

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	6
1.3 Tujuan.....	6
1.4 Manfaat	6
1.5 Ruang Lingkup	7
1.6 Sistematika Penulisan.....	7
BAB II LANDASAN TEORI.....	9
2.1 <i>State of The Art</i>	9
2.2 Varian Genetik Manusia.....	10
2.3 <i>ClinVar Dataset</i>	14
2.4 Pengertian Klasifikasi.....	17
2.5 <i>Gradient Boosting Classifier</i>	17
2.6 <i>Decision Tree Regression</i>	26
2.7 <i>Spearman Correlation Coefficient</i>	28
2.8 Metode Imputasi	29
2.8.1 <i>Simple Imputer</i>	30
2.8.2 <i>K-Nearest Neighbor Imputer</i>	30
2.9 <i>Interquartile Range</i>	31
2.10 <i>Winsorization</i>	34
2.11 Metode <i>Encoding Data</i>	35
2.11.1 <i>Target Encoding</i>	35

2.11.2 <i>Binary Encoding</i>	38
2.11.3 <i>One-Hot Encoding</i>	38
2.11.4 <i>Ordinal Encoding</i>	39
2.12 <i>Robust Scaling</i>	40
2.13 <i>Pembobotan Sampel</i>	41
2.14 <i>Cross-Validation</i>	42
2.14.1 <i>K-Fold Cross-Validation</i>	43
2.14.2 <i>Stratified K-Fold Cross-Validation</i>	44
2.15 <i>Grid Search</i>	45
2.16 <i>Evaluasi Model</i>	46
2.17 <i>Tools dan Library</i>	49
2.17.1 <i>Bahasa Pemrograman Python</i>	49
2.17.2 <i>NumPy</i>	49
2.17.3 <i>Pandas</i>	50
2.17.4 <i>Scikit-learn</i>	50
BAB III METODOLOGI PENELITIAN	51
3.1 <i>Garis Besar Penyelesaian Masalah</i>	51
3.2 <i>Pengumpulan Data</i>	51
3.3 <i>Data Preprocessing</i>	51
3.3.1 <i>Validasi Tipe Data Fitur dan Penanganan Fitur Numerik Redundant</i>	52
3.3.2 <i>Data Splitting</i>	58
3.3.3 <i>Data Imputation</i>	59
3.3.4 <i>Outlier Handling</i>	63
3.3.5 <i>Data Encoding</i>	66
3.3.6 <i>Data Scaling</i>	68
3.4 <i>Optimasi Hyperparameter</i>	70
3.5 <i>Pembobotan Sampel</i>	71
3.6 <i>Pembangunan Model</i>	71
3.7 <i>Pengujian Model</i>	79
3.8 <i>Evaluasi Model</i>	81
BAB IV HASIL DAN PEMBAHASAN	84
4.1 <i>Lingkungan dan Perangkat yang Digunakan untuk Penelitian</i>	84
4.2 <i>Skenario Pelatihan dan Pengujian Model Gradient Boosting Classifier</i>	84

4.2.1 Skenario Pengujian 1: Penentuan <i>Hyperparameter</i> Terbaik untuk <i>Gradient Boosting Classifier</i>	84
4.2.2 Skenario Pengujian 2: Penerapan <i>Feature Importance</i> dari Model Terbaik.....	85
4.3 Hasil dan Analisa Skenario Pelatihan dan Pengujian Model <i>Gradient Boosting Classifier</i>	85
4.3.1 Hasil Pelatihan dan Pengujian Skenario 1	85
4.3.1.1 Hasil Pelatihan Tanpa Bobot Sampel.....	86
4.3.1.2 Hasil Pelatihan dengan Bobot Sampel	87
4.3.2 Hasil Pengujian Model Terbaik dengan <i>Feature Importance</i>	89
4.4 Analisa Hasil Pengujian Skenario	92
4.4.1 Pengaruh <i>N_estimators</i> Terhadap Kinerja Model.....	92
4.4.2 Pengaruh <i>Max_depth</i> Terhadap Kinerja model.....	93
4.4.3 Pengaruh <i>Subsample</i> Terhadap Kinerja Model.....	93
4.4.4 Pengaruh <i>Max_features</i> Terhadap Kinerja Model	94
4.4.5 Pengaruh <i>Sample Weight</i> Terhadap Kinerja Model.....	95
4.4.6 Pengaruh <i>Feature Importance</i> Terhadap Kinerja Model.....	96
BAB V PENUTUP	97
5.1 Kesimpulan.....	97
5.2 Saran.....	97
DAFTAR PUSTAKA.....	98
LAMPIRAN 1. Fitur dan Jenis Data dari <i>Dataset ClinVar</i>	108
LAMPIRAN 2. Hasil Pelatihan Tanpa Bobot Sampel	112
LAMPIRAN 3. Hasil Pelatihan dengan Bobot Sampel.....	114