



SKRIPSI

**PERBANDINGAN OPTIMASI ALGORITMA
RANDOM FOREST MENGGUNAKAN RANDOM
SEARCH DAN GRID SEARCH PADA
KLASIFIKASI KUALITAS AIR BERSIH**

ADI FAJRI FIRMANSYAH
NPM 21081010175

DOSEN PEMBIMBING
Dr. Basuki Rahmat, S.Si., MT.
Muhammad Muharrom Al Haromainy, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

**PERBANDINGAN OPTIMASI ALGORITMA
RANDOM FOREST MENGGUNAKAN RANDOM
SEARCH DAN GRID SEARCH PADA
KLASIFIKASI KUALITAS AIR BERSIH**

ADI FAJRI FIRMANSYAH

NPM 21081010175

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si., MT.

Muhammad Muharrom Al Haromainy, S.Kom., M.Kom.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

SURABAYA

2025

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

PERBANDINGAN OPTIMASI ALGORITMA RANDOM FOREST MENGUNAKAN RANDOM SEARCH DAN GRID SEARCH PADA KLASIFIKASI KUALITAS AIR BERSIH

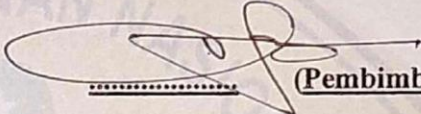
Oleh :

ADI FAJRI FIRMANSYAH
NPM. 21081010175

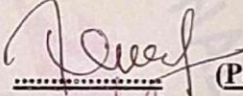
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 15 Juli 2025.

Menyetujui

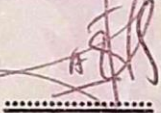
Dr. Basuki Rahmat, S.Si. MT.
NIP. 19690723 202121 1 002


..... (Pembimbing I)

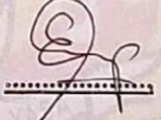
Muhammad Muharrom Al Haromainy,
S.Kom., M.Kom.
NIP. 19950601 202203 1 006


..... (Pembimbing II)

Fetty Tri Anggraeny, S.Kom. M.Kom.
NIP. 19820211 202121 2 005


..... (Ketua Penguji)

Eka Prakarsa Mandyartha, S.T., M.Kom.
NIP. 19880525 201803 1 001


..... (Anggota Penguji)

Mengetahui
Dekan Fakultas Ilmu Komputer


.....
Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

PERBANDINGAN OPTIMASI ALGORITMA RANDOM FOREST MENGUNAKAN RANDOM SEARCH DAN GRID SEARCH PADA KLASIFIKASI KUALITAS AIR BERSIH

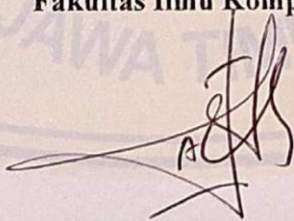
Oleh :

**ADI FAJRI FIRMANSYAH
NPM. 21081010175**

Menyetujui,

Koordinator Program Studi Informatika

Fakultas Ilmu Komputer



Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : ADI FAJRI FIRMANSYAH
NPM : 21081010175
Program Studi : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 15 Juli 2025

Yang Membuat Pernyataan,



Adi Fajri Firmansyah
NPM. 21081010175

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Adi Fajri Firmansyah / 21081010175
Judul Skripsi : Perbandingan Optimasi Algoritma Random Forest
Menggunakan Random Search dan Grid Search pada
Klasifikasi Kualitas Air Bersih
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si., MT.
2. Muhammad Muharrom Al Haromainy, S.Kom.,
M.Kom.

Air bersih layak konsumsi merupakan kebutuhan dasar manusia yang krusial bagi kesehatan. Di Indonesia, tantangan terhadap kualitas dan akses air bersih masih tinggi karena keterbatasan infrastruktur dan pencemaran lingkungan. Penelitian ini bertujuan untuk membandingkan efektivitas optimasi algoritma Random Forest menggunakan Random Search dan Grid Search dalam klasifikasi kualitas air bersih. Dataset yang digunakan adalah “Water Quality and Potability” yang diperoleh dari laman Kaggle, terdiri atas 3.276 data dengan 9 fitur parameter dan 1 variabel target. Seluruh proses klasifikasi dilakukan pada platform Google Colab. Hasil evaluasi menunjukkan bahwa model Random Forest tanpa optimasi menghasilkan akurasi sebesar 84.12% dalam waktu komputasi 3 detik. Sementara itu, model dengan optimasi Random Search menghasilkan akurasi 84.38% dalam waktu 2 jam 2 menit 54 detik, dan Grid Search mencapai akurasi tertinggi 85.62% dengan waktu komputasi 5 jam 12 menit. Berdasarkan evaluasi keseluruhan, optimasi hyperparameter terbukti mampu meningkatkan performa klasifikasi, terutama pada Grid Search. Namun demikian, peningkatan akurasi tersebut diikuti oleh lonjakan waktu komputasi yang signifikan, sehingga perlu pertimbangan efisiensi dalam penerapannya.

Kata Kunci : Klasifikasi, Air Bersih, *Machine Learning*, Random Forest, Random Search, Grid Search.

