

**KLASIFIKASI KUALITAS AIR MINUM DENGAN METODE *LOGISTIC*
REGRESSION BERBASIS *GRID SEARCH OPTIMIZATION***



OLEH :

Tzalvano Syahputra Permana

09011181924014

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2023**

**KLASIFIKASI KUALITAS AIR MINUM DENGAN METODE *LOGISTIC*
REGRESSION BERBASIS *GRID SEARCH OPTIMIZATION***

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



OLEH :

Tzalvano Syahputra Permana

09011181924014

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2023**

LEMBAR PENGESAHAN

**KLASIFIKASI KUALITAS AIR MINUM DENGAN METODE *LOGISTIC*
REGRESSION BERBASIS *GRID SEARCH OPTIMIZATION***

TUGAS AKHIR

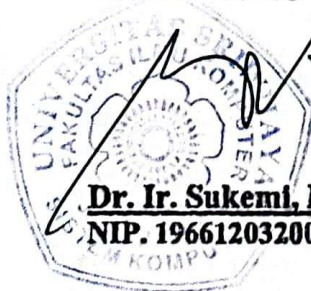
**Program Studi Sistem Komputer
Jenjang S1**

OLEH:

**TZALVANO SYAHPUTRA PERMANA
09011181924014**

**INDRALAYA, 8 JANUARI 2024
MENGETAHUI,**

Ketua Jurusan Sistem Komputer



**Dr. Ir. Sukemi, M.T.
NIP. 196612032006041001**

Pembimbing Tugas Akhir

**Dr. Ahmad Zarkasi, S.T., M.T.
NIP. 197908252023211007**

HALAMAN PERSETUJUAN

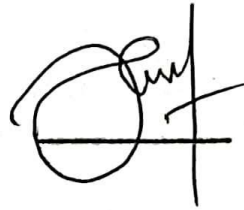
Telah diuji dan lulus pada :

Hari : Jumat

Tanggal : 29 Desember 2023

Tim Penguji :

1. Ketua Sidang : Ahmad Fali Oklilas, S.T., M.T.



2. Sekretaris Sidang : Nurul Afifah, S.Kom., M.Kom.



3. Penguji Sidang : Dr. Firdaus, S.T., M.Kom.



4. Pembimbing : Dr. Ahmad Zarkasi, S.T., M.T.

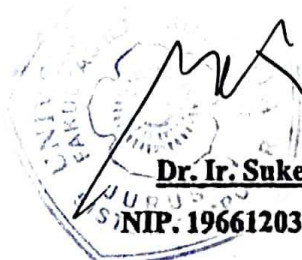


Mengetahui,

Ketua Jurusan Sistem Komputer

Dr. Ir. Sukemi, M.T.

NIP. 196612032006041001



HALAMAN PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Tzalvano Syahputra Permana

NIM : 09011181924014

Judul : *KLASIFIKASI KUALITAS AIR MINUM DENGAN METODE LOGISTIC REGRESSION BERBASIS GRID SEARCH OPTIMIZATION*

Hasil pengecekan *software ithenticate/ Turnitin* : 7%

Menyatakan bahwa Laporan Skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam Laporan Skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya. Demikian, pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.



Indralaya, Januari 2024



Tzalvano Syahputra Permana

NIM. 09011181924014

HALAMAN PERSEMBAHAN

“Hadiah Terbaik Adalah Apa yang Kamu Miliki, dan Takdir Terbaik Adalah Apa yang Sedang Kamu Jalani. Cara Kerja Tuhan Memang Aneh, Salah Satu Cara Memahaminya Hanyalah Dengan Bersyukur”

(Tzalvano Syahputra Permana)

“Angin Tidak Berhembus untuk Menggoyangkan Pepohonan, Melainkan Menguji Kekuatan Akarnya”

(Ali bin Abi Thalib)

Karya ini ku persembahkan untuk:

- Orang tuaku Mama (Sulistiani) tercinta yang tak pernah lelah untuk mengingatkan, memberikan semangat serta memberikan dukungan, perjuangan, dan memberikan motivasi dalam hidup ini.
- Nenek & keluargaku yang selalu mendukung, mengingatkan serta merelakan waktunya untuk memberikan canda dan tawa serta kasih sayang.
- Almamater Universitas Sriwijaya

KATA PENGANTAR

Puji syukur penulis ucapkan kehadiran Allah SWT yang telah memberikan rahmat dan karunianya sehingga penulis dapat menyelesaikan penulisan Skripsi yang berjudul **“Klasifikasi Kualitas Air Minum Dengan Metode *logistic Regression* Berbasis *Grid Search Optimization*”** dengan tepat waktu.

