

DAFTAR ISI

LEMBAR PERSETUJUAN

LEMBAR PENGESAHAN

LEMBAR PERNYATAAN KARYA SENDIRI

LEMBAR PERSEMBAHAN

ABSTRAK i

ABSTARCT..... ii

KATA PENGANTAR..... iii

DAFTAR ISI..... v

DAFTAR GAMBAR..... viii

DAFTAR TABEL ix

DAFTAR RUMUS x

DAFTAR LAMPIRAN xi

BAB I PENDAHULUAN..... 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 6

1.3 Tujuan Penelitian 6

1.4 Manfaat Penelitian 6

1.5 Batasan Masalah..... 7

1.6 Kerangka Pemikiran..... 7

1.7 Sistematika Penulisan 10

BAB II KAJIAN LITERATUR 11

2.1 Tinjauan Pustaka 11

2.2 *Data Mining* 15

2.3 *CRISP-DM* 16

2.3.1 *Business Understanding*..... 17

2.3.2 *Data Understanding*..... 18

2.3.3 *Data Preparation* 19

2.3.4 *Modeling* 19

2.3.5 *Evaluation* 20

2.3.6 Deployment	21
2.4 Machine Learning	21
2.5 Decision Tree	23
2.6 Algoritma C5.0.....	23
2.7 Pseudo-Labeling	25
2.8 Algoritma K-Means	25
2.9 SMOTE.....	26
2.10 Feature Selection	27
2.11 GridSearchCV.....	28
2.12 K-Fold Cross Validation	28
2.13 Explainable AI SHAP	29
2.14 Confusion Matrix	31
2.15 Python	32
BAB III METODOLOGI PENELITIAN	35
3.1 Business Understanding.....	36
3.1.1 Identifikasi Masalah	36
3.1.2 Merancang Solusi dan Merumuskan Tujuan.....	36
3.1.3 Kebutuhan Perangkat	37
3.2 Data Understanding.....	37
3.2.1 Pengumpulan Data	38
3.2.2 Eksplorasi Data	38
3.3 Data Preparation	38
3.3.1 Pembersihan dan Transformasi Data.....	38
3.3.2 Pembagian Data	38
3.3.3 Pseudo-Labeling (K-Means Clustering)	39
3.3.4 Penyeimbangan Data (SMOTE)	40
3.3.5 Seleksi Fitur (Feature Selection).....	40
3.4 Modeling	40
3.4.1 Optimasi Parameter (GridSearchCV& K-Fold Cross Validation) .	41
3.4.2 Decision Tree C5.0	41
3.5 Evaluation	42
3.5.1 Confusion Matrix	42
3.5.2 Explainable AI SHAP	42

BAB IV HASIL DAN PEMBAHASAN	43
4.1 Hasil <i>Business Understanding</i>	43
4.1.1 Hasil Identifikasi Masalah.....	43
4.1.2 Hasil Merancang Solusi dan Merumuskan Tujuan	43
4.1.3 Hasil Kebutuhan Perangkat.....	44
4.2 Hasil <i>Data Understanding</i>	45
4.2.1 Hasil Pengumpulan Data.....	45
4.2.2 Hasil Eksplorasi Data.....	46
4.3 Hasil <i>Data Preparation</i>	47
4.3.1 Hasil Pembersihan dan Transformasi Data	47
4.3.2 Hasil Pembagian Data	49
4.3.3 Hasil <i>Pseudo-Labeling (K-Means Clustering)</i>	50
4.3.4 Hasil Penyeimbangan Data (<i>SMOTE</i>).....	54
4.3.5 Hasil Seleksi Fitur (<i>Feature Selection</i>)	55
4.4 Hasil <i>Modeling</i>	58
4.4.1 Hasil Optimasi Parameter (<i>GridSearchCV& K-Fold Cross Validation</i>).....	58
4.4.2 Hasil <i>Decision Tree C5.0</i>	60
4.5 Hasil <i>Evaluation</i>	63
4.5.1 Hasil <i>Confusion Matrix</i>	63
4.5.2 Hasil <i>Explainable AI SHAP</i>	67
BAB V PENUTUP.....	72
5.1 Kesimpulan	72
5.2 Saran.....	72
DAFTAR PUSTAKA	74
LAMPIRAN.....	80
DAFTAR RIWAYAT HIDUP	94

