

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN	iii
HALAMAN PERSEMBAHAN	iv
HALAMAN MOTTO	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG	xiv
INTISARI	xvi
ABSTRACT	xvii
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	4
1.4 Tinjauan Pustaka	5
1.5 Metode Penelitian	10
1.6 Sistematika Penulisan	11
II DASAR TEORI	12
2.1 Machine Learning	12
2.1.1 Perbedaan Pemodelan Statistik dan <i>Machine Learning</i>	13
2.2 <i>Supervised Learning</i>	14
2.2.1 Cara Kerja <i>Supervised Learning</i>	14
2.2.2 Jenis-Jenis <i>Supervised Learning</i>	15
2.3 <i>Decision Tree</i>	15
2.3.1 Terminologi dan Istilah yang terkait dengan <i>Decision Tree</i>	16
2.3.2 Kelebihan dan Kelemahan <i>Decision Tree</i>	17
2.4 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	18
2.5 <i>Hyperparameter Tuning dengan Grid Search</i>	20
2.6 k-Fold Cross Validation	20
2.7 <i>Gradient Boosting</i>	22
2.7.1 Komponen Utama <i>Gradient Boosting</i>	23

2.7.2	Algoritma <i>Gradient Boosting</i>	24
2.8	Evaluasi Model Klasifikasi	26
2.9	ROC Curves	29
2.10	Stunting	30
2.10.1	Dampak Stunting pada Anak	30
2.10.2	Indikator Stunting	31
III	EXTREME GRADIENT BOOSTING DAN LIGHT GRADIENT BOOSTING MACHINE	35
3.1	<i>Extreme Gradient Boosting</i>	35
3.1.1	Algoritma XGBoost	40
3.2	Light Gradient Boosting Machine (LightGBM)	41
3.2.1	Algoritma LightGBM	43
IV	STUDI KASUS	46
4.1	Deskripsi Data	46
4.2	<i>Preprocessing Data</i>	50
4.2.1	<i>Data Cleaning</i>	50
4.2.2	<i>Label Encoding</i>	50
4.2.3	<i>Handling Imbalance Class</i>	51
4.2.4	<i>Splitting Data</i>	55
4.3	Implementasi Model <i>Extreme Gradient Boosting</i> (XGBoost)	56
4.4	Implementasi Model Light Gradient Boosting Machine (LightGBM)	68
4.5	Hasil Uji Coba dan Evaluasi Model	73
4.5.1	Hasil Uji Coba dan Evaluasi pada Model XGBoost	73
4.5.2	Hasil Uji Coba dan Evaluasi pada Model LightGBM	78
4.5.3	Perbandingan Evaluasi Hasil Penerapan pada Model XGBoost dan LightGBM	83
4.6	ROC dan AUC	88
4.7	Kesesuaian Data Stunting Nusa Tenggara Timur terhadap Peraturan Menteri Kesehatan Republik Indonesia dan Pedoman WHO	93
V	PENUTUP	97
5.1	Kesimpulan	97
5.2	Saran	98
	DAFTAR PUSTAKA	99
A	SKRIP PROGRAM R	103
1.1	Summary dan Preprocessing Data	103
1.2	<i>Hyperparameter</i> dan Evaluasi Model XGBoost	104
1.3	<i>Hyperparameter</i> dan Evaluasi Model LightGBM	109

1.4	ROC dan AUC	114
1.5	Kesesuaian Data Stunting Nusa Tenggara Timur terhadap Peraturan Menteri Kesehatan Republik Indonesia dan Pedoman WHO	118
B	Output	121
2.1	Summary dan Preprocessing Data	121
2.2	Hyperparameter dan Evaluasi Model XGBoost	123
2.3	Hyperparameter dan Evaluasi Model LightGBM	124
2.4	ROC dan Overfitting Check	124
2.5	Kesesuaian Data Stunting Nusa Tenggara Timur terhadap Peraturan Menteri Kesehatan Republik Indonesia dan Pedoman WHO	125
C	DATA STUNTING BALITA DI NUSA TENGGARA TIMUR TAHUN 2024	127

DAFTAR TABEL

2.1	Pemodelan Statistika dan <i>Machine Learning</i>	13
2.2	<i>Confusion matrix</i> untuk mengevaluasi model klasifikasi	29
2.3	Tabel Indikator Stunting	33
4.1	Definisi Variabel Penelitian	47
4.2	Distribusi Data Stunting per Kabupaten di Provinsi NTT	48
4.3	Jumlah Missing Value pada Setiap Variabel	50
4.4	Perbandingan Data Sebelum dan Setelah Label Encoding	51
4.5	Contoh Data Sebelum dilakukan SMOTE	53
4.6	Contoh Data Setelah dilakukan SMOTE	54
4.7	Distribusi Kelas Sebelum dan Sesudah SMOTE	54
4.8	Hasil Pembagian Data <i>Training</i> dan <i>Testing</i>	55
4.9	Sampel Data Fitur Usia dan Stunting	57
4.10	Gradien dan Hessian	59
4.11	Parameter Model XGBoost.	65
4.12	Parameter Klasifikasi XGBoost	65
4.13	Hasil Hyperparameter Tuning Model XGBoost pada Proporsi Data	
	80:20	66
4.14	Hasil Hyperparameter Tuning Model XGBoost pada Proporsi Data	
	70:30	67
4.15	Hasil Hyperparameter Tuning Model XGBoost pada Proporsi Data	
	60:40	67
4.16	Parameter Model LightGBM.	69
4.17	Parameter Klasifikasi LightGBM	70
4.18	Hasil Hyperparameter Tuning Model LightGBM pada Proporsi Da-	
	ta 80:20	71
4.19	Hasil Hyperparameter Tuning Model LightGBM pada Proporsi Da-	
	ta 70:30	71
4.20	Hasil Hyperparameter Tuning Model LightGBM pada Proporsi Da-	
	ta 60:40	72
4.21	Hasil Uji Coba Klasifikasi Model XGBoost	77
4.22	Hasil Uji Coba Klasifikasi Model LightGBM	82
4.23	Perbandingan Evaluasi Hasil Uji Coba Klasifikasi pada Rasio Data	
	80:20	84