Pada kesempatan ini, penulis menyadari bahwa penulisan skripsi ini tidak akan terselesaikan tanpa adanya bimbingan, petunjuk, masukan dan bantuan dari berbagai pihak. Untuk itu, penulis dengan segala kerendahan hati ingin mengucapkan banyak terima kasih kepada pihak – pihak yang telah membantu dalam penyusunan skripsi ini, khususnya kepada :

1. Bapak Prof. Dr. Erwin, S.Si., M.si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya Sriwijaya.
3. Bapak Dr. Ahmad Zarkasi, S.T., M.T. yang telah bersedia memberikan bimbingan, pengarahan, inspirasi, dan motivasi serta saran dengan penuh keikhlasan dan kesabaran sehingga penulis dapat menyelesaikan skripsi ini.
4. Seluruh Dosen dan Staf Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Mama, Nenek, Bunda, Saudaraku dan Ai yang yang selalu mengingatkan dan memberikan motivasi bagi saya untuk menyelesaikan skripsi ini.
6. Rekan-rekan mahasiswa Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
7. Teman – teman kelompok bimbingan, teman – teman magang saya yang selalu berbagi keluh kesah dan saling memberikan motivasi.
8. Semua pihak yang telah memberikan bantuannya kepada penulis dalam menyelesaikan penulisan skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Akhir kata penulis mengucapkan terima kasih banyak pada semua pihak yang telah memberikan bantuan dan bimbingan kepada penulis. Semoga segala bantuan yang telah diberikan mendapatkan limpahan rahmat dan pahala dari Allah SWT. Aamiin.

Indralaya, 1 Januari 2023

A handwritten signature in black ink, consisting of stylized, cursive letters that appear to be 'Rifa'.

Penulis

CLASSIFICATION OF DRINKING WATER QUALITY USING LOGISTIC REGRESSION METHOD BASED ON GRID SEARCH OPTIMIZATION

Tzalvano Syahputra Permana (09011181924014)

Department of Computer Systems, Faculty of Computer Science, Sriwijaya University

Email : tzalvanosyahputrapermana@gmail.com

ABSTRACT

Clean water is a vital factor for human survival. The concept of adequate drinking water emphasizes water quality that meets health standards and can be consumed directly. Monitoring and managing water sources is key in maintaining water availability. These efforts are necessary to ensure the quality and continuity of water sources that meet the health standards required for human consumption. This research aims to improve the accuracy of drinking water quality classification through the implementation of a Logistic Regression model based on Grid-Search Optimization. This research uses a dataset from Kaggle and involves program simulations in Python. Model performance evaluation includes accuracy, sensitivity, and specificity. The main features used are pH, conductivity, total dissolved solids, and turbidity level. The results of this research are that the system performance in classifying drinking water shows a very good recall rate, reaching 98%. The best model performance was obtained with a combination of C parameters 0.001, solver liblinear, and max_iter 100, with an average Recall value of 98.70%, Precision 55.77%, Specificity 70.30%, Accuracy 78.11%, Error 21.89 %, And F1-Score 71.27%.

Keywords : *Logistic Regression, Grid-Search Optimization, Classification, Drinking Water*

**KLASIFIKASI KUALITAS AIR MINUM DENGAN METODE *LOGISTIC*
REGRESSION BERBASIS *GRID SEARCH OPTIMIZATION***

Tzalvano Syahputra Permana (09011181924014)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : tzalvanosyahputrapermana@gmail.com

ABSTRAK

Air bersih adalah faktor vital untuk kelangsungan hidup manusia. Konsep Air Minum yang Layak menekankan kualitas air yang memenuhi standar kesehatan dan bisa langsung dikonsumsi. Pemantauan dan pengelolaan sumber air menjadi kunci dalam menjaga ketersediaan air. Upaya ini diperlukan untuk memastikan kualitas dan kelangsungan sumber air yang memenuhi standar kesehatan yang diperlukan bagi konsumsi manusia. Penelitian ini bertujuan untuk meningkatkan akurasi klasifikasi kualitas air minum melalui implementasi model *Logistic Regression* berbasis *Grid-Search Optimization*. Penelitian ini menggunakan dataset dari Kaggle dan melibatkan simulasi program dengan Python. Evaluasi kinerja model mencakup akurasi, sensitivitas, dan spesifisitas. Fitur-fitur utama yang digunakan adalah pH, konduktivitas, total padatan terlarut, dan tingkat kekeruhan. Hasil dari penelitian ini yaitu kinerja sistem dalam melakukan klasifikasi air minum menunjukkan tingkat recall yang sangat baik, mencapai 98%. Performa model terbaik diperoleh dengan kombinasi parameter C 0.001, solver liblinear, dan max_iter 100, dengan rata-rata nilai Recall 98,70%, Presisi 55,77%, Spesifitas 70,30%, Akurasi 78,11%, Error 21,89%, Dan F1-Score 71,27%.

