

Sistem Prediksi Kualitas Air Konsumsi Machine Learning Menggunakan Algoritma Random Forest

**Fredyo Eltanin Lumban Raja¹, Daniel Purba², Syifa Nur Rakhmah³, Imam Sutoyo⁴,
Findi Ayu Sariasih⁵**

^{1,2,4,5}Program Studi Teknik Informatika, Universitas Bina Sarana Informatika

³Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika

*fredyolumbanraja1033@gmail.com

Abstrak

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem prediksi kualitas air bersih berbasis machine learning dengan algoritma Random Forest. Latar belakang penelitian ini adalah terbatasnya akses publik terhadap informasi cepat mengenai kelayakan air, sementara pengujian laboratorium membutuhkan waktu dan biaya yang signifikan. Data yang digunakan merupakan data sintetis yang disusun berdasarkan rentang nilai dan ambang batas kualitas air pada SNI 3553:2015, SNI 3553:2023, dan Peraturan Menteri Kesehatan No. 2 Tahun 2023, sehingga tetap representatif dan selaras dengan kondisi nyata. Dataset ini dibuat karena data lapangan sulit diperoleh secara lengkap dan dalam bentuk yang terstandarisasi, namun tetap meniru variasi kondisi nyata sesuai standar nasional, dengan label kelayakan yang ditentukan berdasarkan regulasi resmi. Pengembangan sistem menggunakan metode Agile melalui tahapan pembuatan dataset, preprocessing, pelatihan, evaluasi, dan implementasi aplikasi. Model Random Forest digunakan untuk mengklasifikasikan air menjadi kategori layak, cukup layak, dan tidak layak. Hasil pengujian menunjukkan bahwa model dengan tiga parameter fisik mencapai akurasi 96,67%, sedangkan model dengan sepuluh parameter kimia mencapai akurasi 100%, yang menegaskan bahwa penambahan parameter dapat meningkatkan akurasi prediksi. Sistem ini diharapkan dapat membantu masyarakat dan petugas lingkungan dalam melakukan penilaian awal kualitas air secara cepat sebelum pengujian laboratorium dan masih dapat dikembangkan lebih lanjut agar lebih aplikatif di berbagai wilayah.

Kata kunci : Air bersih, Prediksi kualitas air, Random Forest, Machine Learning, SNI

Abstract

This study aims to design and implement a clean water quality prediction system based on machine learning using the Random Forest algorithm. The background of this research is the limited public access to fast information regarding water feasibility, while laboratory testing requires significant time and cost. The data used is synthetic data constructed based on the value ranges and threshold limits of water quality in SNI 3553:2015, SNI 3553:2023, and the Ministry of Health Regulation No. 2 of 2023, so that it remains representative and aligned with real conditions. This dataset was created because field data is difficult to obtain completely and in a standardized form, but it still imitates real condition variations according to national standards, with feasibility labels determined based on official regulations. The system development uses the Agile method through the stages of dataset creation, preprocessing, training, evaluation, and application implementation. The Random Forest model is used to classify water into suitable, moderately suitable, and unsuitable categories. The test results show that the model with three physical parameters achieved an accuracy of 96.67%, while the model with ten chemical parameters achieved an accuracy of 100%, confirming that adding more parameters can improve prediction accuracy. This system is expected to help the public and environmental officers in conducting an initial assessment of water quality quickly before laboratory testing and can still be further developed to become more applicable in various regions.

Keywords : Clean water, Water quality prediction, Random Forest, Machine Learning, SNI.

1. Pendahuluan

Air bersih merupakan kebutuhan yang sangat penting bagi kehidupan manusia. aktivitas manusia seperti minum, memasak , dan kegiatan rumah tangga lainnya bergantung pada ketersediaan air konsumsi yang aman dan layak. Air yang tidak memenuhi standar kualitas dapat menimbulkan gangguan kesehatan, sehingga pemantauan kualitas air menjadi hal yang penting untuk menjaga keselamatan masyarakat [1].

