

KOMPARASI KINERJA ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK KLASIFIKASI KASUS DIABETES

SKRIPSI



**OLEH:
RESA ASTARI BR TARIGAN
2215091018**

**PROGRAM STUDI SISTEM INFORMASI (SI) (S1)
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA
2026**



- 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-E - BSSN, validitas dokumen elektronik ini bisa dicek menggunakan aplikasi mobile VeryDS oleh BSR-E
- Cetak dokumen ini merupakan salinan dari file dokumen bertandatangan elektronik yang keabsahannya dapat diakses melalui scan QRCode yang terdapat pada sertifikat ini.

SKRIPSI

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

Menyetujui

Pembimbing I	Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. NIP.198907112020122004
Pembimbing II	Ir. Gede Surya Mahendra, S.Pd., M.Kom. NIP.199003132022031009



- 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.

Skripsi oleh Resa Astari br Tarigan ini
telah dipertahankan di depan dewan penguji
Pada tanggal 15 Juli 2026

Dewan Penguji

Ketua	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008
Anggota	Ir. Kadek Teguh Dermawan, S.Pd., M.Kom. NIP.199406132025061003
Anggota	Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. NIP.198907112020122004
Anggota	Ir. Gede Surya Mahendra, S.Pd., M.Kom. NIP.199003132022031009



- 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.

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
guna memenuhi syarat-syarat untuk mencapai gelar Sarjana Komputer

Menyetujui

Ketua Ujian	Ir. Made Windu Antara Kesiman, S.T., M.Sc., Ph.D. NIP.198211112008121001
Sekretaris Ujian	Ir. I Made Dendi Maysanjaya, S.Pd., M.Eng. NIP.199005152019031008



PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul “Komparasi Kinerja Algoritma *Random Forest* Dan *Xgboost* Untuk Klasifikasi Kasus Diabetes” beserta seluruh isinya adalah benar-benar karya sendiri dan saya tidak melakukan penjiplakan dan pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran atas etika keilmuan dalam karya saya ini atau ada klaim terhadap keaslian karya saya ini.

Singaraja, 27 Juli 2026

Yang membuat pernyataan,



METERAI
TEMPEL
70050AJX628491359

Br Tarigan

NIM 2215091018

PRAKATA

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul <KOMPARASI KINERJA ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK KLASIFIKASI KASUS DIABETES=.

Penyusunan skripsi ini dimaksudkan untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana Komputer pada Universitas Pendidikan Ganesha. Penulis menyadari bahwa keberhasilan penyelesaian skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang tulus kepada:

1. Bapak Dr. Kadek Rihendra Dantes, S.T., M.T., selaku Dekan Fakultas Teknik dan Kejuruan yang telah memberikan kesempatan dan fasilitas kepada penulis selama menempuh pendidikan.
2. Bapak Putu Hendra Suputra, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika
3. Bapak I Made Dendi Maysanjaya, S.Pd., M.Eng., selaku Koordinator Program Studi Sistem Informasi dan selaku Dosen Penguji I yang telah memberikan nasihat, kritik, serta saran yang membangun dalam penyempurnaan laporan ini.
4. Ir. I Gusti Ayu Agung Diatri Indradewi, S.Kom., M.T. selaku Dosen Pembimbing I telah memberikan waktu, arahan, saran, masukan serta bimbingan bagi Penulis dalam menyelesaikan karya ini.
5. Ir. Gede Surya Mahendra, S.Pd., M.Kom. selaku Dosen Pembimbing II dan Dosen Pembimbing Akademik yang telah memberikan waktu, arahan, saran, masukan serta bimbingan bagi Penulis dalam menyelesaikan karya ini.
6. Ir. Kadek Teguh Dermawan, S.Pd., M.Kom. selaku Dosen Penguji II yang telah memberikan nasihat, kritik, serta saran yang membangun dalam penyempurnaan laporan ini.
7. Kepada Mamak tercinta, yang selalu memberikan doa dan dukungan, kasih sayang, serta pengorbanan yang luar biasa demi keberhasilan dan masa depan penulis. Terima kasih telah menjadi sumber kekuatan terbesar dalam hidup penulis.

8. Kepada seseorang yang kini telah beristirahat di rumah Bapa di surga, yang penulis panggil "Bapak" sosok yang sangat penulis rindukan. Mungkin langkah ini tak sempat Bapak saksikan secara langsung, namun setiap pencapaian ini tak pernah lepas dari jejak doa dan kasihmu yang tertanam kuat sejak awal kehidupan penulis. Terimakasih sudah menjadi Bapak yang semasa hidupnya selalu berusaha memberikan kehidupan yang baik bagi penulis, kasih sayang yang tiada henti. Penulis percaya, meski tak terlihat bapak selalu ada di hati penulis.

