

**TESIS**

**KLASIFIKASI KUALITAS AIR PEMBENIHAN IKAN LELE  
MENGUNAKAN METODE STORET DENGAN  
ALGORITMA XGBOOST**



Disusun oleh:

**Nama : Muhammad Yusuf Widlyanto**  
**NIM : 23.51.2488**  
**Konsentrasi : Business Intelligence**

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**TESIS**

**KLASIFIKASI KUALITAS AIR PEMBENIHAN IKAN LELE  
MENGUNAKAN METODE STORET DENGAN  
ALGORITMA XGBOOST**

**CLASSIFICATION OF WATER QUALITY IN CATFISH  
HATCHERIES USING THE STORET METHOD WITH THE  
XGBOOST ALGORITHM**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Muhammad Yusuf Widlyanto  
NIM : 23.51.2488  
Konsentrasi : Business Intelligence

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

**HALAMAN PERSETUJUAN**

**KLASIFIKASI KUALITAS AIR PEMBENIHAN IKAN LELE  
MENGUNAKAN METODE STORET DENGAN ALGORITMA  
XGBOOST**

**CLASSIFICATION OF WATER QUALITY IN CATFISH HATCHERIES  
USING THE STORET METHOD WITH THE XGBOOST  
ALGORITHM**

yang disusun dan diajukan oleh  
**Muhammad Yusuf Widiyanto**  
23.51.2488

telah disetujui oleh Dosen Pembimbing Tesis  
pada tanggal 09 Oktober 2025

**Dosen Pembimbing,**



**Tonny Hidayat, S.Kom., M.Kom., Ph.D**  
**NIK. 190302182**

**HALAMAN PENGESAHAN**

**KLASIFIKASI KUALITAS AIR PEMBENIHAN IKAN LELE  
MENGUNAKAN METODE STORET DENGAN ALGORITMA  
XGBOOST**

**CLASSIFICATION OF WATER QUALITY IN CATFISH HATCHERIES  
USING THE STORET METHOD WITH THE XGBOOST  
ALGORITHM**

yang disusun dan diajukan oleh

**Muhammad Yusuf Widiyanto**

**23.51.2488**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 09 Oktober 2025

Susunan Dewan Penguji

**Nama Penguji**

**Tanda Tangan**

**Emha Taufiq Luthfi, S.T., M.Kom., Ph.D**  
NIK. 190302125

**Dhani Ariatmanto, S.Kom., M.Kom., Ph.D**  
NIK. 190302197

**Tonny Hidavat, S.Kom., M.Kom., Ph.D**  
NIK. 190302182



Tesis ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Magister Komputer  
Tanggal 09 Oktober 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



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

## HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Muhammad Yusuf Widiyanto**  
NIM : **23.51.2488**  
Konsentrasi : **Business Intelligence**

Menyatakan bahwa Tesis dengan judul berikut:  
**Klasifikasi Kualitas Air Pembentukan Ikan Lele Menggunakan Metode STORET Dengan Algoritma XGBoost**

Dosen Pembimbing Utama : **Tonny Hidayat, S.Kom., M.Kom., Ph.D.**

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 Tim 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, 9 Oktober 2025  
Yang Menyatakan,



Muhammad Yusuf Widiyanto

## HALAMAN PERSEMBAHAN

Halaman Persembahan Ini Saya tujukan Kepada :

1. Diri Saya Sendiri
2. Kedua Orang Tua Saya
3. Universitas Amikom Yogyakarta
4. Masa Depan Saya
5. Seluruh Warga Indonesia



## KATA PENGANTAR

Dalam Proses Pengerjaan Thesis Ini, Terdapat Banyak Sekali Orang - Orang Yang Berperan Penting Dalam Membantu Dan Juga Support. Maka Dari Itu Saya Ucapkan Banyak Terimakasih Sebesar – Besarnya Kepada :

1. Diri Saya Sendiri Yang Telah Menyelesaikan Kewajiban Sebagai Seorang Mahasiswa. Dan Tidak Terasa 1,5 Tahun Perkuliahan Di Universitas Amikom Yogyakarta Telah Diselesai Dengan Baik Sehingga LULUS Dengan Mendapatkan Gelar Magister (M.Kom).
2. Orang Tua Saya yang bernama Agus Supriyanto, S.E. selaku papah saya dan Yuskalita, S.E. selaku Mamah Saya. Yang Telah Memberikan Banyak Sekali Support Yang Mana Saya Tidak Bisa Menuliskannya Satu – Persatu di dalam Kata Pengantar Ini. Karena Doa, Berkah Dan Segalanya, Saya Dapat Menyelesaikan Naskah Thesis Saya Ini Dengan Sempurna.
3. Orang tua saya di yogyakarta yaitu mas yoyon dan mbak tutik yang telah memberikan banyak sekali support sehingga naskah ini juga dapat diselesaikan dengan sempurna.
4. Dosen Pembimbing saya Bapak Tonny Hidayat, S.Kom., M.Kom., Ph.D. yang telah memberikan bimbingan dan waktu luang kepada saya untuk menyelesaikan thesis ini dengan baik.
5. Teman-teman saya, Mas Fajar, Mas Arya, Bang David, Rio, dan juga teman – teman saya di perkuliahan angkatan 31 Reguler.

Yogyakarta, 09 Oktober 2025

Penulis.

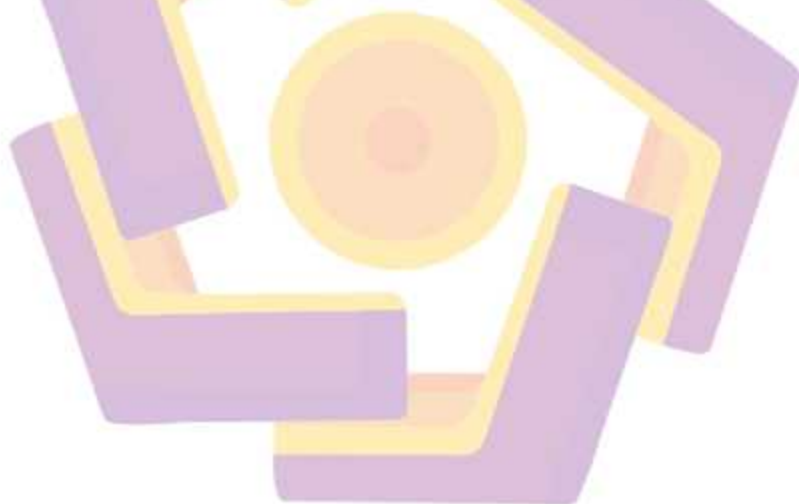
## DAFTAR ISI

HALAMAN JUDUL .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN THESIS.....	v
HALAMAN PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	ix
INTISARI.....	xii
ABSTRACT .....	xiii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
BAB II TINJUAN PUSTAKA .....	7
2.1 Tinjauan Pustaka .....	7
2.2 Keaslian Penelitian .....	10
2.3 Landasan Teori .....	18
2.3.1 Kualitas Air Kolam .....	18
2.3.2 Labelling Data Pada STORET .....	20
2.3.2.1 Pengukuran Metode STORET .....	21
2.3.3 Algoritma Xtreme Gradient Boosting .....	25
2.3.4 Hyperparameter .....	25
2.3.5 Hyperparameter Model .....	25
2.3.6 Hyperparameter Tunning .....	26
2.3.7 Oversampling .....	27
2.3.7.1 SMOTE .....	27
2.3.7.2 Borderline-SMOTE .....	28
2.3.7.3 ADSYN (Adaptive Synthetic Sampling) .....	28
2.3.8 K-Fold.....	29
2.3.9 Confussion Matrix .....	29
BAB III METODE PENELITIAN.....	32
3.1 Jenis, Sifat dan Pendekatan Penelitian .....	32
3.2 Metode Pengumpulan Data .....	32

3.3 Metode Analisis Data .....	33
3.4 Alur Penelitian .....	33
<b>BAB IV Hasil Penelitian dan Pembahasan .....</b>	<b>39</b>
4.1 Oversampling Dengan SMOTE .....	39
4.2 Hasil Skenario Pertama .....	40
4.3 Hyperparameter Tuning Xtreme Gradient Boosting .....	40
4.4 Evaluasi Skenario Penelitian Pertama .....	41
4.4.1 K-Fold 2 .....	46
4.4.2 K-Fold 3 .....	51
4.4.3 K-Fold 4 .....	56
4.4.4 K-Fold 5 .....	61
4.4.5 K-Fold 6 .....	66
4.4.6 K-Fold 7 .....	70
4.4.7 K-Fold 8 .....	74
4.4.8 K-Fold 9 .....	78
4.4.9 k-Fold 10 .....	83
4.4.10 K-Fold 11 .....	88
4.4.11 K-Fold 12 .....	92
4.4.12 K-Fold 13 .....	96
4.4.13 K-Fold 14 .....	100
4.4.14 K-Fold 15 .....	104
4.4.15 K-Fold 16 .....	108
4.4.16 K-Fold 17 .....	112
4.4.17 K-Fold 18 .....	115
4.4.18 K-Fold 19 .....	118
4.4.19 K-Fold 20 .....	121
4.5 Ringkasan Pengujian Setiap K-Fold .....	124
<b>BAB V PENUTUP .....</b>	<b>127</b>
5.1 Kesimpulan .....	127
5.2 Saran .....	129
<b>DAFTAR PUSTAKA .....</b>	<b>130</b>

## DAFTAR TABEL

Tabel 2.1 Matriks Literatur Review dan Posisi penelitian.....	10
Tabel 2.2 Standar Nasional Indonesia Produk Benih Ikan Lele .....	19
Tabel 2.3 Lanjutan .....	20
Tabel 2.4 Pengukuran Kualitas Air Pada Pembenihan Ikan .....	20
Tabel 2.5 Penentuan Status Mutu Air Dengan STORET.....	21
Tabel 2.6. Pengumpulan Data Kualitas Air .....	21
Tabel 2.7 Lanjutan .....	22
Tabel 2.8. Minimal, Maksimal dan Rerata Data .....	22
Tabel 2.9 Nilai Paramater STORET.....	23
Tabel 2.10 Skor Baku Mutu Air.....	23
Tabel 2.11 Lanjutan .....	24
Tabel 2.12 Nilai Environmental Protection Agency.....	24
Tabel 4.13. Preprocessing Hyperparameter Tunning Xtreme Gradient Boosting	41
Tabel 4.14 Ringkasan Pengujian Setiap K-Fold .....	124
Tabel 4.15 Lanjutan .....	125



## DAFTAR GAMBAR

Gambar 2.1 Confusion Matrix .....	30
Gambar 3.2 Dataset Kualitas Air .....	32
Gambar 3.3 Skenario Pertama Preprocessing dan Processing .....	34
Gambar 3.4 Skenario Kedua Preprocessing dan Processing .....	35
Gambar 4.5. Pengecekan Setiap Kelas Pada Data Training .....	39
Gambar 4.6 Oversampling Dengan SMOTE .....	40
Gambar 4.7 Hyperparameter Terbaik .....	41
Gambar 4.8 Hasil Skenario Pertama .....	42
Gambar 4.9. Metrix Evaluasi dan Diagram k-Fold 2 .....	46
Gambar 4.10 Metrik Evluaasi dan Diagram K-Fold 3 .....	52
Gambar 4.11 Metrik Evluaasi dan Diagram K-Fold 4 .....	57
Gambar 4.12 Metrik Evluaasi dan Diagram K-Fold 5 .....	62
Gambar 4.13 Metrik Evluaasi dan Diagram K-Fold 6 .....	67
Gambar 4.14 Metrik Evluaasi dan Diagram K-Fold 7 .....	71
Gambar 4.15 Metrik Evluaasi dan Diagram K-Fold 8 .....	75
Gambar 4.16 Metrik Evluaasi dan Diagram K-Fold 9 .....	79
Gambar 4.17 Metrik Evluaasi dan Diagram K-Fold 10 .....	83
Gambar 4.18 Metrik Evluaasi dan Diagram K-Fold 11 .....	88
Gambar 4.19 Metrik Evaluasi dan Diagram K-Fold 12 .....	92
Gambar 4.20 Metrik evaluasi dan Diagram K-Fold 13 .....	96
Gambar 4.21 Metrik Evluaasi dan Diagram K-Fold 14 .....	100
Gambar 4.22 Metrik Evluaasi dan Diagram K-Fold 15 .....	104
Gambar 4.23 Metrik Evluaasi dan Diagram K-Fold 16 .....	109
Gambar 4.24 Metrik Evluaasi dan Diagram K-Fold 17 .....	112
Gambar 4.25 Metrik Evluaasi dan Diagram K-Fold 18 .....	115
Gambar 4.26 Metrik Evluaasi dan Diagram K-Fold 19 .....	118
Gambar 4.27 Metrik Evluaasi dan Diagram K-Fold 20 .....	121

## INTISARI

Kualitas air merupakan faktor utama yang menentukan keberhasilan dalam kegiatan budidaya ikan lele, khususnya pada fase pembenihan. Pengukuran kualitas air secara manual menggunakan metode STORET memiliki keterbatasan karena tidak dapat menampilkan kondisi air secara real-time dan bersifat statis dalam sistem penilaiannya. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi kualitas air berbasis machine learning dengan algoritma eXtreme Gradient Boosting (XGBoost) yang dilatih menggunakan data hasil penilaian metode STORET sebagai label.

Hasil penilaian metode STORET sebagai label. Penelitian ini menggunakan parameter fisika dan kimia air yang meliputi suhu, pH, oksigen terlarut (DO), amonia, alkalinitas, mangan, dan nitrat. Untuk meningkatkan kualitas data, diterapkan teknik Synthetic Minority Oversampling Technique (SMOTE) guna menyeimbangkan distribusi jumlah data antar kelas. Selain itu, dilakukan proses hyperparameter tuning menggunakan metode Randomized Search untuk memperoleh kombinasi parameter XGBoost yang paling optimal, meliputi learning rate, max depth, n estimators, dan subsample ratio. Evaluasi model dilakukan menggunakan metode K-Fold Cross Validation dengan variasi nilai K dari 2 hingga 20 untuk mengukur kestabilan dan kemampuan generalisasi model.

Hasil pengujian menunjukkan bahwa penerapan SMOTE mampu menyeimbangkan data antar kelas secara proporsional, sementara proses hyperparameter tuning meningkatkan kinerja model secara signifikan dibandingkan konfigurasi default. Berdasarkan hasil evaluasi, performa terbaik diperoleh pada K-Fold = 19, dengan nilai akurasi sebesar 0,9433, precision 0,9428, recall 0,9433, dan F1-score 0,9429. Nilai evaluasi yang konsisten pada berbagai skenario menunjukkan bahwa model XGBoost memiliki tingkat generalisasi yang tinggi dalam mengklasifikasikan kualitas air.

Secara keseluruhan, hasil penelitian ini membuktikan bahwa kombinasi antara metode STORET dan algoritma XGBoost yang dioptimasi melalui teknik SMOTE, hyperparameter tuning, serta validasi K-Fold mampu menghasilkan sistem klasifikasi kualitas air yang akurat, efisien, dan dapat diterapkan pada sistem pemantauan berbasis IoT untuk mendukung pengelolaan budidaya ikan lele secara berkelanjutan.

**Kata kunci: kualitas air, STORET, XGBoost, machine learning, budidaya ikan lele**

## ABSTRACT

Water quality is a crucial factor determining the success of catfish farming, particularly during the hatchery phase. Manual water quality assessments using the STORET method have limitations, as they cannot provide real-time monitoring and are static in their evaluation system. Therefore, this study aims to develop a water quality classification model based on machine learning using the eXtreme Gradient Boosting (XGBoost) algorithm, which was trained using STORET-based water quality assessment data as labels.

The research utilizes physical and chemical parameters of water, including temperature, pH, dissolved oxygen (DO), ammonia, alkalinity, manganese, and nitrate. To improve data quality, the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the class distribution. Furthermore, hyperparameter tuning was performed using the Randomized Search method to determine the optimal configuration of XGBoost parameters, such as learning rate, max depth,  $n$  estimators, and subsample ratio. The model performance was evaluated using K-Fold Cross Validation with K values ranging from 2 to 20 to assess model stability and generalization capability.

The experimental results show that the application of SMOTE successfully balanced the data distribution across classes, while hyperparameter tuning significantly improved model performance compared to the default configuration. The best performance was achieved at K Fold = 19, with an accuracy of 0.9433, precision of 0.9428, recall of 0.9433, and an F1-score of 0.9429. The consistent evaluation metrics across various folds indicate that the XGBoost model achieved high generalization capability in classifying water quality levels.

Overall, this study demonstrates that the integration of the STORET method and the XGBoost algorithm—enhanced through SMOTE, hyperparameter tuning, and K-Fold validation—can produce an accurate, efficient, and reliable water quality classification system. This system has strong potential to be implemented in IoT-based water monitoring to support sustainable catfish hatchery management.

**Keyword:** water quality, STORET, XGBoost, machine learning, catfish aquaculture

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang Masalah

Air Tawar merupakan bagian penting yang digunakan benih ikan lele dalam bertahan hidup. Semua kebutuhan dasar yang meliputi nutrisi hingga lingkungan yang sesuai untuk pertumbuhan, terpenuhi dalam kondisi kualitas air tawar yang baik. Baik atau buruknya kualitas air bergantung pada tujuan penggunaannya, seperti untuk keperluan minum, mencuci, dan budidaya ikan. Issue Perubahan kualitas air tawar yang disebabkan oleh beberapa faktor seperti faktor fisika yang meliputi Suhu dan Kekeruhan dan Manganese dan faktor kimia seperti PH, Oksigen, Alkalinitas dan Ammonia. Apabila faktor-faktor tersebut tidak di tangani dengan baik, maka proses kelangsungan hidup benih ikan akan terganggu dan dapat menimbulkan penyakit, seperti pertumbuhan yang lamban dan tingkat kematian yang tinggi. Sehingga dibutuhkan penanganan lebih lanjut untuk proses pengobatan dan penyembuhan pada ikan. Dalam melakukan budidaya ikan, kualitas air yang berperan penting dalam pengaruh laju pertumbuhan serta kelangsungan hidup benih ikan [1]. Pada penelitian yang dilakukan oleh [2] proses pemantauan kualitas air pembenihan ikan lele dilakukan pada pukul 8 pagi dan pukul 5 sore. Karena pada waktu tersebut fotosintesis alga di kolam menyebabkan perbedaan kualitas air paling signifikan pada saat fajar dan senja. penelitian yang sama dilakukan oleh [3] dimana proses pemantauan kualitas air dilakukan setiap 10 menit untuk mengantisipasi kemungkinan hilangnya paket data atau error sensor. Selain itu,

penelitian yang dilakukan oleh [4] pengukuran kualitas air dilakukan pada pagi hari (early morning) karena suhu air masih rendah (sekitar 29 C), pH relatif stabil (6.5–6.9). Kondisi ini terjadi karena sinar matahari belum cukup kuat memanaskan air. Pada siang hari (Middle) suhu meningkat hingga 35.5 C, pH naik menjadi sekitar 7 akibat fotosintesis tanaman air yang menyerap CO<sub>2</sub>. Hal ini menggambarkan kondisi air paling panas dan metabolisme ikan meningkat. Dan pada sore hari (evening) suhu mengalami penurunan (27–30 C), pH cenderung stabil (6.9–7). Saat matahari terbenam, fotosintesis berkurang dan respirasi organisme mulai menambah CO<sub>2</sub> ke air sehingga pH perlahan turun. Pengukuran kualitas air secara manual dengan STORET mempunyai keterbatasan dalam menghasilkan kondisi air secara langsung. Keterbatasan tersebut berupa jeda waktu yang menyebabkan kondisi air bisa saja berubah dengan cepat yang diakibatkan oleh beberapa faktor yang tidak bisa dihitung secara manual [4]. Selain itu, hilangnya parameter yang dilakukan dalam penilaian dan kesalahan yang ditimbulkan dalam pengukuran kualitas air, dapat mempengaruhi nilai yang dihasilkan dalam pengukuran label kualitas air. Hal ini dapat menyebabkan kesalahan dalam klasifikasi status air seperti Bagus, Tercemar Ringan, Tercemar Sedang atau Tercemar Berat [5]. Sehingga metode manual tidak dapat menghasilkan nilai kualitas air yang cepat dan tepat tanpa bantuan machine learning [6]. Dalam menghasilkan nilai kualitas air pada kolam pembenihan ikan lele dengan Machine Learning. Proses pelatihan model yang digunakan menggunakan dataset yang telah diberi label menggunakan metode STORET dengan kelas seperti Baik, Tercemar Ringan, Tercemar Sedang dan Tercemar Berat. Hasil klasifikasi yang didapatkan menjadi dasar bagi model

dalam mempelajari kualitas air [7]. Meskipun metode ini mempunyai kerentanan seperti hilangnya pengelompokan data berdasarkan kriteria tertentu dalam evaluasi kualitas air yang dapat mengakibatkan penilaian yang tidak sepenuhnya akurat, seperti overestimasi dan underestimasi [8]. Namun, penggunaan metode ini tetap bermanfaat sebagai acuan awal dalam membangun dataset berlabel. Selanjutnya, penerapan machine learning diharapkan dapat menjadi solusi dari keterbatasannya alat yang digunakan dalam memantau kualitas air pembenihan ikan lele. Dataset yang telah dilabeli menggunakan metode STORET berdasarkan kriteria kelas kemudian diproses melalui beberapa tahapan. Pertama, dilakukan 2 penyeimbangan kelas menggunakan SMOTE untuk mengatasi permasalahan data yang tidak seimbang [9]. Selanjutnya, model machine learning dilatih dengan melakukan pencarian hyperparameter menggunakan Random Search untuk memperoleh kombinasi parameter terbaik. Penentuan nilai hyperparameter yang digunakan merujuk pada penelitian sebelumnya yang dilakukan oleh [10] dan [11]. Setelah itu, dilakukan proses evaluasi menggunakan K-Fold Cross Validation guna memastikan performa model tetap konsisten pada data yang berbeda. Untuk melihat seberapa baik evaluasi K-Fold, peneliti melakukan 20 kali pengujian, yang dimulai dari K-Fold 2 – 20. Setelah model terlatih dan tervalidasi, dilakukan evaluasi kinerja menggunakan metrik akurasi, presisi, recall, dan F1-Score untuk menilai sejauh mana model mampu melakukan klasifikasi kualitas air secara optimal [10]. Pada penelitian sebelumnya yang dilakukan oleh [10] yang menggunakan 19 parameter pengujian berdasarkan label Indeks Kualitas Air untuk air minum dengan model algoritma Gradient Boosting mendapatkan nilai akurasi 0,9964 dan XGBoost

mendapatkan nilai akurasi 0,9954. Selain itu pada penelitian yang dilakukan oleh [3] algoritma XGBoost mendapatkan nilai akurasi 0,80 tertinggi daripada algoritma yang lain dalam mendeteksi konsentrasi nitrat air tanah. Pemilihan model XGBoost dipilih dalam penelitian ini karena dianggap lebih baik dalam menghasilkan nilai evaluasi diantaranya nilai Akurasi, Presisi, Recall dan F1-Score dan efisiensi waktu pada proses pelatihan.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang masalah yang telah disampaikan, maka perlu rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana cara mengatasi keterbatasan metode STORET dalam menghasilkan label kualitas air dengan cepat ?
2. Seberapa akurat algoritma Xtreme Gradient Boosting dalam melakukan klasifikasi berdasarkan parameter yang ditentukan ?

### **1.3. Batasan Masalah**

Bagian ini memuat penjelasan tentang:

1. Penggunaan metode STORET dalam menentukan Indeks Kualitas Air
2. Penelitian ini menggunakan Air Tawar sebagai objek penelitian, karena benih ikan lele termasuk dalam kategori ikan air tawar
3. jenis bibit ikan lele yang digunakan adalah Dumbo berdasarkan refrensi dari kementrian Kelautan dan Perikanan.
4. Algoritma Gradient Boosting digunakan dalam melakukan klasifikasi kualitas air pembenihan bibit ikan lele

5. Penelitian ini dilakukan di kolam kecil untuk benih ikan lele.
6. Parameter kualitas air yang digunakan yaitu Suhu, PH, Oksigen, Amonia, Alkalinitas, Manganese, Nitrate.
7. Label tidak menerangkan secara detail pencemaran kualitas air

#### **1.4. Tujuan Penelitian**

Bagian ini memuat penjelasan secara spesifik:

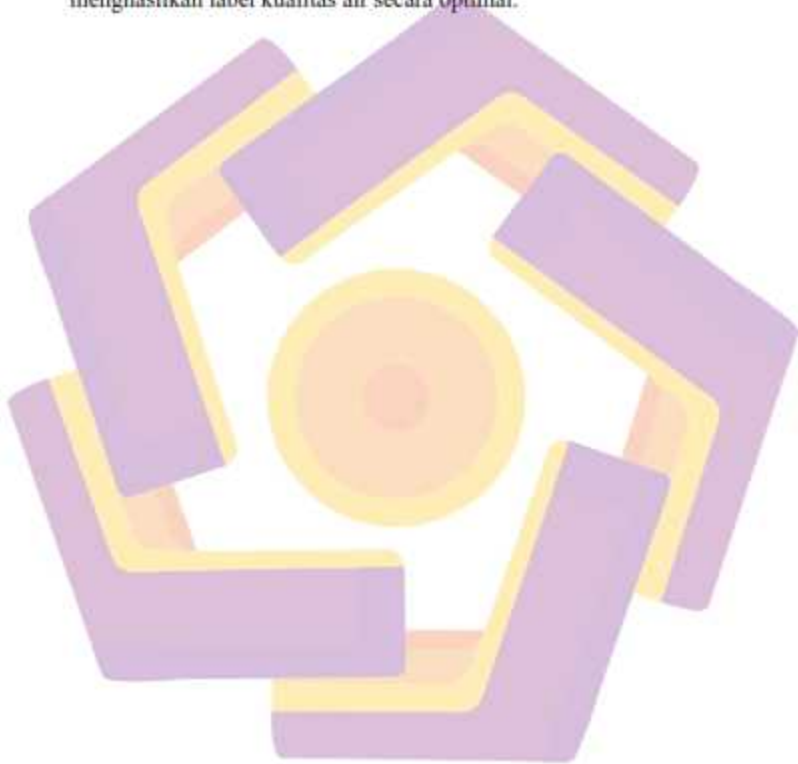
1. Merancang metode yang dapat melakukan prediksi dalam klasifikasi kualitas air dengan machine learning yang telah diujikan.
2. Mengimplementasikan algoritma Xtreme Gradient Boosting (XGBoost) untuk meningkatkan keakuratan dalam penentuan kategori kualitas air.
3. Mengevaluasi performa model menggunakan metrik Akurasi, Presisi, Recall, dan F1-Score, serta membandingkan hasil dengan berbagai skenario (dengan dan tanpa K-Fold).
4. Penelitian ini bertujuan untuk menghasilkan model yang dapat memberikan peringatan dini terkait kualitas air kolam, melalui beberapa label yang dihasilkan.

#### **1.5. Manfaat Penelitian**

Bagian ini memuat penjelasan tentang:

1. Penelitian ini dapat memberikan kontribusi pada ilmu pengetahuan pada bidang perikanan dengan merancang metode klasifikasi kualitas air menggunakan machine learning

2. Penelitian ini dapat menghasilkan label kualitas air berdasarkan tingkat pelabelan, sehingga dapat mengurangi risiko seperti pertumbuhan ikan yang lambat, penyakit, dan tingkat kematian yang tinggi.
3. Penerapan model ini dapat mendukung pengelolaan risiko dengan menghasilkan label kualitas air secara optimal.



