

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mutu air sungai merupakan salah satu indikator penting dalam menjaga keberlanjutan lingkungan hidup, kesehatan masyarakat, serta mendukung aktivitas sosial dan ekonomi. Indonesia memiliki ketersediaan air permukaan yang melimpah berupa ribuan sungai yang tersebar di berbagai wilayah dan berperan sebagai sumber air utama untuk keperluan pertanian, domestik, industri, dan habitat berbagai organisme perairan (Firmansyah dkk., 2021; Sipahutar dkk., 2025). Namun, peningkatan aktivitas manusia, seperti urbanisasi, industrialisasi, dan pertanian intensif, telah menyebabkan penurunan mutu air sungai secara signifikan di berbagai wilayah dunia (Bhaskoro & Ramadhan, 2018). Degradasi mutu air sungai yang tidak terkendali dapat mengakibatkan kerusakan ekosistem perairan dan meningkatkan risiko gangguan kesehatan masyarakat (UNEP, 2022).

Secara global, sekitar 80% air limbah dibuang ke lingkungan tanpa melalui pengolahan yang memadai (UNEP, 2022), sementara *World Health Organization* melaporkan bahwa sekitar 2 miliar penduduk dunia masih menggunakan sumber air yang terkontaminasi. Kondisi global tersebut menegaskan bahwa pemeliharaan status mutu air sungai agar tetap optimal merupakan aspek penting dalam menjaga keberlanjutan ekosistem, mencegah penyakit berbasis air (*waterborne diseases*), dan meningkatkan efisiensi ekonomi melalui pengurangan biaya pengolahan air..

Di Indonesia, permasalahan pencemaran air sungai masih menjadi isu lingkungan yang krusial. Laporan Kementerian Lingkungan Hidup dan Kehutanan (KLHK, 2017) membuktikan bahwa sebagian besar sungai di wilayah perkotaan berada pada kondisi tercemar ringan hingga berat. Kondisi ini mendorong perlunya sistem pemantauan dan evaluasi Mutu air sungai yang lebih efektif, objektif, dan berbasis data (Inayah dkk., 2021). Penilaian dan klasifikasi mutu air sungai menjadi langkah penting dalam pengelolaan sumber daya air, penentuan kebijakan lingkungan, serta perencanaan upaya pengendalian pencemaran. Upaya tersebut sejalan dengan *Sustainable Development Goals* (SDGs), khususnya SDG 6 (Air

Bersih dan Sanitasi Layak) dan SDG 15 (Ekosistem Daratan), yang menekankan pentingnya pengelolaan sumber daya air yang berkelanjutan (United Nations, 2015).



Gambar 1.1 Peta Kualitas Air Sungai berdasarkan IP
Sumber: <https://ppkl.menlhk.go.id/onlimo-2022>

Mutu air sungai dapat dievaluasi secara komprehensif menggunakan Indeks Mutu Air, yang berfungsi merangkum berbagai indikator mutu air ke dalam satu nilai indeks yang merepresentasikan kondisi mutu air secara keseluruhan (Ratnaningsih dkk., 2018). Penggunaan indeks ini bertujuan untuk menyederhanakan interpretasi data mutu air yang bersifat multidimensi sehingga informasi yang dihasilkan lebih mudah dipahami oleh pengambil kebijakan maupun masyarakat (Rasyid dkk., 2025). Di Indonesia, penilaian status mutu air sungai secara resmi menggunakan Indeks Pencemaran (IP), yang mengklasifikasikan mutu air menjadi memenuhi baku mutu, cemar ringan, cemar sedang, dan cemar berat berdasarkan perbandingan hasil pengukuran indikator mutu air dengan baku mutu yang berlaku (Ahmad dkk., 2024). Penentuan status mutu air sungai tersebut mengacu pada Keputusan Menteri Lingkungan Hidup Nomor 115 Tahun 2003 tentang Pedoman Penentuan Status mutu air sungai (Prasetia & Walukow, 2021).