Namun, di berbagai daerah kualitas air konsumsi masih menjadi permasalahan serius akibat pencemaran lingkungan, limbah industri, serta kondisi alam seperti tanah gambut yang memiliki kandungan logam berat serta curah hujan tinggi yang memengaruhi komposisi kimia air. Kualitas air yang buruk dapat menyebabkan penyakit seperti diare, keracunan logam berat, dan infeksi saluran pencernaan. Pemeriksaan laboratorium memang memberikan hasil akurat, tetapi prosesnya membutuhkan waktu, biaya, dan tenaga ahli, sehingga masyarakat sering terlambat mengetahui kelayakan air yang mereka konsumsi [1].

Dengan kemajuan teknologi, perkembangan machine learning dalam bidang lingkungan telah memberikan solusi inovatif terhadap berbagai permasalahan, termasuk pengelolaan kualitas air, prediksi cuaca ekstrem, pengendalian polusi udara, hingga deteksi pencemaran laut. Teknologi ini mampu menganalisis data

lingkungan dalam jumlah besar dengan cepat dan akurat, sehingga membantu pengambilan keputusan lebih efektif dan efisien. Berbagai penelitian telah membuktikan penerapan algoritma machine learning efektif untuk memprediksi dan mengklasifikasikan kualitas air konsumsi. Hadirnya teknologi sekarang dapat membantu mempercepat hasil proses dari kualitas air yang dapat dikonsumsi maupun tidak. Kualitas air dapat digambarkan dalam bentuk parameter yang terdiri dari isi kandungan zat yang ada di dalamnya, dari parameter atau variabel tersebut dapat digunakan untuk memprediksi kondisi air menggunakan metode klasifikasi data dengan memanfaatkan machine learning [2].

Beberapa penelitian sebelumnya telah menggunakan algoritma random forest untuk mengklasifikasikan kualitas air dan terbukti menghasilkan performa yang baik [3], [4]. Namun, penelitian tersebut masih memiliki keterbatasan. Sebagian besar penelitian tersebut hanya berfokus pada pembangunan model tanpa mengembangkan sistem atau aplikasi yang dapat digunakan langsung oleh masyarakat maupun petugas lapangan. Selain itu, beberapa penelitian menggunakan dataset luar negeri atau dataset yang tidak mengikuti standar nasional Indonesia, sehingga kurang relevan untuk kondisi lokal. Berdasarkan kekurangan tersebut, penelitian ini memiliki pendekatan baru dengan merancang

aplikasi prediksi kualitas air berbasis Random Forest yang memungkinkan masyarakat melaporkan kondisi air dan petugas memasukkan data seperti pH, suhu, serta kekeruhan. Sistem ini tidak menggantikan uji laboratorium, tetapi berfungsi sebagai alat bantu pendekripsi awal untuk memberikan informasi cepat mengenai kelayakan air konsumsi menggunakan data simulasi yang disesuaikan dengan standar kualitas air SNI.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada pengembangan sistem prediksi kualitas air bersih berbasis machine learning menggunakan algoritma Random Forest. Sistem ini diharapkan dapat membantu masyarakat dan petugas lingkungan dalam memantau kondisi air tanah atau air konsumsi secara real-time. Petugas dapat memasukkan data hasil pengukuran parameter air, sementara masyarakat dapat melaporkan kondisi air seperti warna, bau, atau rasa. Sistem kemudian akan menampilkan hasil prediksi serta rekomendasi tindakan pencegahan, sehingga kualitas air konsumsi tetap terjaga dan risiko penyakit akibat air tercemar dapat diminimalkan.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Tinjauan penelitian terdahulu menjadi landasan penting dalam memperkuat dasar teori dan kerangka analisis penelitian ini. Penulis mengutip

beberapa penelitian sebelumnya yang relevan sebagai acuan dalam memperdalam dan mendukung pengembangan sistem prediksi kualitas air bersih berbasis machine learning menggunakan algoritma Random Forest.

