

**GRID SEARCH PADA RANDOM FOREST UNTUK PREDIKSI
KUALITAS AIR YANG LAYAK DIKONSUMSI OLEH TUBUH
MANUSIA**

SKRIPSI

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



disusun oleh

IKA SEPTIANA F

20.11.3591

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**GRID SEARCH PADA RANDOM FOREST UNTUK PREDIKSI
KUALITAS AIR YANG LAYAK DIKONSUMSI OLEH TUBUH
MANUSIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

IKA SEPTIANA F

20.11.3591

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**GRID SEARCH PADA RANDOM FOREST UNTUK PREDIKSI
KUALITAS AIR YANG LAYAK DIKONSUMSI OLEH TUBUH
MANUSIA**

yang disusun dan diajukan oleh

Ika Septiana F

20.11.3591

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Agustus 2024

Dosen Pembimbing,



Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276

HALAMAN PENGESAHAN
SKRIPSI
GRID SEARCH PADA RANDOM FOREST UNTUK PREDIKSI
KUALITAS AIR YANG LAYAK DIKONSUMSI OLEH TUBUH
MANUSIA

yang disusun dan diajukan oleh

Ika Septiana F

20.11.3591

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Agustus 2024

Susunan Dewan Penguji

Nama Penguji

Jeki Kuswanto, M.Kom
NIK. 190302456

Bayu Setiaji, M.Kom
NIK. 190302216

Ferian Fauzi Abdullah, M.Kom
NIK. 190302276

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Agustus 2024

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 : Ika Septiana F
NIM : 20.11.3591

Menyatakan bahwa Skripsi dengan judul berikut:

**GRID SEARCH PADA RANDOM FOREST UNTUK PREDIKSI
KUALITAS AIR YANG LAYAK DIKONSUMSI OLEH TUBUH MANUSIA**

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, 26 Agustus 2024

Yang Menyatakan,



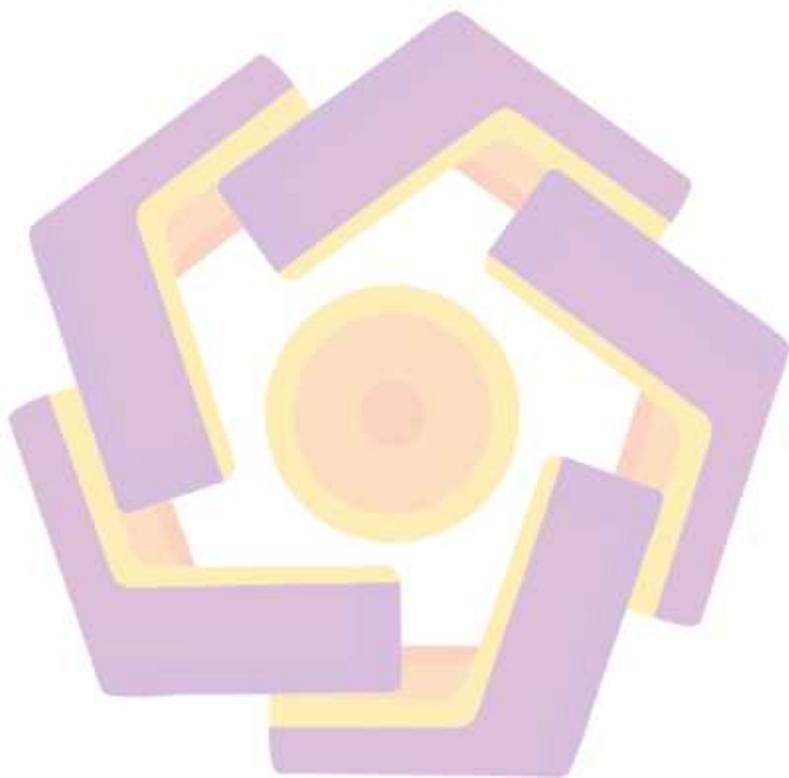

Ika Septiana F

HALAMAN PERSEMBAHAN

Alhamdulillah rabbil'alam, segala puji dan syukur kepada Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, yang telah memberikan hidayah dan kemudahan dalam setiap langkah perjalanan penulisan skripsi ini. Sholawat serta salam selalu tercurah limpahkan kepada baginda Nabi Muhammad SAW. Kehadiran berbagai dukungan dari semua pihak membuat penulis mampu menyelesaikan skripsi ini dengan baik. Ucapan terima kasih yang sebesar-besarnya penulis haturkan kepada:

