

BAB I

PENDAHULUAN

1.1 Latar Belakang

Air bersih merupakan kebutuhan yang sangat penting bagi kehidupan manusia dan makhluk hidup lainnya. Ketersediaan air bersih yang cukup dan berkualitas tinggi memiliki pengaruh langsung terhadap kesehatan masyarakat, dan keberlanjutan lingkungan. Dalam konteks global, akses terhadap air bersih merupakan hak asasi manusia yang harus dipenuhi, mengingat dampak serius yang ditimbulkan oleh kekurangan air bersih, seperti peningkatan penyakit dan kematian [1]. Di Indonesia, tantangan dalam penyediaan air bersih sering kali disebabkan oleh faktor geografis, iklim, dan polusi yang merusak sumber daya air. Kualitas air minum di Indonesia masih menjadi perhatian utama, terutama terkait dengan akses dan pemilihan sumber air minum oleh masyarakat. Berdasarkan hasil analisis Survei Kualitas Air Minum Rumah Tangga di Indonesia Tahun 2021, hanya sekitar 11,4% masyarakat yang menggunakan air ledeng atau air perpipaan sebagai sumber air minum utama, sementara mayoritas masih mengandalkan sarana air minum non-perpipaan seperti depot air minum, sumur galian, sumur bor atau pompa, air kemasan, serta sumber lainnya [2].

Hal ini menunjukkan bahwa masyarakat lebih banyak menggunakan air isi ulang, sumur galian, dan sumur bor dibandingkan air ledeng atau perpipaan, tidak hanya untuk konsumsi air minum tetapi juga untuk keperluan lain seperti mencuci, memasak, dan mandi. Secara regional, di wilayah Jawa dan Bali, sebanyak 15,1% masyarakat menggunakan air galon kemasan, sedangkan 27,5% mengandalkan air isi ulang sebagai sumber utama. Hal tersebut dapat terjadi karena dipengaruhi oleh faktor keterjangkauan ketersediaan sarana air minum (SAM) di lingkungan sekitar, serta kepraktisan dalam memenuhi kebutuhan air sehari-hari [2]. Kondisi ini mencerminkan tantangan dalam pemerataan akses air bersih yang aman dan berkualitas, sehingga diperlukan kajian lebih lanjut mengenai faktor-faktor yang mempengaruhi pemilihan sumber air serta dampaknya terhadap kesehatan dan kesejahteraan masyarakat secara keseluruhan.

Kualitas air yang layak untuk diminum ditentukan oleh beberapa parameter. Air yang aman untuk dikonsumsi harus memenuhi standar tertentu yang ditetapkan oleh badan kesehatan, seperti tidak mengandung zat pencemar kimia dan biologis yang berbahaya [3]. Dalam melakukan pengujian kualitas air yang akurat dan konsisten sangat penting untuk melindungi kesehatan masyarakat. Ciri-ciri air yang layak diminum dapat diidentifikasi melalui analisis beberapa indikator kualitas, seperti pH, kejernihan, dan keberadaan zat-zat berbahaya. Air dengan pH antara 6,5 hingga 8,5 dianggap ideal, sedangkan air yang mengandung polutan seperti arsenik atau mikroorganisme patogen harus dihindari [4]. Meskipun ada beberapa metode pengujian, banyak yang memerlukan waktu dan sumber daya yang signifikan. Oleh karena itu, penting untuk mengembangkan metode deteksi yang lebih cepat dan efisien untuk memastikan kualitas air yang aman untuk konsumsi.

Deteksi kualitas air yang efisien dan akurat sangat diperlukan untuk memastikan bahwa sumber air yang digunakan oleh masyarakat aman untuk dikonsumsi. Berkembangnya teknologi seperti kecerdasan buatan dengan menggunakan algoritma pembelajaran mesin dapat diterapkan untuk prediksi kelayakan air yang dapat diminum. Kecerdasan buatan memungkinkan algoritma meniru kecerdasan manusia dalam menyelesaikan tugas dan membuat kesimpulan, sementara pembelajaran mesin atau *machine learning* (ML) berfokus pada sistem yang dapat menyesuaikan perilakunya berdasarkan data baru selama tahap pelatihan [5]. Hal ini tidak hanya meningkatkan akurasi deteksi tetapi juga memungkinkan respon yang lebih cepat terhadap masalah kualitas air. Penggunaan kecerdasan buatan dalam deteksi kualitas air memiliki kemampuannya untuk memproses dan menganalisis data dalam jumlah besar. Algoritma seperti Random Forest dapat digunakan untuk mengklasifikasikan data kualitas air berdasarkan berbagai parameter yang diukur, sehingga memberikan analisa yang lebih mendalam tentang kondisi air [6]. Algoritma Random Forest merupakan salah satu metode pembelajaran mesin yang paling banyak digunakan untuk klasifikasi dengan berbagai kasus.

