PERBANDINGAN PERFORMA ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MENDETEKSI DIABETES BERDASARKAN DATA REKAM MEDIS

SKRIPSI

Shyehan Rafael Adlinugroho
00000052738

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

PERBANDINGAN PERFORMA ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MENDETEKSI DIABETES BERDASARKAN DATA REKAM MEDIS

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Shyehan Rafael Adlinugroho
00000052738

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Shyehan Rafael Adlinugroho
Nomor Induk Mahasiswa	:	00000052738
Program Studi	:	Informatika

Skripsi dengan judul:

Perbandingan Performa Algoritma Random Forest dan XGBoost dalam Mendeteksi Diabetes Berdasarkan Data Rekam Medis

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 3 Januari 2025

(Shyehan Rafael Adlinugroho)

HALAMAN PENGESAHAN

Skripsi dengan judul

PERBANDINGAN PERFORMA ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MENDETEKSI DIABETES BERDASARKAN DATA REKAM MEDIS

oleh

Nama	:	Shyehan Rafael Adlinugroho
NIM	:	00000052738
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Rabu, 8 Januari 2025
Pukul 13.00 s/s 15.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI))
NIDN: 0315109103

Penguji

(Adhi Kusnadi, S.T, M.Si.)
NIDN: 0303037304

Pembimbing

(David Agustriawan, S.Kom., M.Sc., Ph.D.)
NIDN: 0525088601

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI))
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Shyehan Rafael Adlinugroho
NIM	:	00000052738
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Perbandingan Performa Algoritma Random Forest dan XGBoost dalam Mendeteksi Diabetes Berdasarkan Data Rekam Medis

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

☑ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.

☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.

☐ Lainnya, pilih salah satu:

- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
- Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 3 Januari 2025

Yang menyatakan

Shyehan Rafael Adlinugroho

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

”Discipline is the bridge between goals and accomplishment.”

Jim Rohn

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini yang berjudul "PERBANDINGAN PERFORMA ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MENDETEKSI DIABETES BERDASARKAN DATA REKAM MEDIS". Penulisan tugas akhir ini bertujuan untuk memberikan kontribusi dalam pengembangan teknologi prediktif di bidang kesehatan, khususnya dalam deteksi dini diabetes.

Penulis menyadari bahwa penyelesaian tugas akhir ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan selama proses penelitian ini.

Mengucapkan terima kasih

1. Bapak Dr. Andrey Andoko, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc. (OCA, CEH, CEI), selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak David Agustriawan, S.Kom., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. dr. Ria Novitasari dan dr. Gita Permatasari, MKK, selaku narasumber yang telah memberikan wawasan dan informasi yang sangat berguna untuk mendukung penelitian ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
7. Teman-teman yang telah membantu baik secara langsung maupun tidak langsung dalam proses penyelesaian tugas akhir ini, atas segala dukungan, diskusi, serta motivasinya.

Semoga karya ilmiah ini dapat bermanfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang deteksi dini diabetes, serta menjadi referensi bagi penelitian selanjutnya.

Tangerang, 3 Januari 2025

Shyehan Rafael Adlinugroho

PERBANDINGAN PERFORMA ALGORITMA RANDOM FOREST DAN XGBOOST DALAM MENDETEKSI DIABETES BERDASARKAN DATA REKAM MEDIS

Shyehan Rafael Adlinugroho

ABSTRAK

Diabetes merupakan kondisi kronis dengan kadar gula darah tinggi yang dapat menyebabkan komplikasi serius. Faktor risiko meliputi pola makan tidak sehat, kurang aktivitas fisik, faktor genetik, dan gaya hidup yang buruk. Berdasarkan data WHO tahun 2022, terdapat 537 juta orang dewasa dengan diabetes di dunia, dengan prevalensi yang terus meningkat, terutama di negara berkembang seperti Indonesia. Penelitian ini membandingkan algoritma Random Forest dan XGBoost dalam mendeteksi diabetes berdasarkan data medical check-up dari Rumah Sakit Pusat Pertamina dengan rentang usia 46-65 tahun. Pemilihan algoritma Random Forest dan XGBoost didasarkan pada kemampuan keduanya dalam menganalisis data dengan banyak fitur dan menangani klasifikasi yang kompleks, termasuk dalam kasus data kesehatan yang sering kali memiliki distribusi tidak seimbang. Random Forest menggunakan pendekatan bagging dengan membangun banyak pohon keputusan secara paralel, sedangkan XGBoost menggunakan pendekatan boosting yang membangun pohon secara bertahap dengan memperbaiki kesalahan prediksi sebelumnya. Perbandingan ini penting dilakukan untuk mengevaluasi akurasi dan stabilitas prediksi dalam kasus deteksi diabetes yang kompleks. Hasil percobaan menunjukkan bahwa Random Forest menghasilkan akurasi pengujian sebesar 90% dengan F1-Score 0,87, sedangkan XGBoost mencapai akurasi 88% dengan F1-Score 0,77. Random Forest menunjukkan performa yang lebih baik dalam mendeteksi diabetes dengan distribusi data yang lebih seimbang, sementara XGBoost cenderung lebih selektif pada fitur dengan kontribusi signifikan. Distribusi data yang lebih terfokus dan seimbang terbukti meningkatkan performa model dalam mendeteksi diabetes, yang diharapkan dapat membantu upaya deteksi dini dan pengambilan keputusan klinis yang lebih efektif di masa depan.

Kata Kunci: Deteksi Dini, Diabetes, Medical Check-Up, Random Forest, XGBoost

PERFORMANCE COMPARISON OF RANDOM FOREST AND XGBOOST ALGORITHMS IN DETECTING DIABETES BASED ON MEDICAL RECORD DATA

Shyehan Rafael Adlinugroho

ABSTRACT

Diabetes is a chronic condition characterized by high blood sugar levels, potentially leading to severe complications. Risk factors include unhealthy eating habits, physical inactivity, genetic predisposition, and poor lifestyle choices. According to WHO data in 2022, there are 537 million adults living with diabetes worldwide, with a rising prevalence, particularly in developing countries like Indonesia. This study compares the Random Forest and XGBoost algorithms in detecting diabetes based on medical check-up data from Rumah Sakit Pusat Pertamina for individuals aged 46-65 years. The selection of Random Forest and XGBoost is based on their ability to analyze datasets with numerous features and handle complex classification tasks, including imbalanced health data. Random Forest employs a bagging approach, building multiple decision trees in parallel, whereas XGBoost uses a boosting approach, constructing trees iteratively to correct previous prediction errors. This comparison is crucial for evaluating prediction accuracy and stability in complex diabetes detection cases. Experimental results indicate that Random Forest achieves a testing accuracy of 90% with an F1-score of 0.87, while XGBoost achieves an accuracy of 88% with an F1-score of 0.77. Random Forest demonstrates better performance in detecting diabetes with more balanced data distribution, whereas XGBoost tends to focus on features with significant contributions. A more focused and balanced data distribution has been proven to enhance model performance in detecting diabetes, offering potential support for early detection efforts and more effective clinical decision-making in the future.

Keywords: Diabetes, Early Detection, Medical Check-Up, Random Forest, XGBoost

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN