

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Teluk Jakarta merupakan kawasan pesisir yang memiliki peranan strategis bagi Provinsi DKI Jakarta, baik dari aspek ekologis, sosial, maupun ekonomi. Kawasan ini berfungsi sebagai wilayah perikanan, transportasi laut, pariwisata bahari, serta muara dari berbagai aktivitas perkotaan dan industri di wilayah Jakarta dan sekitarnya. Namun demikian, pesatnya pertumbuhan penduduk, urbanisasi, serta aktivitas industri dan domestik telah memberikan tekanan signifikan terhadap kualitas lingkungan perairan Teluk Jakarta (Edward & Kusnadi, 2023)

Berbagai penelitian menunjukkan bahwa Teluk Jakarta mengalami permasalahan kualitas air yang kompleks, seperti eutrofikasi, peningkatan konsentrasi nutrien, penurunan kadar oksigen terlarut, serta pencemaran oleh limbah domestik dan industri. Kondisi ini berdampak langsung terhadap ekosistem laut dan berpotensi mengganggu keberlanjutan sumber daya pesisir (Damar et al., 2019) Oleh karena itu, pemantauan kualitas air laut secara berkelanjutan menjadi kebutuhan penting dalam upaya pengelolaan lingkungan pesisir Teluk Jakarta.

Pemantauan kualitas air laut di Indonesia umumnya masih mengandalkan pendekatan konvensional berbasis analisis statistik dan perbandingan langsung terhadap baku mutu, seperti yang diatur dalam Keputusan Menteri Lingkungan Hidup Nomor 51 Tahun 2004 tentang Baku Mutu Air Laut. Pendekatan ini efektif sebagai acuan regulasi, namun memiliki keterbatasan dalam menangkap pola nonlinier dan hubungan kompleks antarparameter kualitas air, terutama ketika jumlah parameter dan volume data semakin meningkat (Uddin et al., 2023)

Seiring dengan berkembangnya teknologi komputasi, metode *machine learning* mulai banyak diterapkan dalam analisis dan prediksi kualitas air karena kemampuannya dalam memodelkan hubungan nonlinier

dan kompleks antarvariabel. Berbagai penelitian internasional menunjukkan bahwa algoritma *ensemble learning* seperti *Random Forest*, *LightGBM*, dan *CatBoost* memiliki kemampuan yang baik dalam melakukan klasifikasi maupun prediksi kualitas air (Shah et al., 2024); (Singh et al., 2025). Dibandingkan dengan metode konvensional, pendekatan berbasis *machine learning* dinilai lebih mampu menangani hubungan yang kompleks antarparameter lingkungan. Penelitian di wilayah pesisir juga memperlihatkan bahwa *machine learning* cukup efektif dalam menghadapi variasi spasial dan temporal kualitas air laut yang cenderung dinamis (Bai et al., 2024).

Namun demikian, penerapan *machine learning* pada data lingkungan tidak terlepas dari berbagai tantangan. Permasalahan seperti ketidakseimbangan distribusi kelas, jumlah data yang terbatas, serta karakteristik lingkungan pesisir yang kompleks dapat memengaruhi stabilitas model. Dalam beberapa kasus, model dapat mengalami *overfitting* atau menunjukkan performa yang tidak konsisten ketika diuji pada data yang berbeda. Oleh karena itu, diperlukan proses evaluasi yang lebih menyeluruh, salah satunya melalui pendekatan validasi silang (*cross-validation*), untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik (Nishat et al., 2025).

Selain performa model, aspek interpretasi juga menjadi perhatian penting. Model *machine learning* sering kali dianggap sebagai black box karena proses pengambilan keputusannya sulit dipahami. Padahal, dalam konteks pengelolaan lingkungan, informasi mengenai parameter yang paling berpengaruh terhadap kualitas air sangat dibutuhkan sebagai dasar pengambilan keputusan. Pendekatan *Explainable Artificial Intelligence* (XAI), khususnya metode SHAP (*SHapley Additive exPlanations*), telah banyak digunakan untuk membantu menjelaskan kontribusi masing-masing parameter terhadap hasil prediksi model secara lebih transparan (Lundberg & Lee, 2017); (Makumbura et al., 2024).

