

**SISTEM MONITORING KUALITAS AIR BERBASIS IOT DENGAN
PREDIKSI MACHINE LEARNING DAN SISTEM PAKAR**

Laporan Ujian Akhir Semester

Dosen Pengampu Machine Learning dan sistem pakar: Liptia Venica, S.T., M.T



Kelas 5A

**PROGRAM STUDI MEKATRONIKA DAN KECERDASAN BUATAN
KAMPUS UPI DI PURWAKARTA
UNIVERSITAS PENDIDIKAN INDONESIA**

2025

Nama-nama dan Nomor Induk Mahasiswa (NIM) Kelompok 2:

No.	Nama	NIM
1.	Ardhan Anggana Prasetya	2310051
2.	Ardita Aura Syawaliani	2307386
3.	Edgard Altamerano Ferdinand	2301912
4.	Excel Thrive Valerian Haryanto	2304906
5.	Faiz Yudiana Fathul Hakim	2300196
6.	Indra Nurdiansyah	2307436
7.	Labib Musyari Azis	2300755
8.	Muhammad Hikam Baidawih	2300798
9.	Muhammad Iqbal Ridho	2309141
10.	Muhammad Raihan Ramadhan	2304058
11.	Muthmainnah Indriani Mulyana	2301933
12.	Rafly Furqon Musharaf	2301615
13.	Riyadhil Haqqy Arifin	2311156

DAFTAR ISI

DAFTAR ISI	i
DAFTAR GAMBAR.....	iii
DAFTAR TABEL	iv
BAB I PENDAHULUAN DAN PERENCANAAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan Proyek	2
1.4 Ruang Lingkup.....	2
BAB II TINJAUAN PUSTAKA	3
2.1 Dasar Teori Machine Learning (yang digunakan)	3
2.1.1 Akuisisi Data dan Spesifikasi Sensor.....	3
2.1.2 Preprocessing Data (Data Preprocessing)	4
2.1.3 Pelatihan Model (Model Training)	4
2.2 Konsep Dasar Sistem Pakar dan Logika Fuzzy.	4
2.2.1 Variabel dan Himpunan Fuzzy	5
2.2.2 Basis Aturan (Rule Base)	5
2.2.3 Proses Fuzzifikasi, Inferensi, dan Defuzzifikasi.....	6
BAB III DATA LATIH DAN REKAYASA FITUR.....	7
3.1 Data Eksternal (Luar Sistem)	7
3.2 Feature Engineering.....	8
BAB IV MODEL MACHINE LEARNING KLASIK	11
4.1 Pemilihan Model	11
4.2 Evaluasi Model.....	11
BAB V PEMBENTUKAN KNOWLEDGE BASE	13
5.1 Metode Akuisisi Pengetahuan.....	13
5.2 Basis Aturan.....	15
BAB VI IMPLEMENTASI LOGIKA FUZZY	17
6.1 Fungsi Keanggotaan	17
6.1.1 Fungsi Keanggotaan pH	17
6.1.2 Fungsi Keanggotaan TDS.....	17
6.1.3 Fungsi Keanggotaan Turbidity	18
6.2 Mekanisme Inferensi dan Defuzzifikasi	18

6.2.1 Mekanisme Inferensi	18
6.2.2 Mekanisme Defuzzifikasi	19
6.3 Mekanisme Menyajikan Kesimpulan/Hasil Inferensi	19
BAB VII ANALISIS, PENGUJIAN, DAN KESIMPULAN.....	21
7.1 Pengujian Fitur Machine Learning dan Sistem Pakar	21
7.2 Status Integrasi	22
7.3 Kesimpulan dan Saran	23
LAMPIRAN	24
Link repositori git.....	24
Link video	24
Kontribusi Setiap Tim	24
Screenshot Activity Git	26

DAFTAR GAMBAR

Gambar 5. 1 Parameter pH berdasarkan sumber.....	13
Gambar 5. 2 Parameter TDS berdasarkan sumber	13
Gambar 5. 3 Parameter Kekerusuhan berdasarkan sumber	14
Gambar 5. 4 Dokumentasi wawancara dengan pakar	15

