

SKRIPSI



**ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST* DAN *XGBOOST*
UNTUK KLASIFIKASI PENYAKIT KARDIOVASKULAR**

ALIF FAQIH

2010511032

INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2023

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer



**ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST* DAN *XGBOOST*
UNTUK KLASIFIKASI PENYAKIT KARDIOVASKULAR**

ALIF FAQIH

2010511032

INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2023

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Alif Faqih

NIM : 2010511032

Tanggal : 12 Januari 2024

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 12 Januari 2024

Yang menyatakan,



(Alif Faqih)

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Alif Faqih

NIM : 2010511032

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

**ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST* DAN *XGBOOST*
UNTUK KLASIFIKASI PENYAKIT KARDIOVASKULAR**

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 12 Januari 2024

Yang menyatakan,



(Alif Faqih)

LEMBAR PENGESAHAN**LEMBAR PENGESAHAN**

Skripsi diajukan oleh :

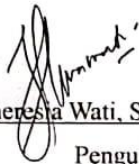
Nama : Alif Faqih

NIM : 2010511032

Program Studi : S1 Informatika

Judul Skripsi : Analisis Perbandingan Algoritma *Random Forest* dan *XGboost* untuk Klasifikasi Penyakit Kardiovaskular.

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Theresia Wati, S.Kom, M.TI.

Penguji 1



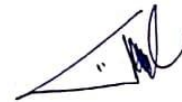
Muhammad Panji Muslim, S.Pd. M.Kom.

Penguji 2



Neny Rosmawarni, M.Kom

Pembimbing 1



Musthofa Galih Pradana, M.Kom

Pembimbing 2



Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM
Dekan Fakultas Ilmu Komputer



Dr. Widya Cholil, M.I.T
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 12 Januari 2024

ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST* DAN *XGBOOST* UNTUK KLASIFIKASI PENYAKIT KARDIOVASKULAR

ALIF FAQIH

ABSTRAK

Jantung adalah salah satu organ berotot yang berfungsi sebagai alat untuk pemompa oksigen dan darah menuju ke seluruh organ tubuh. Berdasarkan laman resmi dari *World Health Organization* (WHO), terdapat 17,9 juta orang meninggal dunia setiap tahunnya disebabkan oleh penyakit *kardiovaskular* yang menyerang bagian jantung manusia. Maka dari itu dibutuhkan model *machine learning* yang memiliki performa yang baik dan dapat melakukan klasifikasi pada penyakit kardiovaskular dengan cepat agar apabila terdapat orang yang terdeteksi berisiko terkena penyakit kardiovaskular dapat langsung menemui dokter untuk dilakukan pemeriksaan lebih lanjut dan penyakit dapat lebih cepat untuk ditangani. Penelitian ini menggunakan *dataset* Cardiovascular Disease Risk Prediction dengan *Random Forest* dan *XGBoost* sebagai model untuk melakukan klasifikasi. Penelitian ini dilakukan dengan tujuan untuk menemukan model dengan performa terbaik antara *Random Forest* dan *XGBoost*. Hasil dari penelitian ini merupakan evaluasi dari *Random Forest* dan *XGBoost* dalam melakukan klasifikasi penyakit kardiovaskular. Didapatkan performa terbaik pada model *Random Forest* dengan nilai akurasi sebesar 0.95, *precision* sebesar 0.96, dan *recall* sebesar 0.93.

Kata Kunci: Penyakit Kardiovaskular, *Random Forest*, *XGBoost*

COMPARATIVE ANALYSIS OF RANDOM FOREST AND XGBOOST ALGORITHMS FOR CARDIOVASCULAR DISEASE CLASSIFICATION

ALIF FAQIH

ABSTRACT

Heart is a muscular organ that functions as a tool to pump oxygen and blood to all organs of the body. Based on the official website of the World Health Organization (WHO), 17.9 million people die every year due to cardiovascular disease which attacks the human heart. Therefore, we need a machine learning model that has good performance and can classify cardiovascular disease quickly so that if someone is detected to be at risk of developing cardiovascular disease, they can immediately see a doctor for further examination and the disease can be treated more quickly. This research uses the Cardiovascular Disease Risk Prediction. *dataset* with Random Forest and XGBoost as models for classification. This research was conducted with the aim of finding the model with the best performance between Random Forest and XGBoost. The results of this study are an evaluation of Random Forest and XGBoost in classifying cardiovascular disease. The best performance was obtained from the Random Forest model with an accuracy value of 0.95, precision of 0.96, and recall of 0.93.

