

**ANALISIS KOMPARASI ALGORITMA C4.5,
RANDOM FOREST, DAN XGBOOST
MENGUNAKAN PENDEKATAN STRATIFIED
CROSS VALIDATION DAN SMOTE UNTUK
KLASIFIKASI KELULUSAN TEPAT WAKTU
PADA IMBALANCED SMALL DATASET**

TESIS



OLEH:

**WIDYA DHARMA SIDI
2329101023**

**PROGRAM STUDI ILMU KOMPUTER (S2)
PROGRAM PASCASARJANA
UNIVERSITAS PENDIDIKAN GANESHA**



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSR - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSR
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



Balai
Sertifikasi
Elektronik

TESIS

DIAJUKAN UNTUK MELENGKAPI TUGAS DAN MEMENUHI SYARAT-SYARAT UNTUK MENCAPAI GELAR MAGISTER KOMPUTER

Menyetujui

Pembimbing I	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Pembimbing II	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.



Balai
Sertifikasi
Elektronik

Tesis oleh Widya Dharma Sidi telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Ilmu Komputer (S2), Program Pascasarjana, Universitas Pendidikan Ganesha
Pada tanggal 21 Juli 2026

Dewan Penguji

Ketua	Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. NIP.198307252008011008
Anggota	Dr. I Gede Aris Gunadi, S.Si. M.Kom. NIP.197703182008121004
Anggota	Dr. I Nyoman Sukajaya, M.T. NIP.196711151993031001
Anggota	Prof. Dr. Ir. Dewa Gede Hendra Divayana, S.Kom., M.Kom., MCE., IPU., ASEAN.Eng., APEC.Eng. NIP.198407242015041002

Mengetahui Direktur Program Pascasarjana
Universitas Pendidikan Ganesha,



Prof. Dr. I Nyoman Jampel, M.Pd.

NIP.195910101986031003



- UU ITE No. 11 Tahun 2008 Pasal 5 Ayat 1 "Informasi Elektronik dan/atau hasil cetaknya merupakan alat bukti hukum yang sah"
- Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan BSrE - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSrE
- Cetakan dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Pendidikan dari Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 23 Juli 2026

Yang memberi pernyataan,



Widya Dharma Sidi

PRAKATA

Puji syukur penulis panjatkan ke hadirat Ida Sang Hyang Widhi Wasa, atas asung kertha wara nugraha-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul **“Analisis Komparasi Algoritma C4.5, Random Forest, dan XGBoost Menggunakan Pendekatan Stratified Cross Validation dan SMOTE untuk Klasifikasi Kelulusan Tepat Waktu pada Imbalanced Small Dataset”**.

Tesis ini disusun sebagai salah satu syarat untuk melanjutkan ke tahap penelitian tesis dalam rangka menyelesaikan studi pada Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Penulis menyadari bahwa selesainya penyusunan tesis ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, melalui kesempatan ini, penulis menyampaikan ucapan terima kasih yang tulus kepada:

1. Bapak Prof. Dr. I Wayan Lasmawan, M.Pd., selaku Rektor Universitas Pendidikan Ganesha yang telah memberikan kesempatan kepada penulis untuk menempuh pendidikan di lembaga ini.
2. Bapak Prof. Dr. I Nyoman Jampel, M.Pd., selaku Direktur Program Pascasarjana Universitas Pendidikan Ganesha.
3. Bapak Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs., selaku Pembimbing I dan Koordinator Program Studi Ilmu Komputer Program Pascasarjana Universitas Pendidikan Ganesha yang dengan sabar telah memberikan bimbingan, arahan, dan motivasi dalam penyusunan tesis ini.
4. Bapak Dr. I Gede Aris Gunadi, S.Si., M.Kom., selaku Pembimbing II yang telah banyak memberikan masukan dan saran konstruktif demi kesempurnaan tesis ini.
5. Para Pejabat dan civitas akademika di lingkungan Universitas PGRI Mahadewa Indonesia (UPMI Bali), yang telah memberikan izin penelitian serta akses data pada Sistem Informasi Akademik (SIKAD) UPMI Bali sebagai lokasi studi kasus.
6. Segenap Dosen dan Staf Administrasi di lingkungan Program Pascasarjana Universitas Pendidikan Ganesha atas ilmu dan pelayanan yang diberikan.