DAFTAR TABEL

Tabel 4. 1 Hasil Evaluasi Model	11
Tabel 5. 1 Aturan kondisi tidak layak minum (single-parameter dominant rule) .	15
Tabel 5. 2 Aturan kondisi cukup layak minum (multi-parameter evaluative rule)	15
Tabel 5. 3 Aturan kondisi layak minum (strict quality rule)	16
Tabel 6. 1 Himpunan Fuzzy dari pH	17
Tabel 6. 2 Himpunan Fuzzy dari TDS	18
Tabel 6. 3 Himpunan Fuzzy dari Turbidity	18
Tabel 7. 1 Hasil Uji Coba Model Terhadap Data Baru	21
Tabel 7. 2 Hasil Uji Coba Sistem Pakar Terhadap Data Baru	21

BAB I

PENDAHULUAN DAN PERENCANAAN

1.1 Latar Belakang

Air merupakan elemen vital dalam keberlangsungan ekosistem dan kehidupan manusia. Namun, degradasi kualitas air akibat pencemaran lingkungan kini menjadi tantangan global yang memerlukan penanganan presisi. Metode pemantauan konvensional yang mengandalkan pengambilan sampel manual dan uji laboratorium seringkali tidak efisien, memakan waktu, dan tidak mampu menyajikan data secara real-time. Dalam perspektif rekayasa modern, khususnya pada domain Mekatronika dan Kecerdasan Buatan, diperlukan pendekatan teknis yang mampu mengotomatisasi proses tersebut dengan akurasi tinggi.

Proyek ini mengusulkan pengembangan sistem cerdas pendeteksi kualitas air yang mengintegrasikan teknologi Internet of Things (IoT) dengan metode komputasi cerdas. Parameter fisis dan kimia yang dijadikan indikator utama meliputi Derajat Keasaman (pH), Kekeruhan (Turbidity), dan Padatan Terlarut Total (Total Dissolved Solids / TDS). Ketiga parameter ini dipilih karena merupakan standar baku dalam menentukan indeks kualitas air.

Urgensi dari penelitian ini terletak pada penerapan hibrida antara Machine Learning dan Sistem Pakar. Data sensor lingkungan seringkali memiliki karakteristik noisy dan ketidakpastian. Oleh karena itu, Machine Learning diterapkan untuk mempelajari pola data historis guna mengklasifikasikan kualitas air secara objektif. Di sisi lain, Sistem Pakar berbasis Logika Fuzzy (Fuzzy Logic) diimplementasikan untuk menangani ambiguitas data (seperti kondisi peralihan antara "keruh" dan "jernih") dengan meniru mekanisme penalaran seorang pakar. Sinergi kedua metode ini diharapkan dapat menghasilkan sistem monitoring yang tidak hanya otomatis, tetapi juga cerdas dalam pengambilan keputusan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, rumusan masalah spesifik yang akan diselesaikan dalam proyek ini adalah:

1. Bagaimana merancang arsitektur sistem IoT yang mampu mengakuisisi data parameter pH, Turbidity, dan TDS secara real-time dan presisi?

2. Bagaimana kinerja model Machine Learning Klasik dalam mengklasifikasikan status kualitas air berdasarkan dataset yang dikumpulkan?
3. Bagaimana merancang Sistem Pakar dengan metode Logika Fuzzy untuk memberikan inferensi keputusan mengenai kelayakan air berdasarkan basis pengetahuan (knowledge base) yang dinamis?

1.3 Tujuan Proyek

Tujuan utama yang ingin dicapai melalui pelaksanaan proyek akhir ini adalah:

1. Terbangunnya prototipe sistem pemantauan kualitas air berbasis IoT yang terintegrasi dengan sensor pH, Turbidity, dan TDS.
2. Menghasilkan model Machine Learning yang terlatih dan teruji akurasi dalam mengklasifikasikan kategori kualitas air.
3. Tersedianya sistem pakar berbasis Fuzzy Logic yang mampu memberikan rekomendasi status kualitas air (misalnya: Layak, Tercemar Ringan, atau Tercemar Berat) yang selaras dengan standar kesehatan.

