

DAFTAR ISI

HALAMAN PENGESAHAN PEMBIMBING	iii
HALAMAN PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR	v
PERNYATAAN BEBAS PLAGIASI.....	vi
ABSTRAK.....	vii
<i>ABSTRACT</i>.....	viii
KATA PENGANTAR	ix
DAFTAR ISI	xii
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Tahapan Penelitian	4
1.7 Sistematika Penulisan	5
BAB II TINJAUAN LITERATUR	7
2.1 Kualitas Air	7
2.2 Status Mutu Air	7
2.3 Indeks Pencemaran (IP).....	8
2.3.1 Perhitungan Rasio Parameter terhadap Baku Mutu.....	8
2.3.2 Perhitungan Nilai Indeks Pencemaran.....	9
2.3.3 Klasifikasi Status Mutu Air	9
2.4 <i>Preprocessing Data</i>	9
2.4.1 <i>Data Cleaning</i>	10
2.4.2 <i>Label Encoding</i>	10
2.4.3 <i>Data Splitting</i>	10
2.4.4 Normalisasi Fitur	10
2.5 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	11
2.6 <i>Random Forest</i>	12

2.6.1	Klasifikasi <i>Random Forest</i> (<i>Majority Voting</i>).....	13
2.6.2	<i>Gini Impurity</i>	13
2.6.3	Kompleksitas Komputasi <i>Random Forest</i>	13
2.6.4	Ilustrasi <i>Random Forest</i>	14
2.7	<i>Particle Swarm Optimization</i> (PSO)	14
2.7.1	Representasi Posisi Partikel.....	15
2.7.2	Pembaruan Kecepatan dan Posisi	15
2.7.3	Fungsi <i>Fitness</i>	15
2.7.4	Seleksi Fitur dengan PSO	16
2.7.5	Konversi <i>Hyperparameter Integer</i>	16
2.7.6	Kompleksitas Komputasi <i>Random Forest</i> menggunakan PSO	16
2.7.7	Ilustrasi PSO	17
2.8	<i>Baseline Model</i>	18
2.9	<i>Stratified K-Fold Cross Validation</i>	18
2.10	Evaluasi Model.....	19
2.11	<i>State of Art</i> (SOTA).....	20
BAB III METODOLOGI PENELITIAN.....		25
3.1	Metodologi Penelitian	25
3.2	Pengumpulan Data.....	26
3.3	Penentuan Status Mutu Air.....	29
3.4	<i>Preprocessing Data</i>	32
3.4.1	Pembersihan Data (<i>Data cleaning</i>)	33
3.4.2	Konversi Label (<i>Label encoding</i>).....	35
3.4.3	Pembagian Data (<i>Data splitting</i>).....	35
3.4.4	Penanganan <i>Missing value</i>	36
3.4.5	<i>Data Normalization</i>	37
3.5	<i>Synthetic Minority Over-sampling Technique</i> (SMOTE).....	39
3.6	Pengembangan Model <i>Random Forest</i>	42
3.6.1	Model Baseline <i>Random Forest</i> dengan Parameter Default.....	43
3.6.2	<i>Particle Swarm Optimization</i> (PSO) dan <i>Stratified K-Fold Cross Validation</i>	43
3.6.3	Ilustrasi Kombinasi PSO dan <i>Random Forest</i> (Hasil Seleksi Fitur) ...	52
3.6.4	<i>Bootstrap Aggreagating</i> (<i>Bagging</i>)	53

3.6.5	Analisis Kompleksitas Komputasi	57
3.7	Evaluasi Model	58
3.8	Metodologi Pengembangan Sistem	60
3.8.1	<i>Requirements Analysis and Definition</i>	60
3.8.2	<i>System and Software Design</i>	61
3.8.3	Implementasi	64
3.8.4	Pengujian Sistem	64
BAB IV HASIL PENGUJIAN DAN PEMBAHASAN.....		65
4.1	Implementasi	65
4.1.1	Pengumpulan Data.....	65
4.1.2	Implementasi <i>Preprocessing Data</i>	65
4.1.3	Implementasi SMOTE.....	68
4.1.4	Implementasi Optimasi PSO dan Pemodelan Random Forest	68
4.1.5	Implementasi Antarmuka Pengguna.....	72
4.2	Hasil.....	77
4.2.1	Hasil <i>Preprocessing Data</i>	77
4.2.2	Hasil Optimasi PSO.....	78
4.2.3	Hasil Seleksi Fitur	80
4.2.4	Hasil Evaluasi Model	81
4.2.5	<i>Confusion Matrix</i>	82
4.2.6	Hasil Implementasi Antaramuka Pengguna	84
4.3	Pembahasan	86
BAB V PENUTUP		88
5.1	Kesimpulan.....	88
5.2	Saran	88
DAFTAR PUSTAKA		90

