

**ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST*
DAN *GRADIENT BOOSTING* DALAM PREDIKSI PENYAKIT
GAGAL GINJAL KRONIS**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

HAMMAN KHADAFI AL HABIBIE

21.11.4164

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST*
DAN *GRADIENT BOOSTING* DALAM PREDIKSI PENYAKIT
GAGAL GINJAL KRONIS**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh

HAMMAN KHADAFI AL HABIBIE

21.11.4164

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN ALGORITMA RANDOM FOREST DAN
GRADIENT BOOSTING DALAM PREDIKSI PENYAKIT GAGAL
GINJAL KRONIS**

yang disusun dan diajukan oleh

Hamman Khadafi Al Habibie

21.11.4164

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 14 Februari 2025

Dosen Pembimbing,


Kusnawi, S.Kom., M.Eng.
NIK. 190302112

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS PERBANDINGAN ALGORITMA RANDOM FOREST DAN
GRADIENT BOOSTING DALAM PREDIKSI PENYAKIT GAGAL
GINJAL KRONIS

yang disusun dan diajukan oleh

Hamman Khadafi Al Habibie

21.11.4164

Telah dipertahankan di depan Dewan Penguji
pada tanggal 14 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ike Verawati, S.Kom., M.Kom.
NIK. 190302237

Rumini, S.Kom., M.Kom.
NIK. 190302246

Kusnawi, S.Kom., M.Eng.
NIK. 190302112



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 14 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hamman Khadafi Al Habibie
NIM : 21.11.4164

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Perbandingan Algoritma *Random Forest* Dan *Gradient Boosting*
Dalam Prediksi Penyakit Gagal Ginjal Kronis**

Dosen Pembimbing : Kusnawi, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 14 Februari 2025

Yang Menyatakan,



Hamman Khadafi Al Habibie

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan kasih-Nya, sehingga penyusunan skripsi ini dapat diselesaikan dengan baik dan tepat waktu. Dengan penuh rasa hormat dan terima kasih, karya ini penulis persembahkan kepada:

1. Tuhan Yang Maha Esa, atas segala karunia dan petunjuk-Nya yang senantiasa mengiringi setiap langkah penulis, hingga akhirnya skripsi ini dapat diselesaikan dengan lancar. Segala puji dan syukur hanya kepada-Nya, Sang Pemilik Kehidupan.
2. Ayahanda Sarwono dan Ibunda Tituk Rahmawati, S.Pd., yang telah menjadi sumber kekuatan, kasih sayang, dan doa yang tiada henti. Terima kasih atas segala pengorbanan, dukungan moril maupun materi, serta cinta yang tulus yang senantiasa mengiringi setiap perjuangan penulis.
3. AM Tsabita Rihadatul' Aisy yang kelak menjadi pendamping hidup, yang telah menjadi sumber semangat, dukungan, dan doa yang tulus. Kehadirannya menjadi penguat dalam setiap proses yang dijalani serta penyemangat dalam meraih keberhasilan.

Semoga karya sederhana ini dapat menjadi wujud syukur dan bentuk kecil dari segala cinta dan dukungan yang telah diberikan.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillah rabbil 'alamin, Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya. Shalawat dan Salam senantiasa tercurah kepada Baginda Rasulullah Muhammad SAW dengan mengucapkan *Allahmma Shalli'ala Muhammad Wa'ala'ih* Syaidina Muhammad yang telah membawa manusia dari alam jahiliyah kepada alam yang terang menerang yang penuh ilmu pengetahuan seperti saat sekarang ini, sehingga penulis dapat menyelesaikan penyusunan skripsi ini dengan judul "Analisis Perbandingan Algoritma *Random Forest* dan *Gradient Boosting* dalam Prediksi Penyakit Gagal Ginjal Kronis" sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika.

Penyusunan skripsi ini tidak terlepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Kusnawi, S.Kom., M.Eng., selaku Dosen Pembimbing yang telah dengan sabar membimbing, memberikan arahan, serta meluangkan waktu dan tenaga selama proses penyusunan skripsi ini.
2. Ibu Ike Verawati, S.Kom., M.Kom. dan Ibu Rumini, S.Kom., M.Kom., selaku Dosen Penguji, atas kritik, saran, dan masukan yang sangat konstruktif demi penyempurnaan skripsi ini serta mengarahkan penulis dengan penuh kesabaran dan dedikasi..
3. Seluruh dosen dan staf pengajar di Program Studi Informatika, yang telah membekali penulis dengan ilmu pengetahuan dan pengalaman selama masa studi.
4. Kedua orang tua penulis yang selalu memberikan doa, dukungan moral, serta semangat yang tak henti-hentinya selama proses perkuliahan hingga penyusunan skripsi ini.
5. Seseorang yang kelak menjadi pendamping hidup, yang telah menjadi sumber semangat, dukungan, dan doa yang tulus. Kehadirannya menjadi

penguat dalam setiap proses yang dijalani serta penyemangat dalam meraih keberhasilan.