1.4 Ruang Lingkup

Agar pengerjaan proyek tetap terarah dan sesuai dengan batasan waktu serta sumber daya, ruang lingkup penelitian dibatasi sebagai berikut:

1. Parameter Fisik: Variabel yang diukur terbatas pada pH, Kekeruhan (NTU), dan TDS (ppm).
2. Akuisisi Data: Menggunakan mikrokontroler berbasis IoT seperti ESP32 sebagai unit pemroses sinyal sensor.
3. Metode Machine Learning: Berfokus pada algoritma Supervised Learning kategori Klasik (Non-Deep Learning) untuk tugas klasifikasi.
4. Metode Sistem Pakar: Menggunakan pendekatan Logika Fuzzy (Metode Mamdani/Sugeno) dengan aturan IF-THEN yang diadopsi dari standar baku kualitas air.
5. Output Sistem: Sistem memberikan indikator status kualitas air tanpa menyertakan mekanisme penjernihan atau filtrasi otomatis (hanya sistem monitoring dan deteksi).

BAB II

TINJAUAN PUSTAKA

2.1 Dasar Teori Machine Learning (yang digunakan)

Machine Learning (Pembelajaran Mesin) adalah cabang dari kecerdasan buatan yang berfokus pada pengembangan sistem yang mampu belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan intervensi manusia yang minimal. Dalam konteks proyek ini, pendekatan yang digunakan adalah Supervised Learning, di mana model dilatih menggunakan data berlabel (data yang sudah memiliki status kualitas air).

Proses pembangunan model Machine Learning dalam penelitian ini didasarkan pada tiga tahapan fundamental:

2.1.1 Akuisisi Data dan Spesifikasi Sensor

Langkah awal dalam Machine Learning adalah pengumpulan data yang valid. Dalam sistem embedded dan IoT, pemahaman terhadap datasheet komponen sensor sangat krusial untuk memastikan data yang diakuisisi merepresentasikan kondisi fisik yang sebenarnya.

1. Sensor pH

Mengukur aktivitas ion hidrogen untuk menentukan tingkat keasaman atau kebasaan air. Rentang nilai standar adalah 0-14.

2. Sensor Turbidity

Mengukur kekeruhan air berdasarkan hamburan cahaya yang disebabkan oleh partikel terlarut. Satuan yang umum digunakan adalah NTU (Nephelometric Turbidity Units).

3. Sensor TDS

Mengukur total padatan terlarut (garam, mineral, logam) dalam air, biasanya dalam satuan ppm (parts per million). Pemahaman terhadap spesifikasi teknis (datasheet) masing-masing sensor diperlukan untuk menentukan rentang operasional, kalibrasi, dan linearitas data sebelum digunakan sebagai input model.

2.1.2 Preprocessing Data (Data Preprocessing)

Data mentah yang diperoleh dari sensor IoT seringkali mengandung noise, nilai yang hilang (missing values), atau perbedaan skala yang signifikan antar parameter (misalnya nilai pH berkisar 0-14, sedangkan TDS bisa mencapai 500 ppm). Oleh karena itu, preprocessing adalah tahapan wajib untuk meningkatkan kualitas data sebelum dilatih. Teknik yang digunakan meliputi:

1. Pembersihan Data: Menghapus atau memperbaiki data yang anomali (outlier) akibat kesalahan pembacaan sensor.
2. Normalisasi/Scaling: Menyamakan skala data (misalnya menggunakan Min-Max Scaler) agar fitur dengan rentang nilai besar (seperti TDS) tidak mendominasi fitur dengan rentang nilai kecil (seperti pH) selama proses pembelajaran model.

2.1.3 Pelatihan Model (Model Training)

Pelatihan model adalah proses di mana algoritma mempelajari pemetaan antara input (pH, Turbidity, TDS) dan output (Status Kualitas Air). Pada tahap ini, dataset dibagi menjadi dua bagian: Data Latih (Training Set) untuk membentuk model, dan Data Uji (Testing Set) untuk mengevaluasi kinerja model. Tujuan dari pelatihan ini adalah meminimalkan fungsi kesalahan (loss function) sehingga model dapat melakukan generalisasi dan memprediksi data baru yang belum pernah dilihat sebelumnya dengan akurasi tinggi.

