

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PENGESAHAN</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN</b>	<b>iii</b>
<b>HALAMAN PERSEMBAHAN</b>	<b>iv</b>
<b>HALAMAN MOTTO</b>	<b>v</b>
<b>PRAKATA</b>	<b>vi</b>
<b>DAFTAR ISI</b>	<b>viii</b>
<b>DAFTAR TABEL</b>	<b>x</b>
<b>DAFTAR GAMBAR</b>	<b>xi</b>
<b>INTISARI</b>	<b>xii</b>
<b>ABSTRACT</b>	<b>xiii</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1. Latar Belakang	1
1.2. Pembatasan Masalah	3
1.3. Tujuan dan Manfaat Penelitian	4
1.4. Tinjauan Pustaka	4
1.5. Metodologi Penelitian	7
1.6. Sistematika Penulisan	8
<b>II LANDASAN TEORI</b>	<b>10</b>
2.1. <i>Data Mining</i>	10
2.2. <i>Machine Learning</i>	10
2.3. Analisis Klasifikasi	12
2.4. <i>Boxen Plot</i>	13
2.5. Data Tidak Seimbang	14
2.6. Evaluasi Performa Klasifikasi	15
2.7. <i>Logistic Regression</i>	18
2.8. SVM	19
2.9. <i>Decision Tree</i>	20
2.10. <i>Ensemble Learning</i>	22
2.11. <i>Cross Validation</i>	24
<b>III IMPLEMENTASI METODE OPTIMASI AMBANG BATAS PADA ANALISIS KLASIFIKASI DATA TIDAK SEIMBANG</b>	<b>26</b>
3.1. Penanganan Data Tidak Seimbang	26

3.1.1. Metode <i>Resampling</i> . . . . .	26
3.1.2. Metode <i>Cost-Sensitive Learning</i> . . . . .	27
3.1.3. Metode <i>Ensemble Learning</i> . . . . .	27
3.1.4. Metode Optimasi Ambang Batas . . . . .	27
3.2. <i>Random Forest</i> . . . . .	28
3.3. <i>Gradient Boosting</i> . . . . .	35
3.4. Optimasi Ambang Batas . . . . .	37
<b>IV STUDI KASUS</b> . . . . .	<b>40</b>
4.1. Deskripsi Data . . . . .	40
4.2. Eksplorasi dan Visualisasi Data . . . . .	41
4.3. Analisis Klasifikasi dengan <i>Logistic Regression</i> . . . . .	44
4.4. Analisis Klasifikasi dengan SVM . . . . .	48
4.5. Analisis Klasifikasi dengan <i>Random Forest</i> . . . . .	52
4.6. Analisis Klasifikasi dengan <i>Gradient Boosting</i> . . . . .	57
4.7. Perbandingan Performa Klasifikasi . . . . .	61
<b>V PENUTUP</b> . . . . .	<b>66</b>
5.1. Kesimpulan . . . . .	66
5.2. Saran . . . . .	68
<b>DAFTAR PUSTAKA</b> . . . . .	<b>69</b>
<b>A Data</b> . . . . .	<b>72</b>
<b>B Syntax Python</b> . . . . .	<b>73</b>

## DAFTAR TABEL

2.1	<i>Confusion Matrix</i> untuk Klasifikasi Biner . . . . .	15
3.1	Data Contoh . . . . .	29
3.2	Data Hasil <i>Bootstrap</i> . . . . .	30
4.1	Variabel Data <i>Drug Consumption (Quantified)</i> . . . . .	42
4.2	<i>Confusion Matrix</i> untuk <i>Logistic Regression</i> . . . . .	46
4.3	Evaluasi Model <i>Logistic Regression</i> . . . . .	47
4.4	<i>Confusion Matrix</i> untuk SVM . . . . .	50
4.5	Evaluasi Model SVM . . . . .	51
4.6	<i>Confusion Matrix</i> untuk <i>Random Forest</i> . . . . .	54
4.7	Evaluasi Model <i>Random Forest</i> . . . . .	55
4.8	<i>Confusion Matrix</i> untuk <i>Gradient Boosting</i> . . . . .	58
4.9	Evaluasi Model <i>Gradient Boosting</i> . . . . .	59

## DAFTAR GAMBAR

2.1	Perbandingan <i>Box Plot</i> dengan <i>Boxen Plot</i> . . . . .	14
2.2	Struktur <i>Decision Tree</i> . . . . .	21
3.1	<i>Decision Tree</i> 1 . . . . .	31
3.2	<i>Decision Tree</i> 2 . . . . .	33
3.3	<i>Decision Tree</i> 3 . . . . .	35
4.1	<i>Bar Plot</i> Variabel <i>heroin</i> . . . . .	41
4.2	<i>Boxen Plot</i> Prediktor Numerik . . . . .	43
4.3	<i>Normalized Bar Plot</i> Prediktor Kategorik . . . . .	43
4.4	<i>Bar Plot</i> Prediktor Kategorik . . . . .	44
4.5	Kurva ROC & PR pada <i>Logistic Regression</i> . . . . .	45
4.6	Kurva ROC & PR pada SVM . . . . .	49
4.7	Kurva ROC & PR pada <i>Random Forest</i> . . . . .	53
4.8	Kurva ROC & PR pada <i>Gradient Boosting</i> . . . . .	57
4.9	Perbandingan Metrik Evaluasi ROC AUC & PR AUC . . . . .	61
4.10	Perbandingan Metrik Evaluasi <i>Accuracy</i> . . . . .	62
4.11	Perbandingan Metrik Evaluasi <i>Recall</i> . . . . .	63
4.12	Perbandingan Metrik Evaluasi <i>F-measure</i> . . . . .	64
4.13	Perbandingan Metrik Evaluasi <i>G-mean</i> . . . . .	65