## BAB II

### TINJAUAN PUSTAKA

#### 2.1. Tinjauan Pustaka

Metode Klasifikasi merupakan suatu pendekatan yang digunakan dalam menentukan kualitas air kolam budidaya. Penggunaan metode ini dapat membantu mengelompokkan kualitas air berdasarkan beberapa parameter yang telah ditentukan [12]. Pada penelitian yang dilakukan oleh [10] dengan judul "Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI" pelabelan kualitas air dengan 5 label yang meliputi Bersih, Tidak Bersih, Polutan, Tinggi Polutan, Extrem Polutan. Penelitian tersebut menggunakan beberapa model yang salah satunya adalah eXtreme Gradient Boosting dan Gradient Boosting dengan hyperparameter yang telah ditentukan dan menggunakan K-Fold untuk meningkatkan kinerja pada model. Selain itu, Pada penelitian yang dilakukan oleh [13] dengan judul "Enhanced Water Quality Control Based on Predictive Optimization for Smart", penelitian ini membahas tentang pengendalian kualitas air berbasis optimasi dalam meningkatkan efisiensi kualitas air pada kolam budidaya. Terdapat 5 parameter yang digunakan dalam penelitian ini seperti Suhu, PH, Konduktivitas Air dan Tinggi air. Penggunaan IOT digunakan dalam melakukan pemantauan kualitas air secara real-time dan algoritma fuzzy digunakan dalam memberikan label kualitas air. Pada penelitian yang sama yang dilakukan oleh [14] parameter kualitas air yang digunakan seperti, oksigen terlarut (DO), pH, suhu, amonia, nitrit, nitrat, dan lainnya memanfaatkan pendekatan

algoritma berbasis machine learning dan teknologi seperti Internet of Things (IoT) untuk memantau dan menganalisis parameter tersebut secara real-time, yang dapat membantu dalam pengambilan keputusan dalam pengelolaan kualitas air akuakultur. Dari kedua penelitian diatas terdapat kelemahan yang dilakukan dalam menentukan label kualitas air. Penentuan 6 kualitas air berdasarkan algoritma Fuzzy terdapat kelemahan yaitu bertambahnya jumlah variabel dan jumlah aturan dalam logika fuzzy dapat menyebabkan peningkatan pada beban komputasi [15]. Sehingga dibutuhkan sebuah machine learning untuk menganggulangi kekurangan tersebut. Terdapat juga penelitian yang sama dengan penulis dalam menentukan kualitas air menggunakan Indeks Kualitas Air atau Water Quality Index (WQI). Penelitian yang digunakan oleh [16] menggunakan pendekatan berdasarkan pembobotan parameter. Hasil akhir yang dikeluarkan dalam penelitian ini adalah nilai numerik yang dikelompokkan kedalam lima kategori yang sudah ditetapkan yang meliputi Sangat Baik, Baik, Sedang, Rendah dan Buruk. Dengan memanfaatkan machine learning yang digunakan, pendekatan ini dapat memungkinkan evaluasi kualitas air yang lebih cepat, efisien, dan akurat. Selain itu penelitian yang dilakukan oleh [17] dengan judul "Water quality prediction using machine learning models based on grid search method" pada metode itu menggunakan 7 parameter inti. Dalam melakukan pelabelan yang digunakan, penggunaan WQI pada jurnal ini menjadi standart utama dalam mengelompokkan kategori berdasarkan labelnya. Terdapat 3 label yang digunakan dalam penelitian ini, seperti Baik, Sedang dan Buruk berdasarkan rentang angka yang di tetapkan. Pada penelitian yang dilakukan oleh [18] dengan judul "Optimizing silt density index prediction in water treatment

systems using pressure-based gradient boosting hybridized with Salp Swarm Algorithm". Penelitian pada jurnal ini menggunakan algoritma Gradient Boosting dalam melakukan klasifikasi dan pelabelan dataset ini menggunakan 3 label yaitu Baik, Buruk dan Tidak Layak berdasarkan rentang angka yang telah ditetapkan. Dalam ketiga jurnal tersebut, terdapat kesamaan pada jurnal yang penulis lakukan yaitu pelabelan yang menggunakan Water Quality Indeks. Namun, dari ketiga jurnal tersebut, penelitian ini memiliki perbedaan dalam pelabelan Indeks Kualitas Air. Penulis menggunakan metode STORET dan mengacu pada standar baku mutu kualitas air untuk pembenihan ikan, sehingga pelabelan data yang digunakan menjadi lebih jelas, terukur, dan sesuai dengan kebutuhan penelitian. 7 Dalam meningkatkan kinerja model Gradient Boosting, penelitian yang dilakukan oleh [18] menggunakan model Particle Swarm Optimization (PSO) dan Salp Swarm Algorithm (SSA) dapat meningkatkan nilai prediksi. Dalam penelitian yang dilakukan oleh [10] Hyperparameter Tuning yang digunakan melibatkan optimasi parameter seperti learning rate, jumlah estimator, kedalaman maksimum pohon (max\_depth), jumlah sampel minimum per daun (min\_samples\_leaf), dan rasio subsample yang menggunakan library RandomizedSearchCV pada processnya. Teknik ini memastikan model dilatih pada berbagai kombinasi parameter untuk mencapai performa terbaik pada data validasi. Penelitian yang penulis gunakan ini, akan membandingkan model algoritma gradient boosting yang tidak menggunakan model tambahan apapun dan membandingkan dengan metode yang digunakan oleh [10]. Dalam mengelompokkan nilai Indeks Kualitas air berdasarkan metode yang diusulkan oleh penulis.

## 2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian

Penelitian Klasifikasi Kualitas Air Pembemihan Ikan Lele Menggunakan Metode STORET dengan Algoritma XGBoost

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
1.	Accuracy assessment of RFerns, NB, SVM, and kNN machine learning classifiers in aquaculture	Yilmaz, Mesut, Mustafa Çakir, Mükerrerem Atalay Oral, Hüseyin Özgür Kazancı, and Okan Oral. 2023. Saudi Journal of Biological Sciences (Q1)	pendekatan machine learning yang digunakan dalam meningkatkan kualitas pada air yang berdampak pada ikan	penggunaan algoritma yang digunakan yaitu SVM dan RF tersebut mendapatkan nilai akurasi mencapai 100%. Sehingga menjadi nilai pencapaian ini yang baik daripada algoritma K-NN dan NB yang	<b>Keunggulan :</b> preprocessing, nilai akurasi mencapai 100%. <b>Kelemahan:</b> kurangnya dataset yang hanya mendapatkan 90 data yang digunakan, sehingga terjadi oversampling pda processing.	Data yang digunakan menggunakan 2 skenario dalam proses pelatihan data dan penggunaan dataset yang besar serta terdapat pendekatan tambahan pada model algoritma Gradient Boosting

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
				mendapatkan nilai 91,3%		
2.	An Intelligent-IoT-Based Data Analytics for Freshwater Recirculating Aquaculture System	Singh, Munesh Sahoo, Kshira Sagar Gandomi, 2024, IEEE Internet of Things Journal (Q1)	mengembangkan sistem IoT berbasis kecerdasan buatan (AI) dengan Gradient Boosting, Random Forest, Artificial Neural Network dan M5 Model Tree	Algoritma M5 memiliki performa terbaik dalam memprediksi DO (Oksigen) dengan Korelasi 0,877 dan MAE 0,963. Sedangkan untuk PH dengan korelasi 0,975 dan MAE 0,0123	<b>Keunggulan :</b> Penerapan IOT secara langsung pada objek penelitian. <b>Kelemahan :</b> Dalam pengambilan dataset masih menggunakan 200 data. Sedangkan dalam penerapannya, algoritma gradient boosting membutuhkan jumlah data yang besar untuk	Penulis menggunakan metode STORET untuk menentukan kualitas air pada air kolam dan untuk dataset yang digunakan mempunyai data yang lebih dari 1000 data

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
					processing agar terhindar dari overfitting	
3.	Evaluation of disease outbreak in terms of physico-chemical characteristics and heavy metal load of water in a fish farm with machine learning techniques	Yilmaz, Mesut Çakir, 2023, Saudi Journal of Biological Sciences (Q1)	memprediksi penyakit pada air menggunakan 3 machine learning yaitu SVM, Logistic Regression dan Naïve Baiyes	Algoritma SVM mendapatkan nilai akurasi tertinggi daripada algoritma lainnya dengan nilai 93,3%.	<b>Keunggulan :</b> Penggunaan ML dalam mendeteksi dan memprediksi wabah penyakit lebih awal, yang dapat mendukung produksi akuakultur berkelanjutan dengan meminimalkan kerugian akibat penyakit.	Penelitian ini melakukan klasifikasi kualitas air dengan Metode STORET sebagai pelabelannya dan algoritma eXtreme Gradient Boosting sebagai modelnya dan penggunaan Hyperparameter Tunning Random Search untuk

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
					<p><b>Kelemahan :</b> Pelabelan yang menggunakan 2 pelabelan saja seperti "terdeteksi penyakit" dan "tidak terdeteksi penyakit". Selain itu disarankan untuk menggunakan algoritma tambahan yang digunakan untuk perbandingan</p>	meningkatkan performa modelnya.
4.	Comprehensive river water quality monitoring using convolutional neural	Geetha T. S, 2024, Journal of Environmental Management (Q1)	menggunakan pendekatan hibrida yang menggabungkan	Pendekatan model CNGRU-WQM yang dipadukan dengan IOT dalam	<p><b>Keunggulan:</b> Algoritma yang diusulkan mampu menganalisis data</p>	Penggunaan parameter yang digunakan berdasarkan SNI

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
	networks and gated recurrent units: A case study along the Vaigai River		Convolutional Neural Networks (CNN) dan Gated Recurrent Units (GRU) atau CNGRU.	mendeteksi kualitas air pada sungai dengan nilai akurasi 97.86%. performa pada model ini dapat melampaui beberapa penelitian terdahulu dalam penelitian ini.	secara real-time dari berbagai parameter kualitas air, sehingga cepat dalam menentukan kondisi air yang buruk dan mendukung pengelolaan sumber daya air secara lebih optimal <b>Kelemahan:</b> Banyaknya parameter yang digunakan yang dapat menyebabkan bias pada pembelajaran dan	dan penggunaan algoritma yang tidak membutuhkan komputasi yang tinggi

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
					membutuhkan komputasi yang tinggi	
5.	Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI	S. Ramya, 2024, Case Studies in Chemical and Environmental Engineering (Q1)	menganalisa parameter kualitas air berdasarkan WQI, menggunakan algoritma Decision Tree, Random Forest, KNN, Logistic Regression, SVM, Naive Baiyes, Gradient Boosting, Bagging, Extra Trees, Multilayer Perception (MLP), Quadratic	Penggunaan Gradient Boosting dalam penelitian ini dapat melampaui beberapa model lainnya dengan nilai akurasi 99,64% yang dipadukan dengan penggunaan Hyperparameter tuning dengan Randomized.	<b>Kelebihan:</b> Dari beberapa algoritma yang diusulkan, gradien boosting mendapatkan nilai akurasi yang tinggi dalam melakukan klasifikasi dan prediksi pada tingkat pencemaran Sungai <b>Kekurangan:</b> Saran dari penelitian ini, algoritma gradient	Penggunaan algoritma extreme gradient boosting untuk menentukan kualitas air kolam dengan menerapkan beberapa parameter saja dan pengujian dengan model-model pendukung lainnya untuk peningkatan akurasi yang baik.

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			Discriminant Analysis, XGBoost, Catboost		boosting dapat digunakan dalam klasifikasi perikanan dan dilakukan beberapa skenario dalam pengujian	
6.	Klasifikasi Kualitas Air Sungai Menggunakan Metode KNN Berbasis Web (Studi Kasus Dinas Lingkungan Hidup dan Kebersihan Aceh Utara)	Muhammad Fathan, eProceeding of TIK, 2024 (Sintu 4)	Menerapkan metode K-Nearest Neighbor (KNN) yang dipadukan dengan metode Storet untuk menentukan tingkat pencemaran air sungai berdasarkan parameter TSS, BOD, COD, PO4,	Hasil akurasi klasifikasi menggunakan KNN di RapidMiner adalah 51,85%, sedangkan dengan evaluasi <i>Confusion Matrix</i> meningkat menjadi 54,28%	<b>Kelebihan:</b> Implementasi secara langsung menggunakan Website. Implementasi ini membantu penulis tersebut dalam memprediksi menggunakan model.	Penulis menggunakan algoritma Extreme Gradient Boosting untuk menentukan kualitas air berdasarkan metode STORET. Hasil penelitian menunjukkan

No	Judul	Nama Peneliti, Tahun, Index	Metode Penelitian	Hasil	Keunggulan dan Kelemahan	Perbandingan
			DO, E. coli, pH, dan NO3-N.	(menggunakan K = 5).	<b>Kekurangan:</b> Penelitian ini menggunakan algoritma K-NN untuk menentukan kualitas air sungai	bahwa algoritma Extreme Gradeint Boosting mendapatkan nilai akurasi yang tinggi daripada penelitian ini dengan nilai akurasi sebesar 94%.

### 2.3. Landasan Teori

Pada landasan teori ini akan mencakup beberapa pembahasan seperti Kualitas Air Kolam, Labelling Data Pada STORET, Pengukuran Metode STORET, Algoritma Extreme Gradient Boosting, Hyperparameter, Hyperparameter Model, Hyperparameter Tunning, Oversampling, K-Fold dan Confusion Matrix

#### 2.3.1 Kualitas Air Kolam

Kualitas air kolam merupakan upaya yang dilakukan oleh peneliti dalam memantau dan menjaga kualitas air kolam agar tetap stabil dan dilakukan secara rutin untuk mencapai status baik bagi ekosistem ikan [19]. Dalam menjaga kualitas air kolam terdapat beberapa parameter yang digunakan dalam penelitian ini.

Diantaranya:

1. Suhu : Laju pelepasan nutrisi dari sedimen ke air dapat meningkat ketika suhu air meningkat, yang berpotensi meningkatkan konsentrasi polutan dalam air. Namun, penurunan suhu air cenderung memperlambat proses ini. Ini karena nutrisi yang dilepaskan dari sedimen dapat menyebabkan pertumbuhan alga dan tanaman air yang berlebihan, yang pada gilirannya dapat menurunkan kualitas air dan membahayakan ekosistem akuatik [20]
2. Ph : pH badan air adalah pengukuran tingkat keasaman atau alkalinitasnya, yang berkisar antara 0 hingga 14. Nilai pH yang lebih rendah menunjukkan lingkungan yang lebih asam, sedangkan nilai pH yang lebih tinggi menunjukkan lingkungan yang lebih basa [21].
3. Oksigen : Oksigen pada air sangat penting dalam kelangsungan hidup ikan,

termasuk dalam menjaga fungsi fisiologisnya dan sistem imun ikan dalam menghadapi patogen [22].

4. Amonia : merupakan zat yang tercipta dari penguraian feses pada ikan dan menumpuknya pakan yang tidak terurai pada kolam. meningkatnya ammonia dapat menyebabkan menurunnya kualitas air dan kematian masal pada ikan [23].
5. Nitrate : merupakan parameter yang dapat membantu dalam meningkatkan kadar oksigen pada air. Nitrate yang berlebih akan mempercepat eutrofikasi dan menyebabkan peningkatan pertumbuhan tanaman air sehingga mempengaruhi kadar oksigen terlarut, suhu, dan parameter lainnya.
6. Kekeruhan : merupakan parameter pada akuakultur yang digunakan dalam mengetahui kadar kekeruhan pada kondisi air.
7. Manganese (Mangan) : merupakan salah satu unsur kimia (Mn) yang dapat berperan dalam proses metabolisme. Selain itu kebutuhan dari unsur ini juga , berperan dalam pertumbuhan dan perkembangan pada ikan.

Terdapat standa Nasional Indonesia No 6484.4:2014 yang Tentang Ikan lele dumbo (*Clarias sp*). Bagian 4: pengukuran kualitas air tersebut dapat dilihat pada Tabel 2.2 dilanjut dengan Tabel 2.3 sebagai Berikut [24].

Tabel 2.2. Standar Nasional Indonesia Produksi Benih Ikan Lele

Parameter Kualitas Air	Satuan	Kisaran
Suhu	C	25 – 30
pH	-	6.5 – 8
Oksigen Terlarut	Mg/L	> 3

Tabel 2.3 Lanjutan

Parameter Kualitas Air	Satuan	Kisaran
Amonia Terlarut	Mg/L	< 0.01
Ketinggian Air	cm	25 – 40
Alkalinitas	Mg/L	> 50

Pada penelitian ini, akan menggunakan dataset yang diambil dari Kaggle, yang mana pada dataset tersebut memiliki 7 parameter dalam mengukur kualitas air. Parameter-parameter tersebut dijelaskan pada tabel 2.3 sebagai berikut.

Tabel 2.4. Pengukuran Kualitas Air Pada Pembenihan Ikan

Parameter Kualitas Air	Satuan	Kisaran
Suhu	C	25 – 30
pH	-	6.5 – 8
Oksigen Terlarut	Mg/L	> 3
Amonia Terlarut	Mg/L	< 0.01
Nitrate	Mg/L	< 50
Alkalinitas	Mg/L	40 - 200
Manganese	Mg/L	< 1

### 2.3.2 Labelling Data Pada STORET

Menurut Keputusan Menteri Negara Lingkungan Hidup Nomor : 115 Tahun 2003 Tentang Pedoman Penentuan Status Mutu Air. Metode STORET merupakan salah satu metode untuk menentukan status mutu air dengan cara membandingkan data pada kualitas air yang didapatkan dengan baku mutu air yang disesuaikan dengan peruntukannya. Cara untuk menentukan status mutu air adalah dengan

menggunakan sistem nilai dari “US-EPA (Environmental Protection Agency)” dengan mengklasifikasikan mutu air dalam empat kelas label, yaitu Baik, Cemar Ringan, Cemar Sedang dan Cemar Berat. Pada table 2.4 diperlihatkan Skor dan Label Indeks Kualitas Air dengan metode Storet [25].

Tabel 2.5. Penentuan Status Mutu Air Dengan STORET

Kelas	Skor	Label
A	0	Baik
B	-1 sd -10	Cemar Ringan
C	-11 sd -30	Cemar Sedang
D	>31	Cemar Berat

#### 2.3.2.1 Pengukuran Metode STORET

Pengukuran metode STORET ini meliputi beberapa langkah seperti:

1. Melakukan pengumpulan data kualitas dan debit air secara berkala sehingga membentuk data dari waktu ke waktu (time series data). Uraian tersebut terdapat pada Tabel 2.6 dan Tabel 2.7 sebagai lanjutan. Penentuan kualitas air, mengambil sample data pada setiap baris dengan berjumlah 3 waktu dengan kategori hasil 1 (h1), hasil 2 (h2) dan hasil 3 (h3).

Tabel 2.6. Pengumpulan Data Kualitas Air

Parameter	Satuan	Baku	h1	h2	h3
Suhu	*C	25-30	24	34	25
pH		6,5 - 8	7,25	9	7
Oksigen	mg/L	> 3	7	2	5
Amonia	mg/L	< 0,01	0,19	0	0,1
Nitrate	mg/L	< 50	20	60	20
Manganese	mg/L	< 0.1	0.01	1	0,1

Tabel 2.7 Lanjutan

Parameter	Satuan	Baku	h1	h2	h3
Alkalinitas	Mg/L	<1	0.1	0.10	0.0.1

2. Mencari nilai terkecil atau minimal (Min) dan mencari nilai terbesar atau maksimal (Max) dari setiap parameter dan Mencari nilai Rerata pada data yang dikumpulkan berdasarkan parameternya. Uraian tersebut terdapat pada Tabel 2.8. Sedangkan untuk mendapatkan nilai rerata rumus tersebut terdapat pada rumus 1.

$$\text{Rerata} = \frac{\text{Maksimal} + \text{Minimal}}{2} \quad (1)$$

Tabel 2.8. Minimal, Maksimal dan Rerata Data

Parameter	h1	h2	h3	Max	Min	Rerata
Suhu	24	34	25	34	24	29
pH	7,25	9	7	9	7	8
Oksigen	7	2	5	7	2	4,5
Amonia	0,19	0	0,1	0,19	0	0,095
Nitrate	20	60	20	60	20	40
Manganese	0.01	1	0,1	1	0,1	0,55
Alkalinitas	0.1	0.10	0.0.1	0	0	0

3. Nilai yang telah didapatkan dengan mencari nilai maksimal, minimal dan rerata, selanjutnya data nilai akan dicocokkan dengan nilai baku pada STORET. Uraian tersebut terdapat pada Tabel 2.9.

Tabel 2.9. Nilai Parameter STORET

Jumlah Contoh	Nilai	Parameter		
		Fisika	Kimia	Biologi
< 10	Maksimum	-1	-2	-3
	Minimum	-1	-2	-3
	Rata - rata	-3	-6	-9
> 10	Maksimum	-2	-4	-6
	Minimum	-2	-4	-6
	Rata - rata	-6	-12	-18

4. Jika hasil pengukuran memenuhi baku mutu air (hasil pengukuran  $\leq$  baku mutu) maka diberi skor 0.
5. Jika hasil pengukuran tidak memenuhi nilai baku mutu air (hasil pengukuran  $>$  baku mutu), maka diberi skor sesuai dengan Tabel 7 Hasil pengukuran tersebut peneliti uraikan pada Tabel 8.
6. Setelah mendapatkan hasil pengukuran dari masing-masing parameter. Maka langkah selanjutnya adalah penjumlahan Skor pada masing-masing skor maksimal, skor minimal dan skor rerata. Penjelasan tersebut dapat dilihat pada tabel 2.10.

Tabel 2.10. Skor Baku Mutu Air

Max	Min	Rerata	Jumlah Skor
-1	-1	0	-2
0	0	0	0

Tabel 2.11 Lanjutan

Max	Min	Rerata	Jumlah Skor
0	-2	0	-2
-2	0	0	-2
-2	0	0	-2
-2	0	-6	-8
Hasil			-16

7. Jumlah negatif dari seluruh parameter dihitung dan ditentukan status mutunya dari jumlah skor yang didapat menggunakan sistem nilai. Rumus penjumlahan tersebut terdapat pada rumus 2.

$$\text{Jumlah Skor} = \text{Maksimal} + \text{Minimal} + \text{Rerata} \quad (2)$$

Pengukuran ini menggunakan sistem nilai dari US-EPA (Environmental Protection Agency) dengan mengelompokkan kelas mutu air dalam empat kelas yang diuraikan pada Tabel 2.12.

Tabel 2.12. Nilai Environmental Protection Agency

Klasifikasi	Nilai	Keterangan
Baik Sekali	0	Memenuhi Baku Mutu
Baik	-1 s/d -10	Tercemar Ringan
Sedang	-11 s/d -30	Tercemar Sedang
Buruk	> 31	Tercemar Berat

### 2.3.3 Algoritma eXtreme Gradient Boosting

EXtreme Gradient Boosting merupakan algoritma yang dapat mengoptimalkan pemakaian memori pada komputer sehingga lebih efisien dan dapat mempercepat waktu pemrosesan data dibandingkan dengan algoritma Gradient Boosting [10]. Algoritma ini termasuk dalam kategori *ensemble learning* yang dapat beroperasi pada data yang kecil [26], di mana beberapa model lemah (*weak learners*), seperti pohon keputusan, digabungkan secara berulang untuk membangun model prediksi yang lebih kuat dan akurat. Proses ini dilakukan dengan cara menyesuaikan model baru untuk memperbaiki kesalahan prediksi yang dibuat oleh model sebelumnya, sehingga dapat meningkatkan kinerja secara bertahap. Kelebihan dari algoritma ini adalah nilai akurasi prediktif yang tinggi dan mampu menangani berbagai jenis data dengan baik [10].

### 2.3.4 Hyperparameter

Hyperparameter dalam klasifikasi adalah cara model mempelajari pola data dan membuat keputusannya dalam menentukan sebuah kategori atau label [27]. Fungsi penggunakan Hyperparameter ini adalah untuk mengoptimalkan kinerja model, sehingga dapat menghasilkan nilai klasifikasi yang lebih akurat. Dalam penggunaannya, terdapat hyperparameter yang terdapat pada model itu sendiri dan hyperparameter tuning yang digunakan untuk optimasi pada model pelatihan.

### 2.3.5 Hyperparameter Model

Setiap model mempunyai Hyperparameternya masing-masing, seperti algoritma Support Vector Machine (SVM) yang mempunyai Cost dan Gamma dan

K-NN yang mempunyai Distance Metrix dan Weighting. Hyperparameter sendiri tidak dipelajari langsung dari data pelatihan, namun nilai hyperparameter tersebut ditentukan sebelum proses pelatihan dimulai. Beberapa hyperparameter tersebut diantaranya seperti  $n\_estimator$ , Max-Depth (Kedalaman Pohon) dan Learning Rate dalam melakukan klasifikasi pada kualitas air pada budidaya.

1.  $N\_estimator$  adalah jumlah total pohon yang akan dibuat dalam model ensemble. Penambahan jumlah pohon bisa meningkatkan kemampuan model dalam menangkap pola data yang kompleks, tetapi jika terlalu banyak bisa menyebabkan overfitting. [11].
2. Learning Rate digunakan untuk memperbarui bobot model selama pelatihan dan memperbaiki kesalahan pada sebelumnya (ensemble).
3. Max Depth digunakan untuk menentukan batas kedalaman maksimum pada pohon keputusan, sehingga dapat membatasi model dalam mempelajari fitur dan target yang terlalu spesifik [10].

### **2.3.6 Hyperparameter Tuning**

Hyperparameter tuning merupakan metode yang digunakan dalam machine learning sebelum dilakukan pelatihan untuk mencari nilai kombinasi optimal dari nilai parameter yang telah ditentukan sebelumnya. Metode ini memiliki beberapa pendekatan yang digunakan untuk meningkatkan performa model pelatihan seperti :

1. Grid Search : merupakan metode yang digunakan dalam mencari nilai parameter terbaik dengan mencoba semua kombinasi hyperparameter yang telah ditentukan sebelumnya [17].

2. Bayesian Optimization : Merupakan metode yang mempertimbangkan ketidakpastian dalam prediksi untuk memilih hyperparameter yang tepat [28].
3. Random Search : Merupakan metode yang digunakan dalam mencari parameter terbaik dengan mencoba sejumlah sampel acak untuk menemukan kombinasi dari beberapa hyperparameter yang optimal.

Pada penelitian ini, penulis menggunakan pendekatan Random Search karena pendekatan ini membutuhkan waktu yang optimal daripada hyperparameter lainnya. Meskipun kelemahan Random Search adalah mengambil sample acak pada hyperparameter, namun pendekatan ini dapat memaksimalkan nilai iterasi untuk mendapatkan nilai hyperparameter yang optimal, seperti yang dilakukan oleh beberapa penelitian sebelumnya.

### **2.3.7 Oversampling**

Oversampling merupakan teknik yang dilakukan pada data yang tidak seimbang dengan menyeimbangkan data minoritas dengan diseimbangkan dengan yang mayoritas sehingga data tersebut dapat seimbang. Teknik tersebut dapat dilakukan setelah pembagian data latih dan data uji dilakukan [28]. Di bawah ini merupakan beberapa teknik oversampling untuk menyeimbangkan data yang tidak seimbang. Beberapa teknik tersebut diantaranya.

#### **2.3.7.1 SMOTE**

Synthetic Minority Oversampling Technique (SMOTE) adalah metode oversampling yang menghasilkan data sintetik baru berdasarkan data minoritas yang dipilih secara acak tanpa menduplikasi data tersebut agar terhindar dari

overfitting dan kemudian membuat titik baru di antara data tersebut dan salah satu tetangga terdekatnya berdasarkan perhitungan jarak [29]

#### **2.3.7.2 Borderline-SMOTE**

Merupakan metode oversampling pada data minoritas, yang sama seperti SMOTE. Namun dalam menghasilkan data baru, metode ini membuat data baru yang posisinya berada dekat dengan batas antara kelas minoritas dan kelas mayoritas. Metode ini menjadi solusi atas kelemahan metode SMOTE, yaitu potensi terjadi tumpang tindih pada data sintetis pada data mayoritas (Kaisar dan Chowdhury 2022). Sehingga apabila itu terjadi, maka data yang tumpang tindih tersebut akan membuat sebuah noise dan berpotensi mengganggu proses pelatihan (training) model.

#### **2.3.7.3 ADASYN (Adaptive Synthetic Sampling)**

ADASYN merupakan metode oversampling yang menggunakan teknik menambahkan data sintesis dengan mempertimbangkan jumlah tetangga yang paling dekat dengan data mayoritas. Tidak seperti SMOTE dan Borderline-SMOTE, ADASYN merupakan teknik oversampling yang digunakan dalam membantu algoritma klasifikasi agar lebih fokus dalam menentukan area yang sulit dalam melakukan klasifikasi dan mengurangi kesalahan dalam menentukan data minoritas [30].

Pada Penelitian ini akan menggunakan Teknik SMOTE Oversampling untuk menyeimbangkan data pada label kualitas air yang tidak seimbang. Sehingga kesalahan dalam melakukan prediksi kualitas air akan sangat kecil.

### 2.3.8 K-Fold

K-Fold merupakan teknik yang digunakan pada model untuk membagi dataset menjadi nilai K bagian, di mana setiap bagian data secara acak akan bergantian digunakan sebagai data pengujian. Pengujian dengan metode K-Fold sangatlah berguna untuk mengurangi kesalahan seperti bias dan overfitting dengan mengevaluasi subset pada data yang telah ditentukan oleh model [31]. Penggunaan nilai K-fold ini berdasarkan beberapa literatur penelitian yang telah dilakukan pada penelitian sebelumnya. Meskipun tidak ada aturan yang pasti untuk menentukan nilai pada K, namun pengujian ini akan menggunakan pengujian menggunakan dari K-Fold 2 sampai K-Fold 20, untuk mendapatkan nilai model yang terbaik [32].

### 2.3.9 Confusion Matrix

Confusion Matrix adalah sebuah tabel yang diperoleh dari hasil pengujian model klasifikasi, baik pada tahap training maupun testing, yang menggambarkan jumlah prediksi benar dan salah dalam klasifikasi, termasuk bagaimana prediksi tersebut didistribusikan ke masing-masing kelas. [33].

Penggunaan Confusion Matrix tersebut digunakan untuk mengukur performa dari model dengan metrik Akurasi, Presisi, Recall dan F1-Score [34]. Dalam penggunaan Confusion Matrix, kita akan melihat jumlah data yang terbagi menjadi 4 bagian.

1. Bagian pojok kiri atas yang diprediksi sebagai data yang benar-benar positif dan diprediksi sebagai positif. (True Positive).
2. Bagian pojok kanan atas data yang sebenarnya negatif tetapi diprediksi sebagai positif (False Positive).

3. Pada bagian kiri bawah kita akan menemukan jumlah data yang sebenarnya positif tetapi diprediksi sebagai negatif. (False Negative)

Pada pojok kanan bawah kita akan menemukan jumlah data yang benar-benar negatif dan diprediksi sebagai negatif. (True Negative).