Kata Kunci : Logistic Regression, Grid-Search Optimization, Klasifikasi, Air Minum

DAFTAR ISI

LEMBAR PENGESAHAN	I
HALAMAN PERSETUJUAN	II
HALAMAN PERNYATAAN	III
HALAMAN PERSEMBAHAN	IV
KATA PENGANTAR	V
DAFTAR ISI	IX
DAFTAR GAMBAR	XII
DAFTAR TABEL	XIII
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	5
1.4. Tujuan	5
1.5. Sistematika Penulisan	5
BAB I PENDAHULUAN	5
BAB II TINJAUAN PUSTAKA	5
BAB III METODOLOGI PENELITIAN	6
BAB IV HASIL DAN ANALISIS	6
BAB V KESIMPULAN	6
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu	7
2.2. Kualitas air	11
2.2.1. Peraturan Menteri Kesehatan	11
2.2.2. Kelayakan air minum	12
2.2.2.1. Konduktivitas	14
2.2.2.2. Kekeruhan	15
2.2.2.3. Total Padatan Terlarut	16
2.3. <i>Machine Learning</i>	16
2.3.1. <i>Logistic Regression</i>	17
2.3.1.1. <i>Logistic Regression Biner</i>	20
2.4. <i>Hyperparameter-Optimization</i>	21
2.4.1. <i>Grid Search Optimization</i>	22
2.5. <i>Confusion Matrix</i>	23
2.5.1. Akurasi	23
2.5.2. Recall.....	23
2.5.3. Spesifisitas.....	23

2.5.4. Presisi	24
2.5.5. F1 Score.....	24
2.6. Arduino Uno	24
2.7. Modul PH-4502C.....	25
2.8. Modul <i>Total Dissolved Solids</i> SEN0244	26
2.9. Modul Kekeruhan MJKDZ	27
BAB III METODOLOGI.....	29
3.1. Pendahuluan.....	29
3.2. Kerangka Kerja	29
3.3. Studi Literatur	31
3.4. Perancangan Sistem	31
3.5. Perancangan Perangkat Lunak.....	32
3.5.1. Pengumpulan data	33
3.5.2. <i>Preprocessing</i>	35
3.5.2.1. <i>Data Cleaning</i>	35
3.5.2.2. Feature Scaler.....	36
3.5.3. Pembentukan Model.....	39
3.5.3.1. Pembagian data latih dan data uji	39
3.5.3.2. <i>Handling Class Imbalance</i>	40
3.5.3.3. <i>Logistic Regression</i>	42
3.5.3.4. <i>GridSearch</i>	42
3.5.4. Evaluasi Model.....	45
3.5.5. <i>Cross Validation</i>	45
3.5.6. <i>Confussion Matrix</i> visualization	46
3.6. Perancangan Hardware (Perangkat Keras)	46
3.7. Pengujian Sistem.....	48
3.7.1. Pengujian Perangkat Lunak.....	49
3.7.2. Pengujian Perangkat Keras.....	50
3.8. Validasi Performa Model.....	51
3.9. Analisa Sistem	52
3.10. Pengambilan Kesimpulan.....	52
BAB IV HASIL DAN ANALISA	53
4.1. Pendahuluan.....	53
4.2. Proses Pengumpulan Data.....	53
4.3. Pengujian Perangkat Lunak	53
4.3.1. <i>Tuning Hyperparameter Grid-Search Optimization</i>	53
4.3.2. Pengujian sensor keasaman dan kalibrasi	61

4.3.3.	Pengujian sensor konduktifitas dan kalibrasi	63
4.3.4.	Pengujian sensor kekeruhan dan kalibrasi.....	64
4.3.5.	Pengujian sensor <i>Total</i> padatan terlarut dan kalibrasi.....	66
4.4.	Pengujian sistem	68
4.5.	Analisa	69
BAB V KESIMPULAN DAN SARAN.....		71
5.1.	Kesimpulan	71
5.2.	Saran	71
DAFTAR PUSTAKA		72
LAMPIRAN		75

