

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIASI	iii
HALAMAN PERSEMBAHAN	iv
HALAMAN MOTTO	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMBANG	xiii
INTISARI	xiv
ABSTRACT	xv
I PENDAHULUAN	1
1.1. Latar Belakang dan Perumusan Masalah	1
1.2. Tujuan dan Manfaat Penelitian	3
1.3. Tinjauan Pustaka	3
1.4. Metodologi Penelitian	4
1.5. Sistematika Penulisan	4
1.5.1. BAB I Pendahuluan	4
1.5.2. BAB II Landasan Teori	5
1.5.3. BAB III Metodologi Penelitian	5
1.5.4. BAB IV Studi Kasus	5
1.5.5. BAB V Penutup	5
II DASAR TEORI	6
2.1. Asuransi	6
2.2. Asuransi Jiwa	7
2.3. Polis	9
2.4. Statistika Deskriptif	9
2.5. Variabel Random	10
2.5.1. Variabel Random Diskrit	10
2.5.2. Variabel Random Kontinu	11
2.6. Vektor	11
2.7. Probabilitas	12

2.7.1. Ekspektasi	12
2.8. Distribusi Bernoulli	13
2.9. Distribusi Binomial	13
2.10. Maximum Likelihood Estimation (MLE)	14
2.11. Regresi Logistik	15
2.11.1. Pembentukan Model Regresi Logistik Biner	15
2.11.2. Estimasi Parameter dengan Maximum Likelihood Estimation	18
2.12. <i>Machine Learning</i>	20
2.13. <i>Decision Tree</i>	23
2.14. <i>One-Hot Encoding</i>	25
2.15. <i>Underfitting</i> dan <i>Overfitting</i>	25
2.16. Standardisasi	26
2.17. Penanganan Ketidakseimbangan Kelas dengan <i>Resampling</i>	26
III METODE XGBOOST	28
3.1. <i>Ensemble Learning</i>	28
3.2. <i>Boosting</i>	31
3.3. <i>Gradient Boosting</i>	32
3.4. <i>Extreme Gradient Boosting</i>	34
3.4.1. Learning Objective	34
3.4.2. <i>Gradient Tree Boosting</i>	35
3.4.3. Shrinkage dan Column Sampling	38
3.4.4. Split Finding Algorithm	39
3.4.5. Hyperparameter	39
3.5. <i>Confusion Matrix</i>	42
3.6. ROC-AUC	43
IV STUDI KASUS	45
4.1. Deskripsi Data	45
4.1.1. Statistik Deskriptif Umur Pemegang Polis saat <i>Effective Date</i>	47
4.1.2. Statistik Deskriptif Premi	48
4.1.3. Statistik Deskriptif Uang Pertanggungan	49
4.1.4. Statistik Deskriptif Premi Tambahan	50
4.1.5. Statistik Deskriptif Durasi Polis	51
4.1.6. Statistik Deskriptif Sisa Lama Pembayaran Premi	52
4.1.7. Distribusi Jumlah Polis Berdasarkan Kebiasaan Merokok	52
4.1.8. Distribusi Jumlah Polis Berdasarkan <i>Billing Channel</i>	53
4.1.9. Distribusi Jumlah Polis Berdasarkan <i>Premium Term</i>	53

4.1.10. Distribusi Jumlah Polis Berdasarkan <i>Billing Frequency</i> . . .	54
4.2. Pra-pemrosesan Data	54
4.2.1. <i>Missing Data</i>	55
4.2.2. <i>One-Hot Encoding</i>	55
4.2.3. Partisi Data	57
4.2.4. <i>Random Sampling</i>	57
4.2.5. Transformasi	58
4.3. Pemodelan dengan Regresi Logistik	58
4.3.1. <i>Saturated Model</i> Regresi Logistik	58
4.3.2. <i>Fitted Model</i> Regresi Logistik	62
4.3.3. Perbandingan <i>Saturated Model</i> dengan <i>Fitted Model</i> Regresi Logistik	65
4.4. Pemodelan dengan XGBoost	67
4.4.1. Model XGBoost dengan Menggunakan Seluruh Variabel . .	68
4.4.2. Model XGBoost dengan <i>Feature Selection</i>	70
4.5. Perbandingan Model Regresi Logistik dan XGBoost	73
4.5.1. Penghitungan <i>Lapse rate</i>	74
4.5.2. <i>Feature Importance</i>	75
V PENUTUP	78
5.1. Kesimpulan	78
5.2. Saran	78
A Syntax Regresi Logistik dan XGBoost	84