

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah

Air memiliki peranan yang sangat penting bagi kehidupan makhluk hidup, khususnya manusia, karena digunakan dalam berbagai aktivitas sehari-hari seperti minum, memasak, dan menjaga kebersihan. Seiring dengan pertumbuhan populasi dan perkembangan zaman, kebutuhan air bersih terus meningkat secara signifikan [1][2]. Menurut *World Health Organization* (WHO), kebutuhan air per kapita bervariasi berdasarkan tingkat perkembangan suatu negara, yaitu sekitar 60–120 liter per orang per hari di negara maju dan 30–60 liter per orang per hari di negara berkembang. Salah satu aspek terpenting dalam pemanfaatan air adalah pemenuhan kebutuhan konsumsi, sehingga air minum yang digunakan harus memenuhi standar kesehatan untuk mencegah risiko penyakit [3]. Namun, peningkatan kebutuhan air ini sering kali tidak diimbangi dengan pengawasan kualitas yang memadai, sehingga masih ditemukan penyedia air minum yang mengabaikan standar kualitas air layak konsumsi.

Berdasarkan Peraturan Menteri Kesehatan (PERMENKES) Republik Indonesia No. 492 Tahun 2010 tentang Persyaratan Kualitas Air Minum, air minum didefinisikan sebagai air yang telah melalui proses pengolahan atau tanpa pengolahan yang memenuhi persyaratan kesehatan dan dapat langsung diminum [4]. Regulasi ini menetapkan bahwa kualitas air minum harus memenuhi parameter fisika, kimia, mikrobiologi, dan radioaktivitas. Secara fisika, air minum harus tidak berwarna, tidak berbau, tidak berasa, dengan tingkat kekeruhan maksimum 5 NTU dan total padatan terlarut (*Total Dissolved Solids/TDS*) tidak melebihi 500 mg/L. Dari sisi kimia, air minum harus memiliki pH dalam rentang 6,5–8,5 serta kandungan zat kimia berbahaya seperti nitrat, nitrit, besi, mangan, dan logam berat berada di bawah ambang batas yang telah ditetapkan. Sementara itu, secara mikrobiologi, air minum dinyatakan layak apabila tidak mengandung bakteri *Escherichia coli* dan total koliform per 100 mL sampel air [4].

Kualitas air minum ditentukan oleh berbagai parameter fisika, kimia, dan biologi. Parameter fisika meliputi suhu, warna, kekeruhan, serta kandungan padatan terlarut (TDS) dan padatan tersuspensi (TSS), yang dapat mengindikasikan tingkat pencemaran air [5]. Parameter kimia mencakup pH, oksigen terlarut (DO), kesadahan, nitrat, nitrit, amonia, sulfat, serta kandungan logam berat yang dapat berdampak langsung terhadap kesehatan manusia apabila melebihi ambang batas yang ditetapkan. Parameter biologi berkaitan dengan keberadaan mikroorganisme patogen yang berpotensi menyebabkan penyakit. Oleh karena itu, analisis kualitas air minum memerlukan pendekatan yang mampu mengolah banyak parameter secara bersamaan dan mengakomodasi hubungan yang bersifat kompleks serta non-linear antarparameter [6].

Dalam konteks analisis kelayakan air minum, pendekatan berbasis *machine learning* dinilai lebih unggul dibandingkan metode manual atau klasifikasi berbasis aturan sederhana (*if-else*) [7]. Metode *machine learning* mampu memproses berbagai parameter kualitas air secara simultan, mengidentifikasi pola tersembunyi, serta menghasilkan keputusan yang lebih objektif dan konsisten. Kemampuan ini menjadikan *machine learning* sangat relevan untuk analisis kualitas air minum, terutama pada data yang bersifat kompleks dan berdimensi tinggi [8].