DAFTAR GAMBAR

Gambar 2.1 Peraturan Pemerintah Nomor 82 tahun 2001	12
Gambar 2.2 kategori dalam <i>Machine Learning</i>	17
Gambar 2.3 contoh rekayasa bentuk <i>Logistic Regression</i>	18
Gambar 2.4 contoh rekayasa bentuk Grid Search	22
Gambar 2.5 Arduino UNO	24
Gambar 2.6 Modul PH-4502C.....	26
Gambar 2.7 SEN0244	27
Gambar 2.8 Probe kekeruhan	28
Gambar 3.1 Kerangka Kerja Penelitian.....	30
Gambar 3.4 membaca data.....	34
Gambar 3.5 Lima data teratas	34
Gambar 3.6 Menampilkan data NaN	36
Gambar 3.7 Menampilkan data NaN	36
Gambar 3.8 data ph sebelum dan sesudah <i>scaling</i>	37
Gambar 3.9 data TDS sebelum dan sesudah <i>scaling</i>	38
Gambar 3.10 data Conductivity sebelum dan sesudah <i>scaling</i>	38
Gambar 3.11 data Turbidity sebelum dan sesudah <i>scaling</i>	38
Gambar 3.12 Pembagian data latih dan data uji.....	39
Gambar 3.13 SMOTE	41
Gambar 3.14 sesudah SMOTE.....	41
Gambar 3.15 model logistic regression.....	42
Gambar 3.16 Pengaplikasian <i>Grid Search</i>	45
Gambar 4.1 Grid Search parameter	54
Gambar 4.26 Screenshot <i>confusion matrix</i>	69
Gambar 4.26 Screenshot Classification Report	69

DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian Terdahulu	9
Tabel 4.1 Hasil percobaan dengan parameter C 0.001, solver 1bfgs	55
Tabel 4.2 Hasil percobaan dengan parameter C 0.001, solver liblinear.	55
Tabel 4.3 Hasil percobaan dengan parameter C 0,001, solver saga.	56
Tabel 4.4 Hasil percobaan dengan parameter C 0.01, solver 1bfgs..	56
Tabel 4.5 Hasil percobaan dengan parameter C 0,01, solver liblinear	56
Tabel 4.6 Hasil percobaan dengan parameter C 0.01, solver saga ...	57
Tabel 4.7 Hasil percobaan dengan parameter C 0,1, solver 1bfgs....	57
Tabel 4.8 Hasil percobaan dengan parameter C 0.1, solver liblinear	57
Tabel 4.9 Hasil percobaan dengan parameter C 0.1, solver saga	58
Tabel 4.10 Hasil percobaan dengan parameter C 1, solver 1bfgs....	58
Tabel 4.11 Hasil percobaan dengan parameter C 1, solver liblinear	58
Tabel 4.12 Hasil percobaan dengan parameter C 1, solver saga	59
Tabel 4.13 Hasil percobaan dengan parameter C 10, solver 1bfgs...	59
Tabel 4.14 Hasil percobaan dengan parameter C 10, solver liblinear	59
Tabel 4.15 Hasil percobaan dengan parameter C 10, solver saga. ...	60
Tabel 4.16 Hasil percobaan dengan parameter C 100, solver 1bfgs.	60
Tabel 4.17 Hasil percobaan dengan parameter C 100, solver saga. .	60
Tabel 4.18 Hasil percobaan dengan parameter C 100, solver saga ..	61
Tabel 4.19 Evaluasi Performa sensor keasaman	62
Tabel 4.20 Evaluasi Performa sensor konduktifitas.....	64
Tabel 4.21 Evaluasi Performa sensor kekeruhan.	65
Tabel 4.22 Evaluasi Performa sensor <i>Total</i> padatan terlarut.....	67

BAB I

PENDAHULUAN

1.1. Latar Belakang

Evaluasi kondisi air melibatkan analisis karakteristik fisik, kimiawi, dan biologis air sebagai parameter. Penilaian ini mencerminkan sejauh mana air memenuhi kebutuhan biota air dan manusia. Kualitas air juga menjadi tolok ukur utama dalam mengevaluasi kesehatan ekosistem air dan dampaknya terhadap kesehatan manusia, terutama dalam konteks air minum. Studi mengenai kualitas air menghadirkan kompleksitas yang mencerminkan beragam metode pengukuran dan indikator air yang digunakan. Penting untuk melakukan pengukuran langsung di lokasi tertentu karena hal ini memastikan bahwa air berada dalam keseimbangan dengan lingkungannya. Hasil pengukuran di lapangan umumnya mencakup data dasar seperti Ph, *Total* zat terlarut, tingkat kekeruhan air, dan sebagainya. [1].

