

**ANALISIS PERBANDINGAN H2O AUTOML DAN GRID
SEARCH UNTUK OPTIMASI HYPERPARAMETER PADA
MODEL RANDOM FOREST DALAM PREDIKSI PENYAKIT
JANTUNG**

SKRIPSI

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



disusun oleh

EGI DIAN ADI PRADANA

22.11.4745

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS PERBANDINGAN H2O AUTOML DAN GRID
SEARCH UNTUK OPTIMASI HYPERPARAMETER PADA
MODEL RANDOM FOREST DALAM PREDIKSI PENYAKIT
JANTUNG**

SKRIPSI

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



disusun oleh
EGI DIAN ADI PRADANA
22.11.4745

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERBANDINGAN H2O AUTOML DAN GRID
SEARCH UNTUK OPTIMASI HYPERPARAMETER PADA
MODEL RANDOM FOREST DALAM PREDIKSI PENYAKIT
JANTUNG**

yang disusun dan diajukan oleh

EGI DIAN ADI PRADANA

22.11.4745

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Januari 2026

Dosen Pembimbing,



Ferian Fauzi Abdulloh M.kom
NIK. 190302276

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS PERBANDINGAN H2O AUTOML DAN GRID SEARCH
UNTUK OPTIMASI HYPERPARAMETER PADA MODEL RANDOM
FOREST DALAM PREDIKSI PENYAKIT JANTUNG

yang disusun dan diajukan oleh

EGIDIAN ADI PRADANA

22.11.4745

Telah dipertabankan di depan Dewan Penguji
pada tanggal 27 Januari 2026.

Susunan Dewan Penguji

Nama Penguji

Hendra Kurniawan M.kom
NIK. 190302244

Dina Maulina M.kom
NIK. 190302250

Ferian Fauzi Abdullah M.kom
NIK. 190302276

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : EGI DIAN ADI PRADANA
NIM : 22.11.4745

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS PERBANDINGAN H2O AUTOML DAN GRID SEARCH UNTUK OPTIMASI HYPERPARAMETER PADA MODEL RANDOM FOREST DALAM PREDIKSI PENYAKIT JANTUNG

Dosen Pembimbing : Ferian Fauzi Abdulloh M.kom

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, 27 Januari 2026



Egi Dian Adi Pradana

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, saya ingin mengucapkan terimakasih terima kasih yang sebesar-besarnya kepada kedua orang tua saya atas segala doa, kasih sayang, semangat, serta dukungan moral dan finansial yang tak pernah putus selama saya menempuh pendidikan ini.

Setiap pengorbanan dan ketulusan yang di berikan telah menjadi sumber kekuatan yang mendorong saya untuk terus berjuang hingga akhirnya dapat mencapai titik pencapaian ini. Tanpa kehadiran, doa, dan dukungan mereka, keberhasilan ini tidak akan pernah terwujud.

Ucapan terima kasih yang tulus saya sampaikan kepada seluruh dosen dan staf Universitas Amikom Yogyakarta atas ilmu, bimbingan, dan fasilitas yang telah mendukung proses belajar saya. Dedikasi dan perhatian yang diberikan menjadi bekal berharga dalam perjalanan akademik saya hingga tugas akhir ini dapat terselesaikan.

Tidak lupa, saya mengucapkan terimakasih kepada teman-teman seperjuangan atas kebersamaan, dukungan, bantuan, dan semangat yang di berikan sepanjang perjalanan perkuliahan hingga penyusunan tugas akhir ini. Kehadiran kalian menjadikan perjalanan ini lebih bermakna dan penuh kenangan.

Terakhir, saya juga menyampaikan penghargaan dan terima kasih kepada semua pihak yang tidak dapat saya sebutkan satu per satu, baik secara langsung maupun tidak langsung. Semoga segala kebaikan yang telah kalian berikan dibalas dengan keberkahan dan balasan yang berlipat ganda.

Dengan penuh rasa syukur dan kerendahan hati, karya ini saya persembahkan

KATA PENGANTAR

Segala puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, sehingga skripsi berjudul "Analisis Perbandingan H2O Automl Dan Grid Search Untuk Optimasi Hyperparameter Pada Model Random Forest Dalam Prediksi Penyakit Jantung" dapat terselesaikan dengan baik.

Saya berharap hasil penelitian ini dapat memberikan manfaat bagi berbagai pihak, khususnya dalam pengembangan ilmu di bidang analisis perbandingan hyperparameter.