Penjelasan lebih rinci dapat dilihat pada Gambar 2.1.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2.1 Confusion Matrix

Dalam praktiknya, selain penggunaan Confusion Matrix dalam penelitian ini. Penulis menggunakan nilai Akurasi dalam mengukur seberapa baik model dalam melakukan klasifikasi. Nilai Presisi merupakan nilai yang digunakan dalam memprediksi benar positif. Nilai Recall merupakan nilai yang digunakan dalam mengukur banyaknya total kasus positif yang berhasil diidentifikasi oleh model. F1-Score merupakan perbandingan hasil rata-rata nilai Presisi dan Recall untuk memberikan hasil metrik yang seimbang, terutama pada dataset dengan pelabelan yang tidak seimbang [35]. Untuk mengetahui bagaimana cara menghitung nilai Akurasi, Presisi, Recall dan F1-Score, dapat dilihat pada rumus 3 - 6.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

$$PRECISION = \frac{TP}{FP + TP} \quad (4)$$

$$RECALL = \frac{TP}{TP + FN} \quad (5)$$

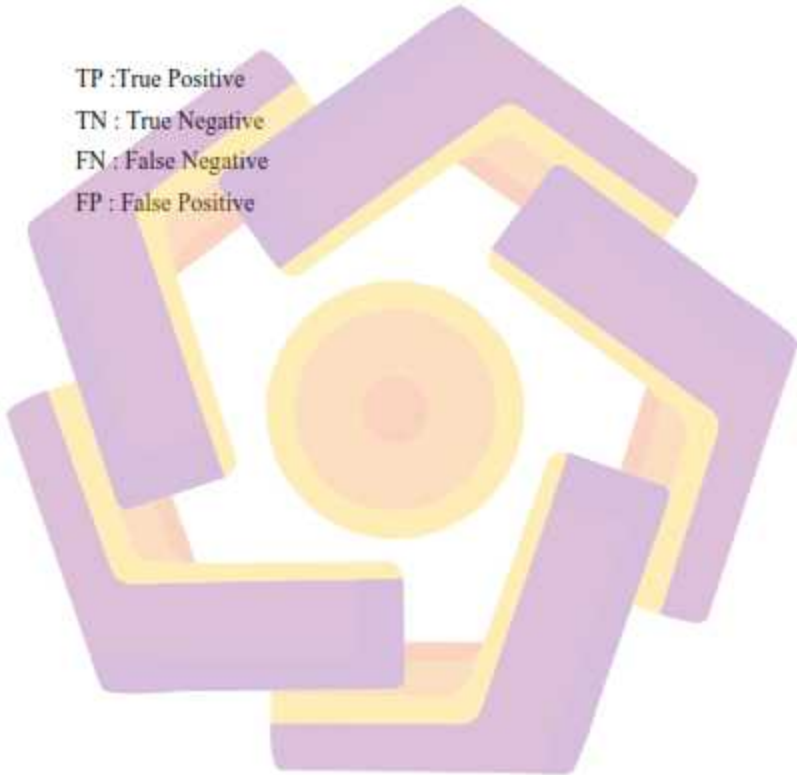
$$F1 = 2 \cdot \frac{Precision + Recall}{Precision + Recall} \quad (6)$$

TP : True Positive

TN : True Negative

FN : False Negative

FP : False Positive



## BAB III METODE PENELITIAN

### 3.1. Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini bersifat kuantitatif, dataset yang digunakan merupakan dataset yang didapatkan dari website Kaggle. Untuk menentukan label kualitas air pada pembenihan ikan lele, penulis akan menggunakan metode STORET pada pelabelan Indeks Kualitas Air. Metode ini merupakan cara untuk menentukan status mutu air menggunakan sistem nilai dari US-EPA (Environmental Protection Agency), yang mengklasifikasikan nilai baku mutu air ke dalam empat kelas, yaitu Baik, Cemar Ringan, Cemar Sedang dan Cemar Berat.

### 3.2. Metode Pengumpulan Data

Dataset yang digunakan berasal dari Kaggle dengan judul Pondsdata. Dataset tersebut terdiri dari 10 data dari hasil monitoring yang memanfaatkan teknologi Internet Of Things, terdapat pada gambar 3.2. Parameter tersebut diantaranya Station, Date, Time, Nitrate, Ph, Ammonia, Suhu, Oksigen, Kekeruhan dan Manganese. Dalam mengolah data nantinya penulis akan menggunakan 7 parameter, seperti yang penulis jelaskan pada landasan teori sebelumnya.

nitrate	ph	ammonia	suhu	oksigen	kekeruhan	manganese	label
5.0	5.1	0.054	23.41	10.5	18.8	0.62	tercemar sedang
18.1	5.2	0.018	23.84	9.4	26.7	0.64	tercemar sedang
1.8	5	0.039	22.85	7.3	18.8	0.67	tercemar sedang
17.7	5.8	0.058	24.21	9.8	26.5	0.7	tercemar sedang
1.4	5	0.008	24.35	8	31.9	0.68	tercemar sedang

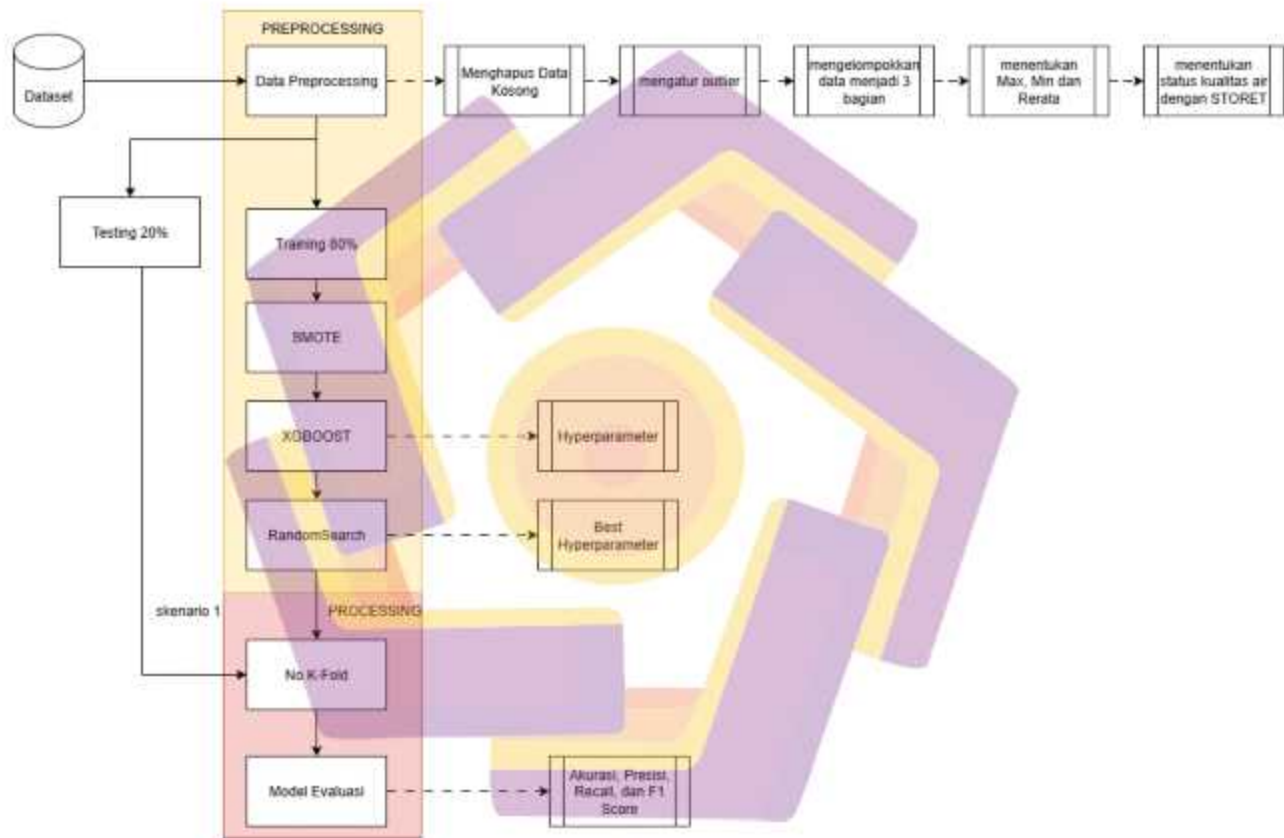
Gambar 3.2. Dataset Kualitas Air

### 3.3. Metode Analisis Data

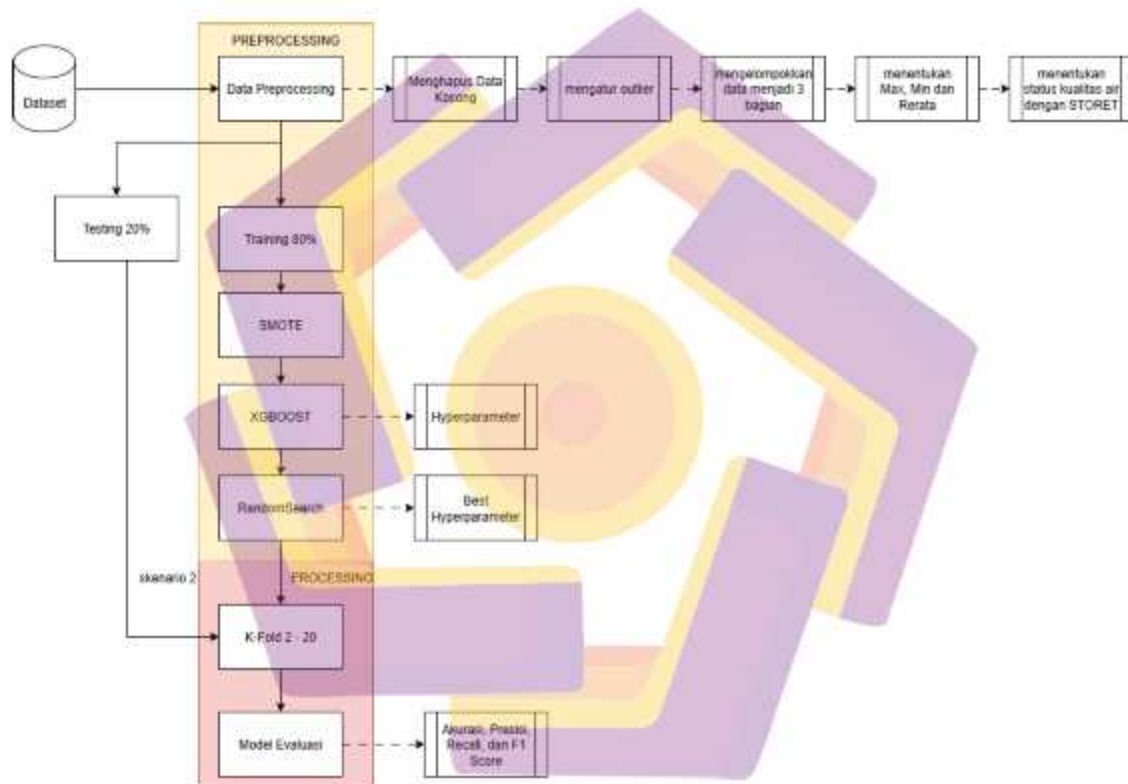
Analisis dalam penelitian ini dapat dilakukan dengan menggunakan beberapa metode, seperti Analisis Performansi dan Analisis Visualisasi. Analisis Performansi merupakan metode yang mengukur seberapa baik machine learning dapat melakukan tugasnya. Dalam analisis performansi ini penulis menggunakan model klasifikasi yang mana pada analisis ini mengukur seberapa akurat nilai Akurasi, Presisi, Recall dan F1-Score pada model. Analisis Visualisasi merupakan metode analisis yang digunakan untuk membandingkan hasil evaluasi algoritma Extreme Gradient Boosting (XGBoost).

### 3.4. Alur Penelitian

Pada bagian ini, penulis memberikan gambaran alur penelitian pada Extreme Gradient Boosting dalam memprediksi kualitas air. Pada penelitian ini penulis membaginya menjadi dua tahap setelah dilakukan tuning pada model



Gambar 3.3. Skenario Pertama Preprocessing dan Processing



Gambar 3.4. Skenario Kedua Preprocessing dan Processing

Alur penelitian secara sistematis dapat dilihat pada Gambar dan akan dijelaskan sebagai berikut:

### **1. Dataset**

Data diambil dari sebuah website yang Bernama Kaggle. Dataset ini sudah penulis jelaskan pada Metode Pengumpulan Data.

### **2. Data Preprocessing**

Tahap preprocessing ini meliputi :

1. menghapus data yang kosong agar dapat dilatih oleh model
2. mengatur outlier nilai pada data dengan mengubahnya sesuai dengan standar nilai data.
3. Mengelompokkan data menjadi tiga bagian. Pengelompokkan data tersebut digunakan untuk mempermudah pelabelan STORET
4. Menentukan nilai Maksimal (max), Minimal (min) dan Rata-rata (Rerata) pada setiap parameter yang selanjutnya akan diberikan nilai skor sesuai dengan nilai STORET. Dan menjumlahkan seluruh skor yang telah didapatkan dari setiap parameter.
5. Menentukan status kelompok status mutu air berdasarkan jumlah keseluruhan skor yang telah dijumlahkan secara keseluruhan

### **3. Model Training**

Tahap ini membagi data menjadi 80% Training dan 20% Testing. Sehingga jumlah pada data Training 26.199 data dan data Testing berjumlah 6.550 data. Data Training mempunyai banyak label kualitas air sebagai berikut.

Kelas 0 (Baik) : 69 data

Kelas 1 (Tercemar Ringan) : 6537 data

Kelas 2 (Tercemar Sedang) : 25420 data

Kelas 3 (Tercemar Berat) : 723 data

#### **4. SMOTE**

Dataset yang tidak seimbang akan lakukan oversampling dengan metode SMOTE. Sehingga data menjadi seimbang dan dapat masuk ke tahap selanjutnya. Rincian kelas yang telah dilakukan oversampling dapat dilihat sebagai berikut.

Kelas 0 (Baik) : 20322 data

Kelas 1 (Tercemar Ringan) : 20322 data

Kelas 2 (Tercemar Sedang) : 20322 data

Kelas 3 (Tercemar Berat) : 20322 data

#### **5. Random Search**

Random Search digunakan untuk pencarian hyperparameter terbaik pada Extreme Gradient Boosting. Penelitian ini menggunakan penggunaan hyperparameter berdasarkan penelitian terdahulu dan tambahan nilai pada penggunaan hyperparameter.

#### **6. Skenario**

Penelitian ini menggunakan dua skenario untuk membandingkan seberapa akurat model dalam mengklasifikasi kualitas air.

### **6.1 K-Fold (Skenario 1)**

Skenario yang kedua, setelah mendapatkan hyperparameter terbaik dengan Random Search, maka data akan dilatih dengan bantuan K-Fold untuk meningkatkan hasil evaluasi dari model. Penelitian ini akan menguji beberapa skenario pada K-Fold dari K-Fold 2 sampai K-Fold 20 dengan menguji seberapa pengujian yang dilakukan pada setiap K-Fold.

### **6.2 Tidak Menggunakan K-Fold (Skenario 2)**

Skenario yang pertama, setelah mendapatkan hyperparameter terbaik dengan Random Search, maka data akan langsung dilatih dan mendapatkan hasil evaluasinya tersebut.

## **7. Model Evaluasi**

Data yang telah dilatih akan dievaluasi dengan melihat nilai akurasi, presisi, recall dan F1-Score.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### 4.1 Oversampling Dengan SMOTE

Pada Pembahasan ini, penulis menggunakan SMOTE untuk menyeimbangkan data yang telah dijelaskan pada BAB III sebelumnya. Sebelum dilakukannya SMOTE, dilakukan pengecekan terhadap jumlah data Training untuk setiap label kelas kualitas air, yaitu 0 (Baik), 1 (Tercemar Ringan), 2 (Tercemar Sedang), dan 3 (Tercemar Berat). Source code tersebut penulis tunjukkan pada gambar 4.5.



```

Sobelum dilakukan SMOTE
# jumlah data training pada label: 0, 1, 2, 3

print("Good ", len(X_train[y_train == 0]))
print("Tercemar Ringan ", len(X_train[y_train == 1]))
print("Tercemar Sedang ", len(X_train[y_train == 2]))
print("Tercemar Berat ", len(X_train[y_train == 3]))

In [ ]: Good 55
        Tercemar Ringan 5239
        Tercemar Sedang 20322
        Tercemar Berat 583
  
```

Gambar 4.5. Pengecekan Setiap Kelas Pada Data Training

Untuk mengatasi permasalahan data yang tidak seimbang, dilakukan oversampling menggunakan metode SMOTE (Synthetic Minority Oversampling Technique) yang berasal dari library `imblearn.over_sampling`. Setelah proses oversampling selesai, dilakukan pengecekan kembali terhadap jumlah data pada masing-masing kelas. Kode program tersebut terdapat pada gambar 4.6.

```

from imblearn.over_sampling import SMOTE

# Buat objek SMOTE
smote = SMOTE(sampling_strategy='auto', random_state=42)

# Lakukan oversampling pada data training
X_train_resampled, y_train_resampled = smote.fit_resample(X_train, y_train)

# Cek distribusi setelah SMOTE
print('kelas 0 (Baik)', len(y_train_resampled[y_train_resampled == 0]))
print('kelas 1 (Tercemar Ringan)', len(y_train_resampled[y_train_resampled == 1]))
print('kelas 2 (Tercemar Sedang)', len(y_train_resampled[y_train_resampled == 2]))
print('kelas 3 (Tercemar Berat)', len(y_train_resampled[y_train_resampled == 3]))

```

kelas 0 (Baik) 20322  
 kelas 1 (Tercemar Ringan) 20322  
 kelas 2 (Tercemar Sedang) 20322  
 kelas 3 (Tercemar Berat) 20322

Gambar 4.6 Oversampling Dengan SMOTE

Setelah dilakukan proses oversampling, setiap kategori kualitas air memiliki jumlah data yang sama, sehingga model dapat melakukan proses pembelajaran dengan lebih optimal dan menghasilkan klasifikasi kualitas air yang lebih akurat dan adil antar kelas

#### 4.2 Hasil Skenario Pertama

Hasil scenario pada penelitian ini tidak menggunakan metode K-Fold dalam mendapatkan hasil uji-evaluasi pada penelitian ini.

#### 4.3 Hyperparameter Tuning Xtreme Gradient Boosting

Dalam penelitian ini, dilakukan proses hyperparameter tuning terhadap algoritma Xtreme Gradient Boosting (XGBoost) dengan menggunakan pendekatan Random Search. Proses ini bertujuan untuk mencari kombinasi parameter terbaik yang dapat meningkatkan kinerja model dalam melakukan klasifikasi kualitas air. Adapun ruang lingkup pencarian hyperparameter yang digunakan ditunjukkan pada tabel 4.13 :

Tabel 4.13. Preprocessing Hyperparameter Tuning Xtreme Gradient Boosting

Nama Hyperparameter	Jumlah Kombinasi
N_estimator	: 10, 60, 100, 200
Max_depth	: 2, 3, 7, 10
Learning_rate	: 1.0, 0.5, 0.01

Proses pencarian hyperparameter terbaik menggunakan metode *Random Search* menghasilkan kombinasi parameter yang optimal, seperti ditampilkan pada Gambar 4.7. Hasil tuning menunjukkan bahwa kombinasi hyperparameter terbaik dan waktu yang ditentukan dalam mendapatkan hyperparameter terbaik, adalah sebagai berikut :

```
Best Hyperparameters: {'n_estimators': 200, 'max_depth': 7, 'learning_rate': 0.5}
```

Gambar 4.7. Hyperparameter Terbaik

1. n\_estimators: 200
2. max\_depth: 7
3. learning\_rate: 0.5

Kombinasi hyperparameter ini memberikan performa terbaik pada data validasi. Oleh karena itu, konfigurasi ini digunakan sebagai model akhir dalam proses klasifikasi kualitas air pada kolam pembenihan.

#### 4.4 Evaluasi Skenario Penelitian Pertama

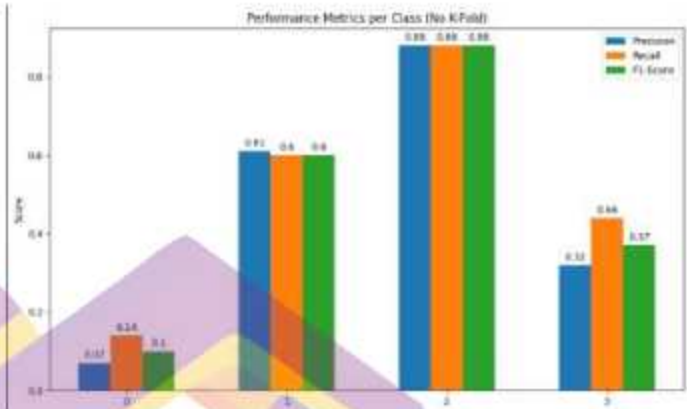
Skenario penelitian pertama ini memperlihatkan hasil processing tanpa menggunakan K-Fold yang ditampilkan pada gambar 4.8.

```

No K-FOLD
-----
Rata-rata Akurasi: 0.8098
Rata-rata Precision: 0.4707
Rata-rata Recall: 0.5152
Rata-rata f1-score: 0.4877
-----
Laporan Klasifikasi:
precision  recall  f1-score
-----
0         0.07   0.14   0.10
1         0.61   0.60   0.60
2         0.88   0.88   0.88
3         0.32   0.44   0.37

accuracy
macro avg  0.47   0.52   0.49
weighted avg 0.82   0.81   0.81

```



Gambar 4.8. Hasil Skenario Pertama

Skenario pertama, pengujian dilakukan tanpa menggunakan metode K-Fold, sehingga proses pelatihan dan pengujian model dilakukan secara langsung tanpa pembagian silang (K-Fold). Hal ini dilakukan untuk melihat performa murni dari model XGBoost terhadap data uji yang tersedia. Berdasarkan hasil evaluasi model yang ditampilkan pada Gambar 6, diperoleh hasil metrik klasifikasi sebagai berikut:

1. Nilai akurasi pada model mencapai 0.8098 dengan nilai akurasi rata-rata 81%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data dengan cukup baik dalam memprediksi secara keseluruhan.
2. Nilai rata-rata precision mencapai 0,4707 yang menunjukkan bahwa model masih memiliki kesalahan dalam melakukan prediksi setiap kelas dengan rincian :
  - 2.1 Kelas 0 (Baik) dengan nilai precision 0.07 yang menunjukkan dari seluruh prediksi model yang menyebut data sebagai "Baik", hanya 7% yang benar-benar

Baik. Dalam hal ini berarti model sering melakukan kesalahan dalam memprediksi data sebagai Baik, padahal sebenarnya bukan.

**2.2** Kelas 1 (Tercemar Ringan) dengan nilai precision 0.61 yang menunjukkan dari seluruh prediksi model yang menyebut data sebagai "Tercemar Ringan", hanya 61% memprediksi kelas ini dan cukup sering benar.

**2.3** Kelas 2 (Tercemar Sedang) dengan nilai precision 0,88 yang menunjukkan dari seluruh prediksi model yang menyebut data sebagai "Tercemar Sedang", hanya 88% yang menunjukkan model sangat akurat saat menyebut data sebagai Tercemar Sedang.

**2.4** Kelas 3 (Tercemar Berat) dengan nilai precision 0,32 yang menunjukkan dari seluruh prediksi model yang menyebut data sebagai "Tercemar Berat", hanya 33% saja sehingga model sering salah ketika menyebut data sebagai "Tercemar Berat".

**3.** Nilai rata-rata recall yang dicapai model adalah sebesar 0,5152. Angka ini menunjukkan bahwa secara keseluruhan, kemampuan model dalam mengenali kelas-kelas yang sebenarnya masih tergolong rendah. Rincian nilai recall untuk masing-masing kelas disajikan pada penjelasan berikut.:

**3.1** Kelas 0 (Baik) mencapai nilai recall 0.14 atau 14 % yang menunjukkan model hampir tidak bisa mengenali data 'Baik' dengan benar. Sebagian besar data kelas ini justru dikira sebagai kelas lain.

**3.2** Kelas 1 (Tercemar Ringan) memiliki nilai recall sebesar 0.60 atau 60%, yang artinya model memiliki kemampuan yang cukup baik dalam mengenali

data pada kelas Tercemar Ringan, meskipun masih terdapat sejumlah kesalahan dalam mendeteksi kelas tersebut..

**3.3 Kelas 2 (Tercemar Sedang)** memiliki nilai recall sebesar 0.88 atau 88%, yang menunjukkan bahwa model memiliki kemampuan yang tinggi dalam mengenali data pada kelas Tercemar Sedang. Dengan demikian, dapat dikatakan bahwa model sangat efektif dalam mendeteksi kelas ini dibandingkan kelas lainnya.

**3.4 Kelas 3 (Tercemar Berat)** memiliki nilai recall sebesar 0.44 atau 44%, yang artinya model hanya mampu mengenali sebagian kecil data yang sebenarnya termasuk dalam kelas ini. Hal ini menunjukkan bahwa banyak data Tercemar Berat yang tidak terdeteksi dengan benar oleh model.

**4.** Nilai rata-rata F1-Score mencapai 0,4877, yang menunjukkan bahwa secara keseluruhan performa model dalam menyeimbangkan antara precision dan recall pada setiap kelas masih tergolong rendah. Hal ini mengindikasikan bahwa model belum mampu melakukan klasifikasi dengan baik terhadap semua kelas secara merata. Rincian dari nilai F1-Score pada masing-masing kelas diuraikan sebagai berikut :

**4.1 kelas 0 (Baik)** Memiliki nilai F1-Score sebesar 0.10 atau 10%, yang menunjukkan bahwa model memiliki kemampuan yang sangat rendah dalam mengklasifikasikan data pada kelas Baik. Nilai ini mencerminkan ketidakseimbangan yang ekstrem antara precision dan recall, serta jumlah data yang sangat sedikit, sehingga model kesulitan dalam mengenali pola yang khas dari kelas ini.

**4.2** Kelas 1 (Tercemar Ringan) Memiliki nilai F1-Score sebesar 0.60 atau 60%, yang menunjukkan bahwa model memiliki kemampuan sedang dalam mendeteksi dan mengklasifikasikan data pada kelas Tercemar Ringan. F1-Score ini menandakan bahwa model mampu menjaga keseimbangan antara precision dan recall pada tingkat yang cukup stabil untuk kelas ini.

**4.3** Kelas 2 (Tercemar Sedang) Memiliki nilai F1-Score sebesar 0.88 atau 88%, yang menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall dalam mengenali data pada kelas Tercemar Sedang. Dengan demikian, dapat dikatakan bahwa model sangat efektif dalam mengklasifikasikan kelas ini dibandingkan kelas lainnya.

**4.4** Kelas 3 (Tercemar Berat) Memiliki nilai F1-Score sebesar 0.37 atau 37%, yang menunjukkan bahwa model memiliki kemampuan yang rendah dalam mengenali data pada kelas Tercemar Berat. Meskipun sedikit lebih tinggi dibanding kelas Baik (kelas 0), nilai ini tetap menunjukkan bahwa model belum mampu mengklasifikasikan kelas ini secara konsisten, kemungkinan disebabkan oleh jumlah data yang minim dan kemiripan fitur dengan kelas lain.

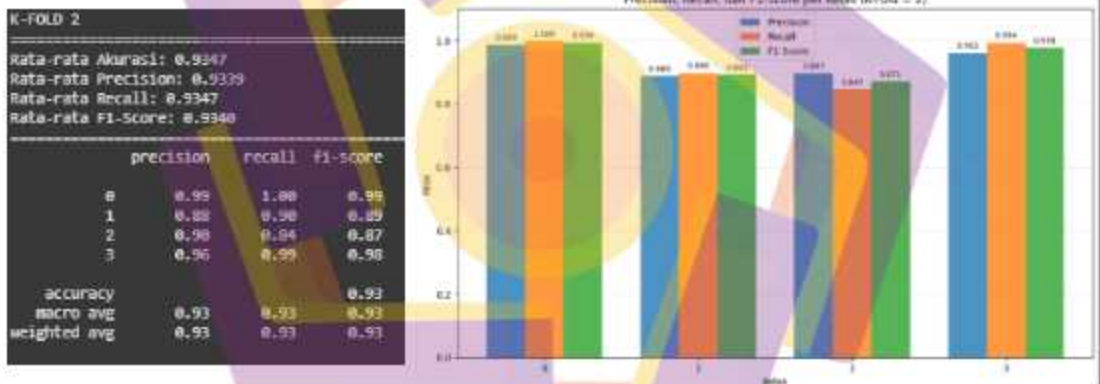
#### **4.4 Hasil Skenario Penelitian Kedua Dengan Pengujian K-Fold**

Skenario penelitian kedua ini menggunakan K-Fold sebagai pembanding antara hasil pengujian pertama yang tidak menggunakan K-Fold dengan hasil pengujian yang menggunakan K-Fold. Pengujian pada K-Fold ini akan menggunakan 20 skenario untuk mengetahui hasil terbaik yang didapatkan dari

setiap K-Fold. Untuk mengetahui setiap K-Fold tersebut, penulis menguraikannya dibawah ini

#### 4.4.1 K-Fold 2

Model pengujian ini menggunakan teknik validasi silang dengan membagi data menjadi dua bagian, data yang dilatih pada satu bagian dan model lainnya sebagai data diuji pada bagian secara bergantian, sehingga diperoleh dalam bentuk metrik akurasi, precision, recall, dan F1-Score pada masing-masing kelas serta nilai rata-rata keseluruhan pada gambar 4.9.



Gambar 4.9 Metrik Evaluasi dan Diagram K-Fold 2

1. Nilai akurasi yang didapatkan pada model sebesar 0,9347 atau 93%. Nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar. Nilai ini mencerminkan tingkat keakuratan yang tinggi dalam keseluruhan proses klasifikasi.
2. Rata-rata pada precision mendapatkan nilai sebesar 0.9339 atau 93.39%, yang berarti dari seluruh prediksi yang dinyatakan positif oleh model, sebanyak 93.39%

merupakan prediksi yang benar. Tingginya nilai precision ini menandakan bahwa model memiliki kemampuan yang sangat baik dalam menghindari kesalahan klasifikasi positif palsu, atau dengan kata lain, mampu mengenali kelas yang tepat dengan tingkat kesalahan minimum.