7. Keluarga tercinta yang senantiasa memberikan doa, dukungan moral, dan materiil yang tak terhingga.
8. Rekan-rekan mahasiswa Program Studi Ilmu Komputer Angkatan 2023 atas kebersamaan dan diskusinya.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Akhir kata, semoga proposal ini dapat bermanfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang *Educational Data Mining*.

Denpasar, Juni 2026

Penulis



DAFTAR ISI

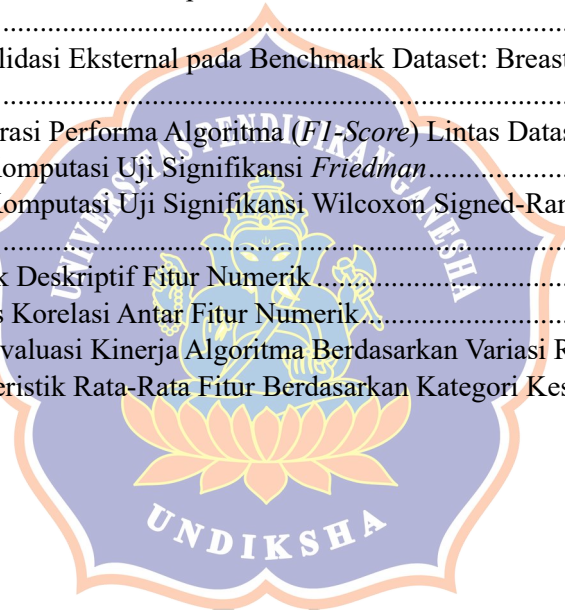
LEMBAR PERSETUJUAN PEMBIMBING.....	iii
LEMBAR PERSETUJUAN DEWAN PENGUJI.....	iv
LEMBAR PERNYATAAN	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
ABSTRAK.....	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	5
1.3 Pembatasan Masalah	6
1.4 Rumusan Masalah	6
1.5 Tujuan Penelitian.....	7
1.6 Manfaat Penelitian	8
BAB II KAJIAN PUSTAKA	9
2.1 Kajian Teori.....	9
2.1.1 Educational Data Mining (EDM)	9
2.1.2 Definisi Kelulusan Tepat Waktu	10
2.1.3 Data Tidak Seimbang dan Paradoks Akurasi	10
2.1.4 Algoritma C4.5	12
2.1.5 <i>Random Forest</i>	13
2.1.6 Extreme Gradient Boosting (XGBoost).....	15
2.1.7 Stratified 10-Fold Cross-Validation	16
2.1.8 Synthetic Minority Over-sampling Technique (SMOTE).....	17
2.1.9 Karakteristik Algoritma pada Small Dataset dan Risiko Overfitting	19
2.1.10 Evaluasi Kinerja Model.....	20
2.1.11 Interpretabilitas Model (<i>Feature Importance</i>)	22
2.1.12 Karakteristik 3 Skenario Pengujian Tambahan dan Dampaknya terhadap Kualitas Prediksi	24
2.2 Kajian Penelitian Relevan	26
2.3 Kerangka Pikir	31