Beberapa algoritma *machine learning* yang banyak digunakan dalam analisis kualitas air minum adalah *Artificial Neural Network* (ANN) dan *Random Forest*. ANN memiliki keunggulan dalam memodelkan hubungan non-linear antarparameter, toleransi terhadap kesalahan data, serta kemampuan adaptasi yang tinggi tanpa memerlukan pemahaman awal terhadap hubungan input dan output. Penelitian oleh Dwi Hartanti dkk. menunjukkan bahwa ANN menghasilkan akurasi tertinggi sebesar 69,54% dibandingkan algoritma lain seperti *Decision Tree*, *Logistic Regression*, dan *Support Vector Machine* dalam klasifikasi kualitas air minum [9][10].

Sementara itu, algoritma *Random Forest* bekerja dengan membangun sejumlah *decision tree* dan menggabungkan hasilnya melalui mekanisme *majority voting*, sehingga mampu meningkatkan akurasi dan mengurangi risiko *overfitting*.

Selain itu, *Random Forest* juga mampu menangani data dengan banyak variabel serta memberikan informasi mengenai tingkat kepentingan masing-masing parameter dalam proses klasifikasi kualitas air [11]. Keunggulan ini menjadikan *Random Forest* sebagai metode yang efektif dan interpretatif dalam analisis kelayakan air minum.

Berdasarkan uraian tersebut, penelitian ini memfokuskan analisis kualitas air layak minum menggunakan metode *Artificial Neural Network* dan *Random Forest* sebagai pendekatan *machine learning* yang andal. Pemanfaatan kedua metode ini diharapkan mampu meningkatkan akurasi klasifikasi kelayakan air minum serta mendukung upaya pemantauan kualitas air secara objektif dan berbasis data guna menjaga kesehatan masyarakat.

1.2 Penelitian Terdahulu

Bagian ini memuat penelitian-penelitian atau kajian riset terdahulu oleh peneliti-peneliti maupun pihak lainnya. Dalam tahap ini, referensi penelitian telah disajikan secara ringkas sebagai langkah untuk memperkuat alasan di balik pelaksanaan penelitian ini yang tertuang pada Tabel 1.1.

Tabel 1.1 Tabel Referensi.

No	Nama Peneliti	Tahun	Judul
1.	Dwi Hartanti, dkk.	2023	"Komparasi Algoritma <i>Machine Learning</i> dalam Identifikasi Kualitas Air"
2.	M Rizky Fajar, dkk.	2024	"Perbandingan Algoritma Random Forest dan Artificial Neural Network untuk Dataset Water Potability"
3.	Agus Mulyanto, dkk.	2022	"Penerapan Artificial Neural Network Pada Prototyping Sistem Monitoring Kualitas Air Di Kota Balikpapan Untuk Mendukung Balikpapan Sebagai Smart City"

4.	Khairul Abdi, dkk	2023	"Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum"
5.	Hmoud Mosleh	2021	"Modelling and Prediction of Water Quality by Using Artificial Intelligence"

Penelitian yang dilakukan oleh Dwi Hartanti dkk. [9] dengan judul "Komparasi Algoritma *Machine Learning* dalam Identifikasi Kualitas Air" bertujuan untuk mengklasifikasikan kualitas air minum menggunakan empat algoritma *machine learning*, yaitu *Decision Tree*, *Logistic Regression*, *Support Vector Machine (SVM)*, dan *Artificial Neural Network (ANN)*. Air merupakan sumber kehidupan yang sangat penting bagi manusia, sehingga diperlukan metode klasifikasi yang akurat untuk memastikan kelayakan air agar aman dikonsumsi. Berdasarkan hasil pengujian yang telah dilakukan, diperoleh tingkat akurasi masing-masing algoritma, yaitu *Decision Tree* sebesar 60,19%, *Logistic Regression* sebesar 62,80%, SVM sebesar 68,59%, dan ANN sebesar 69,54%. Hasil tersebut menunjukkan bahwa algoritma ANN memiliki performa terbaik dibandingkan algoritma lainnya dalam mengklasifikasikan kualitas air minum, sehingga ANN dinilai lebih andal dalam membantu memastikan standar kelayakan air bagi konsumsi manusia.

M. Rizky Fajar dkk. dalam penelitiannya yang berjudul "Perbandingan Algoritma *Random Forest* dan *Artificial Neural Network* untuk Dataset *Water Potability*" [10] membahas klasifikasi kualitas air minum dengan membandingkan dua algoritma *machine learning*, yaitu *Random Forest* dan *Artificial Neural Network (ANN)*. Proses klasifikasi dilakukan berdasarkan beberapa parameter utama, meliputi pH, *hardness*, *solids*, *chloramines*, *sulfate*, *conductivity*, *organic carbon*, *trihalomethanes*, *turbidity*, dan *potability*. Berdasarkan hasil pengujian yang telah dilakukan, algoritma *Random Forest* memperoleh akurasi sebesar 67,82%, sedangkan *Artificial Neural Network (ANN)* mencapai akurasi sebesar 61,01%. Hasil tersebut menunjukkan bahwa algoritma *Random Forest* memiliki performa yang lebih baik dibandingkan ANN dalam mengklasifikasikan kelayakan air minum.

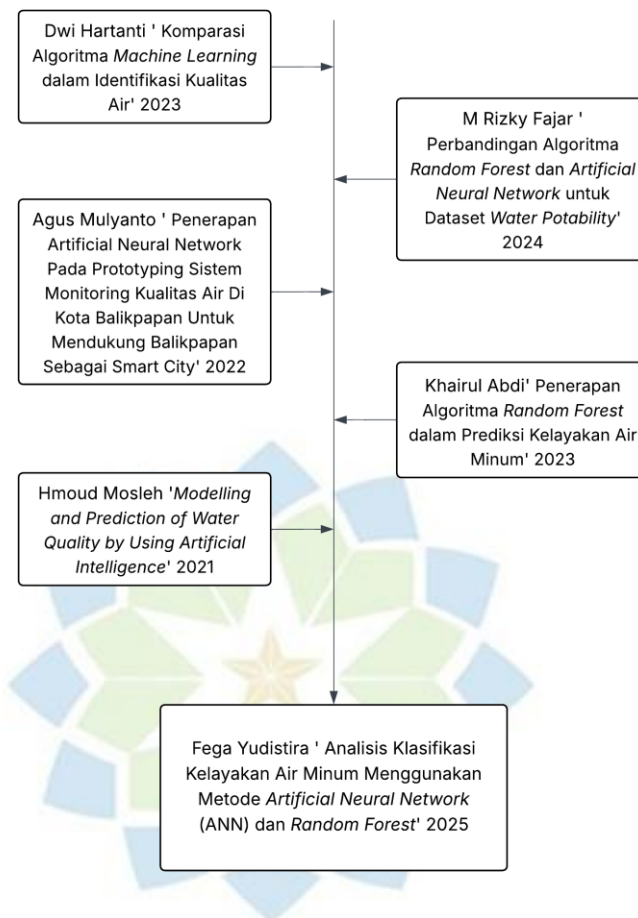
Penelitian yang dilakukan oleh Agus Mulyanto dkk. dengan judul “Penerapan *Artificial Neural Network* pada Prototyping Sistem Monitoring Kualitas Air di Kota Balikpapan untuk Mendukung Balikpapan sebagai *Smart City*” [12] membahas secara komprehensif penerapan *Artificial Neural Networks* (ANN) dalam prediksi kualitas air. Penerapan ANN ini memiliki peran penting dalam pemantauan lingkungan, keberlanjutan ekosistem, serta mendukung kegiatan budidaya perairan (*aquaculture*). Pendekatan tradisional dalam pemantauan kualitas air dinilai kurang mampu menangkap karakteristik data yang bersifat nonlinier dan *non-stationary*. Oleh karena itu, ANN menjadi pendekatan yang menjanjikan karena kemampuannya dalam memodelkan pola data yang kompleks dan dinamis secara lebih akurat.