Adapun juga nilai rata - rata dari setiap kelas pada precision sebagai berikut :

**2.1** Kelas 0 (Baik) memperoleh nilai precision yang hampir sempurna yaitu 0.99 atau 99% yang menjelaskan bahwa pada kelas 0 yang menunjukkan bahwa dari seluruh prediksi yang dilakukan oleh model sebagai kelas Baik. Nilai precision yang sangat tinggi ini mengindikasikan bahwa model memiliki tingkat kesalahan positif palsu (false positive) yang sangat rendah dalam mendeteksi kondisi air yang benar-benar baik. Hal ini menunjukkan bahwa model sangat selektif dan tepat sasaran dalam mengidentifikasi kondisi optimal kualitas air.

**2.2** Kelas 1 (Tercemar Ringan) memperoleh nilai Precision yaitu 0.88 atau 88% Ini berarti bahwa dari semua prediksi yang diklasifikasikan sebagai Tercemar Ringan, sebanyak 88% adalah benar. Meskipun nilai ini tergolong tinggi, masih terdapat 12% prediksi keliru (false positive) yang diklasifikasikan ke kelas ini padahal seharusnya termasuk ke kelas lain. Hal ini menunjukkan bahwa model sudah cukup tepat dalam mendeteksi pencemaran ringan, namun masih terdapat tumpang tindih prediksi dengan kelas lain yang perlu diminimalkan, terutama terhadap kelas 2 yang berdekatan dalam spektrum kualitas air.

**2.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Precision yaitu 0.90 atau 90% model menghasilkan nilai precision sebesar 0.90 atau 90%. Artinya, dari seluruh prediksi yang ditetapkan sebagai kelas ini, sebanyak 90% benar-benar berasal dari

data aktual kelas Tercemar Sedang. Nilai ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik, terutama dalam membedakan kelas ini dari kelas lainnya. Akan tetapi, perlu diperhatikan bahwa meskipun precision-nya tinggi, recall pada kelas ini lebih rendah (84%), yang berarti masih ada data kelas 2 yang belum terdeteksi secara penuh oleh model.

**2.4** Kelas 3 (Tercemar Berat) memperoleh nilai Precision yaitu 0.96 atau 96% yang menunjukkan bahwa sebagian besar prediksi pada kelas ini benar dan hanya sedikit yang salah diklasifikasikan. Nilai ini menandakan bahwa model sangat tepat dalam menentukan kondisi air yang sangat tercemar, dan jarang mengklasifikasikan data dari kelas lain ke dalam kategori ini secara keliru. Dalam konteks budidaya perikanan, hal ini memberikan keuntungan signifikan, karena keputusan kritis seperti penggantian air atau intervensi biologis akan berdasarkan prediksi yang sangat akurat.

**3.** Rata-rata pada nilai Recall diperoleh sebesar 0.9347 atau 93.47%, yang menggambarkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi seluruh dari setiap kelas. Artinya, sebagian besar data aktual dari setiap kelas berhasil dikenali dengan benar oleh model.

Nilai rata-rata pada Recall tersebut, didapatkan dari hasil setiap kelas kualitas air yang dijelaskan sebagai berikut.

**3.1** Kelas 0 (Baik) memperoleh nilai Recall sebesar 1.00 atau 100% yang menunjukkan bahwa seluruh data aktual yang termasuk dalam kelas ini berhasil diidentifikasi dengan benar oleh model. Nilai recall sempurna ini mengindikasikan bahwa tidak ada satupun data pada kelas Baik yang salah klasifikasi. Hal ini

menunjukkan bahwa model memiliki kemampuan sangat tinggi dalam mendeteksi air dengan kualitas terbaik, sehingga dapat diandalkan untuk mendeteksi kondisi optimal dalam budidaya perikanan.

**3.2** Kelas 1 (Tercemar Ringan) memperoleh nilai Recall sebesar 0.90 atau 90% data aktual yang tergolong tercemar ringan berhasil dikenali dengan benar oleh model, sementara sisanya salah diklasifikasikan ke kelas lain. Nilai recall yang cukup tinggi ini menunjukkan bahwa model memiliki kemampuan deteksi yang baik namun masih belum sempurna dalam mengenali kondisi air yang mulai menunjukkan gejala pencemaran ringan. Dalam konteks penerapan di lapangan, hal ini menandakan bahwa masih ada ruang perbaikan agar model lebih sensitif terhadap kelas ini.

**3.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Recall sebesar 0.84 yang menjelaskan sebagian besar data aktual dari kelas ini berhasil dikenali dengan benar, namun masih terdapat 16% data yang salah diklasifikasikan ke kelas lain. Meskipun nilai recall ini cukup tinggi, namun performa model pada kelas ini sedikit lebih rendah dibandingkan kelas lainnya. Hal ini kemungkinan disebabkan oleh kemiripan karakteristik fitur antara kelas Tercemar Sedang dan kelas Tercemar Ringan atau Berat, yang menyebabkan model kesulitan membedakan batas antar kelas.

**3.4** Kelas 3 (Tercemar Berat) memperoleh nilai Recall sebesar 0.99 yang menunjukkan bahwa model memiliki kemampuan hampir sempurna dalam mengenali kondisi air yang sangat tercemar. Hanya terdapat 1% data yang gagal diklasifikasikan ke dalam kelas ini. Tingginya nilai recall ini sangat penting dalam

konteks aplikatif, karena kondisi air yang tercemar berat dapat menimbulkan kerugian besar dalam budidaya ikan. Dengan demikian, performa model pada kelas ini sangat memuaskan dan menunjukkan potensi besar untuk digunakan dalam sistem peringatan dini dalam pengelolaan kualitas air.

4. rata-rata F1-Score sebesar 0.9340 atau 93.40%, yang menunjukkan keseimbangan yang sangat baik antara precision dan recall. Nilai F1-Score yang tinggi ini mengindikasikan bahwa model tidak hanya akurat dalam membuat prediksi, tetapi juga konsisten dalam mengenali semua kelas secara proporsional.

Nilai rata-rata pada F1-Score tersebut, didapatkan dari hasil setiap kelas kualitas air yang dijelaskan sebagai berikut :

**4.1** Kelas 0 (Baik) memperoleh nilai F1-Score sebesar 0.99 atau 99% Nilai ini menunjukkan bahwa model memiliki keseimbangan sangat tinggi antara ketepatan dan kelengkapan dalam mengidentifikasi data kelas Baik. Dengan F1-Score hampir sempurna, dapat disimpulkan bahwa model sangat andal dalam mendeteksi kondisi air yang benar-benar memenuhi standar kualitas tinggi tanpa banyak kesalahan prediksi. Hal ini penting terutama dalam mencegah perlakuan intervensi yang tidak diperlukan pada air yang sebenarnya baik.

**4.2** Kelas 1 (Tercemar Ringan) memperoleh nilai F1-Score sebesar 0.89 atau 89% Nilai ini menunjukkan bahwa model memiliki kinerja yang baik namun tidak sempurna dalam mengenali kelas ini. Meskipun precision dan recall masing-masing bernilai 0.88 dan 0.90, skor F1 menunjukkan adanya sedikit ketidakseimbangan antara jumlah data yang benar terklasifikasi dan kesalahan prediksi. Hal ini bisa diakibatkan oleh tumpang tindih karakteristik data antara kelas Tercemar Ringan

dan kelas Tercemar Sedang, yang menyebabkan model terkadang kesulitan dalam membedakan keduanya secara konsisten.

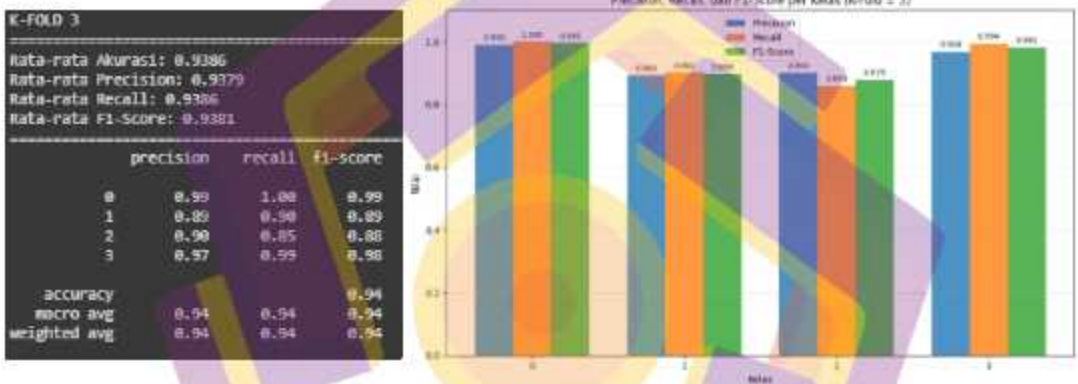
**4.3** Kelas 2 (Tercemar Sedang) memperoleh nilai F1-Score sebesar 0.87 atau 87%. Nilai ini menunjukkan bahwa meskipun precision (0.90) tergolong tinggi, nilai recall-nya (0.84) menurunkan skor F1. Artinya, model cukup efektif dalam menghindari prediksi salah ke kelas ini, namun masih kurang optimal dalam menangkap semua data aktual dari kelas ini. Kondisi ini bisa disebabkan oleh kompleksitas karakteristik air pada kelas Tercemar Sedang yang beririsan dengan kategori lain, terutama dengan kelas 1 dan kelas 3, yang membuat model perlu lebih diperkuat dalam mengenali pola data pada kelas ini.

**4.4** Kelas 3 (Tercemar Berat) memperoleh nilai F1-Score sebesar 0.98 atau 98% yang menunjukkan bahwa model mampu melakukan prediksi dengan tingkat akurasi dan kelengkapan yang sangat tinggi. F1-Score ini merupakan gabungan dari nilai precision 0.96 dan recall 0.99, yang berarti model sangat jarang melewatkan data dari kelas ini dan juga jarang mengklasifikasikan data dari kelas lain secara keliru ke dalam kategori Tercemar Berat. Dengan demikian, model sangat efisien dalam mengenali kondisi pencemaran berat, yang sangat krusial dalam konteks mitigasi dampak terhadap ekosistem dan kegiatan budidaya perikanan.

#### **4.4.2 K-Fold 3**

Evaluasi selanjutnya dilakukan menggunakan validasi silang dengan K-Fold 3, di mana data dibagi menjadi tiga bagian yang akan digunakan secara bergantian sebagai data latih dan data uji. Pendekatan ini memberikan variasi lebih banyak dalam pelatihan dan pengujian model, yang bertujuan untuk meningkatkan

reliabilitas hasil evaluasi. Performa model pada masing-masing lipatan dihitung dan dirata-rata untuk mendapatkan nilai evaluasi yang akurat. Hasil pengukuran metrik ditampilkan untuk menunjukkan seberapa baik model dalam mengklasifikasikan kualitas air pada masing-masing kelas. Untuk penjelasan tersebut, ditunjukkan pada gambar 4.10.



Gambar 4.10 Metrik Evaluasi dan Diagram K-Fold 3

1. Nilai Akurasi yang diperoleh sebesar 0.9386 atau 94%. Nilai akurasi pada K-Fold 3 ini lebih baik daripada nilai K-Fold sebelumnya. Sehingga nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
2. Nilai rata-rata precision yang dicapai model adalah 0.9379 atau setara dengan 93,79%. Nilai ini menunjukkan peningkatan nilai precision daripada nilai sebelumnya. Hasil yang didapatkan menunjukkan dari seluruh prediksi positif yang dihasilkan oleh model, hampir seluruhnya merupakan prediksi yang benar. Tingginya nilai precision ini mencerminkan bahwa model mampu meminimalkan

kesalahan dalam memberikan label kelas yang salah, khususnya untuk kategori pencemaran air yang kritis. Untuk mengetahui nilai precision pada setiap parameter, dapat dilihat dibawah ini.

**2.1** Kelas 0 (Baik) memperoleh nilai Precision 0.99 atau 99% menunjukkan bahwa dari semua prediksi yang diklasifikasikan sebagai air berkualitas baik, hampir seluruhnya benar. Hal ini mencerminkan kemampuan model yang sangat tinggi dalam menghindari kesalahan positif palsu pada kelas ini, sehingga sangat andal dalam mengidentifikasi air yang benar-benar baik tanpa mencampuradukkan dengan kategori lain.

**2.2** Kelas 1 (Tercemar Ringan) memperoleh nilai Precision 0.89 atau 89% Ini berarti bahwa sebagian besar prediksi model terhadap kategori ini memang benar, meskipun terdapat sekitar 11% data dari kelas lain yang keliru diprediksi sebagai tercemar ringan. Meskipun cukup baik, nilai ini masih menyisakan ruang untuk meningkatkan ketepatan model dalam mengenali ciri khas dari pencemaran ringan.

**2.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Precision 0.90 atau 90% yang menunjukkan bahwa dari seluruh prediksi model terhadap kategori tercemar sedang, 90% di antaranya sesuai kenyataan. Dengan demikian, tingkat kesalahan prediksi terhadap kelas ini relatif rendah, namun tetap menunjukkan bahwa ada sejumlah kecil data dari kelas lain yang masih salah dikenali sebagai kelas sedang.

**2.4** Kelas 3 (Tercemar Berat) memperoleh nilai Precision 0.97 atau 97% menandakan bahwa prediksi terhadap kategori ini sangat akurat dan hanya sedikit data dari kelas lain yang keliru diklasifikasikan sebagai tercemar berat. Hal ini sangat penting, mengingat akurasi tinggi pada kategori ini dapat membantu

mendeteksi kondisi air yang sangat membahayakan dengan tingkat kepercayaan yang tinggi.

3. nilai rata-rata pada Recall mendapatkan nilai sebesar 0.9386 atau 93,86%, yang menunjukkan meningkatnya nilai Recall yang didapatkan daripada nilai Recall sebelumnya. Kemampuan ini sangat baik dalam mengidentifikasi seluruh data yang benar-benar termasuk ke dalam masing-masing kelas. Dengan kata lain, model berhasil menangkap mayoritas sampel aktual dari setiap kelas tanpa banyak kehilangan informasi penting. Nilai ini memperlihatkan bahwa model tidak hanya akurat dalam klasifikasi, tetapi juga cukup menyeluruh dalam mengenali seluruh kondisi pencemaran air, dari yang ringan hingga berat. Nilai Recall pada setiap kelas dijelaskan dibawah ini.

**3.1 Kelas 0 (Baik)** memperoleh nilai Recall 1.00 Nilai recall pada kelas 0 (Baik) mencapai 1.00 atau 100%, yang menunjukkan bahwa model mampu mengenali seluruh data yang benar-benar termasuk dalam kategori air berkualitas baik tanpa satu pun terlewat. Hal ini mencerminkan kemampuan maksimal model dalam mendeteksi kondisi kualitas air terbaik, sehingga tidak ada kesalahan dalam mengabaikan air yang sebenarnya tidak tercemar.

**3.2 Kelas 1 (Tercemar Ringan)** memperoleh nilai Recall 0.90 menandakan bahwa model dapat mengenali sebagian besar data yang benar-benar termasuk dalam kategori tercemar ringan, namun masih terdapat sekitar 10% data yang terlewat atau salah diklasifikasikan. Meskipun masih tergolong tinggi, hal ini menunjukkan bahwa ada potensi peningkatan untuk lebih sensitif dalam mendeteksi pencemaran ringan, yang penting dalam upaya pencegahan dini terhadap degradasi kualitas air.

**3.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Recall 0.85 atau 85% nilai ini menunjukkan bahwa model baik dalam mengenali data namun memiliki sedikit kesulitan dalam mengenali seluruh data dengan kategori pencemaran sedang, di mana sekitar 15% data tidak berhasil diidentifikasi dengan benar. Rendahnya nilai recall ini jika dibandingkan kelas lain, mengindikasikan bahwa model perlu perbaikan dalam membedakan karakteristik kelas sedang yang kemungkinan tumpang tindih dengan kelas ringan maupun berat.

**3.4** Kelas 3 (Tercemar Berat) memperoleh nilai Recall 0.99 Ini menandakan bahwa model sangat mampu mendeteksi hampir seluruh data dengan kondisi pencemaran berat, yang sangat penting mengingat kategori ini memerlukan perhatian khusus dan tindakan cepat. Tingginya nilai ini mencerminkan kinerja yang sangat baik dalam mengenali kondisi kualitas air yang paling parah.

**4.** Nilai rata – rata F1-Score yang didapatkan sebesar 0.9381 atau 93,81%, yang merupakan kombinasi seimbang antara precision dan recall. Nilai yang didapatkan ini membuktikan bahwa model memiliki performansi yang stabil dalam ketepatan dan kelengkapan prediksi, serta mampu menjaga keseimbangan antara keduanya. Skor F1 yang tinggi ini sangat penting dalam kasus klasifikasi kualitas air karena dapat menghindarkan konsekuensi salah klasifikasi, baik dalam hal intervensi yang tidak diperlukan maupun pengabaian terhadap pencemaran yang sebenarnya serius

**4.1** Kelas 0 (Baik) memperoleh nilai F1-Score 0.99 atau 99% yang menunjukkan keseimbangan yang sangat baik antara precision dan recall dalam mengenali kondisi air berkualitas baik. Dengan nilai ini, model terbukti sangat optimal dalam

mengklasifikasikan air yang benar-benar berada dalam kategori baik secara konsisten.

**4.2** Kelas 1 (Tercemar Ringan) memperoleh nilai F1-Score 0.89 atau 89% yang menunjukkan bahwa meskipun model cukup andal dalam mengenali pencemaran ringan, masih terdapat perbaikan yang dapat dilakukan baik dari sisi ketepatan (precision) maupun kelengkapan deteksi (recall) untuk mencapai klasifikasi yang lebih stabil pada kategori ini.

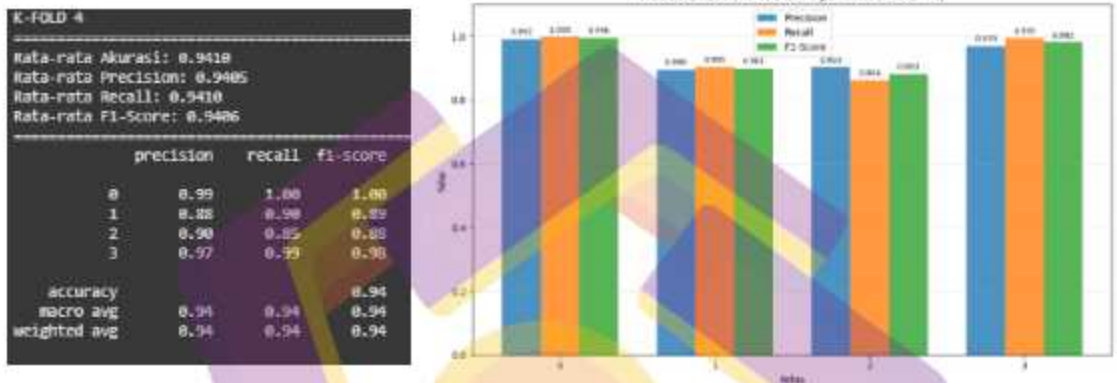
**4.3** Kelas 2 (Tercemar Sedang) memperoleh nilai F1-Score 0.88 atau 88% yang menunjukkan bahwa model cukup efektif dalam mengidentifikasi air yang tercemar sedang, meskipun masih ada potensi salah klasifikasi yang menyebabkan penurunan performa dibanding kelas lainnya. Kombinasi precision dan recall yang sedikit lebih rendah dari kategori lainnya menjadikan kelas ini sebagai salah satu titik fokus evaluasi.

**4.4** Kelas 3 (Tercemar Berat) memperoleh nilai F1-Score 0.98 atau 98% yang menunjukkan bahwa performa yang sangat tinggi dan konsisten dari model dalam mendeteksi kondisi air yang tercemar berat. Tingginya nilai ini menunjukkan bahwa model mampu mendeteksi pencemaran berat dengan tepat.

#### **4.4.3 K-Fold 4**

Pada bagian ini, pengujian model dilakukan menggunakan K-Fold dengan nilai K-4. Dengan membagi dataset menjadi empat bagian, model akan dilatih dan diuji sebanyak empat kali, masing-masing menggunakan fold yang berbeda sebagai data uji. Pendekatan ini bertujuan untuk memastikan bahwa model tidak hanya bekerja baik pada satu subset data, melainkan memiliki performa yang stabil di

berbagai pembagian data. Evaluasi dilakukan berdasarkan metrik akurasi, precision, recall, dan F1-Score pada setiap kelas kualitas air. Penjelasan lebih lanjut dapat dilihat pada gambar 4.11



Gambar 4.11 Metrik Evaluasi dan Diagram K-Fold 4

1. Nilai Akurasi 0.9410 atau 94% menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 4 ini dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.
2. Nilai rata – rata Precision yang didapatkan sebesar 0,9405 atau 94%. Nilai rata-rata yang didapatkan ini menunjukkan model memiliki kemampuan yang sangat baik dalam mengurangi kesalahan positif palsu (false positive). Dengan kata lain, dari seluruh prediksi positif yang dibuat model, sekitar 94% benar-benar sesuai dengan

kondisi kelas aslinya. Nilai yang didapatkan dari setiap kelas, akan dijelaskan dibawah ini.

**2.1** Kelas 0 (Baik) memperoleh nilai Precision 0.99 atau 99% menunjukkan bahwa hampir seluruh prediksi model terhadap kelas “baik” sangat tepat. Artinya, sebagian besar data yang diklasifikasikan sebagai “baik” memang benar-benar berasal dari kelas tersebut. Hal ini mencerminkan minimnya kesalahan prediksi positif palsu (false positive) pada kelas ini.

**2.2** Kelas 1 (Tercemar Ringan) memperoleh nilai Precision 0.88 atau 88% yang menunjukkan bahwa 88% dari data yang diprediksi sebagai “tercemar ringan” memang benar-benar berasal dari kelas tersebut. Nilai ini masih tergolong baik, meskipun menunjukkan adanya sejumlah kesalahan prediksi positif terhadap kelas lain, khususnya pada kondisi dengan distribusi kelas yang berdekatan.

**2.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Precision 0.90 atau 90% model memiliki kemampuan yang sangat baik dalam mengenali dan memprediksi data yang termasuk dalam kategori “tercemar sedang”. Dengan nilai mendekati sempurna, model hanya melakukan sedikit kesalahan dalam mengklasifikasikan data dari kelas lain ke dalam kelas ini.

**2.4** Kelas 3 (Tercemar Berat) memperoleh nilai Precision 0.97 atau 97% , yang menandakan tingkat akurasi prediksi terhadap kelas “tercemar berat” sangat tinggi. Hampir seluruh prediksi yang menyatakan kondisi sebagai “tercemar berat” memang benar adanya, mencerminkan keandalan model dalam mengenali kondisi pencemaran berat secara spesifik.

3. Nilai rata – rata Recall 0.9406 atau 94% yang menunjukkan bahwa model sangat efektif dalam mengenali dan menangkap seluruh kasus yang seharusnya masuk ke masing-masing kelas, dengan tingkat kesalahan negatif palsu (false negative) yang sangat rendah. Nilai yang didapatkan dari masing-masing kelas, akan dijelaskan dibawah ini.

**3.1** Kelas 0 (Baik) memperoleh nilai Recall 1.00 atau 100% yang berarti semua data aktual dari kelas “baik” berhasil dikenali secara sempurna oleh model. Tidak ada satupun data yang tergolong sebagai kelas “baik” yang terlewatkan atau diklasifikasikan ke dalam kelas lain. Hal ini menunjukkan kemampuan deteksi yang sangat tinggi dan ideal terhadap kondisi air yang benar-benar baik

**3.2** Kelas 1 (Tercemar Ringan) memperoleh nilai Recall 0.90 atau 90% yang berarti semua data aktual dari kelas “baik” berhasil dikenali secara sempurna oleh model. Tidak ada satupun data yang tergolong sebagai kelas “baik” yang terlewatkan atau diklasifikasikan ke dalam kelas lain. Hal ini menunjukkan kemampuan deteksi yang sangat tinggi dan ideal terhadap kondisi air yang benar-benar baik

**3.3** Kelas 2 (Tercemar Sedang) memperoleh nilai Recall 0.85 atau 85% model hanya berhasil mengenali 85% dari total data aktual yang termasuk dalam kelas “tercemar sedang”. Ini menunjukkan adanya kelemahan relatif dalam mengidentifikasi kategori ini, yang bisa jadi disebabkan oleh overlap karakteristik parameter air dengan kelas ringan maupun berat.

**3.4** Kelas 3 (Tercemar Berat) memperoleh nilai Recall 0.99 atau 99% yang berarti hampir seluruh data aktual pada kelas “tercemar berat” berhasil diidentifikasi dengan benar oleh model. Hanya sebagian kecil data yang salah diklasifikasikan

sebagai kelas lain, mencerminkan kemampuan tinggi model dalam mengenali kondisi air yang benar-benar tercemar berat.

4. Nilai rata – rata F1-Score 0.9406 atau 94% yang menunjukkan keseimbangan optimal antara precision dan recall. Nilai ini memperkuat bahwa model tidak hanya tepat dalam prediksinya, tetapi juga menyeluruh dalam menangkap semua kasus yang relevan. F1-Score yang tinggi menandakan stabilitas performa model dalam berbagai kondisi pencemaran air.

4.1 Kelas 0 (Baik) memperoleh nilai F1-Score 1.00 atau 100% yang menunjukkan bahwa model memiliki kombinasi yang sangat baik antara presisi dan recall dalam mengidentifikasi kondisi air yang benar-benar baik. Tidak hanya mampu menghindari false positives, tetapi juga secara konsisten mengenali semua instance dari kelas ini. F1-Score ini mencerminkan stabilitas dan keandalan model dalam mendeteksi kualitas air yang ideal.

4.2 Kelas 1 (Tercemar Ringan) memperoleh nilai F1-Score 0.89 atau 89% yang menunjukkan performa model dalam mengenali kelas “tercemar ringan” tergolong cukup baik, meskipun tidak setinggi kelas 0 atau 3. Nilai ini menunjukkan bahwa masih terdapat ketidakseimbangan kecil antara jumlah prediksi benar (true positives) dan kesalahan klasifikasi, yang kemungkinan besar disebabkan oleh karakteristik data yang tumpang tindih dengan kelas “tercemar sedang”.

4.3 Kelas 2 (Tercemar Sedang) memperoleh nilai F1-Score 0.88 atau 88% yang menunjukkan meskipun performa model cukup stabil, kemampuan untuk secara konsisten mengklasifikasikan kelas ini belum optimal. Hal ini bisa jadi karena adanya overlap fitur dengan kelas tetangga (1 dan 3), atau karena distribusi data pada

kelas ini tidak seimbang. Namun demikian, nilai F1- Score ini tetap menunjukkan bahwa model cukup layak digunakan untuk mendeteksi kondisi ini, walaupun perlu peningkatan.

**4.4** Kelas 3 (Tercemar Berat) memperoleh nilai F1-Score 0.98 atau 98% yang menunjukkan model memiliki keseimbangan sangat tinggi antara presisi dan recall dalam mengenali kelas “tercemar berat”. Artinya, sebagian besar prediksi untuk kelas ini benar dan sangat sedikit kesalahan, menjadikan model ini sangat efektif dalam mendeteksi kondisi air yang tercemar berat.

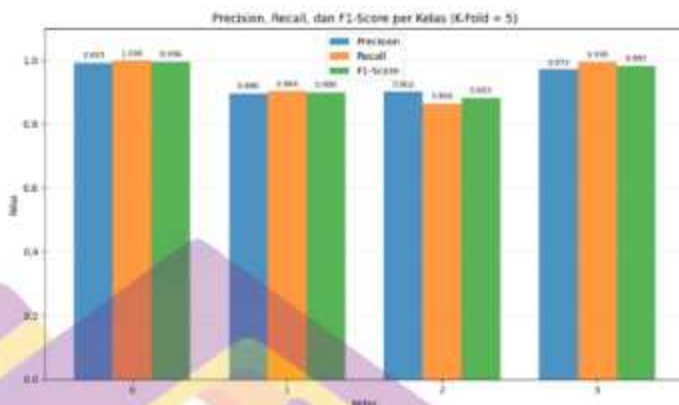
#### **4.4.4 K-Fold 5**

Pengujian akhir dilakukan menggunakan metode K-Fold Cross Validation dengan  $K = 5$ , yang merupakan metode yang membagi data menjadi lima subset, di mana masing-masing subset digunakan sekali sebagai data uji, dan empat lainnya sebagai data latih. Tujuannya adalah untuk menghasilkan evaluasi model yang lebih menyeluruh dan mengurangi pengaruh variasi data. Hasil pengujian ditampilkan dalam bentuk metrik evaluasi pada tiap kelas dan nilai rata-rata keseluruhan model. Penggunaan K-Fold 5 ini sering digunakan oleh beberapa penelitian terdahulu dalam mendapatkan nilai akurasi terbaik. Penjelasan lebih lanjut dapat dilihat pada gambar 4.12

**K-FOLD 5**

Rata-rata Akurasi: 0.9411  
 Rata-rata Precision: 0.9406  
 Rata-rata Recall: 0.9411  
 Rata-rata F1-Score: 0.9407

	precision	recall	f1-score
0	0.99	1.00	1.00
1	0.89	0.98	0.89
2	0.98	0.86	0.88
3	0.97	1.00	0.98
accuracy			0.94
macro avg	0.94	0.94	0.94
weighted avg	0.94	0.94	0.94



Gambar 4.12 Metrik Evaluasi dan Diagram K-Fold 5

1. Akurasi 0.9411 atau 94% menggunakan K-fold 5 ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Akurasi model pada K-Fold 5 mengalami kenaikan tipis sebesar 0.0001 dibandingkan dengan K-Fold 4. Meskipun kenaikan yang dilakukan hanyalah sedikit, model menunjukkan akurasi yang tetap stabil dan konsisten dalam mengklasifikasikan kualitas air
2. Nilai rata-rata Precision sebesar 0.9406 atau 94% Rata-rata precision meningkat sedikit pada K-Fold 5. Artinya, model semakin baik dalam mengurangi false positive dan memberikan prediksi yang lebih tepat terhadap masing-masing kelas.
  - 2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% yang menunjukkan bahwa hampir semua prediksi yang dilakukan model terhadap kelas 0 adalah benar. Artinya, ketika model memprediksi suatu data sebagai "Baik", prediksi tersebut hampir selalu tepat. Ini mencerminkan bahwa model memiliki

kemampuan yang sangat tinggi dalam membedakan data kelas 0 dari kelas lainnya, dengan kesalahan (false positive) yang sangat kecil.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% Nilai ini menunjukkan bahwa model cukup baik dalam mengenali pencemaran ringan, meskipun nilai precision selisih 0.01 daripada kelas 2, kemungkinan kesalahan dalam melakukan prediksi klasifikasi kualitas air masih sedikit. Namun perlu diwaspadai kemungkinan kesalahan dalam membedakannya dengan kelas yang mirip.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% model memiliki kemampuan cukup tinggi dalam memprediksi data yang benar-benar berasal dari kelas 2. Ini berarti bahwa sebagian besar prediksi untuk kelas ini adalah tepat, meskipun ada sedikit kesalahan dalam memasukkan data dari kelas lain sebagai kelas 2. Hal ini menunjukkan bahwa model mampu mengidentifikasi pada label pencemaran sedang.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa hampir seluruh prediksi terhadap kelas 3 adalah benar. Model mampu mengidentifikasi kondisi air yang sangat tercemar dengan sangat baik, dan hanya sedikit data dari kelas lain yang salah diklasifikasikan sebagai kelas 3. Ini menunjukkan bahwa model sangat presisi dalam mengenali kategori ini, yang penting dalam konteks mitigasi pencemaran berat.