Penilaian status mutu air diperlukan untuk memberikan gambaran mengenai tingkat pencemaran yang terjadi pada suatu sungai. Penelitian Kusumaningtyas dkk. (2022) menunjukkan bahwa beberapa titik pengamatan di Sungai Pekalongan telah mengalami pencemaran berdasarkan hasil pengukuran sejumlah indikator

mutu air, antara lain DO, BOD, dan COD. Selanjutnya, Afifah dkk. (2025) menunjukkan bahwa status mutu air Sungai Bengawan Solo selama periode 2020–2022 berada pada kategori cemar berat dan terdapat sejumlah indikator yang tidak memenuhi baku mutu. Temuan tersebut menunjukkan bahwa kondisi status mutu air sungai berkaitan dengan berbagai indikator penyusunnya sehingga diperlukan analisis lebih lanjut untuk mengetahui indikator yang berpengaruh terhadap status mutu air sungai.

Penilaian mutu air sungai umumnya didasarkan pada berbagai indikator fisika, kimia, dan biologi, seperti Nitrat, *dissolved oxygen* (DO), *biochemical oxygen demand* (BOD), *chemical oxygen demand* (COD)(KLHK, 2017). Indikator-indikator tersebut saling berkaitan dan mencerminkan kondisi mutu air yang kompleks (Mestika, 2024). Kompleksitas ini menyebabkan penentuan status mutu air sungai sulit dilakukan secara konvensional dan membutuhkan pendekatan analisis yang mampu mengolah banyak peubah secara simultan (Muniz & Oliveira-Filho, 2023).

Untuk mengetahui faktor-faktor mutu air yang memengaruhi status mutu air sungai, diperlukan pemodelan statistik yang mampu menggambarkan hubungan antara indikator dengan status mutu air sungai (Agresti, 2018). Salah satu metode yang sering digunakan dalam analisis hubungan antar peubah adalah regresi linier klasik. Namun, pemodelan menggunakan regresi linier klasik pada peubah respon kategorik memiliki keterbatasan mendasar, antara lain potensi menghasilkan nilai prediksi di luar rentang probabilitas yang valid, pelanggaran asumsi normalitas dan homoskedastisitas galat, serta interpretasi hubungan antarpeubah yang kurang tepat (Agresti, 2007). Hosmer & Lemeshow (2013) menegaskan bahwa regresi linier tidak dirancang untuk memodelkan peluang kejadian.

Secara konvensional, penentuan status mutu air sungai didasarkan pada perhitungan nilai Indeks Pencemaran (IP) menggunakan nilai maksimum dan rata-rata dari rasio konsentrasi indikator terhadap baku mutunya (KLHK, 2017). Ketika jumlah amatan harian sangat besar dan berasal dari ratusan stasiun pemantau yang tersebar luas, pengelompokan menggunakan metode IP ini menghadapi kendala karena proses evaluasi dan perhitungan manual memerlukan waktu yang relatif

lama. Oleh karena itu, penelitian ini memanfaatkan pendekatan metode klasifikasi statistik sebagai alternatif untuk mempercepat dan mengotomatisasi proses pengelompokan status mutu air sungai berdasarkan data indikator kimia yang tersedia.

Klasifikasi bertujuan untuk memetakan setiap objek ke dalam salah satu kelas berdasarkan pola yang dipelajari dari data berlabel (Nurhalizah & Ardianto, 2024). Proses klasifikasi terdiri atas dua tahap, yaitu tahap pembelajaran, di mana model dibangun menggunakan data dengan kategori kelas yang sudah diketahui, serta tahap pengklasifikasian, di mana model digunakan untuk memprediksi label kategori pada data baru (Nugraha dkk., 2025). Setiap objek direpresentasikan oleh sejumlah atribut yang menggambarkan karakteristik mutu air, sedangkan kelas ditentukan oleh label yang bersifat kategorikal (Jiawei dkk., 2011). Salah satu metode statistik yang banyak digunakan untuk permasalahan klasifikasi multikelas dan memiliki interpretabilitas yang baik adalah regresi logistik ordinal.

Regresi logistik ordinal merupakan metode klasifikasi yang digunakan ketika variabel respon memiliki lebih dari dua kategori yang tersusun secara berurutan. (Agresti, 2018). Metode ini mampu menjelaskan pengaruh variabel prediktor terhadap kecenderungan suatu pengamatan berada pada kategori yang lebih tinggi atau lebih rendah dalam suatu urutan kategori (Purwa dkk., 2025). Karakteristik tersebut menjadikan regresi logistik ordinal sesuai digunakan untuk klasifikasi mutu air sungai yang memiliki lebih dari dua kategori status mutu (Hosmer & Lemeshow, 2013)