Penelitian ini disusun sebagai salah satu syarat untuk menyelesaikan program studi Teknik Komputer di Universitas Amikom Yogyakarta. Dalam proses penyusunan skripsi ini, saya menyadari bahwa pencapaian ini tidak akan terwujud tanpa bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, saya ingin mengucapkan terima kasih kepada:

1. Bapak, Ibu, serta Keluarga Besar saya atas dukungan moral dan kasih sayang yang tak pernah putus tiada hentinya.
2. Bapak Ferian Fauzi Abdulloh M.kom yang memberikan banyak arahan, bimbingan, dan motivasi selama proses penyusunan penelitian ini.
3. Teman-teman terdekat yang telah memberikan semangat, bantuan, dan kebersamaan yang berarti selama proses penyusunan skripsi ini.
4. Tidak lupa, terima kasih juga untuk diri sendiri yang sudah berjuang sejauh ini. Meski sering merasa lelah dan ingin menyerah, akhirnya tetap memilih untuk melangkah dan menyelesaikan perjalanan ini. Demikian kata pengantar ini Terima Kasih untuk semuanya.

Yogyakarta, 27 Januari 2026

Penulis

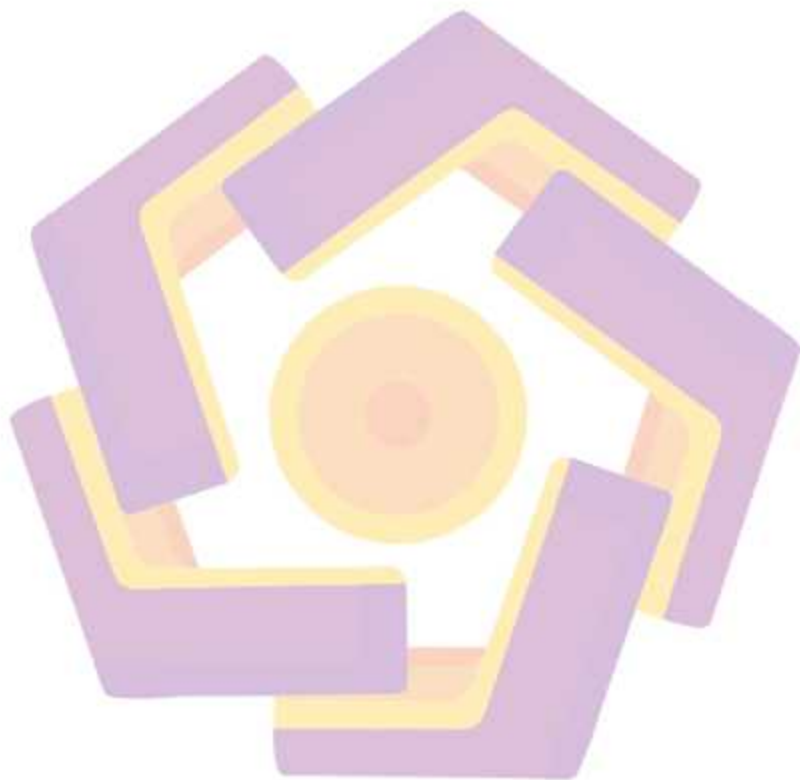
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xiv
INTISARI	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5

2.2	Dasar Teori.....	11
2.2.1	Data Mining	11
2.2.2	Machine Learning	11
2.2.3	Klasifikasi	12
2.2.4	Preprocessing	12
2.2.5	Python	14
2.2.6	Random forest	14
2.2.7	Hyperparameter.....	15
2.2.8	H2o Automl	15
BAB III METODE PENELITIAN		16
3.1	Objek Penelitian.....	16
3.2	Alur Penelitian	17
3.2.1	Pengumpulan Data	18
3.2.2	Preprocessing	18
3.2.3	Split Data (Train-Test Split).....	20
3.2.4	SMOTEENN	20
3.2.5	Feature Selection (Random Forest Feature Importance)	21
3.2.6	Proses Bercabang (Modeling)	21
3.2.7	Evaluasi Dan Perbandingan Model	22
3.3	Alat Dan Bahan	23
3.3.1	Data Penelitian	23
3.3.2	Perangkat Keras	23
3.3.3	Perangkat Lunak	23