**3.** Nilai rata-rata Recall sebesar 0.9411 atau 94% Performa recall juga meningkat sangat tipis. Hal ini menandakan bahwa model menjadi sedikit lebih sensitif dalam

menangkap instance dari setiap kelas secara benar, meskipun peningkatannya sangat kecil.

**3.1 Kelas 0 (Baik)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data yang benar-benar berasal dari kelas 0 berhasil dikenali oleh model tanpa ada yang terlewat. Artinya, tidak ada data dari kelas 0 yang salah diklasifikasikan sebagai kelas lain. Ini mencerminkan performa model yang sempurna dalam mendeteksi kondisi air dengan kualitas baik.

**3.2 Kelas 1 (Tercemar Ringan)** mendapatkan nilai precision sebesar 0.90 atau 90% menandakan bahwa model mampu mengenali 90% dari seluruh data yang benar-benar termasuk dalam kelas 1. Namun, terdapat sekitar 10% data dari kelas 1 yang tidak berhasil dikenali dan justru diklasifikasikan sebagai kelas lain. Ini menunjukkan bahwa model cukup baik dalam mengidentifikasi pencemaran ringan, meskipun masih ada sejumlah kasus yang tidak terdeteksi secara tepat.

**3.3 Kelas 2 (Tercemar Sedang)** mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan dari data yang sebenarnya termasuk dalam kelas 2 berhasil dikenali oleh model, sementara 14% lainnya diklasifikasikan secara keliru ke kelas lain. Hal ini mengindikasikan bahwa model masih menghadapi tantangan dalam membedakan pencemaran sedang dari kategori lain, yang kemungkinan besar karena kemiripan karakteristik data dengan kelas 1 atau 3.

**3.4 Kelas 3 (Tercemar Berat)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data yang benar-benar termasuk dalam kategori tercemar berat berhasil diklasifikasikan dengan tepat oleh model. Tidak ada satu pun data dari kelas 3 yang terlewat. Hal ini sangat penting karena kemampuan

mendeteksi pencemaran berat secara menyeluruh menjadi penting dalam pengambilan keputusan dalam sistem monitoring kualitas air.

4. Nilai rata-rata F1-Score sebesar 0.9407 atau 94% Nilai F1-Score menunjukkan peningkatan serupa dengan precision dan recall, mencerminkan keseimbangan yang sedikit lebih baik antara keduanya dibandingkan dengan fold sebelumnya.

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa model memiliki kesempurnaan baik dalam precision maupun recall untuk kelas ini. Artinya, tidak hanya model dapat mengidentifikasi seluruh data kelas 0 dengan tepat (recall sempurna), tetapi semua prediksi model untuk kelas ini juga benar (precision sempurna). Hal ini menunjukkan bahwa model sangat andal dalam mendeteksi kondisi air yang benar-benar berkualitas baik

4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan keseimbangan yang cukup baik antara precision dan recall untuk kelas 1. Ini berarti model mampu melakukan prediksi yang akurat dan juga tidak melewatkan banyak data pada kategori ini. Namun, ada ruang untuk perbaikan agar deteksi pencemaran ringan dapat mendekati performa optimal

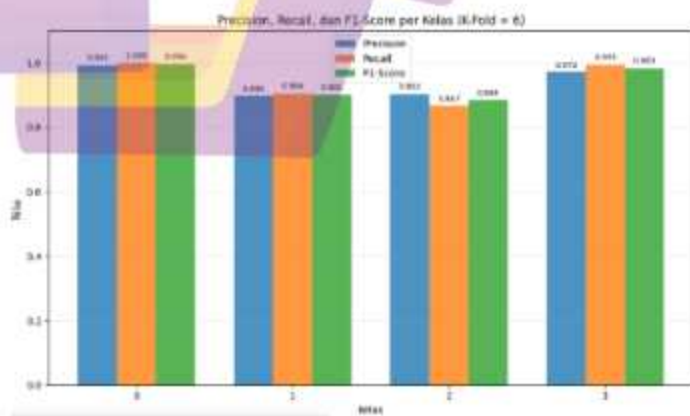
4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% bahwa performa model dalam mendeteksi pencemaran sedang berada pada tingkat yang cukup stabil, meskipun sedikit lebih rendah dibanding kelas lainnya. Ini disebabkan karena nilai precision yang cukup tinggi namun nilai recall pada kelas 2 sedikit lebih rendah. Hal ini mengindikasikan bahwa model lebih sering meleset dalam mengenali seluruh data dari kelas ini, dan kemungkinan juga terjadi kesalahan klasifikasi dengan kelas terdekat seperti kelas 1 atau 3.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan bahwa model nilai pada precision dan nilai recall hampir sempurna dalam mengklasifikasikan kelas 3 (Tercemar Berat). Ini artinya hampir semua data yang diprediksi sebagai tercemar berat memang benar berasal dari kelas tersebut. F1-Score yang tinggi pada kelas ini penting, mengingat kategori ini merupakan tingkat pencemaran yang paling berisiko dan memerlukan tindakan segera.

#### 4.4.5 K-Fold 6

Evaluasi model selanjutnya dilakukan dengan menggunakan K-Fold Cross Validation dengan nilai  $K = 6$ . Pada tahap ini, dataset dibagi menjadi enam subset yang digunakan secara bergantian sebagai data latih dan data uji. Teknik ini memberikan kombinasi pelatihan dan pengujian yang lebih bervariasi, yang bertujuan untuk meningkatkan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Hasil pengujian disajikan dalam bentuk metrik evaluasi pada masing-masing kelas serta nilai rata-rata keseluruhan model. Penjelasan lebih lanjut dapat dilihat pada gambar 4.13.

K-FOLD 6				
rata-rata Akurasi: 0.9415				
rata-rata Precision: 0.9410				
rata-rata Recall: 0.9435				
rata-rata F1-Score: 0.9411				
	precision	recall	f1-score	support
0	0.99	1.00	1.00	2343
1	0.88	0.89	0.89	3388
2	0.98	0.86	0.92	2584
3	0.97	0.99	0.98	3333
accuracy			0.94	12548
macro avg	0.94	0.94	0.94	12548
weighted avg	0.93	0.94	0.94	12548
waktu training : 40.4138951423645 detik				
waktu testing : 0.81 menit				



Gambar 4.13 Metrik Evaluasi dan Diagram K-Fold 6

1. Nilai akurasi yang didapatkan dari K-Fold 6 sebesar 0,9415 atau 94%. Hasil pengujian pada K-Fold ini menunjukkan model pada K-Fold 6 ini dapat melampaui nilai akurasi yang didapatkan dari beberapa nilai K-Fold sebelumnya. Sehingga model berhasil mengklasifikasikan kualitas air dengan benar.

2. Nilai rata-rata Precision sebesar 0.9410 atau 94% menunjukkan bahwa model baik dalam mengurangi false positive dan memberikan prediksi yang lebih tepat terhadap masing-masing kelas. Meskipun nilai precision ini mengalami penurunan sedikit daripada nilai pada K-Fold 5.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% yang menunjukkan nilai bahwa model hampir tidak pernah salah dalam memprediksi kelas 0. Artinya, jika model memutuskan suatu data sebagai "Baik", kemungkinan besar prediksinya benar. Hal ini menunjukkan bahwa model sangat handal dalam mengenali data yang benar-benar memiliki kualitas air baik.

2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% artinya model cukup akurat dalam memprediksi data sebagai "Tercemar Ringan", meskipun masih ada sebagian prediksi yang seharusnya bukan kelas ini. Bisa jadi ada sedikit kebingungan model dengan kelas tetangga seperti kelas 2 (Tercemar Sedang), namun performa ini masih cukup dapat diterima.

2.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan bahwa model memiliki tingkat keyakinan yang cukup tinggi terhadap prediksi untuk kelas 2. Ini artinya sebagian besar data yang diprediksi sebagai kelas ini memang benar kelas 2, meskipun tetap ada sedikit noise dari data kelas lain yang keliru diklasifikasikan

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa ketika model memutuskan sebuah data tergolong dalam pencemaran berat, keputusan tersebut hampir selalu tepat. Ini sangat penting karena kesalahan prediksi pada kelas ini bisa berdampak besar, mengingat kelas ini mencerminkan kondisi air yang paling berbahaya.

**3.** Nilai rata-rata Recall sebesar 0.9415 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mendeteksi setiap kelas berhasil dikenali dengan benar oleh model. Untuk penjelasan terhadap masing-masing kelas, akan diuraikan sebagai berikut

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% Nilai recall sempurna ini menandakan bahwa seluruh data dengan label "Baik" berhasil diprediksi dengan benar oleh model. Tidak ada satu pun data kelas 0 yang terlewat atau salah diklasifikasikan. Ini menunjukkan bahwa model sangat sensitif dalam mengenali ciri-ciri kualitas air yang baik.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan masih terdapat sekitar 11% data pada label kelas 1 (Tercemar Ringan) yang tidak dikenali dengan benar dan kemungkinan diklasifikasikan ke kelas lain. Meskipun begitu, nilai ini tetap menunjukkan performa yang cukup baik

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan Sekitar 14% data dengan label 2 (Tercemar Sedang) gagal dikenali oleh model. Hal ini bisa menunjukkan bahwa fitur dari kelas ini masih sering tumpang tindih dengan kelas lain, terutama kelas 1 (Ringan) atau 3 (Berat), sehingga model agak kesulitan dalam membedakannya secara jelas.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan data yang benar-benar tergolong sebagai pencemaran berat. Ini menunjukkan sensitivitas model yang sangat tinggi dan hampir sempurna dalam mendeteksi kondisi air yang paling buruk, dan ini tentu sangat penting dari sisi dampak dan mitigasi

**4.** Nilai rata-rata F1-Score sebesar 0.9411 atau 94% menunjukkan data yang benar-benar tergolong sebagai pencemaran berat. Ini menunjukkan sensitivitas model yang sangat tinggi dan hampir sempurna dalam mendeteksi kondisi air yang paling buruk, dan ini tentu sangat penting dari sisi dampak dan mitigasi

**4.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa precision dan recall pada kelas ini seimbang dan maksimal. Model tidak hanya tepat dalam mengklasifikasikan kelas "Baik", tetapi juga tidak melewatkan satu pun sampel dari kelas ini. Hasil ini menandakan bahwa fitur dari kelas 0 sangat mudah dikenali oleh model.

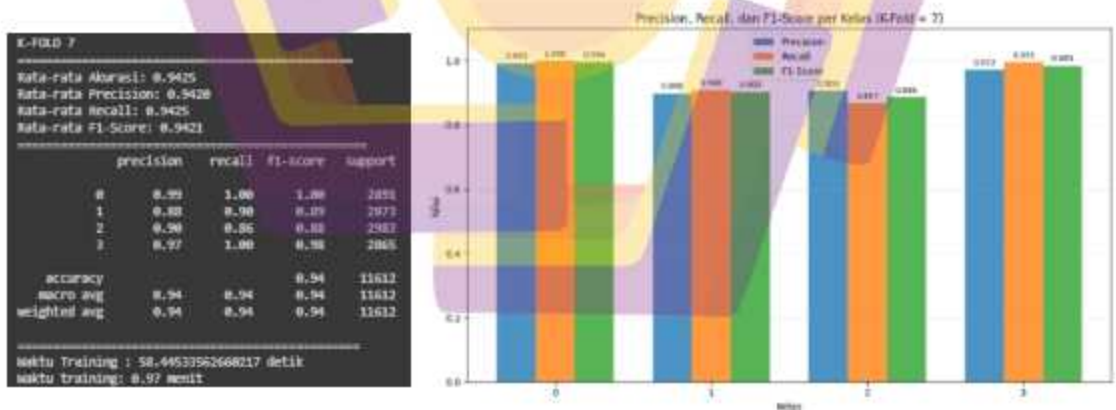
**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% yang menunjukkan bahwa ada sedikit ketidakseimbangan antara precision dan recall pada kelas ini. Walaupun masih tergolong sangat baik, model masih memiliki beberapa kesalahan dalam membedakan kelas ini dengan kelas lain.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% yang menunjukkan bahwa meskipun model cukup baik dalam mengklasifikasikan kelas ini, terdapat sedikit ketidakseimbangan dengan nilai recall yang menyebabkan F1-Score menjadi lebih rendah dibandingkan kelas lainnya

**4.4 Kelas 3 (Tercemar Berat)** mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan bahwa model akurat dalam mendeteksi kualitas air dengan pencemaran berat

#### 4.4.6 K-Fold 7

Pada tahap ini, validasi model dilakukan menggunakan pendekatan K-Fold dengan K = 7. Teknik ini semakin meningkatkan variasi pelatihan dan pengujian karena lebih banyak lipatan yang digunakan. Setiap fold digunakan satu kali sebagai data uji, sedangkan sisanya sebagai data latih. Evaluasi dilakukan untuk mengukur performa model dalam hal akurasi, precision, recall, dan F1-Score, sehingga diperoleh gambaran yang lebih komprehensif terhadap kemampuan model dalam mengklasifikasikan kualitas air. Penjelasan lebih lanjut dapat dilihat pada gambar 4.14.



Gambar 4.14 Metrik Evaluasi dan Diagram K-Fold 7

1. Nilai akurasi yang didapatkan dari K-Fold 7 sebesar 0,9425 atau 94%. Hasil pengujian pada K-Fold ini menunjukkan model pada K-Fold 7 ini dapat melampaui

nilai akurasi yang didapatkan dari beberapa nilai K-Fold sebelumnya. Sehingga model berhasil mengklasifikasikan kualitas air dengan benar

2. Nilai rata-rata Precision sebesar 0.9420 yang menunjukkan bahwa dari seluruh prediksi positif yang dilakukan oleh model, sebanyak 94,20% di antaranya merupakan prediksi yang benar. Nilai ini menandakan bahwa model memiliki kemampuan yang sangat baik dalam menghindari false positive pada seluruh kelas kualitas air. Untuk penjelasan dari masing-masing kelas dengan precision, akan dijelaskan sebagai berikut.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% Model sangat jarang salah dalam mengklasifikasikan data sebagai kelas "Baik". Hampir seluruh prediksi yang diklasifikasikan sebagai kelas 0 adalah benar.

2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% bahwa dari seluruh data yang diprediksi sebagai kelas "Tercemar Ringan", sebesar 88% di antaranya benar-benar berasal dari kelas tersebut. Hal ini menunjukkan bahwa model cukup akurat dalam mengenali data kelas ini, meskipun masih terdapat beberapa kesalahan klasifikasi.

2.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan bahwa model mampu mengenali kelas ini dengan cukup baik, walaupun masih ada sedikit prediksi yang tidak tepat, di mana beberapa data dari kelas lain salah diklasifikasikan ke dalam kelas ini.

2.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa sebagian besar data yang diprediksi sebagai "Tercemar Berat"

memang benar berasal dari kelas tersebut. Ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data pada kategori ini.

3. Nilai rata-rata Recall sebesar 0.9425 atau 94% menunjukkan bahwa secara keseluruhan model mampu menangkap mayoritas data yang benar-benar termasuk ke dalam masing-masing kelas. Nilai tersebut berhasil dikenali dengan benar oleh model karena semakin sedikit data positif yang terlewatkan (false negative).

3.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% Recall sebesar 1.0 atau 100% menunjukkan bahwa model berhasil mengenali seluruh data aktual yang termasuk kategori Baik tanpa ada satu pun yang terlewat. Ini adalah performa sempurna untuk kelas ini

3.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan bahwa model mampu menangkap sebagian besar data aktual kelas Tercemar Ringan, tetapi masih ada sekitar 10% data dari kelas ini yang tidak berhasil dikenali (false negative).

3.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan model cukup baik dalam mengenali kelas Tercemar Sedang. Namun, nilai ini merupakan yang terendah di antara semua kelas, yang mengindikasikan bahwa model sedikit kesulitan membedakan kategori ini dari kelas lain, kemungkinan karena kemiripan karakteristik.

3.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% memperlihatkan bahwa model mampu mengenali semua data aktual kelas Tercemar Berat dengan sempurna, tanpa ada yang terlewat.

4. Nilai rata-rata F1-Score sebesar 0.9421 atau 94% menunjukkan bahwa secara keseluruhan model memiliki keseimbangan yang sangat baik antara precision (ketepatan prediksi positif) dan recall (kemampuan menemukan semua data positif) di semua kelas.

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% Model menunjukkan kinerja sempurna pada kelas ini. Precision dan recall sama- sama tinggi sehingga menghasilkan F1-Score maksimal. Artinya semua prediksi “Baik” benar dan semua data aktual “Baik” berhasil ditemukan.

4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% Meskipun cukup tinggi, ini adalah nilai terendah kedua di antara semua kelas. Artinya ada sedikit ketidakseimbangan antara precision dan recall di kelas ini, kemungkinan disebabkan beberapa data kelas ini diprediksi sebagai kelas lain.

4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% nilai ini merupakan nilai terendah dari semua kelas. Hal ini menunjukkan bahwa model masih agak sulit membedakan kelas 2 (Tercemar Sedang) dengan kelas lain, yang selaras dengan nilai recall terendah yang disebutkan sebelumnya (0.86). Kemungkinan besar kelas ini memiliki karakteristik fitur yang mirip dengan kelas Tercemar Ringan atau Tercemar Berat.

4.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% Performa sangat baik, hanya sedikit di bawah sempurna. Kesalahan prediksi di kelas ini sangat kecil, dan model hampir selalu benar mengenali kondisi Tercemar Berat.

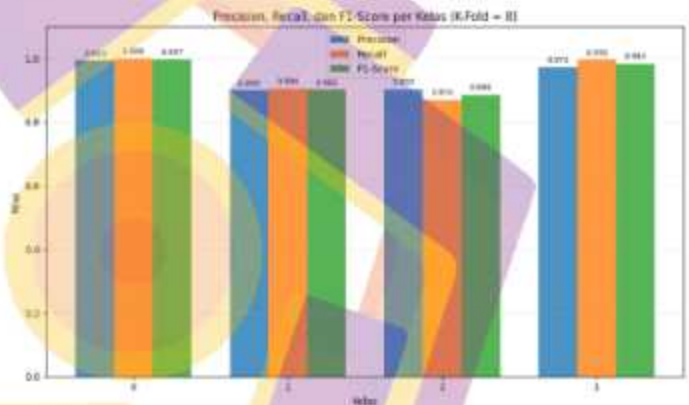
#### 4.4.7 K-Fold 8

Pengujian model dilanjutkan dengan menggunakan K-Fold Cross Validation bernilai  $K = 8$ , yang membagi data menjadi delapan bagian. Metode ini bertujuan untuk menghasilkan evaluasi yang lebih stabil dan representatif, karena setiap bagian data digunakan sebagai data uji secara bergantian. Hasil dari evaluasi ini diharapkan dapat menunjukkan konsistensi performa model, terutama dalam mengenali setiap kelas kualitas air secara adil dan akurat. Penjelasan lebih lanjut dapat dilihat pada gambar 4.15

```

E-FOLD 8
-----
rata-rata Akurasi: 0.9422
Rata-rata Precision: 0.9417
Rata-rata Recall: 0.9422
rata-rata F1-Score: 0.9418
-----
      precision  recall  f1-score  support
0      0.99      1.00      0.99      2542
1      0.89      0.89      0.89      2586
2      0.98      0.96      0.97      2682
3      0.97      1.00      0.98      2511

accuracy 0.94
macro avg 0.94  0.94  0.94  10182
weighted avg 0.93  0.94  0.94  10182
-----
waktu Training : 68.624511231761 detik
waktu training: 1,54 menit
  
```



Gambar 4.15 Metrik Evaluasi dan Diagram K-Fold 8

1. Akurasi 0.9422 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 8, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar

2. Nilai rata-rata Precision sebesar 0.9417 atau 94% menunjukkan bahwa model memiliki ketepatan tinggi dalam melakukan prediksi terhadap setiap kelas. Artinya, sebagian besar prediksi yang diklasifikasikan sebagai positif oleh model memang benar adanya, sehingga kesalahan dalam memberikan label yang tidak sesuai (false positive) relatif kecil. Hal ini menandakan bahwa model cukup andal dalam menghindari kesalahan klasifikasi terhadap kelas yang salah.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir semua prediksi yang diklasifikasikan sebagai "Baik" oleh model memang benar-benar termasuk dalam kelas tersebut. Ini berarti tingkat kesalahan prediksi positif (false positive) sangat kecil, dan model sangat akurat dalam mengenali kondisi air yang tergolong baik.

2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa dari seluruh data yang diprediksi sebagai "Tercemar Ringan", sekitar 89% di antaranya benar-benar tergolong ke dalam kelas tersebut. Meskipun tergolong cukup tinggi, masih terdapat sekitar 11% kesalahan prediksi, yang kemungkinan disebabkan oleh kesamaan karakteristik dengan kelas lain seperti "Tercemar Sedang" atau "Baik".

2.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan bahwa sebagian besar prediksi model terhadap data sebagai "Tercemar Sedang" memang tepat.

2.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model sangat jarang melakukan kesalahan saat memprediksi kondisi air sebagai "Tercemar Berat". Hampir seluruh prediksi yang masuk ke kelas

ini memang benar-benar merupakan data dari kelas tersebut, sehingga performa model dalam mengidentifikasi kondisi terburuk (tercemar berat) tergolong sangat baik dan andal.

**3.** Nilai rata-rata Recall sebesar 0.9422 atau 94% menunjukkan bahwa model mampu mengenali sebagian besar data dari setiap kelas dengan baik. Nilai rata-rata Recall yang tinggi ini memungkinkan model melewatkan data yang seharusnya terdeteksi (*false negative*) sangat rendah.

**3.1** Kelas 0 (Baik) mendapatkan nilai *precision* sebesar 1.0 atau 100% menunjukkan berarti seluruh prediksi model untuk kelas "Baik" benar-benar akurat. Tidak ada data dari kelas lain yang salah diklasifikasikan sebagai "Baik". Ini menunjukkan bahwa model sangat mampu mengenali karakteristik air yang tergolong baik tanpa melakukan kesalahan prediksi.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai *precision* sebesar 0.89 atau 89% menunjukkan bahwa dari seluruh data yang diprediksi sebagai kelas 1 (Tercemar Ringan), sebanyak 11% merupakan prediksi yang keliru (*false positive*). Artinya, beberapa data dari kelas lain, kemungkinan salah dalam mengklasifikasi sebagai kelas 1 (Tercemar Ringan). Ini bisa disebabkan oleh karakteristik data yang tumpang tindih atau ambiguitas antara kelas ringan dan kelas sekitarnya.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai *precision* sebesar 0.86 atau 86% model masih memiliki tingkat kesalahan prediksi yang cukup signifikan, yaitu sekitar 14% dari prediksi untuk kelas ini adalah keliru. Kemungkinan besar, model mengalami kesulitan dalam membedakan data antara kelas "Tercemar Ringan" dan

“Tercemar Berat”, yang bisa jadi memiliki parameter kualitas air yang saling berdekatan

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh prediksi model terhadap kelas 3 (Tercemar Berat) selalu tepat. Yang berarti tidak ada satu pun prediksi kelas lain yang salah dimasukkan ke dalam kategori ini.

**4.** Nilai rata-rata F1-Score sebesar 0.9418 atau 94% menunjukkan bahwa secara keseluruhan model memiliki keseimbangan yang sangat baik antara precision dan recall di semua kelas. Nilai ini tergolong tinggi dan mencerminkan performa model yang kuat dan stabil dalam mengklasifikasikan berbagai tingkat kualitas air.

**4.1** Model sangat akurat dalam mengenali kondisi air yang “Baik”. Hampir semua data aktual “Baik” berhasil terdeteksi dengan benar (recall = 1.00), dan kesalahan prediksi ke kelas lain sangat sedikit (precision = 0.99).

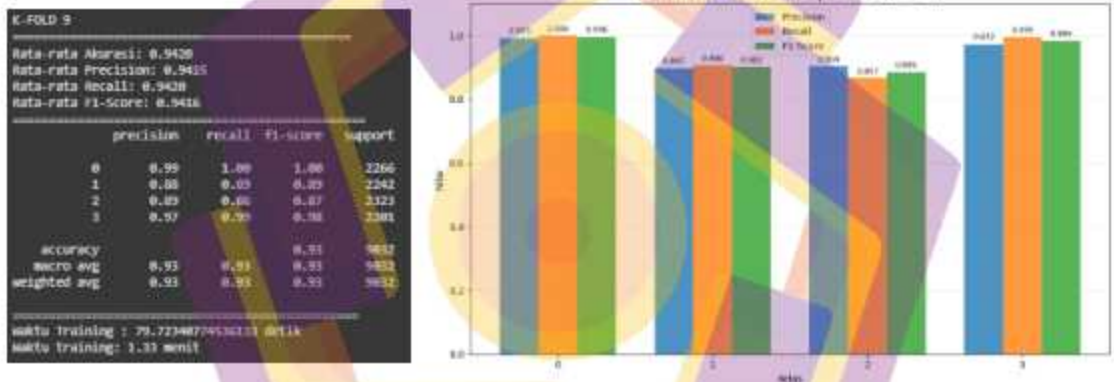
**4.2** Model cukup seimbang antara kemampuan mengenali dan ketepatan dalam memprediksi kelas ini. Namun, masih ada beberapa data “Tercemar Ringan” yang salah diklasifikasikan ke kelas lain.

**4.3** Model sedikit kesulitan membedakan kelas ini. Nilai recall 0.86 menunjukkan bahwa sebagian data “Tercemar Sedang” tidak terdeteksi dengan benar (mungkin terklasifikasi sebagai “Ringan” atau “Berat”).

**4.4** Model sangat baik dalam mengenali air “Tercemar Berat”. Semua data aktual kelas ini berhasil dikenali (recall = 1.00), dan sangat sedikit prediksi salah ke kelas lain.

#### **4.4.8 K-Fold 9**

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 9$ . Semakin banyak jumlah lipatan yang digunakan, maka semakin kecil proporsi data uji dalam setiap iterasi, sehingga model mendapatkan lebih banyak data untuk dilatih. Hal ini memungkinkan model untuk menangkap pola yang lebih kompleks. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.16.



Gambar 4.16 Metrik Evaluasi dan Diagram K-Fold 9

1. Akurasi 0.9420 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 9, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.
2. Nilai rata-rata Precision sebesar 0.9415 atau 94% menunjukkan bahwa model memiliki tingkat ketepatan yang tinggi dalam melakukan klasifikasi kualitas air pada

semua kelas. Artinya, sebagian besar data yang diprediksi oleh model sebagai anggota suatu kelas, benar-benar berasal dari kelas tersebut. Precision yang tinggi ini mencerminkan kemampuan model dalam meminimalkan kesalahan klasifikasi positif palsu (false positive), sehingga hasil prediksi menjadi lebih dapat dipercaya.

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir seluruh data yang diprediksi sebagai "Baik" memang benar berasal dari kelas tersebut. Ini mencerminkan kemampuan model yang sangat akurat dalam mengenali air dengan kualitas baik dan minim kesalahan dalam klasifikasi. Kesalahan klasifikasi pada kelas ini nyaris tidak ada.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa sebagian kecil data dari kelas lain masih salah diklasifikasikan sebagai kelas 1 (Tercemar Ringan). Hal ini mungkin disebabkan oleh kesalahan dalam klasifikasi dengan kelas yang lainnya. Meskipun begitu, nilai precision ini masih berada pada kategori baik dan menunjukkan performa model yang cukup stabil dalam membedakan pencemaran ringan

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa model cukup akurat dalam mendeteksi data yang benar-benar termasuk dalam kategori ini. Namun, kesalahan klasifikasi terhadap kelas ini masih lebih tinggi dibandingkan kelas "Baik" atau "Berat", yang menunjukkan bahwa kelas sedang menjadi area yang menantang bagi model karena kemungkinan tumpang tindih dengan kelas sekitarnya.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model sangat presisi dalam mendeteksi air yang masuk dalam

kategori "Tercemar Berat". Hanya sebagian kecil prediksi yang salah, sehingga model dapat dikatakan sangat andal dalam mengenali kondisi kualitas air yang paling berbahaya. Ini penting secara praktis karena kesalahan dalam mendeteksi pencemaran berat dapat memiliki dampak lingkungan yang serius.