Sejumlah penelitian terdahulu menunjukkan bahwa regresi logistik ordinal merupakan metode yang efektif untuk memodelkan variabel respon yang memiliki kategori berurutan. Yuniarty dkk. (2023) menerapkan regresi ordinal logit dan probit untuk menganalisis faktor yang memengaruhi tingkat kesejahteraan rumah tangga petani tanaman pangan. Hasil penelitian menunjukkan bahwa model regresi ordinal mampu mengidentifikasi variabel-variabel yang berpengaruh terhadap kategori kesejahteraan yang tersusun secara berurutan. Keumala (2022) menunjukkan bahwa regresi logistik ordinal mampu memberikan interpretasi yang baik terhadap pengaruh variabel prediktor pada variabel respon ordinal. Oleh

karena itu, regresi logistik ordinal dipilih dalam penelitian ini untuk memodelkan status mutu air sungai yang terdiri atas kategori memenuhi baku mutu, cemaran ringan, cemaran sedang, dan cemaran berat.

Meskipun memiliki performa yang baik, regresi logistik ordinal masih menghadapi masalah dalam hal ketidakseimbangan kelas, di mana jumlah pengamatan pada beberapa kategori tertentu jauh lebih sedikit dibandingkan kategori lainnya (Irawan, 2025). Kondisi ini umum terjadi, misalnya ketika kategori air dengan kualitas sangat buruk atau sangat baik jumlahnya lebih sedikit dibandingkan kategori sedang atau baik. Untuk mengatasi permasalahan ketidakseimbangan kelas, terdapat dua pendekatan utama yang umum digunakan, yaitu *undersampling* dan *oversampling*.

Undersampling dilakukan dengan mengurangi jumlah data pada kelas mayoritas sehingga distribusi data menjadi lebih seimbang (He dkk., 2008). Namun, pendekatan ini berpotensi menghilangkan informasi penting karena sebagian data mayoritas dibuang. Sementara itu, *oversampling* dilakukan dengan menambah jumlah data pada kelas minoritas tanpa mengurangi data asli, sehingga seluruh informasi tetap terjaga dan lebih sesuai digunakan pada dataset dengan ukuran terbatas (Chawla dkk., 2002). Salah satu metode *oversampling* yang banyak digunakan adalah *Synthetic Minority Over-sampling Technique* (SMOTE) dan *Adaptive Synthetic Sampling* (ADASYN) (Hiola dkk., 2025).

SMOTE bekerja dengan menghasilkan data sintetis pada kelas minoritas melalui proses interpolasi antara suatu observasi dengan beberapa tetangga terdekatnya dalam ruang fitur (Nurdian dkk., 2022). Pendekatan ini memungkinkan penambahan jumlah data pada kelas minoritas tanpa melakukan duplikasi data secara langsung, sehingga struktur distribusi data tetap terjaga dan risiko *overfitting* dapat diminimalkan (Chawla dkk., 2002b). Bertambahnya jumlah observasi kelas minoritas membuat distribusi kelas menjadi lebih seimbang. Kondisi tersebut memberikan kesempatan yang lebih besar bagi model regresi logistik untuk mempelajari pola hubungan antara variabel penjelas dan setiap kategori status mutu air sungai, termasuk kategori dengan jumlah observasi yang sebelumnya sedikit.

Kemampuan SMOTE dalam membantu regresi logistik mempelajari kelas minoritas tersebut telah ditunjukkan dalam beberapa penelitian. Tokandari dkk. (2025) menunjukkan bahwa ketidakseimbangan kelas menyebabkan model cenderung bias terhadap kelas mayoritas sehingga kemampuan dalam mengenali kelas minoritas menjadi rendah. Setelah dilakukan penyeimbangan data menggunakan SMOTE, kinerja model meningkat, terutama pada metrik *Recall* dan *balanced Accuracy*. Pohan dkk. (2023), yang menunjukkan bahwa penerapan SMOTE meningkatkan sensitivitas regresi logistik terhadap kelas minoritas. Secara keseluruhan, kedua penelitian tersebut menunjukkan bahwa regresi logistik tetap relevan digunakan pada permasalahan klasifikasi dengan data tidak seimbang setelah dilakukan penyeimbangan distribusi kelas menggunakan SMOTE.