6. Sahabat seperjuangan: Adam, Arya, Kresna, Novanda, Raffa, Tegar terima kasih atas semangat, tawa, dan motivasi yang tak pernah putus kalian berikan. Kehadiran kalian telah menjadi energi positif dalam setiap langkah perjuanganku.
7. Seluruh pihak yang telah memberikan bantuan, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun guna perbaikan di masa mendatang.

Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta dapat menjadi referensi dalam pengembangan ilmu pengetahuan di bidang informatika, khususnya dalam penerapan algoritma machine learning untuk prediksi penyakit.

Yogyakarta, 14 Februari 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	16
2.2.1 Data Mining	16
2.2.2 Machine Learning	17
2.2.3 Klasifikasi	18
2.2.4 Random Forest	18
2.2.5 Gradient Boosting	19
2.2.6 Preprocessing	21
2.2.7 Evaluasi	22
2.2.8 Deployment Streamlit	23

BAB III METODE PENELITIAN	24
3.1 Objek Penelitian.....	24
3.2 Alur Penelitian	24
3.3 Alat dan Bahan.....	38
3.3.1 Alat :.....	38
3.3.2 Bahan :	38
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Pengumpulan Data	40
4.1.1 <i>Exploratory Data Analyst (EDA)</i>	41
4.2 <i>Data Preprocessing</i>	43
4.2.1 <i>Feature Selection</i>	43
4.2.2 <i>Normalisasi Data dan SMOTE</i>	46
4.3 <i>Split Data</i>	48
4.4 <i>Model Building</i>	50
4.4.1 <i>Model Building Random Forest</i>	50
4.4.2 <i>Model Building Gradient Boosting</i>	52
4.5 Hasil Evaluasi	55
4.6 <i>Deployment</i>	59
BAB V PENUTUP	60
5.1 Kesimpulan	60
5.2 Saran	61
REFERENSI	62
LAMPIRAN.....	66

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 <i>Confusion Matrix</i>	22
Tabel 3. 1 Deskripsi Isi Dataset	26
Tabel 3. 2 Jumlah Nilai Kosong	30
Tabel 3. 3 Hitung <i>Entropy Random Forest</i>	33
Tabel 3. 4 Hitung <i>Information Gain Random Forest</i>	34
Tabel 3. 5 Perhitungan Manual <i>Gradient Boosting</i>	35
Tabel 3. 6 <i>Hardware</i> Spesifikasi	38
Tabel 3. 7 <i>Software</i> Spesifikasi	38
Tabel 3. 8 Fitur-fitur yang digunakan	39
Tabel 4. 1 Membaca File Dataset	40
Tabel 4. 2 Ringkasan Statistik Deskriptif	41
Tabel 4. 3 Data <i>Information</i>	42
Tabel 4. 4 <i>Feature Correlation</i>	43
Tabel 4. 5 <i>Feature Selection</i>	46
Tabel 4. 6 Pembagian Data	46
Tabel 4. 7 Normalisasi Data dan SMOTE	47
Tabel 4. 8 <i>Split Data</i>	48
Tabel 4. 9 <i>Model Building Random Forest</i>	50
Tabel 4. 10 Pengukuran Performa <i>Random Forest</i>	51
Tabel 4. 11 <i>Model Building Gradient Boosting</i>	52
Tabel 4. 12 Pengukuran Performa <i>Gradient Boosting</i>	54
Tabel 4. 13 Evaluasi Model <i>Random Forest</i>	55
Tabel 4. 14 Evaluasi <i>Gradient Boosting</i>	57
Tabel 4. 15 Perbandingan Model <i>Random Forest</i> dan <i>Gradient Boosting</i>	58