- Gafur dkk. Dalam penelitiannya mengenai Pemanfaatan Sumber Air Bersih Yang Sehat Bagi Masyarakat Di Desa Pucak Kec. Tompobulu, Kab. Maros, menjelaskan bahwa air bersih memiliki peran penting dalam menjaga kesehatan masyarakat, karena kualitas air yang buruk dapat menjadi sumber penularan berbagai penyakit seperti diare, kolera, dan disentri [5].

- Yuliani. Dalam penelitiannya Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst, terbukti menjadi metode paling efektif untuk memprediksi kelangsungan hidup pasien gagal jantung. Model ini menunjukkan kinerja yang jauh lebih unggul dibandingkan algoritma lainnya [6].

- Verdiyanto dkk. Dalam penelitiannya mengenai Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest, algoritma ini menunjukkan kinerja yang unggul dalam menangani pola data penjualan yang fluktuatif dan kompleks, sehingga mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi dan konsisten [7].

- Kelin dkk. Dalam penelitiannya Analisis Kualitas Air PAM Layak Minum dengan Metode Random Forest dan Decision Tree, membandingkan beberapa algoritma machine learning untuk memprediksi kelayakan air minum berdasarkan parameter kimia seperti pH, kandungan Klorin, padatan terlarut, dan zat lainnya. Dari hasil pengujian, Random Forest memberikan akurasi terbaik [8].

- Abdi dkk. Dalam penelitiannya Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum, Hasil penelitian menunjukkan model mencapai akurasi sebesar 69%, dengan fitur paling berpengaruh terhadap kualitas air yaitu pH, sulfate, dan hardness. Penelitian ini membuktikan bahwa metode *machine learning* dapat digunakan untuk menganalisis dan memprediksi kualitas air secara efektif [9].

2.2. Landasan Teori

1. Pengertian air bersih

air bersih merupakan kebutuhan dasar manusia yang sangat memengaruhi kualitas hidup. Penelitian ini menekankan bahwa air yang digunakan untuk konsumsi harus bebas dari pencemaran fisik dan kimia agar aman dikonsumsi[10].

2. Algoritma random forest

Random Forest merupakan salah satu metode yang dapat meningkatkan hasil akurasi dalam

membangkitkan atribut untuk setiap node yang dilakukan secara acak. Random Forest terdiri dari sekumpulan decision tree, dimana kumpulan pohon keputusan ini digunakan untuk mengklasifikasi data ke suatu kelas. Pohon keputusan dibuat dengan menentukan node akar dan berakhir dengan beberapa node daun untuk mendapatkan hasil akhir [11].

Random Forest adalah algoritma untuk supervised learning yang bisa digunakan untuk klasifikasi maupun regresi. Algoritma ini sangat fleksibel dan mudah digunakan. Kelebihan Random Forest adalah mampu mengatasi noise dan missing value serta mengatasi data yang besar, kekurangan Random Forest adalah interpretasi yang sulit dan membutuhkan tuning model yang tepat untuk data [12].

3. Standar air konsumsi berdasarkan SNI dan Permenkes

Standar kualitas air di Indonesia diatur dalam SNI 3553:2015 dan 3553:2023 yang diterbitkan oleh BSN serta Permenkes No. 2 Tahun 2023 oleh Kementerian Kesehatan RI. Kedua regulasi tersebut menetapkan parameter fisik dan kimia air seperti pH, TDS, kekeruhan, serta kadar logam berat (Fe dan Mn), dll. Standar ini menjadi acuan dalam penelitian ini untuk menilai tingkat kelayakan kualitas air berdasarkan hasil prediksi model machine learning [13], [14].

Gambar berikut menunjukkan nilai standar kualitas air :

Parameter	Satuan	SNI	Permenkes 2/2023	Keterangan
pH	N/A	6,0 - 8,5	6,0 - 8,5	Tidak berubah dari SNI 3553:2015
TDS	mg/L	maks 500	maks 500	Sama antara SNI dan Permenkes
Kekeruhan	NTU	maks 1,5	maks 1,5	Sama antara SNI dan Permenkes
Suhu	°C	± 3°C	± 3°C	Sama, tetap dari SNI 2015
Besi (Fe)	mg/L	maks 0,1	maks 0,2	Nilai SNI lebih ketat dibanding Permenkes
Mangan (Mn)	mg/L	maks 0,05	maks 0,1	Nilai SNI lebih ketat
Nitrat	mg/L	maks 20	maks 20	Sama antara SNI dan Permenkes
Nitrit	mg/L	maks 0,1	maks 3	Permenkes lebih longgar (beda signifikan)
Klorida	mg/L	maks 250	N/A	Tidak ada di Permenkes, hanya di SNI
Sulfat	mg/L	maks 200	N/A	Tidak ada di Permenkes, hanya di SNI