1. Dengan mengucapkan syukur kepada Allah SWT atas segala nikmat dan pertolongan-Nya. Setiap langkah yang penulis ambil, setiap tantangan yang penulis hadapi adalah wujud kasih sayang dan petunjuk-Nya selama penulis menyusun skripsi.
2. Kedua orang tua penulis Bapak Fadoil dan Ibu Siti Halimah yang tiada henti memberikan doa, kasih sayang, nasihat dan dukungannya, sehingga penulis bisa mencapai titik ini.
3. Teruntuk adik penulis yaitu Ahmad Zul Fahri yang selalu menghibur dan menjadi penyemangat setiap langkah penulis.
4. Kepada Bapak Ferian Fauzi Abdulloh, M. Kom yang telah memberikan arahan dan bimbingan sehingga penulis dapat menyelesaikan skripsi ini.
5. Terima kasih kepada Ibu dan Bapak dosen yang telah memberikan ilmu dan inspirasi yang sangat berharga dalam perjalanan akademik penulis.
6. Untuk teman-teman "SKIRPSHIT KELAR MAKSIMAL SMSTER 8", terima kasih atas kebersamaan baik suka maupun duka, dukungan dan tawa yang selalu menemani penulis dalam menyusun skripsi ini.
7. Teman-teman pengurus HMIF periode 2022/2023, terima kasih atas segala dukungan, kerja sama dan kebersamaan yang telah kita lalui bersama selama menjalankan kepengurusan. Terima kasih telah berjuang bersama dan sukses selalu.

8. Dan yang terakhir, terima kasih untuk diriku sendiri atas segala usaha, kerja keras, semangat juang dan kesabaran selama pengerjaan skripsi ini.



KATA PENGANTAR

Puji syukur kehadiran Allah SWT, atas segala rahmat, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“Grid Search Pada Random Forest untuk Prediksi Kualitas Air yang Layak dikonsumsi oleh Tubuh Manusia”** sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer Program Studi Informatika Universitas AMIKOM Yogyakarta.

Penulis menyadari bahwa dalam penyusunan skripsi ini tidak akan terwujud tanpa adanya bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S. Kom., M. Kom., Ph. D. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D, M. Kom. selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Bapak Ferian Fauzi Abdullah, M. Kom. selaku Dosen Pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi dalam penyusunan skripsi ini.
5. Bapak Fadoil dan Ibu Siti Halimah selaku kedua orang tua penulis yang telah memberikan doa, dukungan, dan motivasi kepada penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini terdapat kekurangan dan keterbatasan pengetahuan dalam penyusunan penelitian ini. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi perbaikan dan kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi penulis khususnya dan pembaca selanjutnya.

Yogyakarta, 26 Agustus 2024

Ika Septiana F

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
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	16

BAB III METODE PENELITIAN	21
3.1 Objek Penelitian	21
3.2 Alur Penelitian.....	21
3.2.1 Preprocessing Data.....	23
3.3 Alat dan Bahan.....	24
3.3.1 Alat Penelitian.....	24
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Persiapan Data.....	26
4.2 Pre-processing Data.....	28
4.2.1 Penanganan Missing Value.....	28
4.2.2 Random Oversampler	30
4.3 Pembagian Data dengan <i>K-Fold Cross-Validation</i>	34
4.4 Evaluasi Model.....	38
BAB V PENUTUP	44
5.1 Kesimpulan.....	44
5.2 Saran.....	44
REFERENSI	46

