

DAFTAR ISI

HALAMAN JUDUL	i
ABSTRAK.....	ii
<i>ABSTRACT</i>	iii
LEMBAR PENGESAHAN	iv
LEMBAR PERNYATAAN ORISINALITAS	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR SIMBOL	xv
DAFTAR SINGKATAN	xvii
DAFTAR LAMPIRAN	xix
Bab I PENDAHULUAN.....	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah	5
I.3 Tujuan Tugas Akhir	6
I.4 Manfaat Tugas Akhir	6
I.5 Batasan dan Asumsi Tugas Akhir	7
I.6 Sistematika Laporan	7
Bab II LANDASAN TEORI	10
II.1 Literatur.....	10
II.1.1 Paru-Paru.....	10
II.1.2 Tuberkulosis Paru	10
II.1.3 Data Rekam Medis.....	12

II.1.4	<i>Data Mining</i>	13
II.1.5	<i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i> ...	14
II.1.5.1	<i>Business Understanding Phase</i>	15
II.1.5.2	<i>Data Understanding Phase</i>	15
II.1.5.3	<i>Data Preparation Phase</i>	15
II.1.5.4	<i>Modeling Phase</i>	17
II.1.5.5	<i>Evaluation Phase</i>	18
II.1.5.6	<i>Deployment Phase</i>	18
II.1.6	<i>Machine learning</i>	18
II.1.7	<i>Extreme Gradient Boosting (XGBoost)</i>	19
II.1.8	<i>Randomized search</i>	22
II.1.9	<i>Confusion Matrix</i>	22
II.1.9.1	<i>Accuracy</i>	23
II.1.9.2	<i>Precision</i>	23
II.1.9.3	<i>Recall</i>	24
II.1.9.4	<i>F1-score</i>	24
II.1.9.5	<i>Kurva ROC-AUC</i>	24
II.1.10	<i>Cross Entropy Loss (Log-Loss)</i>	25
II.2	<i>Pemilihan Metode</i>	27
II.3	<i>Penelitian Terdahulu</i>	28
Bab III	METODE PENYELESAIAN MASALAH	33
III.1	<i>Metode Penelitian</i>	33
III.2	<i>Sistematika Penyelesaian Masalah</i>	34
III.2.1	<i>Business Understanding</i>	34
III.2.2	<i>Data Understanding</i>	35
III.2.3	<i>Data Preparation</i>	38

III.2.4	<i>Modeling</i>	38
III.2.5	<i>Evaluation</i>	39
III.2.6	<i>Feature Importance</i>	40
III.2.7	<i>Deployment</i>	40
Bab IV	PENYELESAIAN PERMASALAHAN	41
IV.1	<i>Business Understanding</i>	41
IV.2	<i>Data Understanding</i>	41
IV.2.1	Pemilihan dan Eksplorasi Data	43
IV.2.2	Karakteristik Data	46
IV.3	<i>Data Preparation</i>	50
IV.3.1	<i>Feature Selection</i>	51
IV.3.2	<i>Data Cleaning</i>	52
IV.3.2.1	<i>Type Correction</i>	52
IV.3.2.2	<i>Handling Missing Values</i>	54
IV.3.2.3	<i>Whitespace Trimming</i>	56
IV.3.3	<i>Data Transformation</i>	57
IV.3.4	<i>Splitting Data</i>	57
IV.3.5	<i>Post-Split Data Transformation</i>	59
IV.3.5.1	<i>Handling Nilai Outlier</i>	59
IV.3.5.2	<i>Standardization</i>	61
IV.4	<i>Modeling</i>	62
IV.5	<i>Evaluation</i>	65
IV.5.1	<i>Confusion Matrix</i>	65
IV.5.2	<i>Accuracy and Loss Validation Curve</i>	66
IV.6	<i>Feature Importance</i>	67
IV.7	Proses Implementasi Model	67

Bab V	VALIDASI, ANALISIS HASIL, DAN IMPLIKASI	69
V.1	Hasil dan Evaluasi Klasifikasi Model	69
V.1.1	<i>Hyperparameter</i> Terbaik Model XGBoost.....	69
V.1.1.1	Rasio 70:30	69
V.1.1.2	Rasio 80:20	70
V.1.1.3	Rasio 90:10	71
V.1.2	<i>Confusion Matrix</i>	71
V.1.2.1	<i>Confusion Matrix</i> Pada Rasio 70:30	72
V.1.2.2	<i>Confusion Matrix</i> Pada Rasio 80:20	73
V.1.2.3	<i>Confusion Matrix</i> Pada Rasio 90:10	75
V.1.3	ROC-AUC.....	77
V.1.3.1	ROC-AUC Pada Rasio 70:30	77
V.1.3.2	ROC-AUC Pada Rasio 80:20	78
V.1.3.3	ROC-AUC Pada Rasio 90:10	79
V.1.4	<i>Accuracy and Loss Validation Curve</i>	80
V.1.4.1	<i>Accuracy and Loss Validation Curve</i> Rasio 70:30.....	81
V.1.4.2	<i>Accuracy and Loss Validation Curve</i> Rasio 80:20.....	81
V.1.4.3	<i>Accuracy and Loss Validation Curve</i> Rasio 90:10.....	82
V.2	Pemilihan Model Terbaik.....	84
V.3	<i>Analisis Feature Importance</i>	85
V.4	Hasil <i>Deployment</i> dan Validasi Model.....	87
V.5	Implikasi Penelitian Tugas Akhir.....	93
Bab VI	KESIMPULAN DAN SARAN	96
VI.1	Kesimpulan.....	96
VI.2	Saran	97
	DAFTAR PUSTAKA.....	98

LAMPIRAN	106
----------------	-----