3. Nilai rata-rata Recall sebesar 0.9420 atau 94% menunjukkan bahwa model memiliki kemampuan sangat baik dalam mengenali data yang benar-benar berasal dari masing-masing kelas. Recall yang tinggi berarti sebagian besar data berhasil dideteksi dengan benar oleh model.

3.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data berhasil dikenali oleh model tanpa ada satu pun yang terlewat. Ini menunjukkan bahwa model tidak melewatkan satupun kondisi air yang benar-benar baik, menjadikan klasifikasi terhadap kelas ini sangat sempurna dari sisi sensitivitas

3.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa sebagian besar data yang benar-benar tergolong "Tercemar Ringan" berhasil dikenali oleh model. Namun, sekitar 11% data aktual dari kelas ini tidak dikenali dan terklasifikasi ke kelas lain. Ini mengindikasikan bahwa model masih memiliki potensi untuk meningkatkan sensitivitas terhadap pencemaran ringan, terutama karena kelas ini memiliki karakteristik yang bisa menyerupai kelas baik atau sedang.

3.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan sekitar 14% data dari kelas 2 (Tercemar Sedang) yang gagal dikenali dengan benar oleh model. Ini bisa disebabkan oleh tumpang tindih

karakteristik antara kelas sedang dan kelas ringan/berat. Nilai recall ini masih cukup baik, namun performa model terhadap kelas ini adalah yang terendah dibanding kelas lain, sehingga dapat menjadi fokus untuk peningkatan performa klasifikasi.

**3.4 Kelas 3 (Tercemar Berat)** mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa model sangat andal dalam mendeteksi kondisi air yang tergolong dalam kategori kelas 3 (Tercemar Berat). Hanya sebagian kecil data aktual dari kelas ini yang tidak dikenali. Performa tinggi pada kelas ini sangat penting, karena keterlambatan atau kegagalan dalam mendeteksi pencemaran berat dapat berdampak kritis.

**4.** Nilai rata-rata F1-Score sebesar 0.9416 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall pada semua kelas. F1-Score digunakan sebagai metrik gabungan untuk mengukur akurasi model dalam kondisi data yang mungkin tidak seimbang antar kelas.

**4.1 Kelas 0 (Baik)** mendapatkan nilai F1-Score sebesar 1.0 atau 100% menunjukkan bahwa model memiliki kinerja optimal dalam mengenali dan mengklasifikasikan air dengan kualitas baik. Tidak ada kesalahan yang signifikan, baik dalam bentuk false positive maupun false negative, sehingga model sangat akurat dan sensitif terhadap kelas ini.

**4.2 Kelas 1 (Tercemar Ringan)** mendapatkan nilai F1-Score sebesar 0.89 atau 89% menunjukkan bahwa model cukup andal dalam mengklasifikasikan pencemaran ringan, namun masih ada keseimbangan yang perlu ditingkatkan antara precision dan recall. Nilai ini menunjukkan bahwa sebagian data dari kelas ini

masih terdeteksi secara kurang akurat, mungkin karena kesamaan karakteristik dengan kelas "Baik" atau "Tercemar Sedang".

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai F1-Score sebesar 0.87 atau 87% menunjukkan bahwa model memiliki tantangan dalam mengidentifikasi kualitas air yang tergolong "Tercemar Sedang". Hal ini bisa disebabkan oleh data yang lebih kompleks atau memiliki kemiripan dengan kelas di sekitarnya, sehingga menyebabkan penurunan baik dari sisi precision maupun recall.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai F1-Score sebesar 0.98 atau 98% menunjukkan bahwa model memiliki kemampuan sangat baik dan stabil dalam mengklasifikasikan kondisi air yang paling tercemar. Performa ini sangat penting karena pencemaran berat perlu dikenali secara akurat agar dapat ditindaklanjuti segera. Nilai ini mencerminkan bahwa baik precision maupun recall berada dalam kondisi optimal untuk kelas ini.

#### **4.4.9 K-Fold 10**

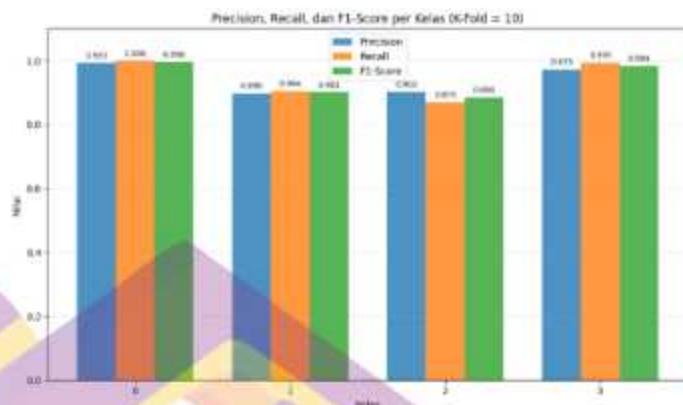
Evaluasi akhir dilakukan dengan menggunakan K-Fold Cross Validation dengan nilai  $K = 10$ , yang merupakan salah satu nilai  $K$  yang paling umum digunakan dalam praktik machine learning. Teknik ini memberikan keseimbangan antara bias dan variansi dalam evaluasi model. Dengan membagi data menjadi sepuluh bagian, setiap iterasi memberikan hasil yang lebih mendetail dan akurat. Hasil evaluasi pada masing-masing kelas disajikan dalam bentuk akurasi, precision, recall, dan F1-Score untuk memberikan pemahaman yang menyeluruh terhadap performa model. Penjelasan lebih lanjut dapat dilihat pada gambar 14.17.

```

K-FOLD 10
-----
Rata-rata Akurasi: 0.9422
Rata-rata Precision: 0.9417
Rata-rata Recall: 0.9422
Rata-rata F1-Score: 0.9418
-----
precision  recall  f1-score  support
-----
0          0.99    1.00    1.00    2950
1          0.88    0.85    0.86    1994
2          0.89    0.88    0.88    2181
3          0.97    1.00    0.98    1983

accuracy          0.93    0.94    0.93    8128
macro avg        0.93    0.94    0.93    8128
weighted avg     0.93    0.93    0.93    8128
-----
waktu training : 36.6442074754056 detik
waktu training : 1.44 menit

```



Gambar 4.17 Metrik Evaluasi dan Diagram K-Fold 10

1. Akurasi 0.9422 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 10, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.

2. Nilai rata-rata Precision sebesar 0.9417 atau 94% menunjukkan bahwa model memiliki tingkat ketepatan yang tinggi dalam memprediksi kelas kualitas air. Precision yang tinggi ini berarti sebagian besar prediksi positif yang dilakukan model benar adanya, sehingga jumlah false positive (prediksi salah terhadap kelas lain) tergolong rendah.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir seluruh data yang diprediksi sebagai "Baik" memang benar berasal dari kelas tersebut. Ini mengindikasikan bahwa model sangat akurat

dalam mengenali kualitas air yang tidak tercemar, dengan risiko kesalahan prediksi yang sangat kecil.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa sebagian besar prediksi untuk kelas “Tercemar Ringan” benar, meskipun masih terdapat sekitar 12% data dari kelas lain yang salah diklasifikasikan sebagai ringan. Ini bisa terjadi karena karakteristik air dengan pencemaran ringan sering kali tumpang tindih dengan kelas “Baik” atau “Sedang”, yang menyebabkan model sesekali membuat kesalahan klasifikasi.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa model cukup baik dalam mengidentifikasi air dengan pencemaran sedang. Meskipun begitu, tingkat kesalahan prediksi positif masih ada, yaitu sekitar 11%, yang mungkin disebabkan oleh kemiripan nilai parameter kualitas air dengan kelas di sekitarnya. Hal ini mengindikasikan bahwa model perlu lebih ditingkatkan dalam membedakan tingkat pencemaran ringan dan sedang..

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model sangat presisi dalam mengidentifikasi air yang tercemar berat. Hampir seluruh data yang diprediksi sebagai “Tercemar Berat” memang benar adanya, yang penting dalam konteks pengawasan kualitas lingkungan karena pencemaran berat memerlukan perhatian khusus dan tindakan segera.

**3.** Nilai rata-rata Recall sebesar 0.9422 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data yang benar-benar berasal dari masing-masing kelas. Dengan nilai recall setinggi ini, berarti

kemungkinan model gagal mendeteksi data dalam kategori (false negative) yang cukup rendah.

**3.1 Kelas 0 (Baik)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data aktual dari kelas “Baik” berhasil dikenali dengan benar oleh model. Tidak ada satupun data kualitas air yang baik yang salah diklasifikasikan ke dalam kelas lain. Ini menunjukkan bahwa model sangat sensitif terhadap kelas ini, dan tidak kehilangan data dari kategori kelas 0 (Baik).

**3.2 Kelas 1 (Tercemar Ringan)** mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa sebagian besar data dengan kelas 1 (Tercemar Ringan) berhasil dikenali oleh model, namun masih terdapat sekitar 11% yang gagal terdeteksi dan mungkin diklasifikasikan ke kelas lain. Hal ini bisa terjadi karena karakteristik kelas ringan tumpang tindih dengan kelas di sekitarnya, sehingga menyulitkan model untuk membedakan secara sempurna.

**3.3 Kelas 2 (Tercemar Sedang)** mendapatkan nilai precision sebesar 0.86 atau 86%. Recall pada kelas ini merupakan yang kelas terendah di antara seluruh kelas, yaitu 86%. Ini berarti ada 14% data aktual dari kelas 2 (Tercemar Sedang) yang tidak dikenali dengan benar oleh model. Kemungkinan besar, model mengalami kesulitan membedakan kelas ini karena karakteristik kualitas air yang berada di antara kelas ringan dan berat, sehingga menyebabkan kebingungan dalam klasifikasi.

**3.4 Kelas 3 (Tercemar Berat)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa semua data aktual dari kelas “Tercemar Berat” berhasil dikenali dengan benar oleh model. Tidak ada data yang luput atau diklasifikasikan

ke kelas lain. Ini sangat penting, karena pendeteksian kondisi air yang sangat tercemar harus dilakukan dengan akurat dan menyeluruh agar bisa segera ditangani secara efektif.

4. Nilai rata-rata F1-Score sebesar 0.9418 atau 94% menunjukkan bahwa model XGBoost memiliki performa yang sangat baik secara keseluruhan dalam hal keseimbangan antara precision dan recall. F1-Score merupakan metrik penting dalam klasifikasi multikelas, terutama ketika terdapat ketidakseimbangan data antar kelas atau adanya kemungkinan salah klasifikasi pada kelas yang saling berdekatan.

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa semua prediksi untuk kelas "Baik" sangat akurat dan konsisten. Model mampu mengidentifikasi seluruh data dari kelas ini dengan tidak hanya akurat (precision tinggi) tetapi juga lengkap (recall tinggi). Ini memperlihatkan keunggulan model dalam mengenali air yang benar-benar layak atau tidak tercemar.

4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa meskipun model cukup baik dalam mengenali kelas ini, terdapat beberapa prediksi yang kurang tepat. Kemungkinan ini disebabkan oleh tumpang tindih karakteristik dengan kelas "Baik" maupun "Tercemar Sedang", sehingga mengurangi presisi maupun kemampuan mengenali semua data aktual (recall) dari kelas ini.

4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa model juga mengalami tantangan serupa dalam membedakan kelas sedang, yang sering kali berada di tengah-tengah karakteristik

ringan dan berat. Hal ini menyebabkan model tidak sepenuhnya akurat dan lengkap dalam memprediksi data pada kelas ini.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan bahwa model sangat andal dalam mengklasifikasikan kualitas air yang termasuk dalam kategori “Tercemar Berat”. Hal ini penting karena air dalam kondisi ini sangat perlu dikenali dengan cepat dan tepat untuk menghindari risiko lingkungan yang serius.

#### **4.4.10 K-Fold 11**

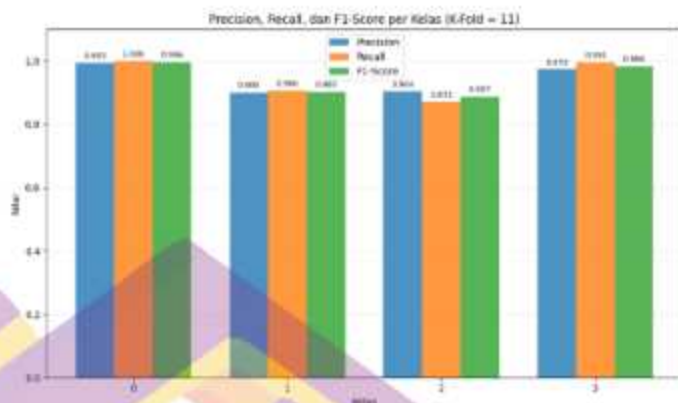
Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 11$ . Semakin banyak jumlah lipatan yang digunakan, maka semakin kecil proporsi data uji dalam setiap iterasi, sehingga model mendapatkan lebih banyak data untuk dilatih. Hal ini memungkinkan model untuk menangkap pola yang lebih kompleks. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.18.

K-FOLD 11

Rata-rata Akurasi: 0.9438  
Rata-rata Precision: 0.9425  
Rata-rata Recall: 0.9438  
Rata-rata F1-Score: 0.9427

	precision	recall	f1-score	support
0	0.99	1.00	1.00	1858
1	0.88	0.85	0.88	1829
2	0.89	0.85	0.87	1909
3	0.97	1.00	0.98	1891
accuracy			0.93	7389
macro avg	0.93	0.93	0.93	7389
weighted avg	0.93	0.93	0.93	7389

waktu Training : 97.8262631883158 detik  
waktu training: 1.63 menit



Gambar 4.18 Metrik Evaluasi dan Diagram K-Fold 11

1. Akurasi 0.9430 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 11, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.

2. Nilai rata-rata Precision sebesar 0.9425 atau 94% menunjukkan bahwa model XGBoost memiliki tingkat akurasi tinggi dalam memberikan prediksi positif yang benar. Precision mengukur seberapa banyak data yang diprediksi ke dalam suatu kelas benar-benar berasal dari kelas tersebut. Semakin tinggi precision, semakin sedikit kesalahan pada prediksi false positive.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan kondisi air berkualitas baik, dengan hanya 1% dari prediksi yang salah diklasifikasikan. Ini menunjukkan bahwa model sangat tepat dalam

mengenali air yang tidak tercemar, dan hampir semua prediksi pada kelas ini benar adanya.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan adanya beberapa kasus di mana air yang diprediksi sebagai tercemar ringan ternyata berasal dari kelas lain, seperti “baik” atau “tercemar sedang”. Hal ini bisa disebabkan oleh karakteristik yang tumpang tindih antara kelas ringan dengan kelas-kelas yang berdekatan.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan adanya prediksi yang kurang tepat. Air dengan kondisi tercemar sedang kadang sulit dibedakan dengan kategori ringan atau berat, sehingga precision tidak sempurna. Namun demikian, nilai 89% tetap menunjukkan performa yang cukup baik.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan air yang tercemar berat, dengan precision sebesar 97%. Ini menandakan bahwa hampir seluruh prediksi untuk kelas ini benar, yang sangat penting mengingat kelas ini merupakan kondisi air paling berbahaya dan memerlukan penanganan segera.

**3.** Nilai rata-rata Recall sebesar 0.9430 atau 94% menunjukkan bahwa model XGBoost memiliki kemampuan yang sangat baik dalam mendeteksi seluruh kasus yang benar dari setiap kelas. Recall mengukur seberapa banyak data dari suatu kelas berhasil dikenali oleh model. Semakin tinggi recall, semakin sedikit kesalahan dalam memprediksi false negative.

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti Model mampu mengenali semua data yang benar-benar masuk kedalam kelas 0 (Baik) tanpa ada yang terlewat. Ini menunjukkan bahwa model sangat andal dalam mendeteksi kondisi air yang masih dalam batas aman atau tidak tercemar.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa masih ada 11% data yang seharusnya dikategorikan sebagai 'tercemar ringan' tetapi gagal dikenali oleh model. Hal ini bisa disebabkan oleh ciri-ciri kelas ini yang menyerupai kelas 'baik' atau 'tercemar sedang'.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision terendah sebesar 0.85 atau 85% Ini berarti sebanyak 15% data dengan kualitas air tercemar sedang tidak berhasil dikenali dengan benar. Bisa jadi, karakteristik kelas ini tumpang tindih dengan kelas lain, sehingga menimbulkan kesalahan klasifikasi.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti Model secara konsisten berhasil mendeteksi seluruh data air yang masuk kategori tercemar berat. Hal ini sangat krusial karena kategori ini merupakan kondisi paling berbahaya, dan akurasi tinggi dalam.

**4.** Nilai rata-rata F1-Score sebesar 0.9427 atau 94% menunjukkan bahwa model XGBoost memiliki performa keseluruhan yang sangat baik dan seimbang antara precision dan recall dalam mengklasifikasikan kualitas air ke dalam kelas-kelas tertentu. F1-Score sangat berguna terutama ketika terdapat ketidakseimbangan antar kelas, karena F1-Score mempertimbangkan kesalahan dari sisi false positive dan false negative secara bersamaan.

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa precision dan recall untuk kelas 'Baik' sama-sama tinggi, artinya hampir tidak ada kesalahan dalam mengklasifikasikan air yang benar-benar masih baik kualitasnya. Model sangat akurat untuk kelas ini.

4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan adanya beberapa kesalahan dalam mengenali atau memprediksi air tercemar ringan, baik karena diklasifikasikan ke kelas yang lebih bersih (kelas 0) atau lebih tercemar (kelas 2). Hal ini mencerminkan kemungkinan adanya overlap fitur antar kelas.

4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.87 atau 87% menunjukkan bahwa model kurang optimal dalam membedakan kualitas air yang termasuk dalam kategori 'tercemar sedang'. Hal ini bisa disebabkan oleh karakteristik kelas ini yang tumpang tindih dengan kelas ringan maupun berat, sehingga menimbulkan ketidakakuratan dalam prediksi.

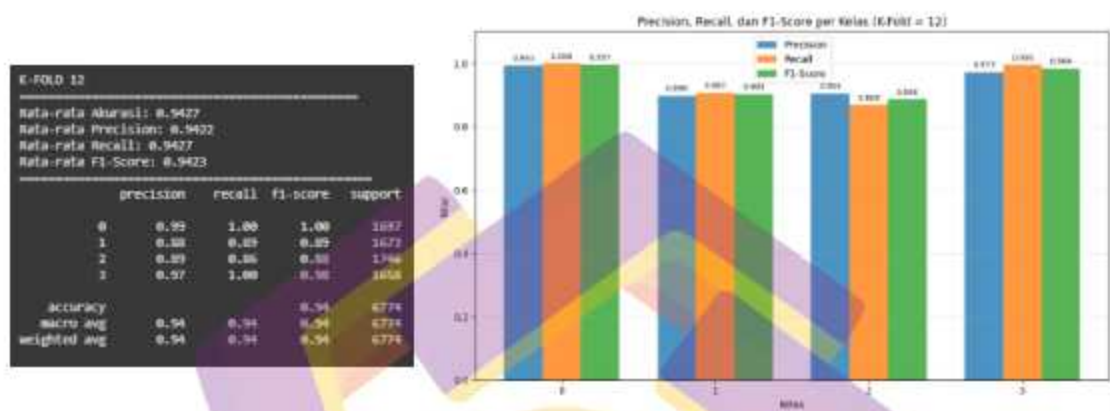
4.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan kemampuan yang sangat baik dalam mengenali kondisi air tercemar berat, dengan nilai F1-Score mendekati sempurna. Artinya, prediksi untuk kategori ini sangat presisi dan jarang meleset, yang penting untuk pencegahan bahaya lingkungan.

#### 4.4.11 K-Fold 12

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 12$ . Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang

digunakan untuk menilai efektivitas model dalam skenario validasi silang ini.

Penjelasan lebih lanjut dapat dilihat pada gambar 4.19.



Gambar 4.19 Metrik Evaluasi dan Diagram K-Fold 12

1. Akurasi 0.9427 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 12, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.
2. Nilai rata-rata Precision sebesar 0.9422 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam memprediksi setiap kelas dengan akurasi tinggi. Artinya, sebagian besar data yang diprediksi ke dalam suatu kelas memang benar berasal dari kelas tersebut, sehingga jumlah false positive relatif rendah secara keseluruhan.

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa prediksi model untuk kelas "Baik" hampir seluruhnya benar. Hanya sekitar 1% data dari kelas lain yang salah diklasifikasikan sebagai "Baik". Ini menunjukkan bahwa fitur-fitur kualitas air yang mencirikan kelas ini sangat dikenali oleh model.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menandakan bahwa sekitar 12% prediksi ke kelas "Tercemar Ringan" merupakan kesalahan atau false positive. Kemungkinan tumpang tindih karakteristik kualitas air dengan kelas "Baik" dan "Sedang" menyebabkan model masih agak sulit membedakan secara akurat kelas ini.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% menandakan bahwa model cukup akurat dalam memprediksi kelas "Tercemar Sedang". Namun, masih ada sekitar 11% prediksi yang salah. Hal ini kemungkinan disebabkan karena fitur parameter kualitas air pada kelas ini memiliki kemiripan dengan kelas "Ringan" dan "Berat", sehingga menyebabkan sebagian data salah klasifikasi.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model sangat akurat dalam memprediksi kelas "Tercemar Berat". Hanya 3% dari prediksi kelas ini yang berasal dari kelas lain. Hal ini menunjukkan bahwa parameter kualitas air untuk kelas ini cukup khas dan mudah dibedakan dari kelas lainnya.

**3.** Nilai rata-rata Recall sebesar 0.9427 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali semua sampel dari

masing-masing kelas. Artinya, sebagian besar data dari setiap kelas berhasil dikenali dan diklasifikasikan dengan benar oleh model (false negative rendah).

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data yang benar-benar termasuk dalam kelas "Baik" berhasil dikenali oleh model tanpa ada yang terlewat. Ini menandakan bahwa fitur atau ciri dari kelas ini sangat konsisten dan mudah dikenali.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa sebagian besar data kelas "Tercemar Ringan" berhasil diprediksi dengan benar oleh model. Namun, masih ada sekitar 11% yang tidak dikenali (false negative), kemungkinan karena fitur kualitas air pada kelas ini tumpang tindih dengan kelas "Baik" atau "Sedang".

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menandakan bahwa ada sekitar 14% data dari kelas "Tercemar Sedang" yang gagal dikenali oleh model dan salah diklasifikasikan ke kelas lain. Ini menunjukkan bahwa fitur kelas ini agak sulit dibedakan secara konsisten, mungkin karena karakteristiknya berada di tengah antara "Ringan" dan "Berat".

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data dari kelas "Tercemar Berat" berhasil diidentifikasi dengan benar oleh model. Hal ini mengindikasikan bahwa fitur kualitas air pada kelas ini sangat khas dan berbeda dari kelas lainnya, sehingga mudah dikenali.

**4.** Nilai rata-rata F1-Score sebesar 0.9427 atau 94% menunjukkan bahwa model memiliki kinerja yang sangat baik dalam hal keseimbangan antara precision dan

recall di seluruh kelas. F1-Score sangat berguna untuk mengukur performa model ketika terdapat ketidakseimbangan antar kelas, karena memperhitungkan baik positif palsu (*false positive*) maupun negatif palsu (*false negative*).

**4.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menandakan bahwa model sangat konsisten dan akurat dalam mengklasifikasikan data yang termasuk dalam kelas ini. Baik precision maupun recall mencapai nilai maksimal, menunjukkan bahwa hampir tidak ada kesalahan dalam prediksi untuk kelas ini.

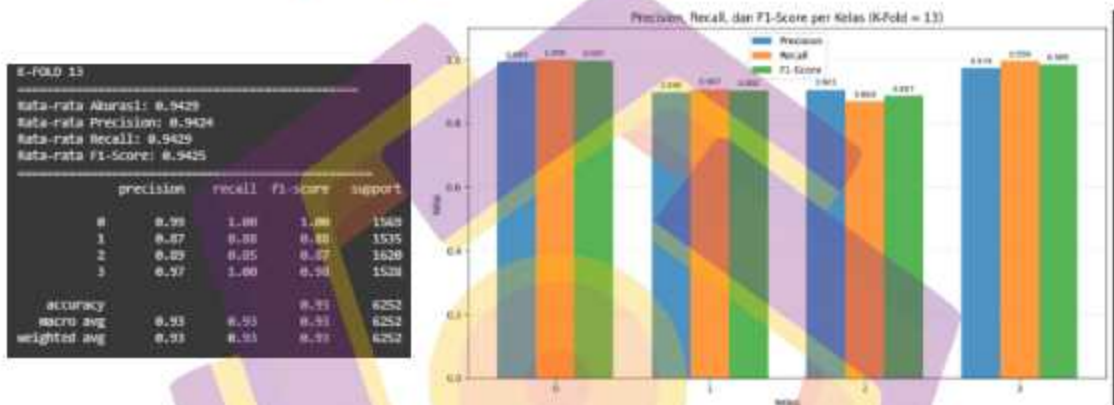
**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menandakan performa yang cukup baik, namun masih terdapat kesalahan klasifikasi. Model mampu mengenali sebagian besar data dari kelas "Tercemar Ringan", tetapi beberapa data mungkin tumpang tindih dengan kelas "Baik" atau "Sedang", menyebabkan adanya kesalahan prediksi baik berupa false positive maupun false negative.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa model cukup mampu membedakan kelas ini, meskipun tidak seakurat kelas lain. Nilai ini mencerminkan tantangan dalam membedakan kelas "Sedang" dari kelas di sekitarnya (Ringan dan Berat), yang mungkin memiliki parameter kualitas air yang saling mendekati.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan bahwa model sangat andal dalam mengenali kondisi air yang benar-benar tercemar berat. Ciri khas parameter kualitas air pada kelas ini kemungkinan besar sangat kuat, sehingga mudah dibedakan dari kelas lain.

#### 4.4.12 K-Fold 13

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation K = 13. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.20.



Gambar 4.20 Metrik Evaluasi dan Diagram K-Fold 13

1. Akurasi 0.9429 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 13, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.
2. Nilai rata-rata Precision sebesar 0.9424 atau 94% menunjukkan bahwa model XGBoost memiliki kemampuan yang sangat baik dalam mengklasifikasikan setiap label dengan tingkat kesalahan minimal dalam bentuk false positive. Semakin tinggi

nilai precision, semakin sedikit prediksi yang salah (prediksi positif yang sebenarnya negatif).

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan Model hampir tidak pernah salah dalam memprediksi kelas "Baik". Artinya, ketika model memprediksi air sebagai kelas 0 (Baik), 99% benar-benar baik, dan hanya sedikit sekali yang salah diklasifikasikan. Ini menunjukkan keakuratan yang sangat tinggi untuk kelas ini.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.87 atau 87%, meskipun nilai yang didapatkan ini tinggi namun nilai pada precision ini menunjukkan bahwa 13% prediksi untuk 'Tercemar Ringan' ternyata salah, biasanya terklasifikasi dari kelas lain. Hal ini bisa disebabkan oleh karakteristik data yang mirip dengan kelas lain, terutama 'Baik' dan 'Tercemar Sedang', sehingga menyebabkan kesalahan model.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa model cukup akurat dalam memprediksi kelas kelas 2 (Tercemar Sedang). Namun, masih ada sekitar 11% prediksi untuk kelas ini yang salah (false positive), kemungkinan karena fitur parameter kualitas airnya mempunyai kemiripan dengan kelas 'Ringan' dan 'Berat'.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa prediksi model untuk kelas 'Tercemar Berat' sangat andal. Hampir semua air yang diprediksi sebagai 'berat' benar-benar berasal dari kelas tersebut, yang penting dalam pemetaan risiko lingkungan atau mitigasi pencemaran.

3. Nilai rata-rata Recall sebesar 0.9429 atau 94% menunjukkan bahwa secara keseluruhan, model mampu mengenali atau menangkap 94% data aktual pada setiap kelas dengan benar.

3.1 Kelas 0 (Baik) mendapatkan nilai Recall sebesar 1.0 atau 100% yang berarti, semua data yang sebenarnya masuk ke kelas 0 (Baik) berhasil dikenali dengan benar oleh model.

3.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai Recall sebesar 0.88 atau 88% menunjukkan bahwa model mampu mengenali sebagian besar sampel dari masing-masing kelas dengan baik. Namun, masih terdapat sebagian kecil data yang seharusnya termasuk ke dalam kelas tertentu namun tidak berhasil dikenali dengan tepat (false negative).

3.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai Recall sebesar 0.85 atau 85% menunjukkan bahwa 85% data dari kelas "Tercemar Sedang" berhasil dideteksi oleh model, sedangkan sisanya (sekitar 15%) tidak terklasifikasi dengan benar. Hal ini bisa terjadi karena fitur air pada kelas ini tumpang tindih dengan kelas "Ringan" dan "Berat", sehingga menyulitkan model untuk membedakannya.

3.4 Kelas 3 (Tercemar Berat) mendapatkan nilai Recall sebesar 1.0 atau 100% menunjukkan bahwa semua sampel dari kelas "Tercemar Berat" berhasil dikenali dengan benar oleh model. Tidak ada satu pun sampel dari kelas ini yang salah diklasifikasikan, yang mengindikasikan bahwa karakteristiknya cukup berbeda dari kelas lainnya.