2.4 Hipotesis Penelitian.....	33
2.4.1 Hipotesis Komparasi Kinerja Algoritma.....	34
2.4.2 Hipotesis Efektivitas Penanganan Level Data (SMOTE)	34
2.4.3 Hipotesis Konsistensi (Validasi Eksternal)	34
2.4.4 Hipotesis Fitur Dominan	34
BAB III METODE PENELITIAN	36
3.1 Jenis Penelitian.....	36
3.2 Rancangan Penelitian	36
3.2.1 Tahap 1: Pengumpulan dan Seleksi Data	38
3.2.2 Tahap 2: Pra-Pemrosesan Data (<i>Preprocessing</i>)	38
3.2.3 Tahap 3: Eksperimen dan Pemodelan	38
3.2.4 Tahap 4: Evaluasi dan Validasi	41
3.2.5 Tahap 5: Hasil Akhir dan Implikasi (<i>Results and Implications</i>).....	42
3.3 Tempat dan Waktu Penelitian	43
3.3.1 Tempat Penelitian.....	43
3.3.2 Waktu Penelitian.....	43
3.4 Populasi dan Sampel Penelitian.....	43
3.4.1 Dataset Utama.....	43
3.4.2 Dataset Pendamping.....	44
3.5 Variabel Penelitian	45
3.5.1 Variabel Dataset Utama (Studi Kasus UPMI).....	45
3.5.2. Variabel Dataset Pendamping 1 (<i>Haberman's Survival</i>)	46
3.5.3 Variabel Dataset Pendamping 2 (<i>Breast Cancer Wisconsin</i>).....	47
3.6 Teknik dan Instrumen Pengumpulan Data.....	48
3.6.1 Teknik Pengumpulan Data.....	48
3.6.2 Instrumen Pengumpulan Data	49
3.7 Validitas dan Reliabilitas Instrumen (Model)	49
3.8 Teknik Analisis Data	51
3.8.1 Pra-Pemrosesan (<i>Preprocessing</i>).....	51
3.8.2 Eksperimen Pemodelan (Modeling)	52
3.8.3 Evaluasi Kinerja (<i>Evaluation Metrics</i>)	54
3.8.4 Uji Signifikansi Statistik (<i>Statistical Significance Testing</i>).....	55
3.8.5 Interpretasi Model	55
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	57
4.1 Tahap Persiapan dan Pra-Pemrosesan Data	57
4.2 Skenario Pelatihan dan Pengujian	58

4.3 Hasil Uji Masing-Masing Skenario	61
4.3.1 Hasil Evaluasi Kinerja Model pada Skenario Utama.....	61
4.3.2 Hasil Uji Ablasi Arsitektur Metodologi (Evaluasi Skenario Tambahan).....	64
4.3.3 Hasil Skenario Evaluasi Kestabilan dan <i>Overfitting</i>	67
4.3.4 Hasil Validasi Eksternal (<i>Cross-Dataset Validation</i>)	68
4.3.5 Hasil Ekstraksi Bobot Fitur Model XGBoost (<i>Feature Importance</i>)	70
4.4 Uji Signifikansi Statistik dan Hipotesis	72
4.4.1 Uji Signifikansi Statistik Kinerja Algoritma (Uji <i>Friedman</i>).....	72
4.4.2 Uji Signifikansi Statistik Efektivitas Resampling SMOTE (Uji Wilcoxon Signed-Rank)	73
4.4.3 Uji Hipotesis	75
4.5 Pembahasan.....	76
4.5.1 Analisis Karakteristik <i>Small Dataset</i> dan Kualitas Ruang Fitur	76
4.5.2 Analisis Studi Ablasi: Signifikansi <i>Stratified CV</i> dan SMOTE pada <i>Small Imbalanced Dataset</i>	79
4.5.3 Analisis Arsitektur Algoritma dan Superioritas XGBoost.....	87
4.5.4 Analisis Kesalahan Klasifikasi (<i>Misclassification Analysis</i>) pada Kondisi Data Ekstrem	90
4.5.5 Analisis Kinerja <i>Early Warning System</i> , Pemaknaan Metrik, dan Biaya Klasifikasi Asimetris	92
4.5.6 Interpretasi Fitur Prediktif terhadap Fenomena Keterlambatan Studi	94
4.5.7 Analisis Validasi Eksternal dan Teorema <i>No Free Lunch</i>	97
4.5.8 Sintesis Uji Hipotesis dan Justifikasi Praktis Pemilihan XGBoost	99
4.5.9 Implikasi Manajerial dan Rekomendasi Kebijakan Institusi.....	100
4.5.10 Kontribusi Penelitian pada Bidang Ilmu Komputer	103
4.5.11 Keterbatasan Penelitian	104
BAB V SIMPULAN DAN SARAN	106
5.1 Simpulan	106
5.2 Saran	107
5.2.1 Saran Operasional bagi Institusi	107
5.2.2 Saran bagi Peneliti Selanjutnya	108
DAFTAR PUSTAKA.....	110
LAMPIRAN.....	114

