

**ANALISIS KUALITAS AIR TELUK JAKARTA
MENGUNAKAN ALGORITMA *LIGHTGBM* DAN
*CATBOOST***

SKRIPSI SARJANA SISTEM INFORMASI

Oleh

Aura Savira

227006516110



**FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI SISTEM INFORMASI
UNIVERSITAS NASIONAL**

2026

**ANALISIS KUALITAS AIR TELUK JAKARTA
MENGUNAKAN ALGORITMA *LIGHTGBM* DAN
*CATBOOST***

SKRIPSI SARJANA SISTEM INFORMASI

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Aura Savira

227006516110



**FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
PROGRAM STUDI SISTEM INFORMASI
UNIVERSITAS NASIONAL**

2026

HALAMAN PENGESAHAN

TUGAS AKHIR

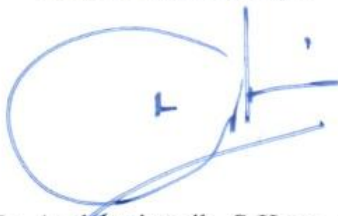
ANALISIS KUALITAS AIR TELUK JAKARTA MENGGUNAKAN
ALGORITMA *LIGHTGBM* DAN *CATBOOST*



Aura Savira

227006516110

Dosen Pembimbing 1



(Dr. Ir. Andrianingsih, S.Kom., MMSI.)

Dosen Pembimbing 2



(Djarot Hindarto, S.Kom., M.Kom.)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS KUALITAS AIR TELUK JAKARTA MENGGUNAKAN ALGORITMA *LIGHTGBM* DAN *CATBOOST*

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, 2 Maret 2026



Aura Savira

227006516110

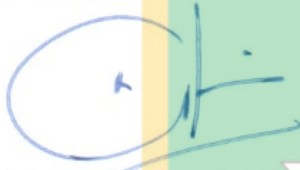
LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

ANALISIS KUALITAS AIR TELUK JAKARTA MENGGUNAKAN ALGORITMA *LIGHTGBM* DAN *CATBOOST*

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2025-2026 pada tanggal 24 Februari Tahun 2026

Dosen Pembimbing I



Dr. Ir. Andrianingsih, S.Kom., MMSI.
0303097902

Dosen Pembimbing 2



Djarot Hindarto, S.Kom., M.Kom.
0317016905

Ketua Program Studi



Dr. Ir. Andrianingsih, S.Kom., MMSI.
0303097902



LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Aura Savira
NPM : 227006516110
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 24 Februari 2026

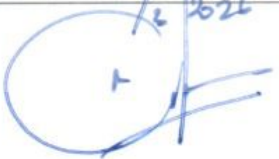


JUDUL DALAM BAHASA INDONESIA :

**ANALISIS KUALITAS AIR TELUK JAKARTA
MENGUNAKAN ALGORITMA *LIGHTGBM* DAN
*CATBOOST***

JUDUL DALAM BAHASA INGGRIS :

**JAKARTA BAY WATER QUALITY ANALYSIS USING
LIGHTGBM AND *CATBOOST* ALGORITHMS**

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 2/3/2026	TGL : 2/3/2026	TGL : 2 Maret 2026
		 Aura Savira

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Aura Savira
NPM : 227006516110
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 24 Februari 2026


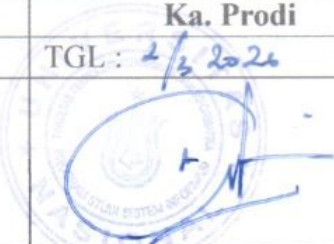
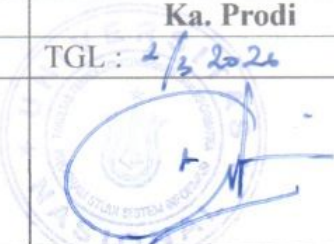

JUDUL DALAM BAHASA INDONESIA :

**ANALISIS KUALITAS AIR TELUK JAKARTA
MENGUNAKAN ALGORITMA *LIGHTGBM* DAN
*CATBOOST***

JUDUL DALAM BAHASA INGGRIS :

**JAKARTA BAY WATER QUALITY ANALYSIS USING
LIGHTGBM AND *CATBOOST* ALGORITHMS**

TANDA TANGAN DAN TANGGAL

Pembimbing 2	Ka. Prodi	Mahasiswa
TGL : 2-3-2026	TGL : 2/3 2026	TGL : 2 Maret 2026
 DJAROT	 	 Aura Savira

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Analisis Kualitas Air Teluk Jakarta Menggunakan Algoritma *LightGBM* dan *Catboost*". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi, di Universitas Nasional.