4. Nilai rata-rata F1-Score sebesar 0.9425 atau 94% menunjukkan bahwa model memiliki kinerja keseluruhan yang sangat baik dalam menyeimbangkan antara

precision dan recall pada setiap kelas. Nilai mendekati 1 (atau 100%) mencerminkan bahwa model cukup andal dalam memprediksi data secara konsisten tanpa terlalu banyak kesalahan prediksi maupun kelolosan data.

**4.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa model benar-benar konsisten dan tepat dalam mengenali kelas “Baik” tanpa terjadi kesalahan prediksi maupun kelolosan data. Artinya, semua data yang seharusnya termasuk kelas ini diprediksi dengan sangat akurat.

**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa model cukup baik dalam mengidentifikasi kelas “Tercemar Ringan”, tetapi masih terdapat sejumlah kesalahan baik dari sisi prediksi yang salah (false positive) maupun data yang terlewat (false negative). Hal ini bisa disebabkan oleh karakteristik parameter kualitas air yang mirip dengan kelas “Baik” dan “Tercemar Sedang”.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.87 atau 87% menunjukkan bahwa model memiliki performa yang cukup baik dalam mengidentifikasi kelas “Tercemar Sedang”. Namun, masih ada sekitar 13% kesalahan yang muncul, baik karena prediksi yang salah maupun karena data dari kelas ini tidak berhasil dikenali. Kemiripan fitur kualitas air dengan kelas lain diduga menjadi penyebabnya.

**4.4** Kelas 3 (tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan bahwa model sangat baik dalam mendeteksi kelas “Tercemar Berat”, dengan sedikit sekali kesalahan prediksi. Artinya, mayoritas data yang tergolong dalam kategori ini dapat dikenali secara tepat oleh model, menunjukkan perbedaan fitur yang cukup jelas dibandingkan kelas lainnya.

#### **4.4.13 K-Fold 14**

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 14$ . Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang

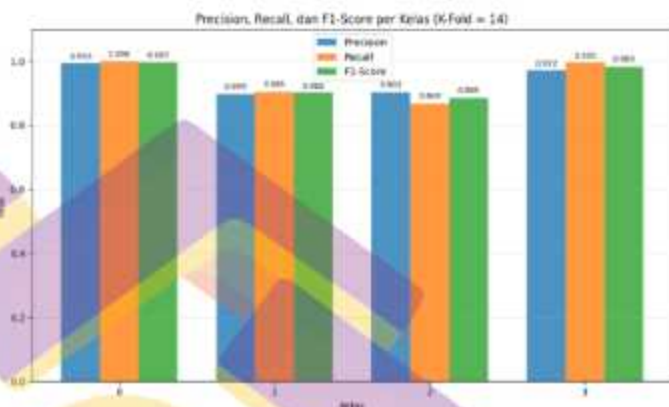
digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.21.

K-FOLD 14

Rata-rata Akurasi: 0.9422  
 Rata-rata Precision: 0.9417  
 Rata-rata Recall: 0.9422  
 Rata-rata F1-Score: 0.9418

	precision	recall	f1-score	support
0	0.99	1.00	1.00	1454
1	0.87	0.88	0.88	1415
2	0.89	0.85	0.87	1513
3	0.97	1.00	0.98	1424
accuracy			0.91	5806
macro avg	0.93	0.93	0.93	5806
weighted avg	0.93	0.93	0.93	5806

Waktu Training : 123.68965746397373 detik  
 Waktu training: 2.66 menit



Gambar 4.21 Metrik Evaluasi dan Diagram K-Fold 14

1. Akurasi 0.9422 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 14, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
2. Nilai rata-rata Precision sebesar 0.9417 atau 94% menunjukkan bahwa secara keseluruhan, model memiliki tingkat akurasi yang tinggi dalam memprediksi kelas dengan benar dibandingkan jumlah seluruh prediksi positif yang dilakukan. Nilai precision sebesar 94% mengindikasikan bahwa sebagian besar prediksi model tepat sasaran, dengan tingkat kesalahan (false positive) yang cukup rendah.

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% bahwa hampir semua data yang diprediksi sebagai kelas “Baik” memang benar-benar berasal dari kelas tersebut. Artinya, model sangat jarang salah dalam mengklasifikasikan data sebagai “Baik”.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.87 atau 87% menunjukkan bahwa masih terdapat sekitar 13% prediksi untuk kelas “Tercemar Ringan” yang salah (*false positive*). Hal ini kemungkinan besar disebabkan oleh karakteristik parameter kualitas air yang tumpang tindih dengan kelas “Baik” dan “Tercemar Sedang”, sehingga membuat model kesulitan membedakannya secara presisi.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% bahwa model cukup akurat dalam memprediksi kelas “Tercemar Sedang”. Namun, masih ada sekitar 11% prediksi untuk kelas ini yang salah, kemungkinan karena fitur parameter kualitas airnya tumpang tindih dengan kelas “Ringan” dan “Berat”.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model mampu dengan baik mengidentifikasi kelas “Tercemar Berat”. Hampir semua data yang diprediksi sebagai kelas ini benar adanya. Artinya, karakteristik kelas ini lebih mudah dibedakan dari kelas lainnya, sehingga model dapat memprediksi dengan akurasi yang tinggi.

**3.** Nilai rata-rata Recall sebesar 0.9417 atau 94% menunjukkan bahwa model cukup andal dalam mengidentifikasi data dari masing-masing kelas secara benar, terutama dalam mengenali semua instance yang sebenarnya termasuk dalam kelas

tersebut (minim kesalahan false negative). Artinya, sebagian besar data yang seharusnya masuk ke suatu kelas berhasil dikenali dengan baik oleh model.

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% bahwa seluruh data yang termasuk dalam kelas “Baik” berhasil dikenali dengan benar oleh model, tanpa ada satu pun yang terlewat atau salah klasifikasi.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan masih terdapat sekitar 12% data dari kelas ini yang tidak dikenali dengan benar oleh model (false negative). Hal ini bisa disebabkan oleh parameter kualitas air pada kelas ini yang tumpang tindih dengan kelas lain, seperti “Baik” atau “Tercemar Sedang”.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.85 atau 85% menunjukkan bahwa model melewatkan sekitar 15% data yang sebenarnya masuk ke dalam kelas “Tercemar Sedang”. Artinya, beberapa data yang seharusnya diklasifikasikan ke kelas ini malah terklasifikasi ke kelas lain, yang bisa jadi karena kemiripan fitur dengan kelas “Ringan” atau “Berat”.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data yang tergolong “Tercemar Berat” berhasil dikenali dan diklasifikasikan dengan benar oleh model. Hal ini menunjukkan bahwa model sangat sensitif terhadap ciri khas dari kelas ini.

**4.** Nilai rata-rata F1-Score sebesar 0.9418 atau 94% menunjukkan keseimbangan yang baik antara precision dan recall dalam performa model. Nilai ini menandakan bahwa model tidak hanya mampu mengidentifikasi data secara benar (recall), tetapi juga akurat dalam memprediksi (precision).

**4.1** Kelas 0 (Baik) mendapatkan nilai F1-Score sebesar 1.0 atau 100% menunjukkan bahwa seluruh prediksi yang diklasifikasikan ke dalam kelas "Baik" memang benar-benar berasal dari kelas tersebut. Artinya, tidak ada data dari kelas lain yang salah diklasifikasikan sebagai "Baik" (false positive = 0). Hal ini menunjukkan bahwa model sangat akurat dalam memprediksi kelas ini.

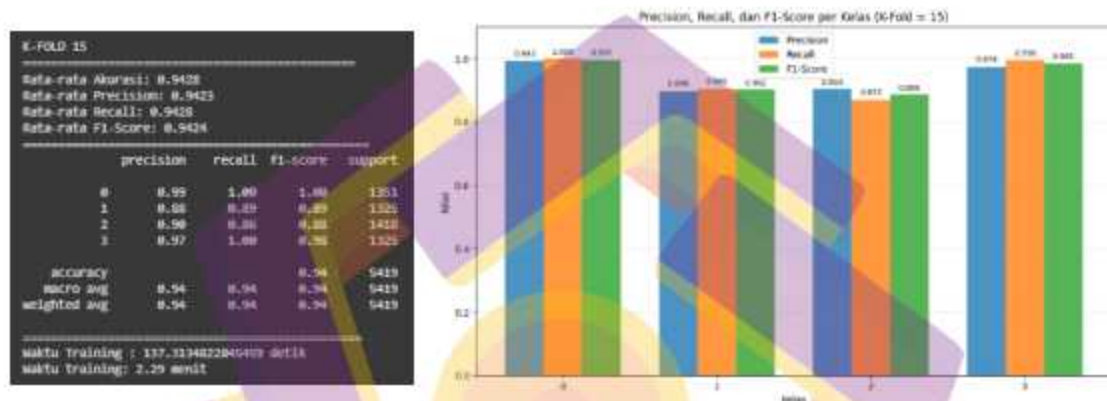
**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai F1-Score sebesar 0.88 atau 88% menunjukkan bahwa sekitar 12% dari prediksi kelas "Tercemar Ringan" merupakan kesalahan (false positive). Kemungkinan besar, data dari kelas lain (seperti "Baik" atau "Tercemar Sedang") masih sulit dibedakan karena adanya kemiripan nilai parameter kualitas air, sehingga terjadi kesalahan klasifikasi ke dalam kelas ini.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai F1-Score sebesar 0.87 atau 87% bahwa 13% dari data yang diprediksi sebagai "Tercemar Sedang" ternyata berasal dari kelas lain. Hal ini menunjukkan bahwa meskipun model cukup baik dalam mengklasifikasikan kelas ini, masih terdapat tumpang tindih antara fitur-fitur pada kelas "Sedang" dengan kelas "Ringan" dan "Berat", yang dapat memengaruhi keakuratan prediksi.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai F1-Score sebesar 0.98 atau 98% menunjukkan bahwa model hampir selalu benar saat memprediksi kelas "Tercemar Berat". Hanya 2% dari data yang diprediksi sebagai kelas ini ternyata berasal dari kelas lain. Hal ini menunjukkan bahwa ciri khas dari kelas "Tercemar Berat" cukup jelas dan mudah dikenali oleh model.

#### **4.4.14 K-Fold 15**

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation K = 15. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.22



Gambar 4.22 Metrik Evaluasi dan Diagram K-Fold 15

1. Akurasi 0.9428 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 15, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
2. Nilai rata-rata Precision sebesar 0.9423 atau 94% menunjukkan bahwa model secara keseluruhan memiliki tingkat ketepatan yang sangat baik dalam mengklasifikasikan kualitas air ke dalam masing-masing kelas. Artinya, sebagian besar prediksi yang dilakukan oleh model adalah benar, dengan jumlah *false positive* yang relatif rendah

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir seluruh data yang diprediksi sebagai "Baik" memang benar-benar termasuk dalam kelas tersebut. Hal ini menandakan bahwa karakteristik kelas "Baik" mudah dikenali oleh model, sehingga kesalahan prediksi sangat minim.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menandakan bahwa sebagian besar prediksi untuk kelas "Tercemar Ringan" benar, namun masih terdapat sekitar 12% prediksi yang keliru (false positive). Kemungkinan ini terjadi karena adanya kemiripan fitur antara kelas "Ringan" dengan kelas "Baik" atau "Sedang", sehingga menyebabkan tumpang tindih dalam klasifikasi.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan bahwa model cukup akurat dalam mendeteksi kelas ini. Meskipun demikian, sekitar 10% dari prediksi kelas ini keliru. Kemungkinan tumpang tindih dengan fitur dari kelas "Ringan" dan "Berat" bisa menjadi penyebabnya.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menandakan bahwa hampir seluruh prediksi untuk kelas ini benar. Artinya, model sangat jarang keliru mengklasifikasikan data yang bukan "Tercemar Berat" ke dalam kelas ini, menunjukkan bahwa ciri-ciri kelas ini cukup khas dan mudah dikenali.

**3.** Nilai rata-rata Recall sebesar 0.9428 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali dan mengidentifikasi

setiap kelas dengan benar. Artinya, sebagian besar data dari tiap kelas berhasil diprediksi secara tepat oleh model, dengan tingkat kesalahan *false negative* yang relatif rendah.

**3.1 Kelas 0 (Baik)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa seluruh data yang memang termasuk dalam kelas ini berhasil dikenali oleh model tanpa ada yang terlewat (*false negative* = 0). Ini menunjukkan bahwa model sangat andal dalam mendeteksi kondisi air yang benar-benar baik.

**3.2 Kelas 1 (Tercemar Ringan)** mendapatkan nilai precision sebesar 0.89 atau 89% menandakan bahwa sebagian besar data yang termasuk dalam kelas “Tercemar Ringan” berhasil dikenali oleh model. Namun, masih ada sekitar 11% data dari kelas ini yang gagal terdeteksi dan justru diprediksi ke kelas lain, kemungkinan karena kemiripan nilai parameter kualitas air dengan kelas “Baik” atau “Sedang”.

**3.3 Kelas 2 (Tercemar Sedang)** mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan bahwa sebagian besar data yang termasuk dalam kelas “Tercemar Sedang” berhasil diprediksi dengan benar, tetapi masih ada sekitar 14% yang tidak terdeteksi dengan tepat. Hal ini bisa disebabkan oleh tumpang tindih karakteristik fitur dengan kelas lain, terutama “Ringan” dan “Berat”, sehingga memengaruhi akurasi model dalam mengenali kelas ini.

**3.4 Kelas 3 (Tercemar Berat)** mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa semua data yang seharusnya masuk dalam kategori ini berhasil diprediksi dengan benar. Ini mengindikasikan bahwa model sangat sensitif terhadap indikator pencemaran berat dan tidak melewatkan satu pun kasus yang tergolong dalam kelas ini.

4. Nilai rata-rata F1-Score sebesar 0.9424 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall dalam melakukan klasifikasi keempat kelas kualitas air. F1-Score merupakan metrik gabungan yang memperhitungkan kedua aspek tersebut, sehingga nilai tinggi menandakan bahwa model tidak hanya tepat dalam memprediksi suatu kelas (precision), tetapi juga konsisten dalam mengenali seluruh data yang seharusnya masuk ke kelas tersebut (recall).

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menunjukkan bahwa model benar-benar akurat dan konsisten dalam memprediksi kondisi air yang baik. Tidak ada data yang salah klasifikasi, baik yang meleset dari kelas ini maupun yang salah diklasifikasikan ke dalamnya.

4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% model sudah cukup baik dalam mengenali dan mengklasifikasi kondisi air yang tercemar ringan. Namun, ada sedikit ketidakseimbangan antara precision dan recall pada kelas ini, yang bisa disebabkan oleh kemiripan parameter dengan kelas "Baik" atau "Sedang".

4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan performa model yang cukup baik, namun masih bisa ditingkatkan. Kemungkinan terdapat tumpang tindih karakteristik fitur dengan kelas lain, sehingga mempengaruhi hasil prediksi model untuk kategori ini.

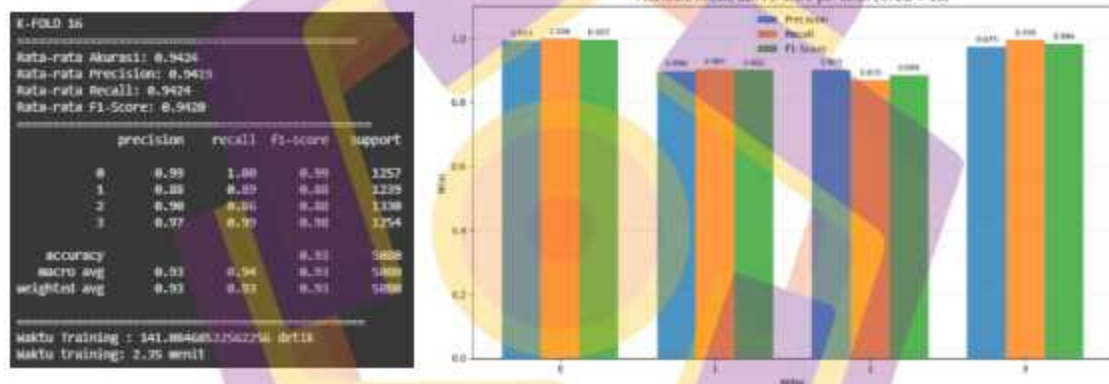
4.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% yang mencerminkan bahwa model hampir selalu tepat dalam mengidentifikasi data yang termasuk dalam kategori pencemaran berat. Ini sangat penting, karena

kesalahan klasifikasi pada kelas ini dapat berdampak besar dalam pengambilan keputusan terkait lingkungan atau kesehatan.

#### 4.4.15 K-Fold 16

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation K = 16.

Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.23.



Gambar 4.23 Metrik Evaluasi dan Diagram K-Fold 16

1. Akurasi 0.9424 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 16, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.

2. Nilai rata-rata Precision sebesar 0.9419 atau 94% kemampuan model yang sangat baik dalam memberikan prediksi yang tepat sesuai dengan kelas targetnya. Nilai ini menunjukkan bahwa secara keseluruhan, kesalahan klasifikasi positif yang dilakukan model tergolong rendah.

2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir semua prediksi yang dikategorikan sebagai kelas 1 (Baik) memang benar-benar berasal dari kelas tersebut.

2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% yang menunjukkan prediksi untuk kelas ini sedikit lebih rentan tercampur dengan kelas lain dibandingkan kelas "Baik".

2.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menunjukkan kemampuan prediksi yang cukup akurat meskipun masih ada potensi salah klasifikasi.

2.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan prediksi pada kelas ini hampir seluruhnya benar, dengan kesalahan yang sangat minim.

3. Nilai rata-rata Recall sebesar 0.9424 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali data yang benar-benar termasuk ke dalam kelasnya. Dengan kata lain, sebagian besar data aktual berhasil terdeteksi sesuai kategorinya, sehingga risiko *false negative* tergolong rendah.

3.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% menandakan bahwa seluruh data aktual dengan kategori "Baik" berhasil dikenali dengan tepat oleh model.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan bahwa sebagian kecil data aktual untuk kategori ini masih terlewat atau salah diklasifikasikan.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menunjukkan tingkat keberhasilan pengenalan yang cukup baik, meskipun masih ada peluang untuk meningkatkan deteksi terhadap kelas ini.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.99 atau 99% yang menunjukkan hampir seluruh prediksi dapat diidentifikasi pada kelas 3 (Tercemar Berat).

**4.** Nilai rata-rata F1-Score sebesar 0.9420 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara *precision* dan *recall* dalam melakukan klasifikasi keempat kelas kualitas air. *F1-Score* merupakan metrik gabungan yang memperhitungkan kedua aspek tersebut, sehingga nilai tinggi menandakan bahwa model tidak hanya tepat dalam memprediksi suatu kelas (*precision*), tetapi juga konsisten dalam mengenali seluruh data yang seharusnya masuk ke kelas tersebut (*recall*).

**4.1** Kelas 0 (Baik) mendapatkan nilai F1-Score sebesar 0.99 atau 99% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara *precision* dan *recall* pada kategori ini. Nilai yang sangat tinggi ini menandakan bahwa model hampir sempurna dalam mengenali data yang termasuk kategori “Baik” dan jarang melakukan kesalahan prediksi.

**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai F1-Score sebesar 0.88 atau 88% menunjukkan bahwa model memiliki keseimbangan yang baik antara

*precision* dan *recall* pada kategori ini, namun masih terdapat peluang perbaikan untuk meningkatkan ketepatan dan konsistensi deteksi data.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai F1-Score sebesar 0.88 atau 88% menunjukkan keseimbangan yang baik antara *precision* dan *recall*, namun serupa dengan kelas 1, model masih berpotensi ditingkatkan agar lebih optimal dalam meminimalkan kesalahan klasifikasi.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai F1-Score sebesar 0.98 atau 98% menunjukkan bahwa model sangat andal dan konsisten dalam mengenali data pada kategori ini, dengan tingkat kesalahan yang sangat rendah.

#### 4.4.10 K-Fold 17

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation K = 17. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.24

```

K-FOLD 17
-----
Rata-rata Akurasi: 0.9417
Rata-rata Precision: 0.9412
Rata-rata Recall: 0.9417
Rata-rata F1-Score: 0.9413
-----

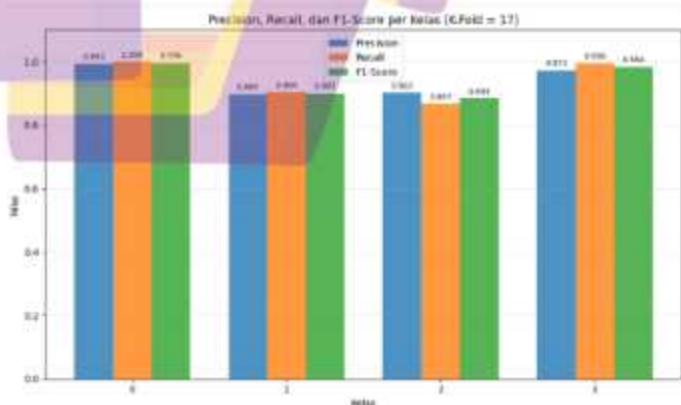
```

	precision	recall	f1-score	support
0	0.99	1.00	1.00	1182
1	0.88	0.88	0.88	1164
2	0.89	0.85	0.87	1244
3	0.97	1.00	0.98	1191
accuracy			0.93	4781
macro avg	0.93	0.93	0.93	4781
weighted avg	0.93	0.93	0.93	4781

```

-----
waktu Training : 157.81725260456084 detik
waktu training: 2.55 menit

```



Gambar 4.24 Metrik Evaluasi dan Diagram K-Fold 17

1. Akurasi 0.9417 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 17, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar.
2. Nilai rata-rata Precision sebesar 0,9412 atau 94% menunjukkan bahwa model memiliki tingkat ketepatan yang sangat baik dalam memprediksi setiap kelas kualitas air. *Precision* mengukur seberapa banyak prediksi positif yang benar-benar relevan, sehingga nilai tinggi menandakan bahwa model jarang melakukan kesalahan dengan memasukkan data yang seharusnya tidak termasuk ke dalam suatu kelas.
  - 2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa model hampir selalu tepat ketika memprediksi kategori “Baik” dan sangat jarang memasukkan data yang tidak sesuai.
  - 2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa mayoritas prediksi pada kelas ini akurat, meskipun masih terdapat beberapa kasus di mana model dapat mengklasifikasikan data yang tidak termasuk sebagai kelas 1 (Tercemar Ringan).
  - 2.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.89 atau 89% yang menandakan menandakan tingkat ketepatan yang baik, namun masih ada sedikit ruang untuk mengurangi kesalahan prediksi terhadap kelas ini.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menunjukkan bahwa model sangat akurat dalam memprediksi kategori ini dan hanya melakukan kesalahan dalam jumlah yang sangat kecil.

**3.** Nilai rata-rata Recall sebesar 0.9417 atau 94% menunjukkan bahwa model mampu mendeteksi sebagian besar data yang benar-benar termasuk ke dalam setiap kelas kualitas air. *Recall* mengukur seberapa banyak data positif yang berhasil dikenali oleh model dibandingkan dengan seluruh data positif yang sebenarnya ada

**3.1** Kelas 0 (Baik) mendapatkan nilai Recall sebesar 1.0 atau 100% menunjukkan seluruh data yang seharusnya termasuk dalam kelas 0 (seluruh data yang seharusnya termasuk kategori "Baik" berhasil diprediksi dengan benar tanpa ada yang terlewat.) berhasil diprediksi dengan benar tanpa ada yang terlewat.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai Recall sebesar 0.88 atau 88% yang berarti sebagian besar data dalam kategori ini berhasil teridentifikasi, namun masih ada sejumlah kecil yang terlewat.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai Recall sebesar 0.85 atau 85% menunjukkan bahwa masih terdapat lebih banyak data yang tidak teridentifikasi dengan benar dibanding kelas lainnya, sehingga kategori ini memiliki potensi untuk ditingkatkan akurasinya.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai Recall sebesar 1.0 atau 100% yang menandakan seluruh data pada kategori ini berhasil dikenali oleh model.

**4.** Nilai rata-rata F1-Score sebesar 0.9413 atau 94% menunjukkan bahwa model memiliki keseimbangan yang baik antara precision (ketepatan) dan recall (kelengkapan) dalam mengklasifikasikan kualitas air. F1-score sangat berguna

untuk mengevaluasi performa model ketika distribusi data antar kelas tidak seimbang, karena :

**4.1** Kelas 0 (Baik) mendapatkan nilai F1-Score sebesar 1.0 atau 100% yang berarti model mampu mengklasifikasikan kategori ini dengan tingkat ketepatan dan kelengkapan maksimal.

**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai F1-Score sebesar 0.88 atau 88% menunjukkan bahwa meskipun performanya baik, masih ada ruang untuk peningkatan baik dari sisi ketepatan maupun kelengkapan.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai F1-Score sebesar 0.87 atau 87% kelas ini mempunyai nilai yang sedikit lebih rendah dibanding kelas sebelumnya, mengindikasikan bahwa kategori ini cenderung lebih sulit diidentifikasi secara konsisten oleh model.

**4.4** Kelas 3 (Tercemat Berat) mendapatkan nilai F1-Score sebesar 0.98 atau 98% menandakan bahwa model hampir selalu tepat dan lengkap dalam mengenali kategori ini.

#### **4.4.17 K-Fold 18**

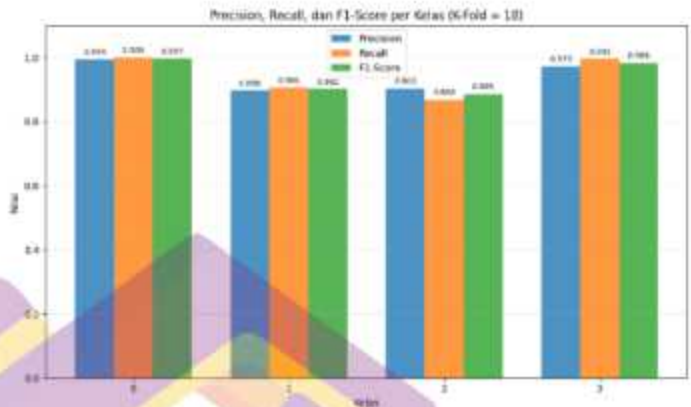
Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 18$ . Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.25.

K-FOLD 18

Rata-rata Akurasi: 0.9423  
 Rata-rata Precision: 0.9418  
 Rata-rata Recall: 0.9423  
 Rata-rata F1-score: 0.9419

	precision	recall	F1-score	support
0	0.99	1.00	1.00	1123
1	0.88	0.89	0.89	1893
2	0.90	0.86	0.88	1175
3	0.97	1.00	0.98	1127
accuracy			0.94	4516
macro avg	0.94	0.94	0.94	4516
weighted avg	0.94	0.94	0.94	4516

Waktu Training : 162.59672881126404 detik  
 waktu training: 2.71 menit



Gambar 4.25 Metrik Evaluasi dan Diagram K-Fold 18

1. Akurasi 0.9423 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 18, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
  - 2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa prediksi kategori ini hampir selalu tepat.
  - 2.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan bahwa meskipun cukup tinggi, masih ada sebagian kecil prediksi kategori ini yang keliru.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% memperlihatkan performa ketepatan yang baik meskipun tidak setinggi kelas "Baik".

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.97 atau 97% menandakan model hampir selalu benar saat memprediksi kategori ini.

**3.** Nilai rata-rata Recall sebesar 0.9423 atau 94% menunjukkan bahwa model mampu mengenali sebagian besar data yang benar-benar termasuk ke dalam masing-masing kelas kualitas air. Artinya, hanya sedikit data yang seharusnya masuk suatu kelas tetapi tidak terdeteksi oleh model.

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% artinya seluruh data kategori ini berhasil dikenali dengan benar oleh model.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan sebagian kecil data kategori ini masih terlewat.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% yang menjadi nilai terendah di antara semua kelas, menandakan masih ada peluang perbaikan dalam mengenali kategori ini.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% menandakan semua data pada kategori ini berhasil diidentifikasi dengan sempurna.

**4.** Nilai rata-rata F1-Score sebesar 0.9419 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall dalam mengklasifikasikan keempat kelas kualitas air. F1-Score adalah metrik gabungan yang memperhitungkan kedua aspek tersebut, sehingga nilai tinggi menandakan bahwa model tidak hanya tepat dalam memprediksi suatu kelas (precision), tetapi

juga konsisten dalam mengenali seluruh data yang seharusnya masuk ke kelas tersebut (recall).

**4.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% yang menandakan kinerja sempurna tanpa kesalahan klasifikasi.

**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan performa yang baik meskipun masih ada beberapa data yang keliru terklasifikasi.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% yang menjadi nilai terendah di antara kelas lainnya, sehingga masih memiliki ruang perbaikan.

**4.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menandakan hampir seluruh data terklasifikasi dengan benar.

#### **4.4.18 K-Fold 19**

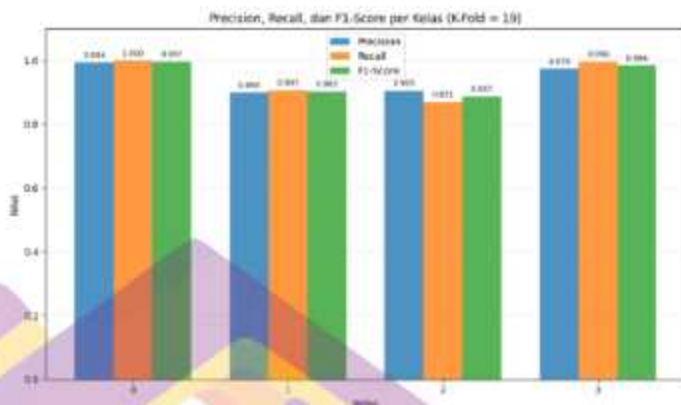
Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation  $K = 19$ . Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. . Penjelasan lebih lanjut dapat dilihat pada gambar 4.26.