Air merupakan kebutuhan vital bagi kelangsungan hidup makhluk di Bumi. Di lingkungan perkotaan, keberadaan air, khususnya air bersih, menjadi sangat krusial. Meskipun kebutuhan akan air terus meningkat setiap tahun, namun terdapat ketidakseimbangan dengan ketersediaan air bersih yang semakin terbatas. Fenomena ini disebabkan oleh pembangunan yang seringkali tidak mempertimbangkan pelestarian daerah resapan air yang semakin sempit. Krisis air bersih melanda berbagai negara di seluruh dunia, di mana hanya sekitar 1% dari *Total* air yang ada yang dapat dikonsumsi oleh manusia. Keterbatasan jumlah air bersih ini mengakibatkan kesulitan bagi penduduk dalam mengakses sumber air yang bersih. Menurut data dari Organisasi Kesehatan Dunia (WHO), sekitar 663 juta penduduk di dunia menghadapi kesulitan dalam mendapatkan akses ke air bersih. [2].

Menurut ketentuan dalam Peraturan Menteri Nomor 492/MENKES/PER/IV/2010, pada Pasal (1) ayat (1), Persyaratan Air Minum merujuk pada air yang telah mengalami proses pengolahan atau bahkan tanpa proses pengolahan, namun tetap memenuhi kriteria kesehatan dan dapat langsung dikonsumsi. Dalam upaya memenuhi standar kualitas air minum, pengawasan di

daerah sumber air menjadi suatu aspek yang sangat penting. Melalui pengawasan tersebut, keberlanjutan dan kualitas sumber air dapat dijaga dengan baik, sehingga menghasilkan air yang memenuhi standar kesehatan yang layak untuk dikonsumsi oleh manusia. [3].

Berbagai penelitian telah dilakukan terkait klasifikasi kualitas air minum dengan menerapkan teknik *Machine Learning*. Salah satunya adalah penelitian yang dilakukan oleh Aldi dan timnya, yang memanfaatkan metode *Naive Bayes*, *Decision Tree*, dan *K-Nearest Neighbours*. Tujuan penelitian ini adalah untuk menentukan metode yang memberikan tingkat akurasi tertinggi dalam mengklasifikasikan kualitas air minum. Hasil penelitian menunjukkan bahwa metode *Decision Tree* mencapai tingkat akurasi paling tinggi dibandingkan metode lainnya. Penelitian ini mencerminkan upaya dalam memanfaatkan kecerdasan buatan untuk meningkatkan klasifikasi kualitas air minum, dengan pemilihan metode *Decision Tree* sebagai pendekatan yang optimal untuk mencapai tingkat akurasi yang maksimal. [4]. Penelitian yang dilakukan oleh Prismahardi dan rekan-rekannya mengeksplorasi klasifikasi kualitas air minum dengan menerapkan metode *Support Vector Machine*, *Decision Tree*, *Naive Bayes*, dan *Artificial Neural Network*. Hasil penelitian tersebut mengungkapkan bahwa tingkat keakuratan tertinggi diperoleh dengan menggunakan metode *Random Forest Classifier*. Dengan demikian, penelitian ini menyoroti pentingnya variasi metode dalam konteks klasifikasi kualitas air minum menggunakan teknik *Machine Learning*. Pemilihan *Random Forest Classifier* sebagai metode dengan tingkat keakuratan paling tinggi menunjukkan potensinya sebagai pendekatan yang efektif dalam mengatasi kompleksitas dan variasi data dalam penilaian kualitas air minum. [5].

Mengidentifikasi kualitas air minum berdasarkan empat parameter yang terdapat dalam dataset *water_potability.csv* dari Kaggle. Penelitian ini fokus pada klasifikasi kualitas air minum, dan untuk itu, penulis menerapkan teknologi *Machine Learning* dengan menggunakan model *Logistic Regression* yang dioptimalkan melalui *Grid Search*. Pendekatan *Supervised Learning* diimplementasikan dengan menggunakan *library Python* untuk manajemen data. Metode ini dirancang untuk secara sistematis menganalisis data dan menghasilkan pemahaman yang lebih mendalam mengenai kualitas air minum berdasarkan

parameter yang diukur. Dengan menerapkan *Logistic Regression* dan mengoptimalkannya melalui *Grid Search*, diharapkan penelitian ini dapat memberikan hasil klasifikasi yang akurat dan dapat diandalkan.