Keyword : Cardiovascular Disease, Random Forest, XGBoost

KATA PENGANTAR

Alhamdulillah, puji serta syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan karunia-Nya, penulis dapat menyelesaikan penulisan tugas akhir skripsi ini dengan judul “Analisis Perbandingan Algoritma *Random Forest* dan *XGBoost* untuk Klasifikasi Penyakit Kardiovaskular”. Tugas akhir skripsi ini ditulis sebagai salah satu syarat untuk memperoleh gelar sarjana program studi Informatika dari Fakultas Ilmu Komputer pada Universitas Pembangunan Nasional Veteran Jakarta. Dalam penyusunan tugas akhir ini, penulis mendapatkan dukungan, bimbingan, dan juga bantuan dari berbagai pihak sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan baik. Maka dari itu, penulis mengucapkan terima kasih kepada:

1. Ibu, Bapak, Kakak, dan seluruh keluarga penulis yang telah memberikan semangat dan juga doa sehingga penulis dapat menyelesaikan penulisan skripsi.
2. Ibu Neny Rosmawarni, M.Kom. dan Bapak Musthofa Galih Pradana, M.Kom. selaku dosen pembimbing yang telah membantu penulis dalam proses penyusunan skripsi.
3. Bapak Jayanta , S.Kom., M.Si, selaku dosen pembimbing akademik penulis.
4. Ibu Dr. Widya Cholil, M.I.T. selaku Kepala Program Studi Informatika UPN Veteran Jakarta.
5. Bapak Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM, selaku Dekan Fakultas Ilmu Komputer UPN Veteran Jakarta.
6. Bapak/Ibu Dosen Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
7. Teman – teman penulis yang selalu memberikan dukungan dan juga selalu mendoakan selama proses penulisan skripsi.

Penulis menyadari bahwa pada skripsi ini masih memiliki banyak kekurangan. Penulis berharap skripsi ini dapat bermanfaat bagi para pembaca dan penulis juga berharap para pembaca bisa mendapatkan pengetahuan yang berharga dari skripsi ini.

Penulis

Alif Faqih

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK.....	vi
ABSTRACT.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	3
1.4. Manfaat Penelitian	3
1.4.1. Bagi Penulis	3
1.4.2. Bagi Peneliti Lain.....	3
1.4.3 Bagi Masyarakat	3
1.5. Batasan Masalah	4
1.6. Luaran Yang Diharapkan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. <i>Machine Learning</i>	5
2.2. <i>Random Forest</i>	5
2.3. <i>XGBoost</i>	6
2.4. <i>Imbalanced data</i>	7
2.5. Evaluasi	8
2.6. Google Colaboratory	9
2.7. Python.....	10
2.8. Penyakit Kardiovaskular	10
2.9. Penelitian Relevan.....	10
BAB III METODE PENELITIAN.....	15
3.1. Tahapan Penelitian	15
3.2. Identifikasi Masalah	15
3.3. Studi Literatur	16