K-FOLD 19

rata-rata Akurasi: 0.9433  
 rata-rata precision: 0.9428  
 rata-rata recall: 0.9433  
 rata-rata f1-score: 0.9428

	precision	recall	f1-score	support
0	0.99	1.00	1.00	1066
1	0.93	0.89	0.91	1077
2	0.90	0.80	0.85	1113
3	0.98	1.00	0.99	1062
accuracy			0.94	4278
macro avg	0.94	0.94	0.94	4278
weighted avg	0.94	0.94	0.94	4278

waktu training : 171.83280040950012 detik  
 waktu training: 2.85 menit



Gambar 4.26 Metrik Evaluasi dan Diagram K-Fold 19

1. Akurasi 0.9433 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 19, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
2. Nilai rata-rata Precision sebesar 0.9428 atau 94% menunjukkan bahwa secara umum model mampu memberikan prediksi yang akurat pada masing-masing kelas kualitas air. Precision mengukur tingkat ketepatan prediksi positif suatu kelas dibandingkan dengan seluruh data yang diprediksi sebagai kelas tersebut. Semakin tinggi nilainya, semakin sedikit kesalahan klasifikasi positif yang dilakukan model.
  - 2.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir seluruh data yang diprediksi sebagai “Baik” benar-benar berasal dari kelas tersebut.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% yang berarti masih terdapat beberapa data dari kelas lain yang keliru diklasifikasikan sebagai "Tercemar Ringan".

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% menandakan tingkat ketepatan yang baik meskipun masih ada sedikit kesalahan klasifikasi.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% yang menunjukkan prediksi kelas ini hampir sepenuhnya tepat.

**3.** Nilai rata-rata Recall sebesar 0.9433 atau 94% menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam menemukan data yang benar-benar termasuk ke dalam masing-masing kelas kualitas air. Recall mengukur seberapa banyak data positif yang berhasil terdeteksi dengan benar dibandingkan dengan seluruh data aktual pada kelas tersebut. Semakin tinggi nilainya, semakin sedikit data aktual yang terlewat oleh model.

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti seluruh data aktual pada kelas ini berhasil teridentifikasi dengan tepat.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menunjukkan masih ada sebagian kecil data aktual kelas ini yang tidak terdeteksi dengan benar.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.86 atau 86% menandakan masih terdapat beberapa data aktual yang terlewat atau diklasifikasikan ke kelas lain.

3.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti seluruh data aktual kelas ini berhasil dikenali oleh model.

4. Nilai rata-rata F1-Score sebesar 0.9429 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara *precision* dan *recall* dalam melakukan klasifikasi keempat kelas kualitas air. *F1-Score* merupakan metrik gabungan yang memperhitungkan kedua aspek tersebut, sehingga nilai yang tinggi menandakan bahwa model tidak hanya tepat dalam memprediksi suatu kelas (*precision*), tetapi juga konsisten dalam mengenali seluruh data yang seharusnya masuk ke kelas tersebut (*recall*).

4.1 Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti model memiliki kinerja sempurna dalam mengklasifikasikan kelas ini tanpa kesalahan.

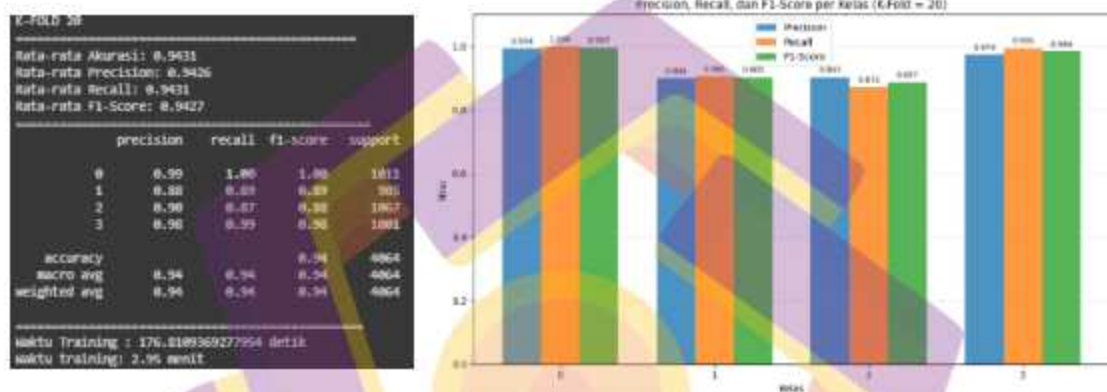
4.2 Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menunjukkan adanya sedikit ketidakseimbangan antara *precision* dan *recall*, meskipun performanya tetap tergolong baik.

4.3 Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.88 atau 88% menandakan model cukup seimbang dalam mengenali kelas ini, walaupun masih ada ruang untuk perbaikan.

4.4 Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan kinerja yang hampir sempurna dalam mendeteksi dan mengklasifikasikan data pada kelas ini.

#### 4.4.19 K-Fold 20

Pada bagian ini, dilakukan evaluasi dengan K-Fold Cross Validation K = 20. Hasil evaluasi mencakup metrik klasifikasi pada tiap kelas kualitas air, yang digunakan untuk menilai efektivitas model dalam skenario validasi silang ini. Penjelasan lebih lanjut dapat dilihat pada gambar 4.27.



Gambar 4.27 Metrik Evaluasi dan Diagram K-Fold 20

1. Akurasi 0.9431 atau 94% ini menunjukkan seluruh prediksi dilakukan oleh model berhasil diklasifikasikan dengan benar. Nilai akurasi yang didapatkan dengan K-Fold 20, dapat melampaui nilai akurasi yang didapatkan daripada nilai – nilai K-Fold sebelumnya. nilai akurasi ini menunjukkan bahwa model berjalan dengan baik dan mayoritas prediksi yang dilakukan oleh model berada pada kelas yang benar
2. Nilai rata-rata Precision sebesar 0.9426 atau 94% menunjukkan bahwa model memiliki tingkat ketepatan prediksi yang sangat baik. Artinya, sebagian besar data yang diprediksi sebagai suatu kelas oleh model memang benar termasuk dalam kelas tersebut. Semakin tinggi nilai *precision*, semakin sedikit kesalahan prediksi positif (*false positive*) yang dihasilkan.

**2.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 0.99 atau 99% menunjukkan bahwa hampir semua data yang diprediksi sebagai “Baik” memang benar-benar berada pada kelas tersebut.

**2.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.88 atau 88% menandakan tingkat ketepatan yang baik meskipun masih terdapat beberapa data yang salah diklasifikasikan sebagai kelas ini.

**2.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.90 atau 90% yang berarti prediksi untuk kelas ini cukup akurat namun masih terdapat peluang untuk mengurangi kesalahan prediksi positif.

**2.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.98 atau 98% menunjukkan ketepatan yang sangat tinggi dalam memprediksi kelas ini.

**3.** Nilai rata-rata Recall sebesar 0.9431 atau 94% menunjukkan bahwa model mampu menangkap sebagian besar data yang sebenarnya termasuk dalam suatu kelas dengan sangat baik. Semakin tinggi nilai recall, semakin sedikit data positif yang terlewatkan (false negative)..

**3.1** Kelas 0 (Baik) mendapatkan nilai precision sebesar 1.0 atau 100% yang berarti semua data yang benar-benar termasuk dalam kelas ini berhasil teridentifikasi tanpa ada yang terlewat.

**3.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai precision sebesar 0.89 atau 89% menandakan sebagian besar data dalam kelas ini berhasil dikenali, meskipun masih ada sebagian kecil yang terlewat.

**3.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai precision sebesar 0.87 atau 87% menunjukkan bahwa kemampuan model mengenali kelas ini masih dapat ditingkatkan agar lebih sedikit data yang terlewat.

**3.4** Kelas 3 (Tercemar Berat) mendapatkan nilai precision sebesar 0.99 atau 99% yang berarti hampir semua data yang benar-benar termasuk kelas 3 (Tercemar Berat) berhasil terdeteksi oleh model.

**4.** Nilai rata-rata F1-Score sebesar 0.9427 atau 94% menunjukkan bahwa model memiliki keseimbangan yang sangat baik antara precision dan recall dalam melakukan klasifikasi keempat kelas kualitas air. F1-Score merupakan metrik gabungan yang memperhitungkan kedua aspek tersebut, sehingga nilai tinggi menandakan bahwa model tidak hanya tepat dalam memprediksi suatu kelas (precision), tetapi juga konsisten dalam mengenali seluruh data yang seharusnya masuk ke kelas tersebut (recall).

**4.1** Kelas 0 (Baik) mendapatkan nilai F1-Score sebesar 1.0 atau 100% menunjukkan bahwa model mampu mengenali dan memprediksi kelas ini secara sempurna.

**4.2** Kelas 1 (Tercemar Ringan) mendapatkan nilai F1-Score sebesar 0.89 atau 89% menunjukkan keseimbangan yang cukup baik antara precision dan recall, meskipun masih ada ruang untuk peningkatan.

**4.3** Kelas 2 (Tercemar Sedang) mendapatkan nilai F1-Score sebesar 0.88 atau 88% yang berarti performa model pada kelas ini masih dapat dioptimalkan agar lebih seimbang dan akurat.

4.4 Kelas 3 (Tercemar Berat) mendapatkan nilai F1-Score sebesar 0.98 atau 98% menandakan bahwa model sangat konsisten dan tepat dalam mengklasifikasikan kelas ini.

#### 4.5 Ringkasan Pengujian Setiap K-Fold

Ringkasan dari setiap pengujian pada K-Fold akan dijelaskan pada uraian pada Tabel 4.14 dan Tabel 4.15 sebagai lanjutan. Pengujian K-Fold sebagai berikut

Tabel 4.14. Pengujian K-Folds

Kelipatan	Akurasi	Precision	Recall	F-1 Score
K-2	0.9347	0.9339	0.9347	0.9430
K-3	0.9438	0.9379	0.9386	0.9381
K-4	0.9410	0.9405	0.9410	0.9406
K-5	0.9411	0.9406	0.9411	0.9407
K-6	0.9415	0.9410	0.9415	0.9411
K-7	0.9425	0.9420	0.9425	0.9421
K-8	0.9422	0.9417	0.9422	0.9418
K-9	0.9420	0.9415	0.9420	0.9416

Tabel 4.15 Lanjutan

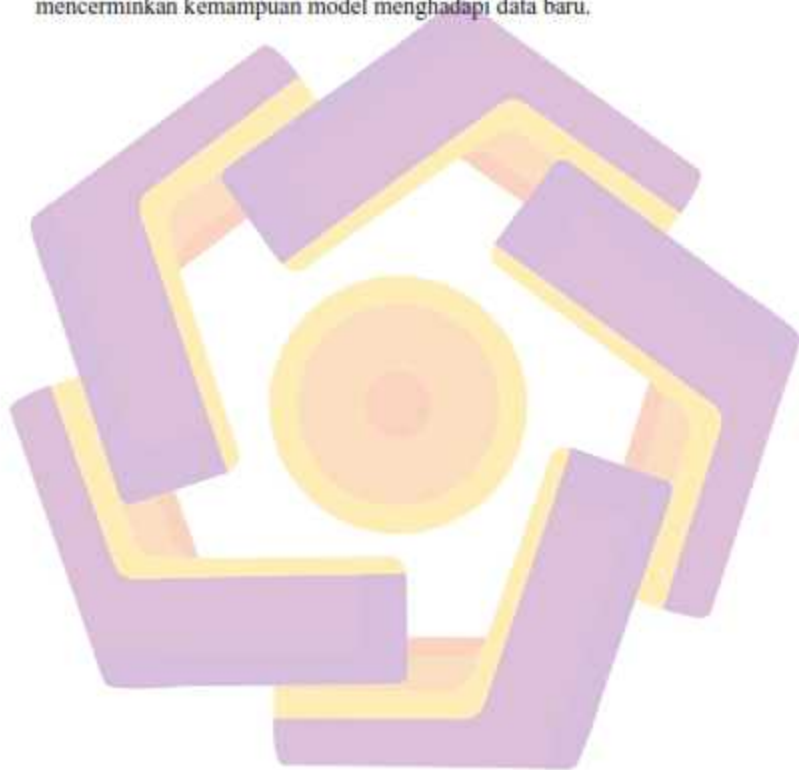
Kelipatan	Akurasi	Precision	Recall	F-1 Score
K-10	0.9422	0.9417	0.9422	0.9418
K-11	0.9430	0.9425	0.9430	0.9427
K-12	0.9427	0.9422	0.9427	0.9423
K-13	0.9429	0.9242	0.9429	0.9425
K-14	0.9422	0.9417	0.9422	0.9418
K-15	0.9428	0.9423	0.9428	0.9424
K-16	0.9424	0.9419	0.9424	0.9420
K-17	0.9417	0.9412	0.9417	0.9413
K-18	0.9423	0.9418	0.9423	0.9419

K-19	0.9433	0.9428	0.9433	0.9429
K-20	0.9431	0.9426	0.9431	0.9427

Berdasarkan hasil penelitian, model XGBoost yang diuji menggunakan metode Holdout menghasilkan akurasi sebesar 81% dengan performa yang belum merata, khususnya pada kelas 0 (Baik) dengan precision 7% dan recall 14%, kelas 1 (Tercemar Ringan) dengan precision 61% dan recall 60%, kelas 2 (Tercemar Sedang) dengan precision 88% dan recall 88% yang merupakan kelas dengan performa terbaik, serta kelas 3 (Tercemar Berat) dengan precision 32% dan recall 44%; sedangkan pengujian menggunakan K-Fold Cross Validation dengan variasi K=2 hingga K=20 secara konsisten memberikan nilai akurasi rata-rata sekitar 94% dengan precision, recall, dan F1-score yang tinggi dan seimbang pada seluruh kelas, yaitu kelas 0 (Baik) dengan precision dan recall hampir sempurna sekitar (99 – 100%), kelas 1 (Tercemar Ringan) dengan nilai precision sekitar 88 – 89% dan nilai recall sekitar 90%, kelas 2 (Tercemar Sedang) dengan precision sekitar 89–90% dan nilai recall sekitar 85–86%, serta kelas 3 (Tercemar Berat) dengan nilai precision sekitar 97% dan nilai recall sekitar 99–100%. Hasil Evaluasi terbaik diperoleh pada K-19 dengan nilai akurasi sebesar 0,9433 atau 94,3%, precision sebesar 0,9428 atau 94,2%, recall sebesar 0,9433 atau 94,3%, dan F1-score sebesar 0,9429 atau 94,2 %. Keempat indikator evaluasi tersebut menggambarkan bahwa model memiliki tingkat keseimbangan yang sangat baik antara kemampuan dalam mengidentifikasi data positif secara benar (recall) dan meminimalkan kesalahan prediksi positif palsu (precision).

Dengan demikian, dapat disimpulkan bahwa penggunaan K-Fold Cross

Validation memberikan hasil evaluasi yang lebih lengkap dan konsisten dibandingkan Holdout, sementara hasil Holdout tetap dijadikan tolak ukur akhir karena menggambarkan performa model ketika diuji pada data uji yang tidak pernah digunakan dalam proses pelatihan maupun validas sehingga mencerminkan kemampuan model menghadapi data baru.



## **BAB V**

### **PENUTUP**

#### **5.1. Kesimpulan**

Penelitian ini berhasil merancang metode klasifikasi kualitas air pada kolam pembenihan ikan lele dengan mengombinasikan metode STORET dan algoritma eXtreme Gradient Boosting (XGBoost). Pendekatan ini menjadi solusi atas permasalahan utama terkait pentingnya pemantauan kualitas air yang dapat berubah secara cepat dan dinamis. Berdasarkan rumusan masalah yang telah disusun, penelitian ini menunjukkan bahwa keterbatasan metode STORET yang hanya mengandalkan pengukuran manual dapat diatasi melalui penerapan algoritma machine learning yang mampu melakukan klasifikasi secara otomatis dan adaptif terhadap data baru.

Meskipun penelitian ini tidak menggunakan alat ukur secara langsung, penggunaan dataset berlabel STORET tetap memadai karena mampu merepresentasikan kondisi kualitas air secara terukur dan menyediakan variasi kelas yang diperlukan dalam proses pelatihan model. Selain itu, penerapan Synthetic Minority Oversampling Technique (SMOTE) terbukti efektif dalam menyeimbangkan jumlah data antar kelas, sehingga mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan generalisasi XGBoost. Hal ini berkontribusi pada peningkatan performa model secara signifikan, terutama dalam mengenali kelas dengan jumlah data yang relatif sedikit seperti kelas Baik dan Tercemar Berat. Meskipun penelitian ini tidak menggunakan alat ukur secara

langsung, penggunaan dataset berlabel STORET tetap memadai karena mampu merepresentasikan kondisi kualitas air secara terukur dan menyediakan variasi kelas yang diperlukan dalam proses pelatihan model. Selain itu, penerapan Synthetic Minority Oversampling Technique (SMOTE) terbukti efektif dalam menyeimbangkan jumlah data antar kelas, sehingga mengurangi bias model terhadap kelas mayoritas dan meningkatkan kemampuan generalisasi XGBoost. Hal ini berkontribusi pada peningkatan performa model secara signifikan, terutama dalam mengenali kelas dengan jumlah data yang relatif sedikit seperti kelas Baik dan Tercemar Berat.

Hasil pengujian menunjukkan bahwa pada skenario tanpa K-Fold (holdout), model XGBoost memperoleh akurasi sebesar 81% dengan keterbatasan pada kelas minoritas. Namun, setelah dilakukan evaluasi menggunakan K-Fold Cross Validation, terjadi peningkatan performa model yang signifikan dengan nilai akurasi tertinggi mencapai 94,33% pada K=19, serta nilai precision, recall, dan F1-score yang relatif seimbang. Hal ini menunjukkan bahwa kombinasi penerapan SMOTE, hyperparameter tuning, dan K-Fold Cross Validation mampu menghasilkan model XGBoost yang stabil dan memiliki kemampuan generalisasi yang baik dalam mengklasifikasikan kualitas air.

Dengan demikian, dapat disimpulkan bahwa metode K-Fold Cross Validation memberikan hasil evaluasi yang lebih lengkap, stabil, dan konsisten dibandingkan metode holdout. Namun demikian, hasil holdout tetap digunakan sebagai tolok ukur akhir karena menggambarkan performa model pada data uji yang

benar-benar baru dan belum pernah digunakan selama proses pelatihan maupun validasi.

Secara keseluruhan, penelitian ini memberikan kontribusi ilmiah terhadap pengembangan metode klasifikasi kualitas air berbasis machine learning. Pendekatan yang diusulkan tidak hanya meningkatkan efektivitas pemantauan kualitas air pada budidaya ikan lele, tetapi juga membuktikan bahwa penerapan machine learning—dengan dukungan teknik SMOTE dan validasi K-Fold—dapat menjadi alternatif yang andal untuk mengatasi keterbatasan metode manual seperti STORET. Hasil penelitian ini diharapkan dapat mendukung pengelolaan ekosistem perairan yang lebih berkelanjutan serta membantu pembudidaya ikan dalam melakukan pengawasan kualitas air secara lebih efisien dan akurat.

## 5.2. Saran

1. Penelitian selanjutnya disarankan untuk menambah jumlah dataset agar model dapat belajar dengan lebih baik dan meningkatkan generalisasi terhadap kondisi nyata.
2. Perlu dilakukan perbandingan dengan algoritma machine learning lain untuk mengetahui performa relatif XGBoost dalam kasus klasifikasi kualitas air.
3. Implementasi model dapat dikembangkan lebih lanjut dalam bentuk sistem monitoring berbasis IoT, sehingga pemantauan kualitas air dapat dilakukan secara real-time di lokasi pembenihan.
4. Untuk keberlanjutan, model ini bisa diintegrasikan dengan aplikasi sederhana berbasis web atau mobile, sehingga mudah digunakan oleh petani ikan lele maupun pihak terkait lainnya.

## DAFTAR PUSTAKA

- [1] M. Y. Abduh, N. I. Asra Aswadi, N. M. A. Husna, S. Syazana, dan N. H. Norazmi-Lokman, "Effects of pH and temperature on striped catfish *Pangasianodon hypophthalmus* juvenile: Data on growth performance and survival rate," *Data Br.*, vol. 52, hal. 0–5, 2024, doi: 10.1016/j.dib.2023.109826.
- [2] A. F. Zambrano, L. F. Giraldo, J. Quimbayo, B. Medina, dan E. Castillo, "Machine learning for manually-measured water quality prediction in fish farming," *PLoS One*, vol. 16, no. 8 August, hal. 1–17, 2021, doi: 10.1371/journal.pone.0256380.
- [3] J. Mahlknecht, J. A. Torres-Martinez, M. Kumar, A. Mora, D. Kaown, dan F. J. Loge, "Nitrate prediction in groundwater of data scarce regions: The futuristic fresh-water management outlook," *Sci. Total Environ.*, vol. 905, no. September, 2023, doi: 10.1016/j.scitotenv.2023.166863.
- [4] N. Ya'acob, N. N. S. N. Dzulkefli, A. L. Yusof, M. Kassim, N. F. Naim, dan S. S. M. Aris, "Water Quality Monitoring System for Fisheries using Internet of Things (IoT)," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1176, no. 1, hal. 012016, 2021, doi: 10.1088/1757-899x/1176/1/012016.
- [5] H. Moeinzadeh, K. T. Yong, dan A. Withana, "A critical analysis of parameter choices in water quality assessment," *Water Res.*, vol. 258, no. April, hal. 121777, 2024, doi: 10.1016/j.watres.2024.121777.
- [6] J. D. Medina *et al.*, "Open-source low-cost design of a buoy for remote water quality monitoring in fish farming," *PLoS One*, vol. 17, no. 6 June, hal. 1–

- 14, 2022, doi: 10.1371/journal.pone.0270202.
- [7] P. Pemerintah, R. Indonesia, P. P. Dan, P. L. Hidup, dan D. K. Lindung, "Hidup yang," no. 097053, 2021.
- [8] A. C. C. Fortes, P. R. G. Barrocas, dan D. C. Kligerman, "Water quality indices: Construction, potential, and limitations," *Ecol. Indic.*, vol. 157, no. July, 2023, doi: 10.1016/j.ecolind.2023.111187.
- [9] B. T. L. I. K. Kitova, "View of Innovative Machine Learning Approaches for Drinking Water Quality Classification\_ Addressing Data Imbalances with Custom SMOTE Sampling Strategy.pdf," Bulgaria, hal. 12, 2025. doi: <https://doi.org/10.30564/jees.v7i3.8195>.
- [10] S. Ramya, S. Srinath, dan P. Tuppada, "Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI," *Case Stud. Chem. Environ. Eng.*, vol. 10, no. April, 2024, doi: 10.1016/j.cscee.2024.100822.
- [11] H. Wijaya, D. P. Hostiadi, dan E. Triandini, "Optimization XGBoost Algorithm Using Parameter Tuning in Retail Sales Prediction," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 3, hal. 769-786, 2024, doi: 10.23887/janapati.v13i3.82214.
- [12] H. M. Sk Injamamul Islam, Foyisal Ahammad, "J World Aquaculture Soc - 2024 - Islam - Cutting-edge technologies for detecting and controlling fish diseases Current.pdf." 2024.
- [13] A. Khudoyberdiev, M. A. Jaleel, I. Ullah, dan D. H. Kim, "Enhanced Water Quality Control Based on Predictive Optimization for Smart Fish Farming,"

- Comput. Mater. Contin.*, vol. 75, no. 3, hal. 5471–5499, 2023, doi: 10.32604/cmc.2023.036898.
- [14] F. M. Yusoff, W. A. D. Umi, N. M. Ramli, dan R. Harun, “Water quality management in aquaculture,” *Cambridge Prism. Water*, vol. 2, 2024, doi: 10.1017/wat.2024.6.
- [15] L. Rathour, V. Singh, M. K. Sharma, N. Dhiman, dan V. N. Mishra, “A review of fuzzy logic analysis in COVID-19 pandemic and a new technique through extended hexagonal intuitionistic fuzzy number in analysis of COVID-19,” *Results Control Optim.*, vol. 17, no. November, 2024, doi: 10.1016/j.rico.2024.100498.
- [16] J. Xu *et al.*, “Assessing and predicting water quality index with key water parameters by machine learning models in coastal cities, China,” *Heliyon*, vol. 10, no. 13, 2024, doi: 10.1016/j.heliyon.2024.e33695.
- [17] M. Y. Shams, A. M. Elshewey, E. S. M. El-kenawy, A. Ibrahim, F. M. Talaat, dan Z. Tarek, “Water quality prediction using machine learning models based on grid search method,” *Multimed. Tools Appl.*, vol. 83, no. 12, hal. 35307–35334, 2024, doi: 10.1007/s11042-023-16737-4.
- [18] B. Ouadi, A. Khatir, E. Magagnini, M. Mokadem, L. Abualigah, dan A. Smerat, “Optimizing silt density index prediction in water treatment systems using pressure-based gradient boosting hybridized with Salp Swarm Algorithm,” *J. Water Process Eng.*, vol. 68, no. October, 2024, doi: 10.1016/j.jwpe.2024.106479.
- [19] M. G. Uddin *et al.*, “Marine waters assessment using improved water quality

- model incorporating machine learning approaches," *J. Environ. Manage.*, vol. 344, no. May, hal. 118368, 2023, doi: 10.1016/j.jenvman.2023.118368.
- [20] S. Huang, J. Xia, Y. Wang, J. Lei, dan G. Wang, "Water quality prediction based on sparse dataset using enhanced machine learning," *Environ. Sci. Ecotechnology*, vol. 20, 2024, doi: 10.1016/j.ese.2024.100402.
- [21] K. Zhang, R. K. Amineh, Z. Dong, dan D. Nadler, "Microwave Sensing of Water Quality," *IEEE Access*, vol. 7, hal. 69481–69493, 2019, doi: 10.1109/ACCESS.2019.2918996.
- [22] S. Yuliana, "PARASIT DAN PENYAKIT IKAN TELEOSTESI," *CV Budi Utama Yogyakarta*, 2021.
- [23] H. Yu, L. Yang, D. Li, dan Y. Chen, "A hybrid intelligent soft computing method for ammonia nitrogen prediction in aquaculture," vol. 8, hal. 64–74, 2021.
- [24] R. A. Hanneman, "PROSEDUR OPERASIONAL PEMBENIHAN LELE (*Clarias spp*)," *Glob. Qual. Stand. Program.*, hal. 2–5, 2009.
- [25] Menteri Negara Lingkungan Hidup, "Pedoman Penentuan Status Mutu Air," *Vasa*, hal. 1–15, 2003, [Daring]. Tersedia pada: <http://medcontent.metapress.com/index/A65RM03P4874243N.pdf>
- [26] M. S. Islam Khan, N. Islam, J. Uddin, S. Islam, dan M. K. Nasir, "Water quality prediction and classification based on principal component regression and gradient boosting classifier approach," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, hal. 4773–4781, 2022, doi: 10.1016/j.jksuci.2021.06.003.

- [27] R. I. Alkanhel, E. S. M. El-Kenawy, M. M. Eid, L. Abualigah, dan M. A. Saeed, "Optimizing IoT-driven smart grid stability prediction with dipper throated optimization algorithm for gradient boosting hyperparameters," *Energy Reports*, vol. 12, no. June, hal. 305–320, 2024, doi: 10.1016/j.egy.2024.06.034.
- [28] S. Albahli, "Efficient hyperparameter tuning for predicting student performance with Bayesian optimization - Shortcut."
- [29] S. B. Belhaouari, A. Islam, K. Kassoul, A. Al-Fuqaha, dan A. Bouzerdoum, "Oversampling techniques for imbalanced data in regression," *Expert Syst. Appl.*, vol. 252, no. PB, hal. 124118, 2024, doi: 10.1016/j.eswa.2024.124118.
- [30] S. Kaiser dan A. Chowdhury, "Integrating oversampling and ensemble-based machine learning techniques for an imbalanced dataset in dyslexia screening tests," *ICT Express*, vol. 8, no. 4, hal. 563–568, 2022, doi: 10.1016/j.ict.2022.02.011.
- [31] G. T. S, G. G, C. C, dan T. Selvan P, "Comprehensive river-water quality monitoring using convolutional neural networks and gated recurrent units: A case study along the Vaigai River," *J. Environ. Manage.*, vol. 365, no. July, 2024, doi: 10.1016/j.jenvman.2024.121567.
- [32] N. Mamat, "Hybrid Support Vector Regression Model and K-Fold Cross Validation for Water Quality Index Prediction in Langat River, Malaysia Naeimah Mamat I\* , Firdaus Mohamad Hamzah I , Othman Jaafar I I," no. Cv, 2021.

- [33] M. Yilmaz, M. Çakir, M. A. Oral, H. Ö. Kazancı, dan O. Oral, "Evaluation of disease outbreak in terms of physico-chemical characteristics and heavy metal load of water in a fish farm with machine learning techniques," *Saudi J. Biol. Sci.*, vol. 30, no. 4, hal. 103625, 2023, doi: 10.1016/j.sjbs.2023.103625.
- [34] M. Heydarian, T. E. Doyle, dan R. Samavi, "MLCM: Multi-Label Confusion Matrix," *IEEE Access*, vol. 10, hal. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [35] D. Valero-Carreras, J. Alcaraz, dan M. Landete, "Comparing two SVM models through different metrics based on the confusion matrix," *Comput. Oper. Res.*, vol. 152, no. April 2022, 2023, doi: 10.1016/j.cor.2022.106131.