Machine Learning (ML) merupakan suatu aplikasi dari kecerdasan buatan (*Artificial Intelligence/AI*) yang berfokus pada pengembangan sistem yang dapat belajar secara mandiri tanpa memerlukan pemrograman berulang. ML bergantung pada data, khususnya data pelatihan, sebagai bagian dari proses pembelajaran sebelum menghasilkan output. Secara sederhana, *Machine Learning* dapat dijelaskan sebagai pengaturan komputer untuk mencapai tujuan atau performa tertentu dengan menggunakan kumpulan data pelatihan atau pengalaman masa lalu. Pendekatan ini memungkinkan komputer untuk mengenali pola, belajar dari data, dan membuat keputusan atau prediksi tanpa adanya instruksi langsung dari pemrogram. Dengan demikian, ML memungkinkan sistem untuk mengadaptasi diri terhadap informasi baru dan mengoptimalkan kinerjanya seiring waktu. [6]

Logistic Regression, kadang disebut model logistik atau model logit, adalah metode dalam statistika yang digunakan untuk memprediksi probabilitas kejadian suatu peristiwa dengan memetakan data ke dalam fungsi logit dari kurva logistik. Metode ini merupakan model linier umum yang khusus digunakan untuk regresi binomial. Seperti halnya regresi pada umumnya, *Logistic Regression* melibatkan variabel prediktor, baik numerik maupun kategorikal. Contohnya, probabilitas seseorang mengalami serangan jantung pada suatu waktu tertentu dapat diprediksi berdasarkan informasi seperti usia, jenis kelamin, dan indeks massa tubuh. *Logistic Regression* juga memiliki penerapan luas dalam berbagai bidang, termasuk kedokteran dan ilmu sosial, serta pemasaran, seperti dalam meramalkan kecenderungan pelanggan untuk membeli suatu produk atau menghentikan langganan. [7].

Optimasi *Hyperparameter* adalah proses pemilihan sekumpulan *Hyperparameter* yang optimal untuk algoritma pembelajaran. *Hyperparameter* ini adalah parameter yang nilainya digunakan untuk mengendalikan proses pembelajaran, sementara parameter lainnya (biasanya bobot node) dipelajari selama pembelajaran. Karena model pembelajaran mesin yang sama mungkin memerlukan pengaturan *Hyperparameter* yang berbeda untuk menggeneralisasi pola data yang

berbeda, penyesuaian *Hyperparameter* menjadi kunci untuk mencapai optimalitas dalam *Machine Learning*. Proses ini melibatkan penyesuaian nilai-nilai *Hyperparameter* agar model dapat menyelesaikan masalah pembelajaran dengan optimal. Pengoptimalan *Hyperparameter* bertujuan untuk menemukan kombinasi *Hyperparameter* yang menghasilkan model optimal dengan meminimalkan fungsi kerugian yang telah ditentukan sebelumnya pada data independen tertentu. Validasi silang sering digunakan untuk memperkirakan kinerja generalisasi, dan pilihan nilai *Hyperparameter* yang menghasilkan hasil terbaik pada validasi silang dianggap sebagai solusi optimal. [8].

Grid Search adalah proses pemindaian data untuk mengonfigurasi parameter optimal untuk model tertentu. Tergantung pada jenis model yang digunakan, parameter tertentu diperlukan. Pencarian *Grid* tidak hanya berlaku untuk satu jenis model. Pencarian *Grid* dapat diterapkan di seluruh pembelajaran mesin untuk menghitung parameter terbaik yang akan digunakan untuk model tertentu. Penting untuk dicatat bahwa pencarian *Grid* bisa sangat mahal secara komputasi dan mungkin memerlukan waktu yang cukup lama untuk menjalankan mesin Anda. *Grid-Search* akan membangun model pada setiap kombinasi parameter yang memungkinkan. Ini mengulangi setiap kombinasi parameter dan menyimpan model untuk setiap kombinasi. Tanpa basa-basi lagi, mari kita bahas beberapa contoh dan implementasinya.