Gambar 1. Standar air menurut SNI 2015, 2023, dan Permenkes NO2 Tahun 2023

Arti setiap parameter :

1. Ph Menunjukkan tingkat keasaman atau kebasaan air.
2. TDS Menunjukkan jumlah zat padat terlarut (seperti garam, mineral, logam).
3. Kekeruhan Mengukur seberapa keruh air akibat partikel tersuspensi.
4. Suhu Mempengaruhi reaksi kimia dan biologi di dalam air.
5. Besi(Fe) Biasanya berasal dari korosi pipa atau tanah.
6. Mangan(Mn) Logam alami dalam tanah.
7. Nitrat Berasal dari pupuk atau limbah domestik.
8. Nitrit Bentuk peralihan antara amonia dan nitrat.
9. Klorida Ditemukan secara alami atau dari limbah industri dan air laut.
10. Sulfat Ditemukan secara alami dalam tanah dan bebatuan.

4. Penggunaan data sintetis

Data sintetis adalah data buatan yang dibuat untuk meniru karakteristik statistik data nyata tanpa harus berasal dari hasil pengukuran langsung. penggunaan data sintetis terbukti valid dan efektif untuk penelitian ilmiah , termasuk prediksi kualitas air. Data sintetis mampu merepresentasikan kondisi nyata saat data lapangan terbatas, dan data buatan yang kami jadikan sebagai dataset untuk penelitian ini juga mengikuti standar SNI dan Permenkes. Dengan Algoritma Random Forest yang mampu menangani variabel kompleks dan menghasilkan akurasi tinggi, pendekatan ini dinilai sah dan andal untuk menghasilkan prediksi kualitas air yang representatif dan ilmiah [15], [16].

no	Bau	Warna	Rasa	pH	TDS	Kekeruhan	Suhu	Besi (Fe)	Mangan (Mn)	Nitrat	Klorida	Sulfat	Kelayakan
1	3	3	3	7.25	248.1	0.05	28.2	0.124	0.06	7.05	0.0	23.75	Cukup Layak
2	3	3	3	6.99	97.0	1.2	24.9	0.165	0.025	0.0	16.74	8.27	Cukup Layak
3	3	3	3	7.31	258.3	3.63	22.5	0.0	0.0	3.53	0.364	12.87	21.64 Cukup Layak
4	2	2	2	7.66	302.0	1.1	23.7	0.096	0.0	6.79	0.428	31.32	10.16 Tidak Layak
5	3	3	3	6.96	217.2	2.05	27.1	0.03	0.067	3.27	0.0	11.15	4.65 Cukup Layak
6	2	2	2	6.96	348.9	2.59	26.4	0.063	0.101	4.69	0.882	18.28	19.07 Tidak Layak
7	4	4	4	7.68	110.4	0.31	24.4	0.098	0.008	5.45	0.593	23.22	23.04 Layak
8	5	5	5	7.36	68.0	0.29	24.1	0.031	0.063	5.51	0.0	4.95	32.12 Cukup Layak
9	4	4	4	6.86	20.0	1.13	24.0	0.016	0.0	6.31	0.0	21.84	5.6 Cukup Layak
10	4	4	4	7.27	314.6	0.64	22.7	0.073	0.011	0.0	0.163	9.56	18.89 Tidak Layak

Gambar2. Contoh data sintetis untuk prediksi kualitas air konsumsi

Parameter bau, warna, dan rasa menggunakan skala 1 hingga 5, di mana angka 1 menunjukkan kualitas yang buruk sedangkan angka 5 kualitasnya bagus. Semakin tinggi nilainya, maka semakin bagus kualitas air tersebut. Sementara itu parameter kimia yang digunakan meliputi pH, TDS, kekeruhan, suhu, besi(Fe), mangan(Mn), nitrat, nitrit, klorida, dan sulfat. Hasil akhir dari

penilaian ini dikategorikan kedalam tiga tingkat kelayakan air yaitu layak, cukup layak, dan tidak layak.