Penelitian yang dilakukan oleh Khairul Abdi dkk. dengan judul “Penerapan Algoritma *Random Forest* dalam Prediksi Kelayakan Air Minum” [12] menerapkan algoritma *Random Forest* untuk mengklasifikasikan kualitas air serta memprediksi kelayakan air minum. Dengan menggunakan teknik *data mining*, khususnya algoritma *Random Forest*, penelitian ini bertujuan untuk mengidentifikasi pola-pola kompleks dalam data serta menentukan faktor-faktor yang memengaruhi kualitas air. Penelitian ini menggunakan dataset *Water Quality* yang diperoleh dari Kaggle sebagai sumber data. Berdasarkan hasil pengujian, model *Random Forest* mencapai tingkat akurasi sebesar 69%, yang menunjukkan kemampuan model dalam melakukan klasifikasi kelayakan air minum secara cukup baik. Selain itu, analisis *Feature Importance Score* digunakan untuk mengungkap kontribusi relatif dari setiap fitur terhadap proses prediksi. Evaluasi performa model juga dilakukan menggunakan kurva ROC yang menunjukkan kinerja klasifikasi model secara optimal, sedangkan *Confusion Matrix* digunakan untuk memberikan gambaran lebih rinci mengenai hasil klasifikasi. Hasil *Confusion Matrix* menunjukkan bahwa model *Random Forest Classifier* berhasil mengklasifikasikan sebanyak 370 instance sebagai “Potabel” dan 84 instance sebagai “Tidak Potabel” secara benar. Namun demikian, masih terdapat 160 instance yang salah diklasifikasikan sebagai “Potabel” serta 42

instance yang salah diklasifikasikan sebagai “Tidak Potabel”.

Penelitian yang dilakukan oleh Hmoud Mosleh dengan judul “*Modelling and Prediction of Water Quality by Using Artificial Intelligence*” [13] mengem-bangkan pendekatan berbasis kecerdasan buatan untuk memprediksi dan mengk-lasifikasikan kualitas air guna mendukung pengelolaan air minum yang berkelan-jutan. Dalam penelitian ini, algoritma *Adaptive Neuro-Fuzzy Inference System* (ANFIS) digunakan untuk memprediksi *Water Quality Index* (WQI), sedangkan algoritma *Feed-Forward Neural Network* (FFNN) dan *K-Nearest Neighbors* (KNN) diterapkan untuk mengklasifikasikan kualitas air. Dataset yang digunakan terdiri dari delapan parameter penting yang berkaitan dengan kualitas air, namun hanya tujuh parameter yang dipilih berdasarkan tingkat signifikansi statistiknya. Berdasarkan parameter-parameter tersebut, metodologi yang diusulkan dibangun untuk memodelkan dan mengklasifikasikan kualitas air secara akurat. Hasil penelitian menunjukkan bahwa model ANFIS memiliki performa yang unggul dalam memprediksi nilai WQI dengan koefisien regresi sebesar 96,17% pada tahap pengujian. Sementara itu, model FFNN mencapai tingkat akurasi tertinggi sebesar 100% dalam klasifikasi kualitas air (*Water Quality Classification*) serta menun-jukkan tingkat *robustness* yang lebih tinggi dibandingkan model lainnya dalam proses klasifikasi.

Berbagai penelitian dilakukan dalam bidang klasifikasi dengan menggunakan pendekatan *machine learning*. Terdapat beberapa perbedaan yaitu metode *machine learning* yang digunakan. Penelitian ini secara khusus menargetkan klasifikasi kelayakan air minum, ini bertujuan untuk memberikan hasil yang lebih relevan terhadap kondisi lokal. Selain itu, dibandingkan dengan penelitian terdahulu yang menguji banyak algoritma sekaligus, penelitian ini akan secara khusus memband-ingkan dua algoritma yaitu ANN dan Random Forest, untuk mengetahui kelayakan air di masyarakat. Adapun urutan penelitian terdahulu ditujukan oleh Gambar 1.1



Gambar 1.1 Hubungan Penelitian.

1.3 Rumusan Masalah

Dari latar belakang masalah yang sudah diuraikan maka rumusan masalah yang perlu dirumuskan adalah sebagai berikut :

1. Bagaimana perancangan dan implementasi *Artificial Neural Network* dan *Random Forest* untuk mengklasifikasikan kelayakan air minum?
2. Bagaimana kinerja algoritma *Artificial Neural Network* dan *Random Forest* dalam mengklasifikasi kelayakan air minum?