BAB IV HASIL DAN PEMBAHASAN	24
4.1 Deskripsi Dataset	24
4.2 Tahapan Preprocessing Data	25
4.2.1 Pemeriksaan Missing Value	25
4.2.2 Penanganan Outlier Dengan Metode IQR	26
4.2.3 Transformasi Logaritmik Untuk Data Skewed	26
4.2.4 Normalisasi Data	26
4.2.5 Penghapusan Fitur Dengan Korelasi Tinggi	28
4.2.6 Penyeimbangan Data Dengan SMOTEENN	28
4.3 Pembagian Data	29
4.4 Seleksi Fitur Menggunakan Random Forest	29
4.5 Hasil Optimasi dan Evaluasi Random Forest	30
4.5.1 Hasil Grid Search Random Forest	30
4.5.2 Confusion Matrix Random Forest	31
4.5.3 Evaluasi Kinerja Random Forest	31
4.6 Hasil Optimasi dan Evaluasi H2O DRF	32
4.6.1 Confusion Matrix H2O DRF	32
4.6.2 Evaluasi Kinerja H2O DRF	33
4.7 Hasil Optimasi Dan Evaluasi Random forest Default	33
4.7.1 Confusion Matrix Random Forest Default	34
4.7.2 Evaluasi Kinerja Random Forest Default	34
4.8 Perbandingan Kinerja Model	35
4.8.1 Perbandingan Waktu komputasi masing masing model	36
BAB V PENUTUP	37
5.1 Kesimpulan	37

5.2 Saran	38
REFERENSI	39
LAMPIRAN	47



DAFTAR TABEL

Tabel 2.1. Tabel Keaslian	8
Tabel 4.1. Dataset	24
Tabel 4.2. Jumlah Missing Value Setelah Imputasi.....	25
Tabel 4.3. Jumlah Outlier Perfitur	26
Tabel 4.4. Hasil Rata-Rata Nilai Setelah Normalisasi	27
Tabel 4.5. Hasil Standar Deviasi Setelah Normalisasi.....	28
Tabel 4.6. Nilai Sebelum SMOTEENN.....	28
Tabel 4.7. Nilai Setelah SMOTEENN.....	29
Tabel 4.8. Hasil Split Data.....	29
Tabel 4.9. Dimensi Setelah Selection Feature	30
Tabel 4.10. Hasil Best Parameter Grid Search	30
Tabel 4.11. Hasil Evaluasi Random Forest.....	31
Tabel 4.12. Hasil Best Parameter H2O	32
Tabel 4.13. Hasil Evaluasi H2O DRF.....	33
Tabel 4.14. Hasil Best Random Forest Default	33
Tabel 4.15. Hasil Evaluasi Random Forest Default.....	34
Tabel 4.16. Hasil Evaluasi Random Forest Grid Search Dan H2o DRF	36