2.3. Tahapan penelitian.

Tahapan penelitian diawali dengan pengumpulan dataset kualitas air melalui identifikasi permasalahan dan penyusunan data sintetis yang mengacu pada standar SNI 3553:2015, 2023, serta Permenkes No. 2 Tahun 2023. Selanjutnya, dilakukan pembersihan data dan normalisasi untuk memastikan konsistensi format sebelum masuk ke tahap pelatihan model Random Forest. Setelah model terbentuk, dilakukan evaluasi model dengan mengukur tingkat akurasi untuk menguji performa prediksi. Model yang telah diuji kemudian masuk ke tahap implementasi sistem prediksi ke dalam aplikasi, dan diakhiri dengan visualisasi hasil agar pengguna dapat melihat status kelayakan air dan mudah dipahami.



Gambar 3. Alur penelitian

3. Metode Penelitian

3.1. Metode Pengumpulan Data

Metode penelitian yang digunakan adalah Agile, yang menekankan pada progress pengembangan sistem secara iteratif, adaptif, dan fleksibel. Metode ini memungkinkan peneliti untuk melakukan pengujian dan perbaikan model secara berulang sampai hasil yang diperoleh mencapai tingkat akurasi yang optimal.

1. Pengumpulan data (Preparation data)

Tahap ini merupakan langkah awal penelitian yang dilaksanakan di Universitas Bina Sarana Informatika. Karena keterbatasan ketersediaan data lapangan yang terstruktur, teknik pengumpulan data dilakukan melalui dua pendekatan :

- Studi pustaka : Peneliti mengumpulkan data sekunder berupa parameter baku mutu air yang mengacu pada standar nasional, yaitu SNI 3553 : 2015, 2023, dan Permenkes No.2 Tahun 2023.
- Data sintetis : Berdasarkan parameter standar tersebut, dilakukan pembuatan data buatan yang merepresentasikan tiga kategori kualitas air (Layak, CukupLayak, TidakLayak). Data disusun mengikuti batas mutu yang ketat agar tetap representatif terhadap kondisi nyata.

2. Pembersihan data (Preprocessing data). Pada tahap ini dilakukan pembersihan data dari duplikasi dan kesalahan parameter fisik warna, bau, dan rasa diubah ke dalam skala penilaian 1

– 5, di mana nilai yang lebih tinggi menunjukkan kualitas air yang lebih baik. Kemudian dilakukan normalisasi agar nilai setiap parameter memiliki rentang yang seimbang.

3. Pelatihan model Random Forest

Pada tahap ini, dataset dibagi menjadi data latih (training) dan data uji (testing) dengan rasio pembagian tertentu (misalnya 80 : 20). Analisis dilakukan menggunakan algoritma Random Forest , yaitu metode ensemble learning yang membangun model prediksi dari kumpulan pohon keputusan (Decision Tree).

Prinsip kerja algoritma ini didasarkan pada teknik bootstrap aggregating (bagging), di mana setiap pohon dilatih menggunakan sampel data acak. Penentuan keputusan akhir (Klasifikasi layak, cukup layak, tidak layak) dilakukan melalui voting mayoritas dari seluruh pohon yang terbentuk. Pendekatan ini dipilih karena terbukti mampu menghasilkan akurasi yang lebih tinggi dan stabil dibandingkan pohon keputusan tunggal dan akan disesuaikan ulang apabila hasil evaluasi belum mencapai target akurasi yang ditentukan.

