



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
HALAMAN PERSEMBAHAN	iv
HALAMAN MOTTO	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMBANG	xv
INTISARI	xvi
ABSTRACT	xvii
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	5
1.5. Sistematika Penulisan	6
II DASAR TEORI	8
2.1. Asuransi	8
2.2. Risiko	8
2.3. Polis	9
2.4. Variabel Random	9
2.4.1. Variabel Random Diskrit	10
2.4.2. Variabel Random Kontinu	10
2.5. Peluang	11
2.6. <i>Machine Learning</i>	11
2.7. <i>Baseline Model</i>	13
2.8. Regresi Logistik	14
2.8.1. Model Logistik	14
2.8.2. Estimasi Parameter	16
2.8.3. Pemilihan Model Terbaik	17
2.9. <i>Decision Tree</i>	18



2.9.1. <i>Entropy</i>	20
2.9.2. <i>Gini Index</i>	21
2.10. <i>Variance Inflation Factor (VIF)</i>	21
2.11. <i>Encoding</i>	21
2.12. <i>Overfitting dan Underfitting</i>	22
III MODEL CATBOOST	23
3.1. <i>Ensemble Learning</i>	23
3.2. <i>Boosting</i>	25
3.3. <i>Gradient Boosting</i>	27
3.4. <i>Categorical Boosting</i>	30
3.4.1. <i>Learning Objective</i>	30
3.4.2. <i>Penanganan Fitur Kategorikal</i>	32
3.4.3. <i>Ordered Boosting</i>	33
3.4.4. <i>Oblivious Decision Tree</i>	35
3.4.5. <i>Hyperparameter</i>	37
3.5. <i>Metrik Evaluasi</i>	39
IV STUDI KASUS	43
4.1. <i>Deskripsi Data</i>	43
4.2. <i>Eksplorasi dan Pra-pemrosesan Data</i>	48
4.2.1. <i>Data Cleaning</i>	48
4.2.2. <i>Feature Engineering</i>	50
4.2.3. <i>Partisi Data</i>	56
4.2.4. <i>Resampling</i>	56
4.2.5. <i>Transformasi Data</i>	57
4.3. <i>Modeling</i>	58
4.4. <i>Klasifikasi dengan Regresi Logistik Data eudirectlapse</i>	60
4.4.1. <i>Data Asli</i>	60
4.4.2. <i>Data dengan Oversampling</i>	61
4.4.3. <i>Data dengan Undersampling</i>	62
4.4.4. <i>Perbandingan Model Regresi Logistik</i>	63
4.5. <i>Klasifikasi dengan CatBoost Data eudirectlapse</i>	64
4.5.1. <i>Data Asli</i>	64
4.5.2. <i>Data dengan Oversampling</i>	66
4.5.3. <i>Data dengan Undersampling</i>	68
4.5.4. <i>Perbandingan Model CatBoost</i>	70
4.6. <i>Klasifikasi dengan Regresi Logistik Data Home Insurance</i>	71



4.6.1.	Data Asli	71
4.6.2.	Data dengan <i>Oversampling</i>	72
4.6.3.	Data dengan <i>Undersampling</i>	73
4.6.4.	Perbandingan Model Regresi Logistik	75
4.7.	Klasifikasi dengan CatBoost Data <i>Home Insurance</i>	76
4.7.1.	Data Asli	76
4.7.2.	Data dengan <i>Oversampling</i>	78
4.7.3.	Data dengan <i>Undersampling</i>	80
4.7.4.	Perbandingan Model CatBoost	82
4.8.	Perbandingan Metrik Evaluasi Model Regresi Logistik dan CatBoost	83
4.8.1.	Data <i>eudirectlapse</i>	83
4.8.2.	Data <i>Home Insurance</i>	84
4.9.	<i>Risk Scoring</i>	85
4.9.1.	Data <i>eudirectlapse</i>	86
4.9.2.	Data <i>Home Insurance</i>	87
V	PENUTUP	88
5.1.	Kesimpulan	88
5.2.	Saran	90
	DAFTAR PUSTAKA	91
A	Kode Python Data <i>eudirectlapse</i>	95
B	Kode Python Data <i>Home Insurance</i>	116