j.hnm.2025.200340>

- Mgomezulu, W. R., Thangata, P., Mkandawire, B., & Amoah, N. (2025b). Advancing predictive analytics in child malnutrition: Machine, ensemble and deep learning models with balanced class distribution for early detection of stunting and wasting. *Human Nutrition and Metabolism*, 42. <https://doi.org/10.1016/j.hnm.2025.200340>
- Mumuni, A., & Mumuni, F. (n.d.). *Automated data processing and feature engineering for deep learning and big data applications: a survey*.
- MySkill.id. (2023). *XGBoost Hyperparameter Tuning*. Medium.
- Nababan, J. F. (2021). *Klasifikasi Penderita Stunting dengan Metode Support Vector Machine (Studi Kasus: Lima Puskesmas di Kota Bandar Lampung)*. Universitas Lampung.
- Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, 13(1), 911–921. <https://doi.org/10.12785/ijcds/130172>
- Palinggik Allorerung, P., Erna, A., Bagussahrir, M., & Alam, S. (2024). Analisis Performa Normalisasi Data untuk Klasifikasi K-Nearest Neighbor pada Dataset Penyakit. In *Jurnal Informatika Sunan Kalijaga* (Vol. 9, Number 3).
- Petrak, S., Rastovac, I., & Mahnić Naglič, M. (2023). Dynamic Anthropometry – Research on Body Dimensional Changes. *Tekstilec*, 66(3), 240–248. <https://doi.org/10.14502/tekstilec.66.2023031>
- Powers, D. M. W., & Ailab. (2020). *EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION*.

- Prendergast, A. J., & Humphrey, J. H. (2014). The stunting syndrome in developing countries. *Paediatrics and International Child Health*, 34(4), 250–265. <https://doi.org/10.1179/2046905514Y.0000000158>
- Pudji Andjar, Mak'ruf Muhammad Ridha, & Nugraha Priyambada Cahya. (2024). *Web-Based Infant Weight and Height Measurement for Early Stunting Detection at Krembangan Posyandu, Surabaya*. <https://doi.org/10.35882/ficse.v2i1.86>
- Salvador, E. L. (2024). *Use of Boosting Algorithms in Household-Level Poverty Measurement: A Machine Learning Approach to Predict and Classify Household Wealth Quintiles in the Philippines*. <http://arxiv.org/abs/2407.13061>
- Santoso, & Priyadi. (2024). *Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI*. 17(2). <https://doi.org/10.51903/elkom.v17i2.2311>
- scikit-learn. (2024). *Cross-validation: evaluating estimator performance*. Scikit-Learn.Org.
- Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ Computer Science*, 10. <https://doi.org/10.7717/PEERJ-CS.2245>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Ssentongo, P., Ssentongo, A. E., Ba, D. M., Ericson, J. E., Na, M., Gao, X., Fronterre, C., Chinchilli, V. M., & Schiff, S. J. (2021). Global, regional and national epidemiology and prevalence of child stunting, wasting and underweight in low- and middle-income countries, 2006–2018. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-84302-w>

- Sugihartono, T., Wijaya, B., Marini, Alkayes, A. F., & Anugrah, H. A. (2025a). Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN. *Journal of Applied Data Sciences*, 6(1), 667–682. <https://doi.org/10.47738/jads.v6i1.494>
- Sugihartono, T., Wijaya, B., Marini, Alkayes, A. F., & Anugrah, H. A. (2025b). Optimizing Stunting Detection through SMOTE and Machine Learning: a Comparative Study of XGBoost, Random Forest, SVM, and k-NN. *Journal of Applied Data Sciences*, 6(1), 667–682. <https://doi.org/10.47738/jads.v6i1.494>
- Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Tharwat, A. (2020). *Behavioral analysis of support vector machine classifier with Gaussian kernel and imbalanced data*. <http://arxiv.org/abs/2007.05042>
- Tim Penyusun SSGI 2024; Asnawi Abdullah; Dwi Puspasari; Iwan Ariawan; A. Razak Thaha; Nur Indrawati Lipoeto; Firman Witoelar; Ahmad Syafiq; Ahmad Avenzora. (2024). *SURVEI STATUS GIZI INDONESIA 2024*.
- Tuays Almuqati, M., Sidi, F., Nurulain, S., Rum, M., Zolkepli, M., & Ishak, I. (2024). *Challenges in Supervised and Unsupervised Learning: A Comprehensive Overview*. 14(4).
- Varoquaux, G., Colliot, O., & Varoquaux, G. (n.d.). *Evaluating machine learning models and their diagnostic value*. Retrieved <https://hal.science/hal-03682454v4>
- Wahyudi, R.; P. E. (2023). Classification of Child Stunting Based on Antropometric Data Using CatBoost Algorithm. *Bulletin of Computer Science and Research*, 5(1).

- Wang, J., & Dong, Y. (2020). Measurement of text similarity: A survey. In *Information (Switzerland)* (Vol. 11, Number 9, pp. 1–17). MDPI AG. <https://doi.org/10.3390/info11090421>
- World Health Organization. (2020, April). *Malnutrition: Key Facts*. Who.Int.
- Yao, F., Sun, J., & Dong, J. (2022). Estimating Daily Dew Point Temperature Based on Local and Cross-Station Meteorological Data Using CatBoost Algorithm. *CMES - Computer Modeling in Engineering and Sciences*, 130(2). <https://doi.org/10.32604/cmcs.2022.018450>
- Zhang, J., Cheng, L., Yang, Z., Xiao, Q., Khan, S., Liang, R., Wu, X., & Guo, Y. (2024). An enhanced semi-supervised learning method with self-supervised and adaptive threshold for fault detection and classification in urban power grids. *Energy and AI*, 17. <https://doi.org/10.1016/j.egyai.2024.100377>
- Zhao, C. ; L. Q. ; Z. Y. (2022). Comparative evaluation of boosting algorithms for clinical data classification. *Journal of Biomedical Informatics*, 127.