1.4 Tujuan

Dari latar belakang dan rumusan masalah maka tujuan dari penelitian ini adalah sebagai berikut :

1. Merancang dan mengimplementasikan algoritma Artificial Neural Network dan Random Forest untuk klasifikasi air layak minum.
2. Mengetahui kinerja algoritma Artificial Neural Network dan random Forest untuk klasifikasi air layak minum

1.5 Manfaat

Pada penelitian ini terdapat dua manfaat yang ingin di capai yaitu:

1. Manfaat Akademis

Adapun manfaat akademis yang didapatkan dari penelitian ini diharapkan dapat berkontribusi dalam memperkaya khasanah keilmuan pada bidang Machine Learning.

2. Manfaat Praktis

Manfaat praktis yang dihasilkan dari penelitian ini diharapkan dapat membantu dalam menentukan kelayakan air minum berdasarkan klasi-fikasi kualitas air. Hasil klasifikasi dapat digunakan oleh masyarakat untuk mengetahui kualitas air yang mereka konsumsi. Selain itu, instansi kesehatan dan pemerintah dapat memanfaatkan penelitian ini dalam pengawasan kualitas air serta pengambilan keputusan terkait perbaikan standar air minum.

1.6 Batasan Masalah

Batasan yang berhubungan dengan masalah ini sangat luas, maka dari itu perlu adanya batasan maslaah dalam penelitian ini, agar yang akan didapat lebih spesifik dan terarah. Batasan masalah ini menitik beratkan pada:

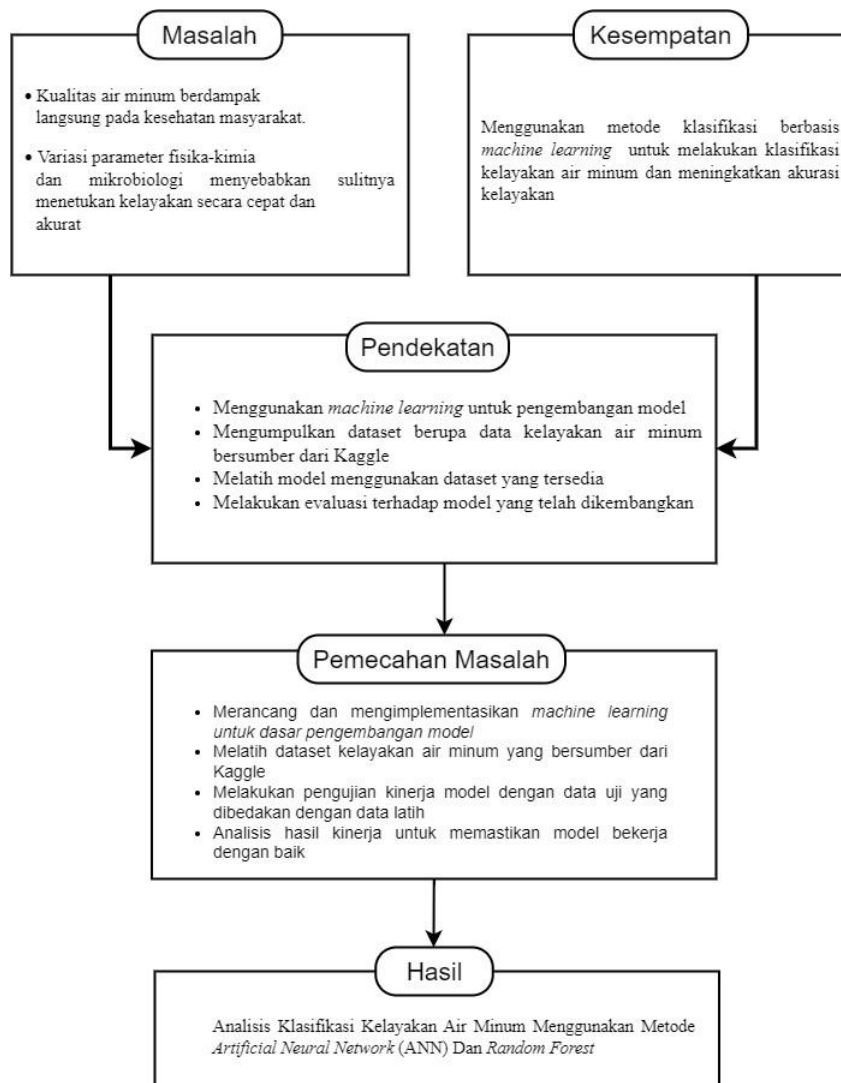
1. Dataset berasal dari Kaggle dengan total data latih sebanyak 3.276 data dengan format csv.
2. Dataset dalam penelitian ini terdiri dari sampel air yang berasal dari beberapa sumber, seperti air sumur dan air sulingan.
3. Analisis kualitas air dibatasi pada parameter fisika dan kimia yang tersedia dalam dataset dan mengacu pada standar PERMENKES