Halaman ini sengaja dikosongkan

ABSTRACT

Nama Mahasiswa / NPM : Adi Fajri Firmansyah / 21081010175
Judul Skripsi : Comparison of Random Forest Algorithm
Optimization Using Random Search and Grid Search
in Potable Water Quality Classification
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si., MT.
2. Muhammad Muharrom Al Haromainy, S.Kom.,
M.Kom.

Potable water suitable for consumption is a fundamental human need that plays a crucial role in health. In Indonesia, challenges regarding water quality and accessibility remain significant due to infrastructure limitations and environmental pollution. This study aims to compare the effectiveness of Random Forest algorithm optimization using Random Search and Grid Search in clean water quality classification. The dataset used is “Water Quality and Potability” obtained from the Kaggle page, consisting of 3,276 data with 9 parameter features and 1 target variable. The entire classification process is carried out on the Google Colab platform. Evaluation results show that the default Random Forest model achieved an accuracy of 84.12% with a computation time of 3 seconds. The Random Search-optimized model reached 84.38% accuracy in 2 hours 2 minutes and 54 seconds, while the Grid Search-optimized model achieved the highest accuracy of 85.62% in 5 hours 12 minutes. Overall, hyperparameter optimization significantly improved classification performance, particularly with Grid Search. However, this improvement was accompanied by a substantial increase in computation time, which should be considered when selecting optimization methods.

Keywords : *Classification, Water Quality, Machine Learning, Random Forest, Random Search, Grid Search.*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji dan syukur selalu penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul **"PERBANDINGAN OPTIMASI ALGORITMA RANDOM FOREST MENGGUNAKAN RANDOM SEARCH DAN GRID SEARCH PADA KLASIFIKASI KUALITAS AIR BERSIH"** sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa Timur.

Penyelesaian skripsi ini tentunya tidak lepas dari peran, dukungan, serta bimbingan dari berbagai pihak. Oleh sebab itu, penulis mengucapkan rasa terima kasih yang mendalam kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas Pembangunan Nasional Veteran Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
3. Ibu Fetty Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur dan dosen penguji satu saya yang senantiasa meluangkan waktu dan tenaga dalam penyempurnaan skripsi
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom. selaku koordinator skripsi Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.
5. Ibu Yisti Vita Via, S.ST., M.Kom. selaku dosen wali saya.
6. Bapak Dr. Basuki Rahmat, S.Si. M.T. selaku dosen pembimbing satu saya yang senantiasa sabar dalam membimbing dan memberi arahan selama pengerjaan skripsi hingga selesai.
7. Bapak Muhammad Muharrom Al Haromainy, S.Kom., M.Kom. selaku dosen pembimbing dua saya yang senantiasa sabar dalam membimbing dan memberi arahan selama pengerjaan skripsi hingga selesai.

8. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom. selaku dosen penguji dua saya yang senantiasa meluangkan waktu dan tenaga untuk memberikan penilaian serta masukan penyempurnaan skripsi.
9. Ibu Anita Turisia, Bapak Budi Harja, dan Nesya selaku orang tua dan adik saya tercinta yang selalu memberikan dukungan yang sangat luar biasa kepada saya.
10. Teman-teman dekat saya yang memberi dukungan selama pengerjaan skripsi hingga selesai.
11. Seluruh keluarga besar Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan, sehingga penulis sangat mengharapkan masukan dan saran yang membangun untuk penyempurnaan di masa mendatang. Besar harapan penulis agar karya ini dapat memberikan manfaat bagi pembaca serta pihak-pihak yang berkepentingan.

Surabaya, 15 Juli 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	v
ABSTRAK	vii
ABSTRACT	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Pendahulu	7
2.2 Landasan Teori.....	13
2.2.1 Air bersih dapat diminum	13
2.2.2 Klasifikasi	15
2.2.3 Algoritma Random Forest.....	17
2.2.4 Algoritma Random Search.....	19
2.2.5 Algoritma Grid Search	22
2.2.6 Confusion Matrix	23
2.2.7 Python	25
BAB III METODOLOGI PENELITIAN	27
3.1 Tahapan Penelitian	27
3.2 Pengumpulan Dataset.....	29
3.3 Prapemrosesan Data	30

3.4	Pemodelan dengan Random Forest	37
3.5	Optimasi Random Forest	44
3.5.1	Optimasi Model Random Forest dengan Random Search	44
3.5.2	Optimasi Model Random Forest dengan Grid Search	50
3.6	Evaluasi dan Perbandingan Model	52
BAB IV HASIL DAN PEMBAHASAN		57
4.1	Pengumpulan Dataset	57
4.2	Prapemrosesan Data	58
4.2.1	Pembersihan Data	58
4.2.2	Transformasi Data	60
4.2.3	Penyeimbangan Data	61
4.2.4	Pembagian Data	63
4.3	Pemodelan Random Forest	64
4.4	Optimasi Model Random Forest	65
4.4.1	Optimasi Model Random Forest dengan Random Search	65
4.4.2	Optimasi Model Random Forest dengan Grid Search	66
4.5	Evaluasi dan Perbandingan Model	68
4.5.1	Model Random Forest	68
4.5.2	Model Random Forest Optimasi Random Search	70
4.5.3	Model Random Forest Optimasi Grid Search	72
4.5.4	Perbandingan Model Klasifikasi	75
BAB V PENUTUP		81
5.1	Kesimpulan	81
5.2	Saran	82
DAFTAR PUSTAKA		83
LAMPIRAN		87