4.24 Perbandingan Evaluasi Hasil Uji Coba Klasifikasi pada Rasio Data	
70:30	86
4.25 Perbandingan Evaluasi Hasil Uji Coba Klasifikasi pada Rasio Data	
70:30	87
4.26 Contoh Dataset dengan Kolom <i>Stunting_WHO</i>	93
4.27 Distribusi Kesesuaian Data Stunting	94
4.28 Contoh Data yang Tidak Sesuai	94
4.29 Tabel Perbandingan Data Stunting Dinas Kesehatan dan WHO	95
3.1 Data Stunting	127
3.2 Data Stunting	128
3.3 Data Stunting	129

DAFTAR GAMBAR

2.1	Intersection of computer science, engineering, and statistics.	12
2.2	Cara Kerja <i>Supervised Learning</i> .	14
2.3	Ilustrasi SMOTE	19
2.4	Ilustrasi <i>Oversampling</i> Menggunakan Teknik SMOTE	19
2.5	Metode <i>k-fold cross validation</i> dengan himpunan data D	21
2.6	ROC <i>curved</i>	29
3.1	Ilustrasi Model <i>Extreme Gradient Boosting</i> .	35
3.2	Konstruksi <i>Level-Wise Tree</i>	43
3.3	Konstruksi <i>Leaf-Wise Tree</i>	43
4.1	Grafik Distribusi Sebaran Stunting di Provinsi Nusa Tenggara Timur	49
4.2	Grafik Distribusi Kelas pada Dataset Sebelum SMOTE	52
4.3	Grafik Distribusi Kelas pada Dataset Setelah SMOTE	55
4.4	Grafik Distribusi Kelas pada Dataset Sebelum SMOTE	60
4.5	Grafik Distribusi Kelas pada Dataset Sebelum SMOTE	61
4.6	<i>Confusion Matrix</i> Klasifikasi Model XGBoost pada Rasio Data 80:20	74
4.7	<i>Confusion Matrix</i> Klasifikasi Model XGBoost pada Rasio Data 70:30	75
4.8	<i>Confusion Matrix</i> Klasifikasi Model XGBoost pada Rasio Data 60:40	76
4.9	<i>Confusion Matrix</i> Klasifikasi Model LightGBM pada Rasio Data 80:20	79
4.10	<i>Confusion Matrix</i> Klasifikasi Model LightGBM pada Rasio Data 70:30	80
4.11	<i>Confusion Matrix</i> Klasifikasi Model LightGBM pada Proporsi Data 60:40	81
4.12	ROC Curve untuk Model XGBoost pada Rasio Data 80:20	89
4.13	ROC Curve untuk Model XGBoost pada Rasio Data 70:30	89
4.14	ROC Curve untuk Model XGBoost pada Rasio Data 60:40	90
4.15	ROC Curve untuk Model LightGBM pada Rasio Data 80:20	90
4.16	ROC Curve untuk Model LightGBM pada Rasio Data 70:30	91
4.17	ROC Curve untuk Model LightGBM pada Rasio Data 60:40	91

DAFTAR LAMBANG

\arg	: Argumen yang meminimalkan suatu fungsi
AUC	: Area Under the Curve (luas di bawah kurva ROC)
$CV_{(k)}$: Nilai rata-rata <i>cross-validation</i>
FN	: <i>False Negative</i>
FP	: <i>False Positive</i>
$f_0(x)$: Model awal sebelum pelatihan
$f_t(x_i)$: Model pada iterasi ke- t untuk data x_i
$f_M(x)$: Model akhir setelah M iterasi
γ	: Parameter regulasi tambahan
g_i	: Gradien pada iterasi ke- i
h_i	: Hessian pada iterasi ke- i
I	: Himpunan sampel pada suatu simpul
I_L	: Himpunan sampel di sisi kiri pohon
I_R	: Himpunan sampel di sisi kanan pohon
k	: Jumlah lipatan dalam <i>cross-validation</i>
$l(y_i, \hat{y}_i)$: Fungsi <i>log-loss</i>
λ	: Koefisien regulasi
$\mathcal{L}(\Phi)$: Fungsi objektif pada parameter Φ
MSE_i	: Mean Squared Error pada iterasi ke- i
$\Omega(f_k)$: Fungsi regulasi
R_{jm}	: <i>Terminal region</i> dari pohon ke- m
S_L	: Similarity score untuk sisi kiri
S_R	: Similarity score untuk sisi kanan
\mathbf{x}_{syn}	: Vektor fitur sampel sintetis
\mathbf{x}_{kn}	: Sampel tetangga terdekat
η	: <i>Learning rate</i>
$\ w\ ^2$: Norma kuadrat dari bobot parameter

- $\hat{f}(x)$: Prediksi akhir dari model
- $error_i$: Residual atau kesalahan pada iterasi ke- i
- TP : *True Positive*
- TN : *True Negative*
- w_j^* : Bobot optimal
- w : Bobot parameter dalam model
- y_i : Nilai aktual pada data observasi ke- i
- \hat{y}_i : Nilai prediksi pada data observasi ke- i