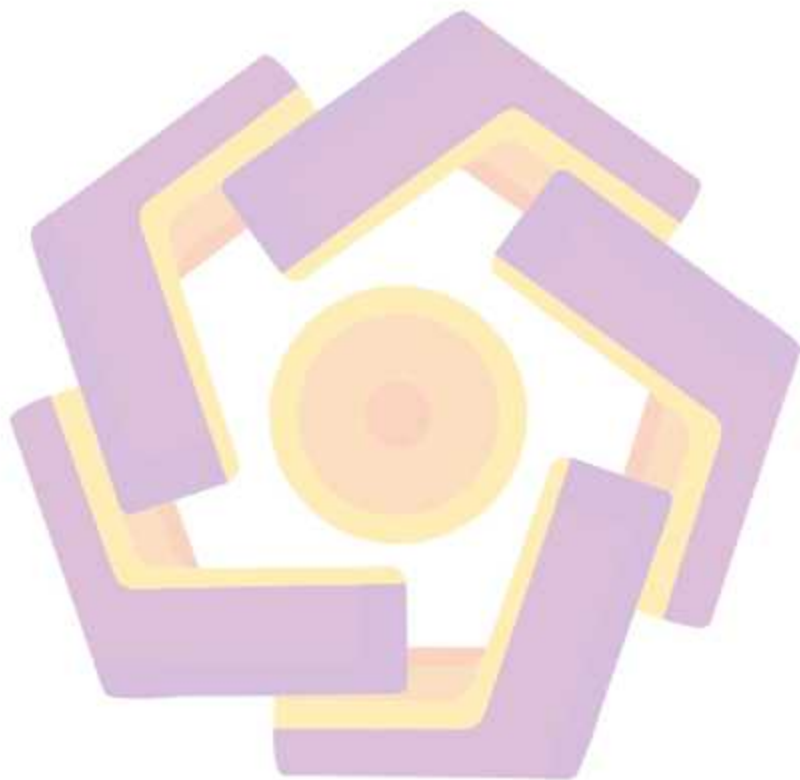
DAFTAR GAMBAR

Gambar 2.1. Pohon Keputusan	14
Gambar 2.2. Model H2O	15
Gambar 3.1. Alur Penelitian	17
Gambar 3.2. IQR.....	19
Gambar 4.1. Heatmap Korelasi Fitur	28
Gambar 4.2. Confusion Matrix Random Forest.....	31
Gambar 4.3. Confusion Matrix H2O DRF.....	32
Gambar 4.4. Confusion Matrix Random Forest Default.....	34
Gambar 4.5. Hasil CURVE RF (Grid Search), H2O DRF Dan RF Default	35
Gambar 4.6. Hasil Waktu Komputasi Dari Ketiga Model	36

DAFTAR LAMPIRAN

Lampiran 1. Link Drive Uji Machine Learning

44



DAFTAR LAMBANG DAN SINGKATAN



ML	Machine Learning
RF	Random Forest
IQR	interquartile Range
GR	Grid Search
SMOTE	synthetic minority over-sampling technique
ENN	Edited Nearest Neighbors
SMOTENN	Smote+Enn
FN	False Negative
FP	False Positif
TN	True Negative
TP	True Positif
AUC	Area Under CURVEa
ROC	Receiver Oprating Characteristic
H2O	H2o.ai
DRF	Distributed Random Forest

DAFTAR ISTILAH



Diagnosa	proses mengidentifikasi penyakit, kondisi, atau cedera dari tanda dan gejalanya.
Diagram	representasi simbolis informasi dalam bentuk geometri 2 dimensi sesuai teknik visualisasi.
Data	sekelompok informasi atau fakta mentah yang dapat berupa simbol, angka, kata-kata, atau citra.
Dokumen	sesuatu yang tertulis atau tercetak yang dapat dipergunakan sebagai bukti atau keterangan.

INTISARI

Penelitian ini bertujuan untuk membandingkan kinerja algoritma Random Forest yang dioptimasi menggunakan dua opsi yang berbeda, yaitu Grid search dan H2O, dalam hal melakukan klasifikasi pada resiko penyakit jantung. Dataset yang akan di gunakan terdiri dari 794 data dengan 14 fitur dengan 1 sebagai variable target.

Proses pre-processing mencakup penanganan missing value, standarisasi, serta pemilihan 10 fitur yang paling berpengaruh menggunakan metode Recursive Feature Elimination (RFE).

Model pertama optimasi yang digunakan Grid Search CV, sedangkan model ke dua menggunakan H2O DRF dengan tuning otomatis.

Hasil akhir penelitian ini ialah melakukan perbandingan terhadap kedua optimasi hyperparameter pada model Random Forest

Dari penelitian ini menunjukan hasil sebagai berikut hasil Grid Search Random Forest accuracy 0.8452, F1-Score 0.8514, AUC 0.9364 dan H2o Accuracy 0.8452, F1-Score 0.8490, AUC 0.9368

Disimpulkan bahwa masing masing model memiliki kemampuan yang sangat baik dalam klasifikasi penyakit jantung. Random Forest dengan Grid Search memberikan keunggulan fleksibilitas pengaturan hyperparameter, sedangkan H2o DRF sangat unggul dalam segi kemudahan dan automasi proses pemodelan

Kata kunci: Random Forest, Grid Search, H2O AutoML, Klasifikasi Penyakit Jantung, Hyperparameter Optimization.

ABSTRACT

This study aims to compare the performance of the Random Forest algorithm optimized using two different options, Grid Search and H2O, in classifying heart disease risk. The dataset used consists of 794 data points with 14 features, with 1 as the target variable. The preprocessing process includes handling missing values, standardization, and selecting the 10 most influential features using the Recursive Feature Elimination (RFE) method.

The first optimization model uses Grid SearchCV, while the second uses H2O DRF with automatic tuning.

The final result of this study was a comparison of the two hyperparameter optimizations in the Random Forest model. This study showed the following results: Grid Search Random Forest accuracy 0.8452, F1-Score 0.8514, AUC 0.9364, and H2o accuracy 0.8452, F1-Score 0.8490, and AUC 0.9368.

It was concluded that each model demonstrated excellent performance in classifying heart disease. Random Forest with Grid Search offered the advantage of flexible hyperparameter settings, while H2o DRF excelled in terms of simplicity and automation of the modeling process.

Keyword: *Random Forest, Grid Search, H2O AutoML, Heart Disease Classification, Hyperparameter, Optimization.*