4. Evaluasi model

Setelah model selesai dilatih, dilakukan evaluasi untuk mengukur tingkat akurasi model dalam melakukan prediksi. Beberapa metrik yang umum digunakan antara lain Accuracy, Precision, Recall, dan F1-Score, yang diperoleh dari hasil perbandingan antara data uji dan prediksi model. Contohnya rumus akurasi:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP = True Positive, TN = True Negative,

FP = False Positive, dan FN = False Negative.

5. Implementasi sistem prediksi.

Tahap ini merupakan proses penerapan model yang sudah dilatih ke dalam bentuk sistem aplikasi prediksi. Dalam sistem ini, pengguna dapat memasukkan nilai parameter air seperti pH, TDS, suhu, dan sebagainya. Model Random Forest kemudian memproses data tersebut dan memberikan hasil prediksi tingkat kelayakan air (Layak, Cukup Layak, atau Tidak Layak).

6. Visualisasi hasil

Tahap terakhir adalah menyajikan hasil prediksi dan evaluasi model dalam bentuk visual agar lebih mudah dipahami. Visualisasi dapat berupa grafik akurasi, diagram perbandingan hasil prediksi dengan data aktual, atau tampilan hasil prediksi dalam antarmuka aplikasi. Visualisasi ini membantu dalam menganalisis performa sistem dan melihat pola kualitas air berdasarkan parameter-parameter yang diuji.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Hasil prediksi dan model.

Model prediksi kualitas air dibangun menggunakan algoritma Random Forest. Proses pembentukan model diawali dengan pengumpulan dan pembersihan data, kemudian

dibagi menjadi 80% data latih dan 20% data uji.

Model dilatih untuk mengenali tiga kategori air, yaitu Cukup Layak, Layak, Tidak Layak.

Terdapat dua jenis model yaitu :

- Model petugas menggunakan 11 fitur yang berasal dari parameter kimia air.
- Model masyarakat menggunakan 3 fitur yang bersumber dari penilaian subjektif masyarakat terhadap air, yaitu bau, warna, dan rasa.

Dalam penelitian ini terdapat dua model yaitu :

1. Hasil prediksi model masyarakat.

Hasil evaluasi model petugas dapat dilihat pada Gambar 4.1. Model ini menunjukkan kinerja sempurna dengan nilai akurasi sebesar 100%, serta nilai precision, recall, dan f1-score sebesar 1.00 pada seluruh kelas (Cukup Layak, Layak, dan Tidak Layak).

--- Model MASYARAKAT: Pelatihan (3 Fitur) ---

AKURASI MASYARAKAT (20% Test): 0.9667				
	precision	recall	f1-score	support
Cukup layak	0.90	1.00	0.95	71
Layak	1.00	0.89	0.94	75
Tidak layak	1.00	1.00	1.00	94
accuracy				1.00
macro avg	0.97	0.96	0.96	240
weighted avg	0.97	0.97	0.97	240

Gambar 4. Model masyarakat.

Hasil evaluasi model petugas dapat dilihat pada Gambar 4.1. Model ini menunjukkan kinerja sempurna dengan nilai akurasi sebesar 100%, serta nilai precision, recall, dan f1-score sebesar 1.00 pada seluruh kelas (Cukup Layak, Layak, dan Tidak Layak).

2. Hasil perediksi model petugas

Hasil pengujian Model Masyarakat ditampilkan pada Gambar 4.2. Model ini memperoleh nilai akurasi sebesar 96,67%, dengan precision dan recall berada pada rentang 0,89–1,00.

--- Model PETUGAS: Pelatihan (11 Fitur) ---

AKURASI PETUGAS (20% Test): 1.0000				
	precision	recall	f1-score	support
Cukup layak	1.00	1.00	1.00	71
Layak	1.00	1.00	1.00	75
Tidak layak	1.00	1.00	1.00	94
accuracy				1.00
macro avg	1.00	1.00	1.00	240
weighted avg	1.00	1.00	1.00	240

Gambar 5. Model petugas.

2. Implementasi aplikasi

Aplikasi yang dikembangkan memiliki dua mode utama, yaitu mode petugas dan mode masyarakat, dengan tampilan antarmuka yang sederhana dan mudah digunakan.