DAFTAR TABEL

Tabel 2.1 Klasifikasi Status Mutu Air Berdasarkan Nilai IP (PermenLHK No.27, 2021).....	9
Tabel 2.2 Contoh Confusion Matrix (Branco et al., 2015)	19
Tabel 2.3 State of The Art.....	21
Tabel 2.3 State of The Art (Lanjutan).....	22
Tabel 3.1 Deskripsi Data Parameter Kualitas Air (Ariesta et al., n.d.; Saputra et al., n.d.).....	26
Tabel 3.1 Deskripsi Data Parameter Kualitas Air Lanjutan (Ariesta et al., n.d.; Saputra et al., n.d.).....	27
Tabel 3.2 Perhitungan Rasio Dasar.....	29
Tabel 3.3 Penentuan Nilai C_i/Lij Setelah Koreksi.....	30
Tabel 3.4 Rekapitulasi Perhitungan IP pada Sampel Cemar Sedang.....	31
Tabel 3.5 Rekapitulasi Perhitungan IP pada Sampel Cemar Berat.....	32
Tabel 3.6 Identifikasi <i>Missing Values</i> pada Dataset	35
Tabel 3.7 <i>Label encoding</i>	35
Tabel 3.8 Ringkasan Proporsi Data Latih & Uji.....	36
Tabel 3.9 Nilai Minimum dan Maksimum Data Latih untuk Normalisasi	38
Tabel 3.10 Hasil Normalisasi Data Sampel 1-3.....	39
Tabel 3.11 Parameter Default Model Random Forest	43
Tabel 3.12 Set Parameter Model Random Forest	45
Tabel 3.13 Konfigurasi Parameter PSO.....	45
Tabel 3.14 Representasi Dimensi Ruang Pencarian PSO.....	45
Tabel 3.15 Tabel Data Latih Ilustrasi (3 Sampel Hasil Normalisasi).....	46
Tabel 3.16 Tabel Inisialisasi Posisi Partikel PSO ($t=0$).....	46
Tabel 3.17 Tabel kecepatan awal partikel PSO ($t=0$)	46
Tabel 3.18 Rekap Gini Impurity Fitur Aktif P1 {BOD, COD}	47
Tabel 3.19 Rekap Gini Impurity Fitur Aktif P2 {TC, FC, AM}	48
Tabel 3.20 Ringkasan Fitness dan Seleksi Fitur Seluruh Partikel ($t = 0$)	50
Tabel 3.21 Tabel Hasil Pembaruan Posisi P1 dan Keputusan Seleksi Fitur ($t=1$)..	51
Tabel 3.22 Perbandingan Konfigurasi P1 Sebelum dan Sesudah Iterasi $t = 1$	52
Tabel 3.23 Konversi Nilai $gBest$ menjadi Konfigurasi Random Forest Final.....	52
Tabel 3.24 Dataset Hasil Normalisasi Min-Max Parameter Kualitas Air	53
Tabel 3.25 Bootstrap Sample 1 (fitur terpilih: TC, FC).....	53
Tabel 3.26 Bootstrap Sample 2 (fitur terpilih: FC, AM)	53
Tabel 3.27 Kandidat Threshold dari Rata-rata Antar Dua Baris pada Bootstrap Sample 2	54
Tabel 3.28 Proses Perhitungan Skor Gini pada Threshold $FC < 0.0020$	54
Tabel 3.29 Rekap Skor Gini Impurity Seluruh Threshold Kandidat (Bootstrap Sample 2).....	55
Tabel 3.30 Evaluasi dengan <i>Confusion Matrix</i>	58