Dalam proses penyusunan skripsi ini, penulis banyak mendapatkan dukungan, bimbingan, serta bantuan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Orang tua dan seluruh keluarga tercinta yang selalu memberikan doa, semangat, serta dukungan moral dan material kepada penulis.
2. Bapak Dr. Agung Triayudi, S.Kom., M.Kom, selaku Dekan Fakultas Teknologi Komunikasi dan Informatika (FTKI) Universitas Nasional (UNAS).
3. Ibu Dr. Andrianingsih, S.Kom., MMSI., selaku Dosen Pembimbing I sekaligus Ketua Program Studi Sistem Informasi, dan Bapak Djarot Hindarto, S.Kom., M.Kom., selaku Dosen Pembimbing II, yang telah meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, arahan, serta motivasi hingga terselesaikannya penelitian dan penyusunan skripsi ini.
4. Maisaan Muzakki Ramadhan, atas kasih sayang, dukungan, kesabaran, serta semangat yang tak henti diberikan kepada penulis selama proses penyusunan skripsi ini.
5. Teman-teman seperjuangan Nanda, Nazwa, Rizka, Heni, dan Alya, serta teman seangkatan berbagai angkatan yang telah memberikan dukungan dan semangat.
6. Seluruh dosen dan staf akademik Universitas Nasional yang telah membekali penulis dengan ilmu dan pengalaman selama masa perkuliahan.

7. Seluruh pihak yang telah memberikan doa, semangat, dan dukungan kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan saran dan kritik yang membangun untuk penyempurnaan karya ini. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi referensi bagi penelitian selanjutnya.

Akhir kata, semoga Tuhan Yang Maha Esa senantiasa memberikan rahmat dan hidayah-Nya kepada kita semua.

Jakarta, 19 Februari 2026



Aura Savira



ABSTRAK

Teluk Jakarta merupakan kawasan pesisir yang memiliki peran penting secara ekologis dan ekonomi, namun menghadapi tekanan pencemaran akibat aktivitas domestik, industri, dan urbanisasi. Evaluasi kualitas air laut umumnya dilakukan melalui pendekatan konvensional berbasis baku mutu yang belum sepenuhnya mampu menangkap hubungan nonlinier antarparameter lingkungan. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *machine learning* berbasis *ensemble learning*, yaitu *Random Forest*, *LightGBM*, dan *CatBoost*, dalam mengklasifikasikan kualitas air Teluk Jakarta berdasarkan KEPMEN LH No. 51 Tahun 2004.

Data yang digunakan merupakan data *in-situ* tahun 2024 dari 53 stasiun pemantauan pada dua periode pengambilan sampel, yang setelah preprocessing menghasilkan 104 sampel lengkap dengan tiga kategori kelas: Baik, Cemar Ringan, dan Cemar Sedang. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* dengan pendekatan *Stratified K-Fold Cross Validation*. Analisis interpretabilitas menggunakan metode SHAP untuk mengidentifikasi parameter dominan, serta visualisasi spasial menggunakan QGIS untuk menggambarkan distribusi geografis kualitas air.

Hasil penelitian menunjukkan bahwa *CatBoost* memberikan performa paling stabil dan optimal dibandingkan model lainnya, sehingga efektif dalam mendukung analisis kualitas air Teluk Jakarta secara akurat dan interpretatif.

Kata Kunci: Kualitas air laut, Teluk Jakarta, LightGBM, CatBoost, SHAP

ABSTRACT

Jakarta Bay is a coastal area with significant ecological and economic importance, yet it faces increasing pollution pressures from domestic, industrial, and urban activities. Conventional water-quality assessment methods based on regulatory threshold comparisons often fail to capture complex nonlinear relationships among environmental parameters. This study aims to analyze and compare the performance of ensemble-based machine learning algorithms Random Forest, LightGBM, and CatBoost in classifying Jakarta Bay water quality according to KEPMEN LH No. 51 of 2004.