Beberapa penelitian sebelumnya telah menggunakan algoritma Random Forest untuk mendeteksi atau mengklasifikasi kualitas air. Penelitian [7]

menggunakan Random Forest untuk klasifikasi kelayakan air minum. Penelitian ini menggunakan dataset Water Quality and Potability. Penelitian [8] membandingkan algoritma Logistic Regression, Support Vector Machine (SVM), Random Forest Classifier, K-Nearest Neighbor(KNN), dan XGBoost Classifier pada kasus klasifikasi kualitas air minum. Hasil penelitian tersebut menunjukkan bahwa algoritma Random Forest memiliki tingkat akurasi paling tinggi sebesar 78%. Penelitian [9] mengungkapkan bahwa algoritma Random Forest dapat dioptimasi dengan menyesuaikan nilai-nilai hyperparameternya (*hyperparameter tuning*). Beberapa hyperparameter Random Forest yang umumnya dioptimasi antara lain: jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah fitur pada setiap split (*max_features*).

Beberapa penelitian terdahulu telah menggunakan berbagai teknik *hyperparameter tuning* untuk mengoptimasi algoritma Random Forest. Penelitian [10] mendeteksi kegagalan jantung menggunakan algoritma Random Forest yang dioptimasi dengan teknik Grid Search, Random Search, dan Bayesian Search. Hasil penelitiannya menunjukkan bahwa penggunaan *hyperparameter tuning* mampu meningkatkan akurasi model. Menghasilkan nilai keakurasian sebesar 77,93% menjadi 85,63% pada optimasi Random Search dan 82,94% pada optimasi Grid Search. Penelitian ini juga mengungkapkan bahwa tingkat akurasi model yang dioptimasi dengan Random Search lebih baik dibandingkan dengan Grid Search dan Bayesian Search. Namun, penelitian [11] terkait klasifikasi penyakit jantung untuk mengoptimasi Random Forest dengan Random Search dan Grid Search. Penelitian tersebut menunjukkan bahwa metode Random Search dan Grid Search memberikan peningkatan akurasi lebih dibandingkan dengan Random Forest *default*. Menghasilkan nilai keakurasian sebesar 82,5% menjadi 81,67% pada optimasi Random Search dan 83,33% pada optimasi Grid Search. Hasil perbandingan model optimasi tersebut menunjukkan bahwa Grid Search memiliki tingkat akurasi lebih tinggi dibandingkan Random Search. Hal ini terjadi karena Grid Search mampu menjelajahi seluruh ruang parameter, sehingga dapat mencoba semua kombinasi parameter yang ada dan paling optimal namun membutuhkan waktu komputasi lebih lama dibandingkan dengan Random Search [12].

Berdasarkan fenomena yang telah diuraikan, penelitian ini bertujuan untuk mengoptimasi algoritma Random Forest menggunakan Random Search dan Grid Search dalam klasifikasi kualitas air bersih. Diharapkan penelitian ini dapat memberikan manfaat dalam pengembangan ilmu pengetahuan, khususnya dalam pembuktian peningkatan akurasi algoritma Random Forest yang dioptimalkan dengan Random Search dan Grid Search dalam klasifikasi air bersih.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang disusun adalah sebagai berikut :

1. Bagaimana mengoptimasi algoritma Random Forest dengan Random Search untuk klasifikasi kualitas air bersih ?
2. Bagaimana mengoptimasi algoritma Random Forest dengan Grid Search untuk klasifikasi kualitas air bersih ?
3. Bagaimana perbandingan hasil evaluasi optimasi algoritma Random Forest dengan Random Search dan Grid Search untuk klasifikasi kualitas air bersih ?

1.3 Batasan Masalah

Diterapkan batasan masalah dalam penelitian ini untuk menghindari penyimpangan yang dibahas dalam upaya mencapai kesimpulan yang diinginkan sebagai berikut :

1. Penelitian ini terbatas pada klasifikasi data tentang kelayakan air bersih untuk potensi dapat diminum dan tidak dapat diminum.
2. Dataset yang digunakan pada penelitian ini yaitu “Water Quality and Potability” yang didapatkan melalui laman Kaggle dengan jumlah dataset sebanyak 3276 data.
3. Dataset memiliki 9 fitur : pH, *hardness*, *solids*, *chloramines*, *sulfate*, *conductivity*, *organic carbon*, *trihalomethanes*, dan *turbidity* dan 1 target : potabilitas.

4. Pendekatan yang digunakan pada penelitian ini adalah algoritma Random Forest dengan optimasi Random Search dan Grid Search.
5. Pengimplementasian algoritma dilakukan menggunakan bahasa pemrograman Python dengan Google Colab sebagai *code editor*.
6. Evaluasi model dilakukan dengan *Confusion Matrix* untuk menghitung akurasi dari setiap pemodelan.

1.4 Tujuan Penelitian

Berangkat dari rumusan masalah yang telah dikemukakan, tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan optimasi algoritma Random Forest dengan Random Search untuk klasifikasi kualitas air bersih.
2. Menerapkan optimasi algoritma Random Forest dengan Grid Search untuk klasifikasi kualitas air bersih.
3. Mengetahui perbandingan hasil evaluasi optimasi algoritma Random Forest dengan Random Search dan Grid Search untuk klasifikasi kualitas air bersih.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi teoretis dalam pengembangan ilmu pengetahuan, khususnya dalam pembuktian empiris efektivitas algoritma Random Forest yang dioptimalkan dengan Random Search dan Grid Search dalam klasifikasi air bersih.
2. Penelitian ini diharapkan dapat berkontribusi secara praktis dengan memperluas wawasan dan meningkatkan kemampuan peneliti, serta dapat menjadi referensi bagi penelitian berikutnya.