1.2. Rumusan Masalah

Berdasarkan uraian latar belakang, perumusan masalah yang akan diangkat adalah:

1. Tingkat recall klasifikasi kualitas air minum yang rendah.
2. Implementasi model untuk klasifikasi kualitas air belum teroptimisasi.
3. Diperlukan peningkatan kinerja model agar dapat mencapai tingkat akurasi sesuai harapan.

1.3. Batasan Masalah

Batasan masalah dalam Tugas Akhir ini mencakup :

1. Penelitian ini memanfaatkan dataset yang diambil dari platform *Kaggle*.
2. Data yang dipakai hanya data bersifat lengkap
3. Penelitian ini melibatkan simulasi program yang dikembangkan dengan menggunakan bahasa pemrograman *Python*.
4. Hasil penelitian hanya menampilkan nilai recall, akurasi, sensitivitas, dan spesifisitas sebagai evaluasi kinerja model dalam mengklasifikasikan data.
5. Fitur yang dipakai adalah ph, konduktifitas, *Total* padatan terlarut, dan kekeruhan

1.4. Tujuan

Tujuan yang akan dicapai melalui penelitian ini mencakup:

1. Meningkatkan recall klasifikasi kualitas air minum.
2. Implementasi model *Logistic Regression* berbasis *Grid-Search Optimization*.
3. Meningkatkan performa model *Logistic Regression*

1.5. Sistematika Penulisan

Adapun alur kepenulisan yang digunakan dalam menulis tugas akhir yaitu sebagai berikut:

BAB I PENDAHULUAN

Bab pertama akan memaparkan sistematis mengenai latar belakang, tujuan penelitian, rumusan masalah, serta bentuk sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab kedua akan menjelaskan teori-teori dasar yang akan menjadi landasan dari penelitian ini. Dasar teori yang akan dibahas pada bab ini adalah

literatur mengenai kualitas air, peraturan pemerintah, *Logistic Regression*, *Optimization*, *Grid-Search* dan performa validasi.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan proses dan rangkaian kegiatan dalam penelitian. Penelitian akan dimulai dari persiapan data, normalisasi dan standarisasi, dan *Hyperparameter Tuning*.

BAB IV HASIL DAN ANALISIS

Bab ini akan memaparkan hasil pengujian yang diperoleh dan menjelaskan analisa terhadap hasil penelitian yang telah dilakukan.

BAB V KESIMPULAN

Bab ini berisi simpulan yang dapat disimpulkan dari penelitian

DAFTAR PUSTAKA

- [1] M. Wibowo and R. Arief Rachman, “Kajian Kualitas Perairan Laut Sekitar Muara Sungai JelitikKecamatan Sungailiat – Kabupaten Bangka,” *J. Presipitasi*, vol. 17, no. 1, pp. 29–37, 2020.
- [2] M. H. D. Barang and S. K. Saptomo, “Analisis Kualitas Air pada Jalur Distribusi Air Bersih di Gedung Baru Fakultas Ekonomi dan Manajemen Institut Pertanian Bogor,” *J. Tek. Sipil dan Lingkung.*, vol. 4, no. 1, pp. 13–24, 2019, doi: 10.29244/jsil.4.1.13-24.
- [3] Permenkes RI, “Peraturan Menteri Kesehatan Republik Indonesia Nomor 492/Menkes/Per/IV/2010 Tentang Persyaratan Kualitas Air Minum,” *Peraturan Mentri Kesehatan Republik Indonesia*. p. MENKES, 2010.
- [4] A. Tangkelayuk, “The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1109–1119, 2022, doi: 10.35957/jatisi.v9i2.2048.
- [5] M. M. Mutoffar and A. Fadillah, “Klasifikasi Kualitas Air Sumur Menggunakan Algoritma *Random Forest*,” *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 4, no. 2, pp. 138–146, 2022, doi: 10.53580/naratif.v4i2.160.
- [6] M. Aminullah, *Perbandingan Performa Klasifikasi Machine Learning dengan Teknik Resampling pada Dataset Tidak Seimbang*. 2021.
- [7] T. S. Wahyuditia, “Penerapan *Logistic Regression* Biner Faktor-Faktor yang Mempengaruhi Berat Badan Bayi Usia 6 Bulan,” 2017.
- [8] A. Toha, P. Purwono, and W. Gata, “Model Prediksi Kualitas Udara dengan Support Vector *Machines* dengan Optimasi *Hyperparameter GridSearch CV*,” *Bul. Ilm. Sarj. Tek. Elektro*, vol. 4, no. 1, pp. 12–21, 2022, doi: 10.12928/biste.v4i1.6079.
- [9] L. Savitri and R. Nursalim, “Klasifikasi Kualitas Air Minum menggunakan