The dataset consists of 2024 in-situ measurements collected from 53 monitoring stations across two sampling periods. After preprocessing, 104 complete samples were obtained and categorized into three classes: Good, Lightly Polluted, and Moderately Polluted. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics with a Stratified K-Fold Cross Validation approach to ensure stability and generalization. SHapley Additive exPlanations (SHAP) were applied to identify dominant contributing parameters, while spatial visualization using QGIS illustrated the geographic distribution of water-quality classes.

The results indicate that CatBoost achieved the most stable and optimal performance among the evaluated models, demonstrating its effectiveness for accurate and interpretable water-quality classification in Jakarta Bay.

Keywords: *Water Quality, Jakarta Bay, LightGBM, CatBoost, SHAP*

DAFTAR ISI

HALAMAN PENGESAHAN	i
PERNYATAAN KEASLIAN TUGAS AKHIR	ii
LEMBAR PERSETUJUAN REVIEW AKHIR	iii
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	iv
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	4
1.4 Batasan Masalah	5
1.5 Manfaat Penelitian	6
1.6 Kontribusi Penelitian	8
BAB II TINJAUAN PUSTAKA	10
2.1 Landasan Teori	10
2.1.1 Kualitas Air Laut dan Dinamika Lingkungan Pesisir	10
2.1.3 Regulasi Baku Mutu Air Laut sebagai Acuan Evaluasi Kualitas Air.....	15
2.1.4 <i>Machine Learning</i> untuk Analisis Lingkungan Pesisir	17
2.1.5 <i>Ensemble Learning</i> dan <i>Boosting</i>	20
2.1.6 Algoritma <i>Light Gradient Boosting Machine (LightGBM)</i> .	23
2.1.7 Algoritma <i>Categorical Boosting (CatBoost)</i>	25
2.1.8 Perbedaan Algoritma <i>LightGBM</i> dan <i>CatBoost</i>	27
2.1.9 <i>Hyperparameter Tuning</i>	29

2.1.10	Evaluasi Model Klasifikasi	31
2.1.11	<i>Explainable AI</i> (SHAP).....	34
2.1.12	<i>Streamlit</i>	35
2.1.13	<i>Quantum Geographic Information System (QGIS)</i>	36
2.2	Studi Literatur	37
2.3	<i>Research Positioning</i>	47
2.4	<i>Fishbone Diagram</i>	50
BAB III	METODOLOGI PENELITIAN	55
3.1	Desain Penelitian	55
3.1.1	Pendekatan Penelitian	55
3.1.2	Tahapan Alur Penelitian.....	56
3.2	Fokus Penelitian.....	59
3.3	Objek Penelitian.....	60
3.4	Variabel Penelitian.....	61
3.5	Hipotesis Penelitian	62
3.6	Lokasi dan Waktu Penelitian	63
3.7	Teknik Pengumpulan Data.....	66
3.8	Teknik Analisis Data	68
3.8.1	<i>Business Understanding</i>	69
3.8.2	<i>Data Understanding</i>	69
3.8.3	<i>Data Preparation</i>	70
3.8.4	<i>Modelling</i>	72
3.8.5	<i>Evaluation</i>	80
3.8.6	<i>Deployment</i>	82
BAB IV	HASIL DAN PEMBAHASAN	84
4.1	Gambaran Umum Data Penelitian	84
4.2	Analisis Deskriptif Data Kualitas Air Laut Teluk Jakarta	85
4.4.1	Statistik Deskriptif Parameter Kualitas Air	85
4.4.2	Distribusi Nilai Parameter Kualitas Air	88
4.3	Hasil <i>Data Preparation</i> dan <i>Preprocessing</i>	90
4.3.1	<i>Cleaning Data</i>	90