Di sisi lain, penyajian hasil analisis dalam bentuk visualisasi spasial juga memiliki peran penting. Integrasi hasil klasifikasi dengan Sistem Informasi Geografis (SIG) memungkinkan kondisi kualitas air ditampilkan dalam bentuk peta, sehingga pola sebaran pencemaran dapat diamati secara lebih jelas dan mudah dipahami oleh pemangku kepentingan (Ardyan, 2025); (Candra & Andrianingsih, 2025).

Berdasarkan uraian tersebut, masih terdapat ruang penelitian dalam analisis kualitas air Teluk Jakarta, khususnya terkait perbandingan performa dan stabilitas beberapa algoritma *ensemble learning*, interpretasi hasil menggunakan pendekatan XAI, serta penyajian hasil klasifikasi dalam bentuk visualisasi spasial yang informatif. Oleh karena itu, penelitian ini dilakukan untuk menganalisis dan meningkatkan akurasi klasifikasi kualitas air laut Teluk Jakarta dengan membandingkan algoritma *Random Forest*, *LightGBM*, dan *CatBoost*, serta mengintegrasikan interpretasi model berbasis SHAP dan visualisasi spasial menggunakan GIS sebagai bentuk dukungan terhadap pengelolaan lingkungan pesisir yang lebih efektif.

## 1.2 Rumusan Masalah

Berdasarkan penjelasan pada latar belakang, kualitas air laut di Teluk Jakarta dipengaruhi oleh banyak parameter yang saling berhubungan. Selama ini penilaian kualitas air lebih banyak dilakukan dengan membandingkan nilai parameter terhadap baku mutu yang berlaku. Cara tersebut memang penting sebagai acuan, namun belum tentu dapat menggambarkan kondisi secara menyeluruh, terutama jika hubungan antarparameter bersifat kompleks. Penggunaan *machine learning* dapat menjadi alternatif untuk membantu proses klasifikasi, tetapi masih perlu dikaji model mana yang paling sesuai dan stabil ketika diterapkan pada data kualitas air. Selain itu, hasil dari model juga perlu dapat dipahami dengan jelas agar dapat memberikan gambaran yang bermanfaat dalam konteks pengelolaan lingkungan.

Berdasarkan hal tersebut, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana hasil kinerja algoritma *Random Forest*, *LightGBM*, dan *CatBoost* dalam mengklasifikasikan kualitas air laut Teluk Jakarta berdasarkan KEPMEN LH No. 51 Tahun 2004?
2. Dari ketiga algoritma tersebut, model mana yang memiliki performa paling baik dan paling stabil berdasarkan hasil evaluasi menggunakan metode *K-Fold Cross Validation*?
3. Parameter kualitas air apa saja yang paling memengaruhi hasil klasifikasi berdasarkan analisis menggunakan metode SHAP?

### 1.3 Tujuan Penelitian

Penelitian ini dilakukan untuk melihat bagaimana penerapan *machine learning* dalam mengklasifikasikan kualitas air laut Teluk Jakarta berdasarkan baku mutu yang berlaku. Selain itu, penelitian ini juga bertujuan untuk membandingkan performa beberapa algoritma yang digunakan serta memahami faktor-faktor yang memengaruhi hasil klasifikasi yang diperoleh. Dengan demikian, penelitian ini tidak hanya melihat hasil dari sisi akurasi saja, tetapi juga memperhatikan kestabilan model serta memahami parameter apa saja yang berpengaruh terhadap hasil klasifikasi.

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui bagaimana kinerja algoritma *Random Forest*, *LightGBM*, dan *CatBoost* dalam mengklasifikasikan kualitas air laut Teluk Jakarta berdasarkan baku mutu KEPMEN LH No. 51 Tahun 2004.
2. Membandingkan performa ketiga algoritma tersebut untuk melihat model mana yang memiliki tingkat akurasi dan kestabilan terbaik berdasarkan evaluasi menggunakan metode *K-Fold Cross Validation*.
3. Mengidentifikasi parameter kualitas air laut yang paling berpengaruh terhadap hasil klasifikasi kualitas air laut Teluk Jakarta menggunakan metode interpretabilitas model SHAP (*SHapley Additive exPlanations*).

Tujuan penelitian ini diharapkan dapat memberikan dasar ilmiah dalam pemilihan model *machine learning* yang efektif dan transparan untuk analisis kualitas air laut, serta mendukung pengambilan keputusan dalam pengelolaan lingkungan pesisir berbasis data.