DAFTAR GAMBAR

Gambar 1. 1 Persentase Status Kelulusan Mahasiswa	2
Gambar 1. 2 Kerangka Pemikiran	8
Gambar 2. 1 Ilustrasi <i>Data Mining</i> [32]	16
Gambar 2. 2 Siklus <i>CRISP-DM</i> [33]	17
Gambar 2. 3 <i>Business Understanding Phase</i> [34].....	17
Gambar 2. 4 <i>Data Understanding Phase</i> [34].....	18
Gambar 2. 5 <i>Data Preparation Phase</i> [34]	19
Gambar 2. 6 <i>Modeling Phase</i> [34].....	20
Gambar 2. 7 <i>Evaluation Phase</i> [34]	20
Gambar 2. 8 <i>Deployment Phase</i> [34].....	21
Gambar 2. 9 <i>Confusion Matrix</i> [35]	31
Gambar 3. 1 Langkah-langkah metode <i>CRISP-DM</i>	35
Gambar 3. 2 Alur Kerja <i>Pseudo-labeling</i> Algoritma <i>K-Means</i>	39
Gambar 4. 1 Hasil Pemeriksaan Nilai Kosong.....	49
Gambar 4. 2 Visualisasi Pengelompokan <i>K-Means</i>	53
Gambar 4. 3 Parameter Terbaik.....	59
Gambar 4. 4 Hasil Pengujian Rasio 60:40.....	64
Gambar 4. 5 Evaluasi Hasil Pengujian Rasio 60:40.....	64
Gambar 4. 6 Hasil Pengujian Rasio 70:30.....	65
Gambar 4. 7 Evaluasi Hasil Pengujian Rasio 70:30.....	65
Gambar 4. 8 Hasil Pengujian Rasio 80:20.....	66
Gambar 4. 9 Evaluasi Hasil Pengujian Rasio 80:20.....	67
Gambar 4. 10 Visualisasi <i>SHAP</i> Skenario 80:20.....	70

DAFTAR TABEL

Tabel 1. 1 Rekapitulasi Status Kelulusan Mahasiswa Teknik Informatika Angkatan 2017-2022	1
Tabel 2. 1 <i>State of The Art</i>	11
Tabel 3. 1 Kebutuhan Perangkat.....	37
Tabel 3. 2 Pembagian Data.....	39
Tabel 4. 1 Kebutuhan Perangkat.....	44
Tabel 4. 2 <i>Dataset</i> Sekunder Hasil Kuesioner.....	45
Tabel 4. 3 Distribusi Data Skor Kuesioner.....	46
Tabel 4. 4 Penyesuaian Nama Atribut	47
Tabel 4. 5 Penyesuaian Nama Atribut dalam <i>Dataset</i>	48
Tabel 4. 6 Distribusi Pembagian Data	50
Tabel 4. 7 Nilai Sentroid pada Rasio 60:40	51
Tabel 4. 8 Nilai Sentroid pada Rasio 70:30	52
Tabel 4. 9 Nilai Sentroid pada Rasio 80:20	52
Tabel 4. 10 Distribusi Kelas pada Data Latih.....	54
Tabel 4. 11 Distribusi Data Latih Sebelum dan Sesudah <i>SMOTE</i>	55
Tabel 4. 12 <i>Information Gain Feature Selection</i>	56
Tabel 4. 13 Hasil <i>GridSearchCV Feature Selection</i>	57
Tabel 4. 14 Hasil <i>Feature Selection</i>	57
Tabel 4. 15 Hasil Perhitungan Pemilihan Simpul Akar Algoritma C5.0	61
Tabel 4. 16 Nilai <i>SHAP</i> Mentah	68
Tabel 4. 17 Tingkat Kepentingan Keseluruhan Fitur	70

DAFTAR RUMUS

Rumus (1) <i>Entropy</i>	24
Rumus (2) <i>Gain</i>	24
Rumus (3) <i>Split Info</i>	24
Rumus (4) <i>Gain Ratio</i>	25
Rumus (5) <i>Euclidean Distance</i>	26
Rumus (6) <i>SMOTE</i>	26
Rumus (7) <i>Conditional Expected Value</i>	29
Rumus (8) <i>SHAP</i>	30
Rumus (9) <i>Recall</i>	32
Rumus (10) <i>Precision</i>	32
Rumus (11) <i>Accuracy</i>	32
Rumus (12) <i>F1-Score</i>	32



DAFTAR LAMPIRAN

1.	<i>Dataset Kuesioner</i>	80
2.	Visualisasi Pohon Keputusan C5.0	85
3.	<i>Source Code Python Library</i>	86
4.	<i>Source Code Data Preparation</i>	86
5.	<i>Source Code Modeling</i>	89
6.	<i>Source Code Evaluation</i>	89
7.	Surat Perizinan Penelitian pada Program Studi Teknik Informatika	91
8.	Surat Perizinan Pengambilan Data Program Studi Teknik Informatika	92
9.	Surat Perizinan Pengambilan Data Hasil Kuesioner	93