2.2 Konsep Dasar Sistem Pakar dan Logika Fuzzy.

Sistem pakar merupakan sistem berbasis komputer yang dirancang untuk mengadopsi pengetahuan dan pola pengambilan keputusan seorang pakar ke dalam bentuk aturan logis. Sistem ini bekerja dengan memanfaatkan basis pengetahuan dan mesin inferensi untuk menghasilkan keputusan berdasarkan data yang diberikan. Pada laporan tugas ini, sistem pakar digunakan sebagai kerangka pengambilan keputusan dalam menentukan kelayakan air minum.

Sistem pakar yang dikembangkan menggunakan metode inferensi forward chaining. Metode ini melakukan penalaran dengan memulai proses dari fakta awal berupa data hasil pengukuran parameter kualitas air, kemudian mencocokkannya dengan aturan-aturan yang tersimpan dalam basis pengetahuan hingga diperoleh kesimpulan akhir. Forward chaining dipilih karena sesuai untuk sistem berbasis

sensor yang bekerja berdasarkan kondisi aktual dan memungkinkan evaluasi beberapa aturan secara paralel.

Untuk menangani ketidakpastian dan rentang nilai parameter kualitas air, sistem pakar ini dikombinasikan dengan logika fuzzy. Logika fuzzy memungkinkan suatu nilai input memiliki derajat keanggotaan pada lebih dari satu himpunan, sehingga lebih fleksibel dibandingkan pendekatan logika tegas. Parameter kualitas air seperti pH, Total Dissolved Solids (TDS), dan kekeruhan tidak memiliki batas klasifikasi yang sepenuhnya kaku, sehingga pendekatan fuzzy dianggap lebih sesuai.

2.2.1 Variabel dan Himpunan Fuzzy

Sistem fuzzy pada laporan ini menggunakan tiga variabel input dan satu variabel output. Variabel input terdiri dari pH, Total Dissolved Solids (TDS), dan kekeruhan (turbidity). Variabel pH dibagi ke dalam lima himpunan fuzzy, yaitu asam, sedikit asam, netral, sedikit basa, dan basa. Variabel TDS dan kekeruhan masing-masing dibagi ke dalam lima himpunan fuzzy, yaitu excellent, good, fair, poor, dan unacceptable.

Variabel output adalah status kelayakan air minum yang dibagi ke dalam tiga himpunan fuzzy, yaitu tidak layak minum, cukup layak minum (perlu waspada), dan layak minum.

2.2.2 Basis Aturan (Rule Base)

Basis aturan merupakan representasi pengetahuan pakar yang digunakan sebagai dasar pengambilan keputusan. Aturan disusun dalam bentuk IF–THEN dengan mengombinasikan kondisi dari variabel input untuk menghasilkan status kelayakan air minum.

Secara umum, aturan fuzzy pada sistem ini dikelompokkan ke dalam tiga kategori utama, yaitu aturan yang menghasilkan keputusan tidak layak minum, aturan yang menghasilkan keputusan cukup layak minum, dan aturan yang menghasilkan keputusan layak minum. Contoh struktur aturan fuzzy yang digunakan adalah sebagai berikut:

**IF pH Netral AND TDS Good AND Turbidity Good
THEN Status Kelayakan = Layak Minum**

Aturan-aturan tersebut dievaluasi secara paralel menggunakan operator logika fuzzy, yaitu operator MIN untuk operasi AND dan operator MAX untuk operasi OR.

2.2.3 Proses Fuzzifikasi, Inferensi, dan Defuzzifikasi

Proses fuzzifikasi dilakukan dengan mengubah nilai input tegas (crisp) menjadi derajat keanggotaan pada masing-masing himpunan fuzzy berdasarkan fungsi keanggotaan yang telah ditentukan. Nilai input dapat memiliki lebih dari satu derajat keanggotaan sesuai dengan karakteristik logika fuzzy.