DAFTAR TABEL

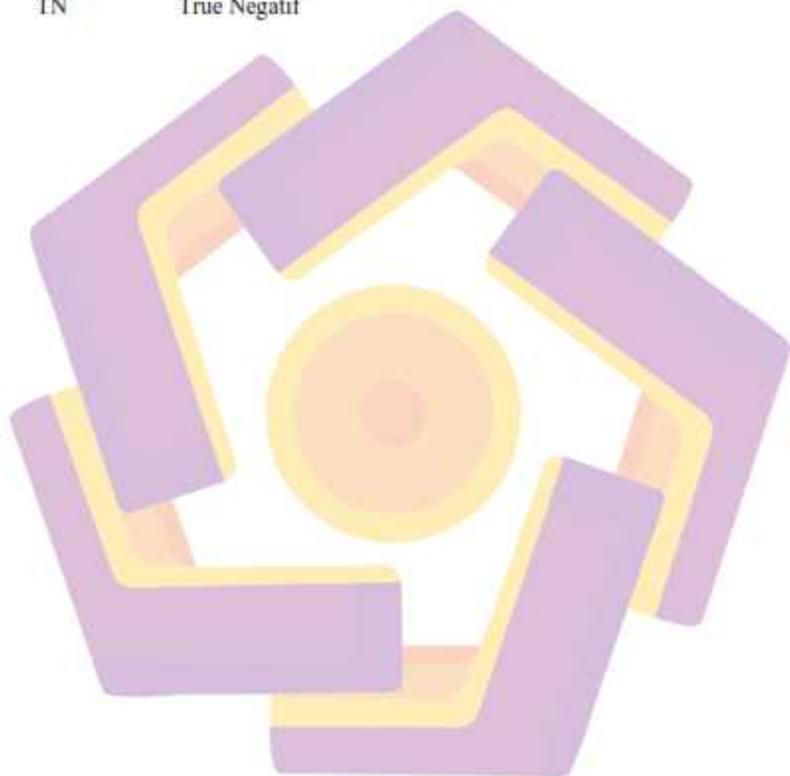
Tabel 2. 1 Keaslian Penelitian	8
Tabel 2. 2 Confusion Matrix	19
Tabel 3. 1 Parameter Dataset Kualitas Air.....	21
Tabel 3. 2 Tampilan Dataset	25
Tabel 4. 1 Tampilan 5 Data Teratas	26
Tabel 4. 2 Tipe Data.....	27
Tabel 4. 3 Data Sebelum Missing Value.....	29
Tabel 4. 4 Dataset Sesudah Missing Value.....	30
Tabel 4. 5 Hasil 5 K-Fold Cross-Validation Sebelum Metode Grid Search.....	36
Tabel 4. 6 Hasil 5 K-Fold Cross-Validation Setelah Metode Grid Search.....	37
Tabel 4. 7 Classification Report Model Random Forest.....	40
Tabel 4. 8 Classification Report Optimasi Grid Search.....	43
Tabel 4. 9 Perbandingan Model Random Forest Sebelum dan Setelah Penggunaan Metode Grid Search	43

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	22
Gambar 4. 1 Distribusi Kualitas Air	28
Gambar 4. 2 Cek Missing Value	28
Gambar 4. 3 Menangani Missing Value	29
Gambar 4. 4 Sebelum Penanganan Oversampling	31
Gambar 4. 5 Data Sebelum Oversampling	32
Gambar 4. 6 Setelah Penanganan Oversampling	33
Gambar 4. 7 Data Setelah Oversampling	33
Gambar 4. 8 Membuat Dataset Sintesis	34
Gambar 4. 9 Normalisasi Data	34
Gambar 4. 10 Inisiasi Pembagian Data K-fold	35
Gambar 4. 11 Pembagian Data Sebelum Grid Search	35
Gambar 4. 12 Pembagian Data Setelah Grid Search	37
Gambar 4. 13 Confusion Matrix Sebelum Optimasi	38
Gambar 4. 14 Confusion Matrix Sesudah Optimasi	41