#### 1.4 Batasan Masalah

Agar penelitian ini memiliki arah yang jelas dan terfokus, diperlukan pembatasan ruang lingkup penelitian. Batasan masalah ditetapkan untuk memastikan bahwa pembahasan dan hasil penelitian tetap relevan dengan tujuan utama, yaitu evaluasi dan peningkatan akurasi klasifikasi kualitas air laut Teluk Jakarta menggunakan pendekatan *machine learning*. Adapun batasan masalah dalam penelitian ini adalah sebagai berikut:

##### 1. Data Penelitian

Penelitian ini menggunakan data kualitas air laut Teluk Jakarta yang diperoleh dari portal Satu Data Jakarta dan/atau data pemantauan resmi Dinas Lingkungan Hidup (DLH) Provinsi DKI Jakarta. Data yang digunakan merupakan hasil pengukuran langsung (*in situ*) pada beberapa titik pengamatan di wilayah Teluk Jakarta. Parameter yang dianalisis dibatasi pada parameter fisika dan kimia yang tersedia dalam dataset dan sesuai dengan penilaian mutu air laut berdasarkan KEPMEN LH No. 51 Tahun 2004.

##### 2. Cakupan Wilayah Penelitian

Penelitian ini hanya membahas wilayah perairan Teluk Jakarta yang berada dalam batas administratif Provinsi DKI Jakarta. Titik pengamatan yang berada di luar wilayah tersebut tidak dimasukkan dalam analisis. Pembatasan ini dilakukan agar data yang digunakan tetap sesuai dengan fokus penelitian dan hasil klasifikasi dapat dianalisis secara lebih jelas berdasarkan wilayah yang sama.

##### 3. Fokus Kajian Penelitian

Penelitian ini berfokus pada klasifikasi kualitas air laut menggunakan algoritma *machine learning*. *Random Forest* digunakan

sebagai model pembanding (*baseline*), sedangkan *LightGBM* dan *CatBoost* digunakan sebagai model utama dalam proses analisis. Penelitian ini tidak membahas secara rinci mengenai penyebab pencemaran, dampak ekologis lanjutan, maupun perubahan kualitas air berdasarkan musim atau waktu.

#### 4. Metode Analisis dan Evaluasi Model

Metode analisis dalam penelitian ini meliputi tahap preprocessing data, pelatihan model *machine learning*, serta evaluasi performa model menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Evaluasi stabilitas dan kemampuan generalisasi model dilakukan menggunakan metode *K-Fold Cross Validation*. Penelitian ini hanya berfokus pada klasifikasi dan tidak mencakup prediksi jangka panjang maupun pemodelan tren waktu.

#### 5. Interpretabilitas dan Visualisasi Hasil

Interpretasi hasil klasifikasi dibatasi pada analisis kontribusi parameter kualitas air menggunakan metode SHAP (*SHapley Additive exPlanations*). Visualisasi hasil klasifikasi dilakukan dalam bentuk peta spasial menggunakan QGIS untuk menampilkan distribusi kualitas air laut pada titik-titik pengamatan. Visualisasi spasial digunakan sebagai sarana pendukung interpretasi hasil dan bukan sebagai fokus utama metode penelitian.

#### 6. Keluaran Penelitian

Keluaran penelitian ini dibatasi pada hasil evaluasi kinerja model *machine learning* dalam klasifikasi kualitas air laut Teluk Jakarta, penentuan model dengan performa terbaik, serta visualisasi spasial distribusi kualitas air laut. Penelitian ini tidak menghasilkan sistem operasional atau aplikasi layanan publik berbasis *real-time*.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoritis, praktis, maupun akademik, khususnya dalam pengembangan analisis

kualitas air laut berbasis *machine learning* dan visualisasi spasial pada wilayah pesisir Teluk Jakarta.

### 1. Manfaat Teoritis

Secara teoritis, penelitian ini memberikan kontribusi dalam pengembangan penerapan *machine learning* pada bidang lingkungan pesisir, khususnya dalam klasifikasi kualitas air laut. Penelitian ini memperkaya kajian ilmiah terkait penggunaan algoritma *ensemble* berbasis *boosting*, yaitu *LightGBM* dan *CatBoost*, dalam meningkatkan akurasi klasifikasi data lingkungan yang bersifat multiparameter. Selain itu, penerapan metode interpretabilitas SHAP memberikan kontribusi terhadap pengembangan pendekatan *Explainable Artificial Intelligence* (XAI) dalam analisis kualitas air laut yang mengacu pada baku mutu KEPMEN LH No. 51 Tahun 2004.