Tahap inferensi dilakukan dengan mengevaluasi seluruh aturan fuzzy yang relevan berdasarkan nilai keanggotaan hasil fuzzifikasi. Setiap aturan menghasilkan nilai keluaran fuzzy yang mencerminkan tingkat aktivasi aturan tersebut.

Tahap akhir adalah defuzzifikasi, yaitu proses mengubah keluaran fuzzy menjadi nilai tegas yang merepresentasikan status kelayakan air minum. Nilai hasil defuzzifikasi kemudian digunakan untuk menentukan kategori kelayakan air minum sesuai dengan rentang nilai yang telah ditetapkan.

Dengan pendekatan ini, sistem pakar berbasis logika fuzzy yang digunakan pada laporan tugas ini mampu memberikan keputusan kelayakan air minum secara sistematis, terstruktur, dan mudah dianalisis.

BAB III

DATA LATIH DAN REKAYASA FITUR

3.1 Data Eksternal (Luar Sistem)

Penelitian ini mengintegrasikan dua sumber data utama untuk membangun model prediksi kelayakan air minum yang *robust* dan adaptif. Sumber pertama berasal dari *dataset Water Potability* yang diperoleh melalui repositori *Kaggle*, yang menyediakan data historis komprehensif mencakup berbagai *parameter* kualitas air dari ribuan sampel yang telah diuji di berbagai lokasi geografis.

Dataset ini dipilih karena memiliki variasi kondisi air yang luas dengan representasi beragam skenario kualitas air mulai dari air bersih hingga air terkontaminasi dan telah tervalidasi melalui pengujian laboratorium standar, sehingga mampu memberikan *baseline* yang kuat dan *ground truth* yang *reliable* untuk proses pembelajaran mesin. Sumber kedua adalah data sensor real-time yang diakuisisi langsung menggunakan sensor pH (mengukur tingkat keasaman/alkalinitas), sensor turbidity (mengukur kejernihan air), dan sensor TDS atau Total Dissolved Solids (mengukur konsentrasi padatan terlarut).

Pembacaan sensor dilakukan secara kontinu dengan interval waktu 5 menit untuk menangkap fluktuasi dinamis kualitas air, kemudian data mentah disimpan secara otomatis ke Firebase Realtime Database melalui modul WiFi ESP32 yang terintegrasi dengan sistem IoT. Selanjutnya, data di-export menjadi file JSON dan diolah menggunakan Google Colab untuk preprocessing, cleaning, feature engineering, dan validasi guna memastikan kualitas serta konsistensi pengukuran sebelum digunakan dalam training model.

3.2 Feature Engineering

Proses feature engineering dalam penelitian ini melibatkan integrasi sistematis antara fitur-fitur yang berasal dari data sensor IoT real-time dengan fitur-fitur dari dataset eksternal Kaggle Water Potability, yang dirancang untuk menghasilkan feature space yang komprehensif dan informatif bagi model Random Forest Classifier. Tahapan pertama dimulai dengan ekstraksi dan preprocessing data IoT, dimana data mentah dari sensor pH, turbidity, dan TDS yang tersimpan dalam format JSON di Firebase diimpor ke Google Colab menggunakan Firebase Admin SDK. Data sensor kemudian dikonversi menjadi Pandas DataFrame dengan struktur kolom yang konsisten, mencakup timestamp pengukuran, nilai pH, nilai turbidity (NTU), nilai TDS (ppm). Proses cleaning dilakukan untuk menangani missing values yang timbul akibat gangguan koneksi atau malfungsi sensor, menggunakan metode forward-fill atau interpolasi linear untuk mempertahankan kontinuitas time-series data. Outlier detection diterapkan menggunakan metode Interquartile Range (IQR) atau Z-score untuk mengidentifikasi dan menangani pembacaan sensor yang anomali yaitu nilai pH di luar range fisik 0-14 atau turbidity negatif dengan pendekatan winsorization atau removal tergantung severity anomali.