No. 492 Tahun 2010.

4. Parameter kualitas air yang digunakan dalam penelitian ini terdiri dari beberapa faktor seperti pH, tingkat kekeruhan, kandungan zat padat terlarut (TDS), kesadahan (*Hardness*), kloramin, sulfat, konduktivitas, total organic carbon (TOC), trihalomethanes (THM)s, dan turbiditas. Parameter tersebut dipilih karena merupakan indikator utama kualitas air menurut standar WHO dan US EPA, mudah diukur, serta langsung mencerminkan kondisi fisika-kimia air yang paling berpengaruh terhadap kesehatan dan kelayakan air minum.
5. Pelatihan dan pengujian model dilakukan menggunakan dataset yang tersedia, tanpa adanya pengambilan sampel air secara berkala untuk validasi hasil klasifikasi dalam kondisi nyata.
6. Algoritma yang digunakan adalah *Artificial Neural Network* (ANN) dan *Random forest*.

1.7 Kerangka Berpikir

Kerangka pemikiran adalah sebuah rancangan yang digagas oleh peneliti yang merupakan penjelasan sementara terhadap gejala yang menjadi objek dalam penelitian. Adapun kerangka pemikiran pada penelitian kali ini dapat dilihat pada Gambar 1.2



Gambar 1.2 Kerangka Berpikir.

1.8 Sistematika Penulisan

Sistematika Penulisan tugas akhir ini terdiri dari enam bab berikut penjabarannya:

BAB I PENDAHULUAN

Pada bab ini menjelaskan latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kerangka pemikiran, metodologi penelitian serta sistematika penulisan.

BAB II TEORI DASAR

Pada bab ini menguraikan tentang teori-teori yang dipakai untuk memperkuat pernyataan yang berhubungan dengan judul skripsi yang terhubung dengan teori yang telah ada. Hal lainnya yang dijelaskan pada bab ini yakni mengenai teori

teori pendukung untuk dipakai penyusunan skripsi ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan metode dan tahapan-tahapan yang dilakukan ketika melakukan penelitian dan jadwal penelitian Analisis Klasifikasi Kelayakan Air Minum Menggunakan Metode Artificial Neural Network (ANN) Dan Random Forest

BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI

Bab ini membahas perancangan sistem dan alur kerja analisis kualitas air minum berbasis *machine learning*. Tahapan yang dijelaskan meliputi pengolahan dataset kualitas air, pemilihan parameter berdasarkan standar PERMENKES No. 492 Tahun 2010, serta implementasi model klasifikasi menggunakan algoritma *Artificial Neural Network* dan *Random Forest*.

BAB V PENGUJIAN DAN ANALISIS

Bab ini menyajikan hasil pengujian dan analisis kinerja model dalam mengklasifikasikan kelayakan air minum. Evaluasi dilakukan dengan membandingkan performa algoritma *Artificial Neural Network* dan *Random Forest* berdasarkan hasil prediksi dan metrik evaluasi yang digunakan.

BAB VI KESIMPULAN DAN SARANN

Bab ini berisi kesimpulan dari hasil perancangan dan pengujian sistem analisis kualitas air minum. Selain itu, disampaikan saran untuk pengembangan penelitian selanjutnya guna meningkatkan akurasi dan cakupan analisis kualitas air.