4.3.2	Transformasi Data (<i>Long to Wide Format</i>).....	91
4.3.3	Pembagian Data (<i>Train-Test Split</i>)	92
4.4	Hasil Pemodelan Klasifikasi Kualitas Air	92
4.4.1	Hasil Pelatihan Model <i>Random Forest</i>	93
4.4.2	Hasil Pelatihan Model <i>LightGBM</i>	93
4.4.3	Hasil Pelatihan Model <i>CatBoost</i>	94
4.4.4	Perbandingan Awal Hasil Pemodelan.....	94
4.5	Hasil Validasi Model Menggunakan <i>K-Fold</i>	94
4.6	Evaluasi Performa Model	97
4.6.1	Evaluasi Performa Model Berdasarkan Metrik Klasifikasi .	97
4.6.2	<i>Confusion Matrix</i> Model Klasifikasi.....	99
4.6.3	Ringkasan Evaluasi dan Pemilihan Model Terbaik	102
4.7	Interpretasi Model Menggunakan SHAP.....	103
4.7.1	<i>Feature Importance</i>	103
4.7.2	Implikasi Parameter Dominan terhadap Kualitas Air Laut	105
4.8	Visualisasi Spasial Kualitas Air Teluk Jakarta	106
4.9	<i>Deployment</i>	108
BAB V KESIMPULAN DAN SARAN		110
5.1	Kesimpulan	110
5.2	Saran	110
DAFTAR PUSTAKA		112
LAMPIRAN.....		115
	Lampiran 1. Letter of Acceptance (LoA).....	115
	Lampiran 2. AI Checker	115
	Lampiran 3. Hasil Turnitin.....	116
	Lampiran 4. Dokumentasi Bimbingan	116

DAFTAR TABEL

Tabel 2. 1 Baku Mutu Air Laut untuk Biota Laut.....	16
Tabel 2. 2 Perbandingan <i>LightGBM</i> dan <i>CatBoost</i>	28
Tabel 2. 3 Penelitian Terdahulu	38
Tabel 2. 5 Perbandingan Penelitian Sebelumnya dan Penelitian Ini.....	48
Tabel 3. 1 Kategori Variabel Dependen (X) Parameter Air Laut	61
Tabel 3. 2 Kategori Variabel Dependen (Y) Kualitas Air Laut	62
Tabel 3. 3 Timeline Penelitian	66
Tabel 3. 4 Sumber Data Penelitian.....	67
Tabel 4. 1 Statistik Deskriptif Parameter Kualitas Air	86
Tabel 4. 2 Hasil Validasi <i>K-Fold</i> Model <i>Random Forest</i>	95
Tabel 4. 3 Hasil Validasi <i>K-Fold</i> Model <i>LightGBM</i>	95
Tabel 4. 4 Hasil Validasi <i>K-Fold</i> Model <i>CatBoost</i>	96
Tabel 4. 5 Hasil Evaluasi Performa Model Berdasarkan Kelas Kualitas Air	97



DAFTAR GAMBAR

Gambar 2. 1 Logo QGIS	37
Gambar 2. 2 Diagram <i>Research Positioning</i>	49
Gambar 2. 3 <i>Fishbone Diagram</i>	51
Gambar 3. 1 Tahap Alur Penelitian.....	57
Gambar 3. 2 Peta Administratif DKI Jakarta	64
Gambar 3. 3 Peta Sebaran Titik Sampling Kualitas Air Laut Teluk Jakarta	65
Gambar 3. 4 Diagram Alur <i>CRISP-DM</i>	68
Gambar 4. 1 Sample Dataset Kualitas Air di Laut dan Teluk Jakarta 2024	84
Gambar 4. 2 <i>Histogram</i> Distribusi Beberapa Parameter Utama Kualitas Air	88
Gambar 4. 3 <i>Boxplot</i> Parameter Kualitas Air Teluk Jakarta	89
Gambar 4. 4 Distribusi Nilai TSS Berdasarkan Kategori Mutu Air	89
Gambar 4. 5 Tampilan Dataset Kualitas Air Laut Teluk Jakarta Setelah Proses Cleaning Data.....	91
Gambar 4. 6 <i>Confusion Matrix</i> Model <i>Random Forest</i>	100
Gambar 4. 7 <i>Confusion Matrix</i> Model <i>LightGBM</i>	100
Gambar 4. 8 <i>Confusion Matrix</i> Model <i>CatBoost</i>	101
Gambar 4. 9 Perbandingan <i>Feature Importance</i> pada Model <i>Random Forest</i> , <i>LightGBM</i> , dan <i>CatBoost</i>	104
Gambar 4. 10 Visualisasi SHAP <i>Bar Plot</i> pada Model <i>CatBoost</i>	105
Gambar 4. 11 Visualisasi Spasial Kualitas Air Teluk Jakarta	107
Gambar 4. 12 <i>Deployment</i>	108