9. Kedua saudara, Ebram Tarigan dan Marta Claudia Br Tarigan yang telah memberikan doa, dukungan moral maupun material, serta pengorbanan selama perjalanan pendidikan penulis.

10. Teruntuk seseorang dengan NIM 123121017, di balik ketidaksengajaan yang mempertemukan kita, terima kasih sudah berkontribusi banyak dalam penulisan skripsi ini, baik waktu, pemikiran dan ilmu kepada penulis. Terima kasih sudah menjadi pendengar yang sabar dalam setiap keluh kesah yang penulis hadapi.

11. Rekan-rekan mahasiswa Program Studi Sistem Informasi, atas kebersamaan, semangat, dan bantuan yang diberikan selama proses perkuliahan dan penyusunan skripsi ini.

12. Kepada Kakak, Abang, serta saudara-saudari terkasih dalam Persekutuan Kristen yang senantiasa memberikan dukungan, motivasi, doa, dan berbagai masukan yang berharga kepada penulis selama proses penyusunan skripsi ini."

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan karya ini di masa mendatang. Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan.

DAFTAR ISI

PRAKATA	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
ABSTRAK	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.2 Tujuan Penelitian	7
1.3 Ruang Lingkup Penelitian	7
1.4 Manfaat Penelitian	8
BAB II KAJIAN PUSTAKA	9
2.1 Kajian Pustaka	9
2.2 Landasan Teori	112
2.2.1 Machine learning	112
2.2.2 Klasifikasi	113
2.2.3 Diabetes	114
2.2.4 Data	20
2.2.5 Analisis Data Eksplorasi (EDA)	21
2.2.6 IQR (Interquartile Range)	22
2.2.7 Random Forest	234
2.2.8 XGBoost	225
2.2.9 Hyperparameter Tuning	228
2.2.10 K-Fold Cross Validation	229
2.2.11 Confusion matrix	30
2.2.12 Software Pendukung	332
BAB III METODE PENELITIAN	34
3.1 Tahapan Penelitian	34
3.1.1 Studi Literatur	37

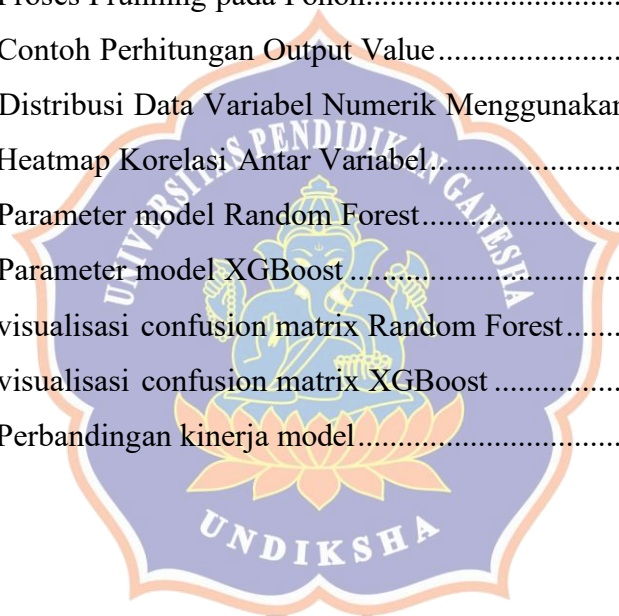
3.1.2 Data Selection	37
3.1.3 Pra-Pemrosesan Data	40
3.1.4 Tranformasi Data	42
3.1.5 Data Mining	43
3.1.6 Evaluasi Model	60
3.2 Spesifikasi Perangkat.....	62
3.2.1 Perangkat Keras	62
3.2.2 Perangkat Lunak	62
BAB IV HASIL DAN PEMBAHASAN	64
4.1 Data Selection	64
4.2 Pra-Pemrosesan Data	66
4.2.1 Analisis Deskriptif	66
4.2.2 Analisis Korelasi dan Identifikasi Redundansi.....	69
4.3 Data Cleaning	70
4.3.1 Missing Value	70
4.3.2 Ketidakkonsistenan Data.....	71
4.3.3 Data Duplikat	71
4.3.4 Data Outlier.....	72
4.3.5 Feature Filtering.....	74
4.4 Transformasi Data	74
4.5 Pembagian Data	75
4.6 Pemodelan	76
4.6.1 Random Forest.....	76
4.6.2 XGBoost	78
4.7 Model Akhir	79
4.7.1 Evaluasi Model	79
4.7.2 Interpretasi Model.....	81
BAB V KESIMPULAN.....	84
5.1 Kesimpulan.	84
5.2 Saran.....	84
DAFTAR PUSTAKA	86