1. Halaman utama



Gambar 6. Tampilan mode pengguna.

Menampilkan pilihan mode pengguna (petugas atau masyarakat). Pengguna dapat memilih mode sesuai kebutuhan. Mode masyarakat direkomendasikan bagi pengguna umum,

sedangkan mode petugas diperuntukkan bagi pengguna yang memiliki alat ukur kualitas air.

2. Halaman input petugas



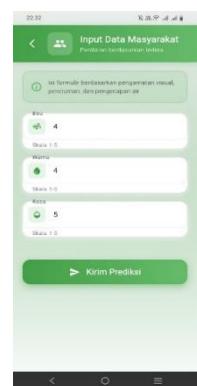
Gambar 7. Tampilan input gambar.

Petugas memasukkan nilai parameter kimia seperti pH, TDS, suhu, dan nitrit. Aplikasi kemudian menampilkan hasil prediksi.



Gambar 8. Tampilan hasil prediksi petugas.

3. Halaman input masyarakat



Gambar 9. Tampilan input masyarakat.

Pengguna cukup memberikan penilaian subjektif terhadap air dengan skala 1–5. Nilai tersebut kemudian diproses oleh model Random Forest untuk menghasilkan prediksi kelayakan air konsumsi.



Gambar 10. Tampilan hasil prediksi masyarakat.

4.2 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, algoritma Random Forest terbukti sangat efektif dalam mengklasifikasikan kelayakan air konsumsi kerena kemampuannya yang signifikan antara dua skenario pengujian. Model untuk masyarakat, yang hanya menggunakan tiga parameter fisik sederhana (bau, rasa, dan warna), berhasil mencapai akurasi sebesar 96,67%. Meskipun fiturnya terbatas, hasil ini menunjukkan bahwa indikator fisik sudah cukup bagus untuk deteksi. Sementara itu, model untuk petugas yang memanfaatkan sepuluh parameter lengkap (termasuk parameter kimia seperti pH, TDS, suhu dll) menghasilkan performa sempurna dengan akurasi 100%. Hal ini

mengonfirmasi bahwa semakin lengkap variabel data yang digunakan sesuai standar SNI 3553:2015,2023, dan Permenkes No.2 Tahun 2023 maka kemampuan model dalam memprediksi kategori kelayakan (layak, cukup layak, tidak layak) akan semakin presisi.

Dampak dari implementasi sistem ini sangat signifikan bagi efisiensi pemantauan lingkungan. Bagi masyarakat, aplikasi ini berfungsi sebagai alat deteksi awal yang cepat dan tanpa harus menunggu uji laboratorium yang memakan waktu. Bagi petugas lapangan, sistem ini mempercepat proses validasi data kualitas air di berbagai titik lokasi secara real-time. Dengan demikian, aplikasi ini tidak hanya menawarkan solusi teknis, tetapi juga memberikan dampak edukatif dalam meningkatkan solusi teknis, tetapi pentingnya standar kualitas air layak konsumsi.

5. Kesimpulan

Penelitian ini menyimpulkan bahwa implementasi algoritma random forest berbasis standar nasional (SNI dan Permenkes) berhasil menghasilkan model prediksi kualitas air yang akurat, sebagai alat deteksi dini bagi masyarakat maupun instrumen validasi presisi bagi petugas. Sistem ini hadir sebagai solusi praktis untuk mengatasi keterbatasan akses uji laboratorium konvesional. Guna meningkatkan efisiensi dan jangkuan pemantauan di masa depan, sistem ini berpotensi dikembangkan melalui integrasi

teknologi Internet of Things (IoT) untuk pengambilan data real-time, serta penambahan fitur GIS dan notifikasi otomatis untuk memetakan persebaran anomali kualitas air secara lebih luas..