### 2. Manfaat Praktis

Secara praktis, penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi Dinas Lingkungan Hidup (DLH) dan instansi terkait dalam melakukan pemantauan kualitas air laut Teluk Jakarta secara lebih berbasis data. Model klasifikasi yang dihasilkan dapat membantu mengidentifikasi lokasi perairan dengan kondisi kualitas air yang berbeda secara lebih cepat dan objektif. Selain itu, visualisasi spasial menggunakan QGIS memberikan gambaran distribusi kualitas air laut yang mudah dipahami, sehingga dapat digunakan sebagai dasar awal dalam penentuan wilayah prioritas pengelolaan dan pengendalian pencemaran.

### 3. Manfaat Akademik

Secara akademik, penelitian ini diharapkan dapat menjadi salah satu referensi bagi mahasiswa atau peneliti yang tertarik mempelajari penerapan *machine learning* dalam analisis kualitas lingkungan, khususnya di wilayah pesisir. Melalui penelitian ini, ditunjukkan bagaimana proses pengolahan data, pembuatan model klasifikasi, analisis menggunakan SHAP, hingga visualisasi spasial dapat

dilakukan dalam satu rangkaian penelitian. Penelitian ini juga diharapkan dapat memberikan gambaran mengenai tahapan analisis berbasis data, sehingga dapat menjadi bahan pertimbangan atau acuan untuk penelitian selanjutnya di bidang *data science* maupun sistem informasi lingkungan.

## 1.6 Kontribusi Penelitian

Penelitian ini memberikan kontribusi dalam analisis kualitas lingkungan pesisir dengan memanfaatkan metode *machine learning* untuk mengklasifikasikan kualitas air laut Teluk Jakarta. Pada penelitian ini digunakan algoritma *LightGBM* dan *CatBoost* untuk mengelompokkan kondisi kualitas air berdasarkan parameter fisika dan kimia yang mengacu pada baku mutu KEPMEN LH No. 51 Tahun 2004. Melalui pendekatan tersebut, proses klasifikasi dilakukan dengan mempertimbangkan banyak parameter kualitas air yang saling berkaitan, sehingga analisis dapat dilakukan secara lebih terstruktur dan berbasis data. Pendekatan ini memberikan alternatif metodologis yang lebih adaptif dalam menangani data lingkungan yang bersifat multiparameter dan memiliki karakteristik nonlinier.

Selain itu, penelitian ini berkontribusi secara ilmiah melalui analisis perbandingan kinerja antara model *baseline Random Forest* dan *algoritma boosting modern*. Hasil penelitian menunjukkan bagaimana *LightGBM* dan *CatBoost* mampu memberikan performa klasifikasi yang lebih stabil dan akurat dibandingkan model *baseline*, serta memperlihatkan kemampuan model *boosting* dalam menangkap pola kompleks pada data kualitas air laut. Hasil penelitian ini dapat menjadi tambahan referensi dalam kajian evaluasi kinerja algoritma *machine learning* pada bidang lingkungan pesisir.

Selain itu, penelitian ini juga menerapkan metode *Explainable Artificial Intelligence (XAI)* melalui SHAP (*SHapley Additive exPlanations*) untuk melihat pengaruh masing-masing parameter kualitas air terhadap hasil klasifikasi. Melalui pendekatan ini, model yang digunakan tidak hanya

menghasilkan prediksi, tetapi juga memberikan penjelasan mengenai faktor-faktor yang memengaruhi hasil tersebut sehingga lebih mudah dipahami dalam konteks pengelolaan lingkungan.

Dari sisi visualisasi, penelitian ini menyajikan peta distribusi kualitas air laut Teluk Jakarta menggunakan QGIS. Peta tersebut membantu menggambarkan kondisi kualitas air pada setiap titik pengamatan secara lebih jelas. Dengan adanya visualisasi spasial ini, hasil analisis menjadi lebih mudah diinterpretasikan dan dapat digunakan sebagai informasi pendukung dalam pemantauan kualitas air di wilayah pesisir.

Secara keseluruhan, penelitian ini diharapkan dapat memberikan gambaran mengenai penerapan *machine learning* dalam analisis kualitas air laut yang tidak hanya fokus pada akurasi, tetapi juga memperhatikan aspek interpretasi dan penyajian hasil yang sesuai dengan kondisi perairan pesisir Indonesia.