Adaptive Synthetic Sampling (ADASYN) merupakan metode *oversampling* yang dikembangkan untuk mengatasi ketidakseimbangan data dengan menghasilkan data sintetis pada kelas minoritas secara adaptif. Metode ini bekerja dengan memberikan bobot lebih besar pada observasi minoritas yang sulit diklasifikasikan, khususnya yang berada di sekitar batas keputusan kelas (*decision boundary*). He dkk. (2008) menunjukkan bahwa ADASYN efektif dalam meningkatkan kemampuan model dalam mengenali kelas minoritas.

Kelebihan ADASYN dibanding SMOTE adalah kemampuannya secara adaptif memfokuskan sintesis data baru pada titik-titik data minoritas yang sulit diklasifikasikan. Berbeda dengan SMOTE yang memperlakukan seluruh data minoritas secara relatif sama, ADASYN menghasilkan lebih banyak data sintetis pada area data yang dianggap sulit dipelajari oleh model, yaitu observasi yang berada di sekitar batas keputusan kelas (*decision boundary*) (Gusti dkk., 2025). Pendekatan ini bertujuan untuk meningkatkan sensitivitas model terhadap pola-pola kompleks pada kelas minoritas, sehingga kemampuan klasifikasi kelayakan mutu air sungai dapat ditingkatkan secara lebih efektif.

Admassu dkk. (2024) membuktikan bahwa kombinasi ADASYN dengan regresi logistik meningkatkan sensitivitas dan performa prediksi dibandingkan tanpa *resampling*. Di Indonesia, Pamuji dkk. (2021) menunjukkan bahwa penerapan ADASYN pada regresi logistik meningkatkan nilai *G-mean*, sehingga

menghasilkan klasifikasi yang lebih seimbang. Temuan-temuan tersebut menunjukkan bahwa ADASYN merupakan pendekatan yang potensial untuk meningkatkan kinerja regresi logistik ordinal pada kasus klasifikasi dengan ketidakseimbangan kelas, seperti prediksi kelayakan mutu air sungai.

Berdasarkan uraian tersebut, penelitian ini berfokus pada penerapan teknik *Adaptive Synthetic Sampling* (ADASYN) pada model regresi logistik ordinal untuk klasifikasi mutu air sungai. Penelitian ini bertujuan untuk mengevaluasi efektivitas metode ADASYN dalam menangani ketidakseimbangan kelas guna meningkatkan kinerja klasifikasi multikelas, khususnya dalam mengenali kategori mutu air sungai yang jarang terjadi, sehingga dapat mendukung pengelolaan dan pengendalian mutu air sungai secara lebih tepat dan berkelanjutan.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana performa regresi logistik ordinal dalam mengklasifikasikan status mutu air sungai ke dalam empat kategori status mutu air sungai?
2. Bagaimana efektivitas metode *Adaptive Synthetic Sampling* (ADASYN) dalam meningkatkan performa klasifikasi regresi logistik ordinal pada data yang tidak seimbang?
3. Variabel apa saja yang berpengaruh signifikan terhadap klasifikasi status mutu air sungai berdasarkan model regresi logistik ordinal?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan yang ingin dicapai dalam penelitian ini adalah untuk:

1. Menganalisis performa model regresi logistik ordinal dalam mengklasifikasikan status mutu air sungai ke dalam empat kategori status mutu air sungai.
2. Menganalisis efektivitas metode *Adaptive Synthetic Sampling* (ADASYN) dalam meningkatkan performa model regresi logistik ordinal pada data yang tidak seimbang.

3. Mengidentifikasi variabel yang berpengaruh signifikan terhadap klasifikasi status mutu air sungai berdasarkan hasil estimasi model regresi logistik ordinal.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini, yaitu:

1. Pendugaan parameter dilakukan menggunakan *Maximum Likelihood Estimation* (MLE) dengan metode iterasi Newton-Raphson.
2. Variabel yang digunakan terbatas pada indikator kimia (Amonia, BOD, COD, DO, Nitrat) tanpa mempertimbangkan faktor eksternal lainnya.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memperkaya kajian ilmiah di bidang statistika dan data mining, khususnya terkait penerapan regresi logistik ordinal pada permasalahan klasifikasi mutu air sungai dengan data tidak seimbang. Selain itu, penelitian ini memberikan kontribusi teoretis dalam membandingkan efektivitas teknik ADASYN dalam meningkatkan kinerja model klasifikasi multikelas pada konteks lingkungan perairan di Indonesia