6. Daftar Pustaka

- [1] E. Alfin, Rahmatulloh, and M. Suendarti, "sentri, Infrastruktur Air dan Tantangan Di Indonesia (2)," *SENTRI J. Ris. Ilm.*, vol. 1, no. 2, pp. 382–391, 2022.
- [2] M. M. Mutoffar, M. Naseer, and A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan," vol. 04, no. 02, pp. 138–146, 2022.
- [3] D. A. Dewi, A. S. Wei, L. C. Lin, and C. D. Heng, "Water Quality Prediction using Random Forest Algorithm and Optimization," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1354–1362, 2024, doi: 10.47738/jads.v5i3.348.
- [4] S. Kelin, M. Ampak, A. Eka, I. K. Arya, and G. Wiguna, "Analisis Kualitas Air PAM Layak Minum dengan Metode Random Forest dan Decision Tree," vol. 3, pp. 925–932, 2025.
- [5] N. S. Abd Gafur, Wardiah Hamzah, "Penerbit: Pusat Kajian dan Pengelola Jurnal Fakultas Kesehatan Masyarakat UMI. Sumber Air Bersih Yang Sehat Bagi Masyarakat Di Desa Pucak Kec Tompobulu, Pemanfaatan Kesehatan Masyarakat, Prodi Kesehatan Masyarakat, Fakultas Kajian Dan Pengelola Jurnal , Wind. Community Dedication J., vol. 3, no. 1, pp. 186–195, 2022.
- [6] Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," *Infotek J. Inform. dan Teknol.*, vol. 5, no. 2, pp. 298–306, 2022, doi: 10.29408/jit.v5i2.5896.
- [7] R. Verdiyanto, D. Hartanti, and E.

- Purwanto, "Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 128–139, 2025, doi: 10.29408/jit.v8i1.28386.
- [8] U. R. Zalti, D. R. Darmakusuma, M. Ridwansyah, E. Ismanto, and E. Ismanto, "Analisis dan Prediksi Kelayakan Air Minum Menggunakan Algoritma Random Forest," vol. 15, no. 2, pp. 312–317, 2025.
- [9] K. Abdi, A. Warjaya, I. Muthmainnah, and P. H. Pahutar, "Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum," *J. Ilmu Komput. dan Inform.*, vol. 3, no. 2, pp. 81–88, 2024, doi: 10.54082/jiki.81.
- [10] L. G. I. Fatristya, Wardatun Saimah, Islamul Hadi, and Evi Aryanti, "Peran Air Bersih dan Sanitasi dalam Meningkatkan Kualitas Hidup: Tinjauan Literatur terhadap Pencapaian Tujuan SDGs 2030," *J. Pendidikan, Sains, Geol. dan Geofis. (GeoScienceEd Journal)*, vol. 6, no. 1, pp. 596–602, 2025, doi: 10.29303/goescienceed.v6i1.598.
- [11] Suci Amaliah, M. Nusrang, and A. Aswi, "Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variansium31.
- [12] N. Novianti, S. P. A. Alkadri, and I. Fakhruzi, "Klasifikasi Penyakit Hipertensi Menggunakan Metode Random Forest," *Progresif J. Ilm. Komput.*, vol. 20, no. 1, p. 380, 2024, doi: 10.35889/progresif.v20i1.1663.
- [13] K. K. R. Indonesia, "Peraturan Menteri Kesehatan Republik Indonesia Nomor 2 Tahun 2023 tentang Persyaratan Kualitas Air Minum," Jakarta, 2023
- [14] B. S. Nasional, "SNI 3553:2023 dan 2015 Air Mineral dan Air Minum dalam Kemasan," Jakarta, 2023. [Online]. Available: <https://akses-sni.bsn.go.id/home>
- [15] J. Jordon et al., "Synthetic Data -- what, why and how?," 2022, [Online]. Available: <http://arxiv.org/abs/2205.03257>
- [16] Muhammad Adha Zidane Syahputra Andry, Devito Syachputra, Hanif Zulfi Fauzan, and Yudi Prastyo, "Literatur Review : Pemanfaatan Artificial Intelligence Dalam Predictive Maintenance Untuk Meningkatkan Keandalan Industri," *J. Manag. Innov. Entrep.*, vol. 2, no. 4, pp. 2384–2395, 2025, doi: 10.70248/jmie.v2i4.2405.