DAFTAR TABEL

Tabel 2.1 Matriks Kajian Penelitian Relevan	28
Tabel 3. 1 Tabel Dataset Pendamping.....	44
Tabel 3.2 Daftar Variabel Independen (Fitur).....	46
Tabel 3.3 Instrumen Penelitian.....	49
Tabel 4. 1 Ruang Pencarian dan Hasil Optimasi Hiperparameter (<i>GridSearchCV</i>)	60
Tabel 4. 2 Rincian Metrik Evaluasi pada 10-Lipatan Validasi Silang.....	62
Tabel 4. 3 Hasil Evaluasi Kinerja Algoritma pada Dataset Utama (UPMI Bali)	63
Tabel 4. 4 Hasil Evaluasi Kinerja Algoritma pada Skenario 1 (Tanpa SMOTE)	65
Tabel 4. 5 Hasil Evaluasi Kinerja Algoritma pada Skenario 2 (Tanpa Stratifikasi dan Tanpa SMOTE).....	65
Tabel 4. 6 Hasil Evaluasi Kinerja Algoritma pada Skenario 3 (Train-Test Split 80:20)....	66
Tabel 4. 7 Hasil Uji Generalization Gap Algoritma.....	67
Tabel 4. 8 Hasil Validasi Eksternal pada Benchmark Dataset: Haberman (Dimensi Ekstrem: 3 Fitur).....	68
Tabel 4. 9 Hasil Validasi Eksternal pada Benchmark Dataset: Breast Cancer (Dimensi Tinggi: 30 Fitur).....	69
Tabel 4. 10 Komparasi Performa Algoritma (<i>F1-Score</i>) Lintas Dataset	70
Tabel 4. 11 Hasil Komputasi Uji Signifikansi <i>Friedman</i>	72
Tabel 4. 12 Hasil Komputasi Uji Signifikansi Wilcoxon Signed-Rank (Tanpa SMOTE vs Dengan SMOTE)	74
Tabel 4. 13 Statistik Deskriptif Fitur Numerik	77
Tabel 4. 14 Matriks Korelasi Antar Fitur Numerik	78
Tabel 4. 15 Hasil Evaluasi Kinerja Algoritma Berdasarkan Variasi Rasio SMOTE	85
Tabel 4. 16 Karakteristik Rata-Rata Fitur Berdasarkan Kategori Kesalahan Klasifikasi ..	90



DAFTAR GAMBAR

Gambar 2.1 Bagan Kerangka Pikir Penelitian.....	31
Gambar 3.1 Diagram Alur Rancangan Penelitian.....	37
Gambar 4. 1 Akumulasi Confusion Matrix Model XGBoost.....	64
Gambar 4. 2 Visualisasi Rata-rata Bobot Fitur (Feature Importance) Model XGBoost....	71
Gambar 4. 3 Ilustrasi Konseptual Pergeseran Batas Keputusan (Decision Boundary) Sebelum dan Sesudah Penerapan SMOTE.....	81



DAFTAR LAMPIRAN

Lampiran 1. <i>Souce Code</i> Pengujian Model	114
Lampiran 2. Cuplikan Dataset Utama dan Pendamping	123