Tabel 3.31 Contoh Evaluasi dengan <i>Confusion Matrix</i>	58
Tabel 3.32 Proses Perhitungan Skor <i>Confusion Matrix</i> Kelas Cemar Ringan ...	59
Tabel 3.33 Rancangan Perbandingan Performa Pengujian pada Model.....	59
Tabel 3.34 Spesifikasi Kebutuhan Perangkat Keras	61
Tabel 3.35 Spesifikasi Kebutuhan Perangkat Lunak	61
Tabel 3.36 Skenario Pengujian Sistem	64
Tabel 4.1 Ringkasan Hasil Preprocessing Data	77
Tabel 4.2 Hasil Deteksi Missing Value	77
Tabel 4.3 Nilai Median Hasil Imputasi Data Latih.....	78
Tabel 4.4 Distribusi Kelas Sebelum dan Sesudah SMOTE.....	78
Tabel 4.5 Hasil Optimasi Hyperparameter PSO	80
Tabel 4.6 Fitur Terpilih oleh PSO (Skenario 3).....	80
Tabel 4.7 Perbandingan Performa Model pada Tiga Skenario	81
Tabel 4.8 Classification Report - Skenario 1 (Baseline).....	81
Tabel 4.9 Classification Report - Skenario 2 (PSO HP).....	81
Tabel 4.10 Classification Report - Skenario 3 (PSO HP+FS)	82
Tabel 4.11 Ringkasan Hasil Pengujian Batch (10 Sampel)	86

DAFTAR GAMBAR

Gambar 2.1 Cara Kerja Metode SMOTE (Sagaceta-Mejía et al., 2025)	12
Gambar 2.2 Alur Kerja Algoritma <i>Random Forest</i> (Ramdani et al., 2025)	14
Gambar 2.3 Ilustrasi <i>Particle Swarm Optimization</i> (Li et al., 2023)	17
Gambar 2.4 Ilustrasi <i>Stratified K-Fold Cross Validation</i> (Duan, 2023)	19
Gambar 3.1 Tahapan Penelitian	25
Gambar 3.2 Data Kualitas Air Sungai Kota Yogyakarta (Sumber: Dinas Lingkungan Hidup Kota Yogyakarta, 2023–2024)	28
Gambar 3.3 Data Kualitas Air Sungai Kota Yogyakarta Lanjutan (Sumber: Dinas Lingkungan Hidup Kota Yogyakarta, 2023–2024)	28
Gambar 3.4 <i>Flowchart Data Preprocessing</i>	33
Gambar 3.5 <i>Flowchart Data Cleaning</i>	34
Gambar 3.6 Distribusi Kelas Sebelum SMOTE	40
Gambar 3.7 Grafik kelas sebelum SMOTE	41
Gambar 3.8 Grafik Kelas setelah SMOTE	41
Gambar 3.9 <i>Flowchart Modelling Train dan Test</i>	42
Gambar 3.10 <i>Flowchart Particle Swarm Optimizaiton (PSO)</i>	44
Gambar 3.11 Struktur Decision Tree P1	48
Gambar 3.12 Struktur Decision Tree P1	49
Gambar 3.13 Percabangan pada Node $FC < 0.0020$	56
Gambar 3.14 Ilustrasi Random Forest	56
Gambar 3.15 Diagram Metode <i>Waterfall</i> (Bagoes Satria & Ardiansyah, 2023) ...	60
Gambar 3.16 Arsitektur Sistem	62
Gambar 3.17 <i>Flowchart</i> Perancangan Sistem	62
Gambar 3.18 <i>Wireframe</i> Halaman Klasifikasi Input Numerik Data Kualitas Air. 63	
Gambar 3.19 <i>Wireframe</i> Halaman Klasifikasi Input File data Kualitas Air	63
Gambar 4.1 Konvergensi PSO - Skenario 2 (<i>Hyperparameter</i>)	79
Gambar 4.2 Konvergensi PSO - Skenario 3 (<i>Hyperparameter</i> + Seleksi Fitur) ...	79
Gambar 4.3 <i>Confusion Matrix</i> - Skenario 1 (<i>Baseline</i>)	82
Gambar 4.4 <i>Confusion Matrix</i> - Skenario 2 (PSO HP)	83
Gambar 4.5 <i>Confusion Matrix</i> - Skenario 3 (PSO HP+FS)	83
Gambar 4.6 <i>Interface Tab Input</i> Data Numerik	84
Gambar 4.7 Hasil Prediksi Tab Input Data Numerik	84
Gambar 4.8 Antarmuka Perhitungan IP dan Model	85
Gambar 4.9 <i>Interface Tab Upload CSV</i>	85