DAFTAR LAMBANG DAN SINGKATAN

TP	True Positif
FN	False Negatif
FP	False Positif
TN	True Negatif



DAFTAR ISTILAH



Data Training	Data yang digunakan untuk melatih model
Data Testing	Kumpulan sampel data yang digunakan untuk menguji kemampuan model.
<i>Oversampling</i>	Teknik pemrosesan data yang memiliki ketidakseimbangan kelas.
<i>K-Fold Cross-Validation</i>	Teknik validasi silang dalam mengevaluasi kinerja model
<i>Hyperparameter Tuning</i>	Proses mengatur nilai <i>hyperparameter</i> sehingga mencapai kinerja model yang lebih baik.
<i>Grid Search</i>	Metode yang digunakan untuk mengevaluasi kinerja model dengan mencoba semua kombinasi dari <i>hyperparameter</i> .

INTISARI

Kualitas air merupakan faktor penting dalam kesehatan tubuh manusia. Air yang tidak layak dikonsumsi dapat menyebabkan berbagai macam penyakit. Penelitian ini menggunakan dataset yang diambil dari *platform kaggle* berisi beberapa parameter yang penting untuk mengukur apakah air yang dikonsumsi atau tidak mencakup informasi seperti *pH*, *Hardness*, *Solids*, *Chloramine*, *Sulfate*, *Conductivity*, *Organic_carbon*, *Trihalomethanes*, *Turbidity* dan *Potability*. Penelitian ini bertujuan untuk mengembangkan model prediksi kualitas air yang layak dikonsumsi menggunakan algoritma pembelajaran mesin, khususnya *Random Forest* serta penggunaan optimasi metode *Grid Search* untuk mengoptimalkan kinerja model agar bekerja dengan baik. Penelitian ini menerapkan beberapa tahapan *preprocessing data*, termasuk menangani *missing value*, *oversampling* untuk mengatasi ketidakseimbangan kelas, dan normalisasi data. Selanjutnya, dilakukan pembagian data menggunakan *k-fold cross-validation*. Hasil dari evaluasi model secara keseluruhan menggunakan model *random forest* tanpa optimasi metode *grid search* untuk memprediksi kualitas air yang layak dikonsumsi menghasilkan akurasi mencapai 88,2%, presisi 86,28%, recall 90,78% dan *f1-score* sebesar 88,47%. Sedangkan, penggunaan optimasi metode *grid search* menghasilkan akurasi mencapai 98,2%, lalu presisi sebesar 98,58%, recall mencapai 97,79% dan *F1-score* mencapai 98,18%.

Kata kunci: Kualitas Air, Random Forest, Grid Search, Oversampling, K-fold Cross-Validation

ABSTRACT

Water quality is an important factor in human health. Water that is not suitable for consumption can cause various diseases. This research uses a dataset taken from the kaggle platform containing several parameters that are important to measure whether water is consumable or not including information such as pH, Hardness, Solids, Chloramine, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, Turbidity and Potability. This research aims to develop a prediction model of water quality that is suitable for consumption using machine learning algorithms, specifically Random Forest and the use of Grid Search method optimisation to optimise model performance to work well. This research applies several stages of data preprocessing, including handling missing values, oversampling to overcome class imbalance, and data normalisation. Furthermore, data sharing was performed using k-fold cross-validation. The results of the overall model evaluation using the random forest model without optimisation of the grid search method to predict the quality of water suitable for consumption resulted in an accuracy of 88.2%, precision of 86.28%, recall of 90.78% and f1-score of 88.47%. Meanwhile, the use of grid search method optimisation resulted in accuracy reaching 98.2%, then precision of 98.58%, recall reaching 97.79% and F1-score reaching 98.18%.

Keywords: Water Quality, Random Forest, Grid Search, Oversampling, K-fold Cross-Validation