DAFTAR LAMPIRAN	92
-----------------------	----



DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi K-Fold Cross Validation.....	30
Gambar 2. 2 Confusion matrix.....	31
Gambar 3. 1 Diagram Alir tahap Penelitian.....	36
Gambar 3. 2 Contoh dari desicion tree	48
Gambar 3. 3 Contoh Membangun Pohon XGBoost	51
Gambar 3. 4 Contoh Perhitungan Similarity dan Gain.....	52
Gambar 3. 5 Contoh Split pada Pohon XGBoost.....	53
Gambar 3. 6 Contoh Split pada Turunan Percabangan.....	53
Gambar 3. 7 Contoh Perhitungan Similarity dan Gain pada Split Lanjutan	54
Gambar 3. 8 Proses Prunning pada Pohon.....	54
Gambar 3. 9 Contoh Perhitungan Output Value.....	55
Gambar 4. 1 Distribusi Data Variabel Numerik Menggunakan Bloxpot.....	69
Gambar 4. 2 Heatmap Korelasi Antar Variabel.....	70
Gambar 4. 3 Parameter model Random Forest.....	77
Gambar 4. 4 Parameter model XGBoost	78
Gambar 4. 5 visualisasi confusion matrix Random Forest.....	79
Gambar 4. 6 visualisasi confusion matrix XGBoost	80
Gambar 4. 7 Perbandingan kinerja model.....	81



DAFTAR TABEL

Tabel 2. 1 Parameter dalam Metode Random Forest	28
Tabel 2. 2 Parameter dalam Metode XGBoost.....	29
Tabel 3. 1 Data set yang akan digunakan.....	40
Tabel 3. 3 Data sebelum Encoding	44
Tabel 3. 4 Data sesudah Encoding	44
Tabel 3. 5 Pembagian data Tes dan Train.....	46
Tabel 3. 6 Dataset untuk membangun model Random Forest	47
Tabel 3. 7 Dataset untuk membangun model XGBoost	51
Tabel 3. 8 Perhitungan Nilai Error atau Residual	52
Tabel 3. 9 Perhitungan Nilai Prediksi pada Model-1	56
Tabel 3. 10 penggunaan hyperparameter Random Forest	57
Tabel 3. 11 Penggunaan hyperparameter XGBoost.....	59
Tabel 3. 12 Contoh perhitungan Confusion matrix.....	63
Tabel 3. 13 Spesifikasi Perangkat Keras.....	63
Tabel 3. 14 Spesifikasi Perangkat Lunak.....	64
Tabel 4. 1 Distribusi kelas awal dataset.....	64
Tabel 4. 2 Karakteristik dataset Diabetes	65
Tabel 4. 3 Statistik Deskriptif Variabel Numerik.....	66
Tabel 4. 4 Hasil Standardisasi Data Kategorikal	71
Tabel 4. 5 Jumlah Data Duplikat	71
Tabel 4. 6 Hasil Deteksi Outlier Menggunakan Metode IQR.....	72
Tabel 4. 7 Rincian Eksekusi Eliminasi Outlier	73
Tabel 4. 8 Tranformasi Data	75
Tabel 4. 9 Pembagian data train dan data test.....	76
Tabel 4. 10 Perbandingan Model	79

DAFTAR LAMPIRAN

Lampiran. 1 Kode Library yang Dibutuhkan	92
Lampiran. 2 Kode Data Selection	92
Lampiran. 3 Kode Analisis Data Eksplorasi (EDA).....	93
Lampiran. 4 Kode Data Pre-Processing	94
Lampiran. 5 Kode Tranformasi Data.....	97
Lampiran. 6 Kode Data Splinting	97
Lampiran. 7 Kode membangun Model Random Forest	97
Lampiran. 8 Kode membangun model XGBoost.....	99
Lampiran. 9 Kode Prediksi Model Random Forest dan XGBoost.....	100
Lampiran. 10 Kode Evaluasi Model Random Forest dan XGBoost	100
Lampiran. 11 Kode confusion matrix XGBoost dan Random Forest	102
Lampiran. 12 Jadwal Penelitian.....	104