3.4.	Pengumpulan Data	16
3.5.	Pra-Proses Data.....	16
3.6.	Pembagian Data	17
3.7.	Pembuatan Model Klasifikasi	17
3.8.	Pengujian Model	17
3.9.	Evaluasi Model	17
3.10.	Analisis Model	19
3.11.	Alat Bantu Penelitian	19
3.11.1.	Perangkat Keras	19
3.11.2.	Perangkat Lunak	19
3.12.	Jadwal Penelitian	19
BAB IV HASIL DAN PEMBAHASAN.....		21
4.1.	Pengumpulan Data	21
4.2.	Pra-Proses Data.....	22
4.2.1.	Memeriksa Duplicate Data	22
4.2.2.	Mengatasi <i>Outlier</i>	23
4.2.3.	Transformasi Data.....	24
4.2.4.	Mengatasi Imbalanced Data	27
4.3.	Pembagian Data	27
4.4.	Pembuatan Model Klasifikasi.....	28
4.4.1	Klasifikasi <i>Random Forest</i>	28
4.4.2	Klasifikasi <i>XGBoost</i>	30
4.5.	Evaluasi Model	32
4.5.1.	Evaluasi Model Menggunakan Rasio Pembagian Data Sebesar 80:20.....	33
4.5.2.	Evaluasi Model Menggunakan Rasio Pembagian Data Sebesar 70:30.....	35
4.5.3.	Evaluasi Model Menggunakan Rasio Pembagian Data Sebesar 60:40.....	38
4.6.	Analisis Hasil.....	40
BAB V PENUTUP.....		43
5.1.	Kesimpulan.....	43
5.2.	Saran.....	43
DAFTAR PUSTAKA.....		45
RIWAYAT HIDUP.....		48

DAFTAR TABEL

Tabel 2. 1 Tabel confusion matrix	8
Tabel 2. 2 Penelitian Relevan	11
Tabel 3. 1 Contoh tabel confusion matrix	17
Tabel 3. 2 Jadwal Penelitian	20
Tabel 4. 1 Sampel <i>Dataset</i>	21
Tabel 4. 2 Data sebelum dilakukan transformasi	24
Tabel 4. 3 Data setelah dilakukan transformasi	25
Tabel 4. 4 Tipe data pada <i>dataset</i>	26
Tabel 4. 5 Jumlah data.....	27
Tabel 4. 6 Pembagian Data Train dan Data Test	27
Tabel 4. 7 confusion matrix random forest dengan rasio pembagian data 80:20	33
Tabel 4. 8 confusion matrix XGboost dengan rasio pembagian data 80:20.....	34
Tabel 4. 9 confusion matrix random forest dengan rasio pembagian data 70:30	35
Tabel 4. 10 confusion matrix XGboost dengan rasio pembagian data 70:30.....	36
Tabel 4. 11 confusion matrix random forest dengan rasio pembagian data 60:40	38
Tabel 4. 12 confusion matrix XGboost dengan rasio pembagian data 60:40.....	39
Tabel 4. 13 Hasil Performa Model.....	40
Tabel 4. 14 Kategori klasifikasi berdasarkan skor AUC (Sihombing & Arsani, 2021).....	42

DAFTAR GAMBAR

Gambar 2. 1 Metode Random Forest (S et al. 2023).....	5
Gambar 3. 1 Tahapan Penelitian	15
Gambar 4. 1 Jumlah missing value pada dataset	22
Gambar 4. 2 Pemeriksaan duplicate data.....	23
Gambar 4. 3 Menghapus duplicate data	23
Gambar 4. 4 Boxplot fitur yang memiliki outlier	23
Gambar 4. 5 Tahap 1 Klasifikasi Random Forest	28
Gambar 4. 6 Tahap 2 Klasifikasi Random Forest	29
Gambar 4. 7 Tahap 3 Klasifikasi Random Forest	29
Gambar 4. 8 Tahap 4 Klasifikasi Random Forest	29
Gambar 4. 9 Tahap 5 Klasifikasi Random Forest	30
Gambar 4. 10 Tahap 6 Klasifikasi Random Forest	30
Gambar 4. 11 Tahap 1 Klasifikasi XGBoost	31
Gambar 4. 12 Tahap 2 Klasifikasi XGBoost	31
Gambar 4. 13 Tahap 3 Klasifikasi XGBoost	31
Gambar 4. 14 Tahap 4 Klasifikasi XGBoost	32
Gambar 4. 15 Tahap 5 Klasifikasi XGBoost	32
Gambar 4. 16 ROC CURVES True Positive dan False Positive Rates.....	41