- Penerapan Algoritma *Machine Learning* dengan Pendekatan Supervised *Learning*,” *Diophantine J. Math. Its Appl.*, vol. 2, no. 01, pp. 30–36, 2023, doi: 10.33369/diophantine.v2i01.28260.
- [10] I. G. Vidiastanta, N. Hidayat, and R. K. Dewi, “Komparasi Metode K-Nearest Neighbors (K-NN) Dengan Support Vector *Machine* (SVM) Untuk Klasifikasi Status Kualitas Air,” *Publikasi.Dinus.Ac.Id*, vol. 4, no. 1, pp. 312–319, 2020, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [11] R. Marten, S. Tumangger, and N. Hidayat, “Komparasi Metode Data Mining Support Vector *Machine* dengan Naive Bayes untuk Klasifikasi Status Kualitas Air,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9614–9619, 2019.
- [12] N. E. Helwig, S. Hong, and E. T. Hsiao-wecksler, “Analisis struktur kovarians indikator terkait kesehatan pada lansia yang tinggal di rumah, dengan fokus pada rasa subjektif terhadap kesehatan.”
- [13] S. W. Al Idrus, “Analisis Pencemaran Air Menggunakan Metode Sederhana Pada Sungai Jangkuk, Kekalik Dan Sekarbela Kota Mataram,” *Paedagoria / FKIP UMMat*, vol. 5, no. 2, p. 8, 2018, doi: 10.31764/paedagoria.v5i2.85.
- [14] B. Permana *et al.*, “Analisis Sifat Fisika dan Derajat Keasaman terhadap Kualitas Air Minum Isi Ulang 20 Rumah RW 01 di Kampung Cilember Desa Jogjogan Kecamatan Cisarua Kabupaten Bogor,” *Risenologi*, vol. 5, no. 1, pp. 64–69, 2020, doi: 10.47028/j.risenologi.2020.51.82.
- [15] P. lumban Toruan, , R., and A. A. Setiawan, “Konduktivitas Listrik Ion Terlarut: Studi Kasus di Air Sumur TPA Sukawinatan Palembang,” *J. Redoks*, vol. 7, no. 1, pp. 48–54, 2022, doi: 10.31851/redoks.v7i1.6760.
- [16] D. T. Suhendar, I. S. Sachoemar, and A. B. Zaidy, “Hubungan Kekeruhan Terhadap Materi Partikulat Tersuspensi (MPT) Dan Kekeruhan Terhadap Klorofil Dalam Tambak Udang,” *Fish. Mar. Res.*, vol. 4, no. 3, pp. 332–338, 2020, [Online]. Available: <http://jfmr.ub.ac.id>.
- [17] M. A. REVANSYAH, L. K. MEN, S. SETIANTO, F. F. L. SAFRIANI,

- and A. APRILIA, “Analisis Tds, Ph, Dan Cod Untuk Mengetahui Kualitas Air Di Desa Cilayung,” *J. Mater. dan Energi Indones.*, vol. 12, no. 02, p. 43, 2023, doi: 10.24198/jme.v12i02.41305.
- [18] I. M. Rahayu, A. Yusuf, and M. Ridwan, “Prediksi Kesiapan Sekolah Menggunakan *Machine Learning* Berbasis Kombinasi Adam dan Nesterov Momentum,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 6, p. 1273, 2022, doi: 10.25126/jtiik.2022955442.
- [19] J. L. Alzen, L. S. Langdon, and V. K. Otero, “A *Logistic Regression* investigation of the relationship between the *Learning Assistant* model and failure rates in introductory STEM courses,” *Int. J. STEM Educ.*, vol. 5, no. 1, pp. 1–12, 2018, doi: 10.1186/s40594-018-0152-1.
- [20] F. Pedregosa *et al.*, “Scikit-learn: *Machine Learning* in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [21] D. Wirasmita and E. Anisa, “Analisis Sentiment Twitter Berbasis *Grid Search* Algorithm (GSA) Dengan Metode Support Vector *Machine* (SVM),” *J. Asimetri J. Ilm. Rekayasa Inov.*, vol. 5, pp. 35–42, 2023, doi: 10.35814/asiimetrik.v5i1.3789.