Tahapan kedua adalah preprocessing dataset Kaggle, dimana dataset Water Potability yang berisi 13.276 sampel dengan 9 parameter kualitas air (pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes, Turbidity, dan Potability) dimuat dan diinspeksi. Karena penelitian ini fokus pada tiga parameter utama (pH, turbidity, TDS), dilakukan feature selection untuk mengekstrak kolom yang relevan. Parameter 'Solids' dalam dataset Kaggle yang merepresentasikan total dissolved solids disamakan konsepnya dengan TDS dari sensor IoT. Missing values pada tabel pH yang mencapai 491 ditangani menggunakan mean imputation. Label target 'Potability' (0 = tidak layak minum, 1 = layak minum) diverifikasi untuk memastikan balance atau ketidakseimbangan kelas, dengan penerapan SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi imbalanced dataset yang dapat bias model.

Tahapan ketiga adalah feature harmonization dan standardization, dimana unit pengukuran dan skala dari kedua sumber data diselaraskan. Data pH dari kedua sumber biasanya sudah dalam skala yang sama (0-14), namun turbidity dan TDS perlu dipastikan menggunakan unit konsisten NTU untuk turbidity dan ppm (mg/L) untuk TDS. Feature scaling diterapkan menggunakan StandardScaler dari scikit-learn untuk menormalisasi distribusi setiap fitur, penting karena Random Forest meskipun tidak sensitive terhadap scaling seperti SVM atau Neural Networks, tetap benefit dari normalized features terutama saat melakukan feature importance analysis. Proses harmonization juga mencakup timestamp alignment, dimana data IoT yang memiliki temporal resolution rendah perdetik untuk dipadukan dengan data Kaggle yang bersifat snapshot measurement.

Tahapan terakhir adalah data merging dan validation, dimana dataset IoT yang telah diproses digabungkan dengan dataset Kaggle menggunakan vertical concatenation (`pd.concat`) untuk memperbesar training set, dengan memastikan kolom-kolom yang digunakan konsisten antar kedua sumber. Alternatifnya, dapat dilakukan stratified split dimana 80% data Kaggle digunakan untuk training awal, 20% untuk validation, sementara data IoT seluruhnya digunakan untuk testing dan continuous evaluation simulasi deployment scenario.

Feature engineering process divalidasi melalui evaluasi performa model menggunakan metrik komprehensif dari classification report. Model Random Forest Classifier ditraining dengan fitur-fitur hasil engineering, kemudian dievaluasi menggunakan accuracy score untuk mengukur proporsi prediksi yang benar secara keseluruhan, precision untuk mengukur ketepatan prediksi positif (berapa persen air yang diprediksi layak minum benar-benar layak), recall untuk mengukur kemampuan model mendeteksi semua air layak minum (sensitivitas), F1-score sebagai harmonic mean antara precision dan recall yang memberikan balance metric, dan support yang menunjukkan jumlah sampel aktual untuk setiap kelas. Validasi tambahan dilakukan dengan confusion matrix untuk visualisasi distribusi true positive, true negative, false positive, dan false negative.

Proses validasi diperkuat melalui testing dengan file dummy yang berisi simulasi data sensor dengan berbagai skenario kondisi air mulai dari air bersih ideal (pH=7.0, Turbidity=1 NTU, TDS=150 ppm) hingga air terkontaminasi ekstrem (pH=4.5, Turbidity=50 NTU, TDS=800 ppm) untuk memverifikasi bahwa model dapat memprediksi dengan benar pada edge cases dan kondisi ekstrem. Feature importance analysis menggunakan Random Forest built-in feature_importances_ attribute dilakukan untuk memverifikasi bahwa engineered features memang kontributif dan untuk interpretability model, memastikan bahwa parameter kritis seperti pH, turbidity, dan TDS memiliki bobot prediksi yang sesuai dengan domain knowledge standar kualitas air WHO/EPA.

BAB IV

MODEL MACHINE LEARNING KLASIK

4.1 Pemilihan Model

Model machine learning yang digunakan pada laporan ini adalah Random Forest, yang termasuk ke dalam kategori machine learning klasik (non-deep learning). Pemilihan Random Forest didasarkan pada karakteristik data dan tujuan sistem yang lebih menekankan pada performa klasifikasi yang stabil serta kemudahan interpretasi hasil.