DAFTAR GAMBAR

Gambar 2. 1 <i>Visualisasi Algoritma Random Forest</i>	19
Gambar 2. 2 <i>Visualisasi Algoritma Gradient Boosting</i>	20
Gambar 3. 1 Alur Penelitian	24
Gambar 4. 1 Menampilkan 10 Data Teratas	41
Gambar 4. 2 Ringkasan Statistik Deskriptif	41
Gambar 4. 3 Menampilkan <i>missing values</i>	42
Gambar 4. 4 Analisis Korelasi	45
Gambar 4. 5 Pembagian Data Sebelum SMOTE	47
Gambar 4. 6 Pembagian Data Setelah SMOTE	48
Gambar 4. 7 Ukuran <i>data training</i> dan <i>testing</i>	49
Gambar 4. 8 Melatih Model <i>Random Forest</i>	51
Gambar 4. 9 Pengukuran Performa <i>Random Forest</i>	52
Gambar 4. 10 Melatih Model <i>Gradient Boosting</i>	53
Gambar 4. 11 Pengukuran Performa <i>Gradient Boosting</i>	54
Gambar 4. 12 <i>Confusion Matrix Random Forest</i>	56
Gambar 4. 13 <i>Confusion Matrix Gradient Boosting</i>	57

INTISARI

Penyakit gagal ginjal kronis (PGGK) merupakan salah satu penyakit tidak menular yang berdampak signifikan terhadap kualitas hidup pasien serta membebani sistem layanan kesehatan. Deteksi dini PGGK sangat penting untuk mencegah perkembangan penyakit yang lebih parah. Namun, tantangan dalam diagnosis dini masih menjadi kendala, terutama karena gejala awal yang tidak spesifik dan kurangnya kesadaran masyarakat terhadap penyakit ini. Oleh karena itu, diperlukan metode prediksi yang akurat untuk membantu tenaga medis dalam mendeteksi PGGK lebih cepat dan efektif. Penelitian ini membandingkan kinerja dua Algoritma *Machine Learning*, yaitu *Random Forest* dan *Gradient Boosting*, dalam memprediksi PGGK menggunakan dataset *Chronic Kidney Disease data* dari Kaggle. Langkah penelitian meliputi eksplorasi data, *Preprocessing* menggunakan teknik normalisasi dan *Synthetic Minority Over-sampling Technique* (SMOTE), serta pembagian data menggunakan metode *Stratified Shuffle Split*. Model diuji menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk menilai efektivitas masing-masing Algoritma dalam melakukan klasifikasi penyakit. Hasil penelitian menunjukkan bahwa Algoritma *Random Forest* memiliki akurasi sebesar 98,36%, lebih unggul dibandingkan dengan *Gradient Boosting* yang mencapai akurasi 96,56%. Dengan demikian, *Random Forest* terbukti lebih efektif dalam memprediksi PGGK dan dapat menjadi dasar pengembangan sistem pendukung keputusan medis. Hasil penelitian ini diharapkan dapat membantu tenaga medis dalam meningkatkan akurasi diagnosis dini, sehingga perawatan dapat dilakukan lebih cepat dan efisien.

Kata kunci: Penyakit Gagal Ginjal Kronis, *Machine Learning*, Klasifikasi, *Random Forest*, *Gradient Boosting*.

ABSTRACT

Chronic Kidney Disease (CKD) is a non-communicable disease that significantly impacts patients' quality of life and burdens healthcare systems. Early detection of CKD is crucial to prevent disease progression. However, challenges in early diagnosis remain due to non-specific initial symptoms and low public awareness. Therefore, an accurate prediction method is needed to assist medical professionals in detecting CKD more quickly and effectively. This study compares the performance of two Machine Learning algorithms, Random Forest and Gradient Boosting, in predicting CKD using the Chronic Kidney Disease data dataset from Kaggle. The research steps include data exploration, Preprocessing using normalization and Synthetic Minority Over-sampling Technique (SMOTE), and data splitting using the Stratified Shuffle Split method. The models were evaluated using performance metrics such as accuracy, precision, recall, and F1-score to assess the effectiveness of each algorithm in disease classification. The results show that the Random Forest algorithm achieved an accuracy of 98.36%, outperforming Gradient Boosting, which reached 96.56%. Thus, Random Forest proves to be more effective in predicting CKD and can serve as a foundation for developing clinical decision support systems. This study's findings are expected to help healthcare professionals improve early diagnosis accuracy, enabling faster and more efficient treatment.

Keywords: Chronic Kidney Disease, Machine Learning, Classification, Random Forest, Gradient Boosting.