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Random Forest	19
Gambar 2. 2 Mekanisme kerja Random Search	21
Gambar 2. 3 Mekanisme kerja Grid Search	23
Gambar 3. 1 Diagram Alur Penelitian.....	27
Gambar 3. 2 Diagram alur pengumpulan dataset.....	29
Gambar 3. 3 Dataset <i>Water Quality and Potability</i>	29
Gambar 3. 4 Diagram alur prapemrosesan data	30
Gambar 3. 5 Diagram alur Random Forest	37
Gambar 3. 6 Pohon 1 Random Forest	41
Gambar 3. 7 Pohon 2 Random Forest	42
Gambar 3. 8 Diagram alur optimasi RF-RS	44
Gambar 3. 9 Diagram alur RF-GS	50
Gambar 3. 10 Diagram alur evaluasi model.....	52
Gambar 4. 1 Dataset	57
Gambar 4. 2 Jumlah <i>missing value</i> saat <i>imputasi</i>	58
Gambar 4. 3 Dataset setelah imputasi mean	59
Gambar 4. 4 Dataset setelah transformasi	61
Gambar 4. 5 Distribusi dataset	61
Gambar 4. 6 Distribusi data seimbang	62
Gambar 4. 7 Dataset setelah penyeimbangan data	63
Gambar 4. 8 Jumlah data latih dan uji	64
Gambar 4. 9 Kombinasi hyperparameter terbaik random search.....	66
Gambar 4. 10 Kombinasi hyperparameter terbaik grid search	67
Gambar 4. 11 Confusion matrix RF	69
Gambar 4. 12 Laporan klasifikasi RF	69
Gambar 4. 13 Confusion matrix RF-RS.....	71
Gambar 4. 14 Laporan klasifikasi RF-RS	72
Gambar 4. 15 Confusion matrix RF-GS	73
Gambar 4. 16 Laporan klasifikasi RF-GS.....	74

Gambar 4. 17 Perbandingan akurasi model.....	76
Gambar 4. 18 Perbandingan waktu komputasi.....	78

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	24
Tabel 3. 1 Dataset terdapat missing value.....	31
Tabel 3. 2 Dataset hasil imputasi	31
Tabel 3. 3 Dataset sebelum normalisasi	32
Tabel 3. 4 Dataset hasil normalisasi.....	33
Tabel 3. 5 Dataset sebelum penyeimbangan data	35
Tabel 3. 6 Dataset hasil penyeimbangan data	35
Tabel 3. 7 Pembagian dataset.....	36
Tabel 3. 8 Dataset latih.....	36
Tabel 3. 9 Dataset uji	36
Tabel 3. 10 Parameter default random forest	39
Tabel 3. 11 Dataset latih random forest	39
Tabel 3. 12 Hyperparameter default dalam contoh.....	40
Tabel 3. 13 Hasil <i>aggregating</i> RF <i>default</i>	43
Tabel 3. 14 Hasil evaluasi data latih RF <i>default</i>	43
Tabel 3. 15 Confusion matrix data latih RS <i>default</i>	43
Tabel 3. 16 Hyperparameter untuk optimasi random forest	44
Tabel 3. 17 Kombinasi hyperparameter	46
Tabel 3. 18 Hasil <i>aggregating</i> Kombinasi 1 RF-RS	47
Tabel 3. 19 Hasil evaluasi data latih Kombinasi 1 RF-RS.....	47
Tabel 3. 20 Confusion matrix Kombinasi 1 RF-RS.....	47
Tabel 3. 21 Hasil aggregating Kombinasi 3 RF-RS.....	48
Tabel 3. 22 Hasil evaluasi data latih Kombinasi 3 RF-RS.....	49
Tabel 3. 23 Confusion matrix Kombinasi 1 RF-RS.....	49
Tabel 3. 24 Perbandingan model optimasi Random Search	49
Tabel 3. 25 Kombinasi hyperparameter RF-GS.....	51
Tabel 3. 26 Model Akurasi Grid Search	52
Tabel 3. 27 Skenario model klasifikasi	52
Tabel 3. 28 <i>Confusion matrix</i>	53

Tabel 3. 29 Metrik evaluasi	53
Tabel 3. 30 Dataset uji	54
Tabel 3. 31 <i>Confusion matrix</i> random forest	54
Tabel 3. 32 Confusion matrix RF-RS	55
Tabel 3. 33 <i>Confusion matrix</i> RF-GS	55
Tabel 3. 34 Perbandingan evaluasi	56
Tabel 4. 1 Kombinasi hyperparameter terbaik	75
Tabel 4. 2 Perbandingan nilai matrik evaluasi	76
Tabel 4. 3 Hasil waktu komputasi	77