Random Forest merupakan metode ensemble learning yang menggabungkan beberapa pohon keputusan untuk menghasilkan prediksi akhir. Pendekatan ini efektif dalam menangani hubungan non-linear antar fitur serta mampu mengurangi risiko overfitting yang sering terjadi pada model pohon keputusan tunggal.

Penggunaan Random Forest dipandang lebih sesuai dibandingkan metode deep learning karena jumlah fitur yang relatif terbatas dan struktur data yang tidak memerlukan representasi kompleks. Selain itu, Random Forest memiliki keunggulan dalam efisiensi komputasi, kemudahan implementasi, serta performa yang konsisten pada berbagai kondisi data.

Berdasarkan pertimbangan tersebut, Random Forest dipilih sebagai model machine learning klasik yang digunakan dalam laporan ini karena mampu memberikan keseimbangan antara akurasi, stabilitas, dan kompleksitas model.

4.2 Evaluasi Model

Evaluasi kinerja model Random Forest dilakukan menggunakan data uji (test set) untuk menilai kemampuan generalisasi model terhadap data baru. Hasil evaluasi ditunjukkan melalui beberapa metrik utama, yaitu akurasi, presisi, recall, dan F1-score.

Tabel 4. 1 Hasil Evaluasi Model

Metrik	Nilai	Interpretasi Singkat
Accuracy	77.5%	Model mampu mengklasifikasikan sebagian besar data uji dengan benar

Metrik	Nilai	Interpretasi Singkat
Precision	81%	Prediksi positif yang dihasilkan model sebagian besar akurat
Recall	69%	Model mampu mengenali sebagian besar data positif yang sebenarnya
F1-Score	74.5%	Keseimbangan yang baik antara precision dan recall

Berdasarkan Tabel 4.1, model Random Forest menunjukkan performa yang cukup baik dan konsisten pada seluruh metrik evaluasi. Nilai akurasi yang diperoleh menunjukkan kemampuan klasifikasi yang memadai, sementara nilai precision dan recall yang relatif seimbang mengindikasikan bahwa model tidak bias terhadap salah satu kelas.

Nilai F1-score yang mendekati nilai akurasi menunjukkan bahwa model memiliki stabilitas dalam melakukan prediksi. Dengan mempertimbangkan hasil evaluasi tersebut serta kompleksitas model yang relatif rendah, Random Forest dinilai layak digunakan sebagai model machine learning klasik pada laporan ini.

BAB V

PEMBENTUKAN KNOWLEDGE BASE

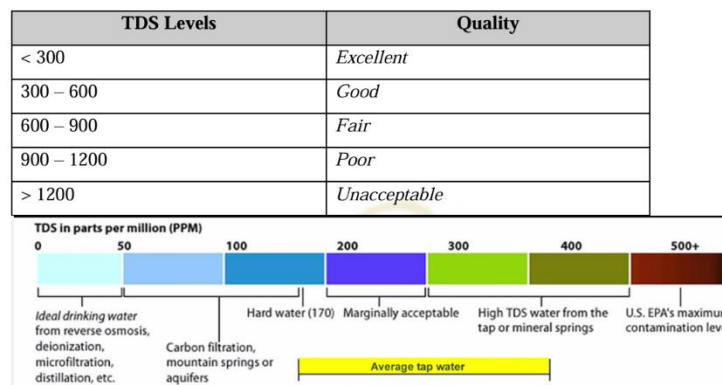
5.1 Metode Akuisisi Pengetahuan

Berikut ini merupakan dokumentasi hasil studi literatur dari ketiga parameter input yang digunakan:

pH air secara alami adalah 7. Sedangkan pH untuk air yang aman untuk diminum berkisar antara 6,5 hingga 9,5. Angka pH yang rendah (0) adalah untuk larutan asam dan tinggi (14) adalah untuk larutan basa. Menurut WHO dan EPA sendiri nilai pH yang aman untuk dikonsumsi penggunaan sehari-hari adalah 6,5 – 8,5 pH. Dengan pH 7 adalah nilai aman dimana air tidaklah bersifat asam ataupun basa. pH lebih besar dari 8,5 atau air alkaline tidak selalu berarti 'tidak dapat dikonsumsi', hanya saja rasa dan baunya akan cenderung tidak menyenangkan. Terlebih lagi pH dengan nilai sebesar itu dapat merusak pipa dan perabotan lainnya. Sedangkan untuk pH lebih kecil dari 6,5, selain dapat merusak pipa dan perabotan rumah, karena sifatnya yang asam lebih berkemungkinan terkontaminasi oleh polusi, membuatnya tidak bisa dikonsumsi (Cirino, 2019).

Gambar 5. 1 Parameter pH berdasarkan sumber

Table 2.9. Parameter Nilai TDS Menurut WHO



Gambar 5. 2 Parameter TDS berdasarkan sumber

Sensor Turbidity ditunjukkan dengan satuan pengukuran yaitu *Nephelometric Turbidity Unit* (NTU). Berdasarkan ketentuan dari badan kesehatan dunia (WHO), batas maksimum tingkat kekeruhan air yang dinyatakan adalah kurang dari 5 NTU, dengan rekomendasi dibawah 1 NTU untuk alasan kesehatan. (Sidik & Siswanto, 2020). Adapun menurut *Environmental Protection Agency* (EPA) (Epa, 2022), tingkatan kualitas air dapat dibagi menjadi kelompok berikut:

Table 2.7. Kualitas Air Berdasarkan Level Kekeruhan

NTU Levels	Quality
0,1 – 1,0	<i>Excellent</i>
1,0 – 5,0	<i>Good</i>
5,0 – 25,0	<i>Fair</i>
25,0 – 100,0	<i>Poor</i>
> 100,0	<i>Unacceptable</i>

Gambar 5. 3 Parameter Kekeruhan berdasarkan sumber

Selain mencari studi literatur, kelompok kami juga validasi dari pakar langsung yaitu Vina Selviana dengan profesi staff bidang kesehatan puskesmas yang berlokasi di Kecamatan Pedes, kota Karawang. dengan latar belakang pendidikan D3 Kebidanan Universitas Padjadjaran.

Berdasarkan hasil wawancara, pakar menyampaikan bahwa penilaian kualitas air minum memang perlu memperhatikan parameter pH, TDS, dan kekeruhan secara bersamaan karena ketiganya saling memengaruhi tingkat keamanan air. pH ideal berada pada rentang 6,5–8,5, di mana nilai di bawah rentang tersebut bersifat asam dan berpotensi menyebabkan iritasi, sedangkan nilai di atasnya bersifat basa dan dapat menimbulkan gangguan kesehatan tertentu.

Untuk TDS, nilai yang terlalu tinggi menunjukkan banyaknya zat terlarut sehingga air disarankan menggunakan sistem filtrasi seperti Reverse Osmosis apabila melebihi batas aman. Sementara itu, kekeruhan yang tinggi menandakan adanya partikel tersuspensi atau kontaminan, sehingga air perlu disaring, diendapkan, atau bahkan dihentikan penggunaannya jika nilainya sangat tinggi.

Pakar juga menegaskan bahwa pada kondisi air yang masih cukup layak, perebusan dan pemantauan berkala sudah memadai, sedangkan pada kondisi layak minum, pemeriksaan rutin tetap diperlukan untuk menjaga kualitas air tetap sesuai standar kesehatan.



Gambar 5. 4 Dokumentasi wawancara dengan pakar

5.2 Basis Aturan

Daftar lengkap aturan IF-THEN dengan total 18 aturan dengan 3 kondisi adalah sebagai berikut:

Tabel 5. 1 Aturan kondisi tidak layak minum (single-parameter dominant rule)

Rule	pH	TDS	Turbidity	Status
R1	Asam	-	-	Tidak Layak Minum
R2	Basa	-	-	Tidak Layak Minum
R3	-	Buruk	-	Tidak Layak Minum
R4	-	Tidak Diterima	-	Tidak Layak Minum
R5	-	-	Buruk	Tidak Layak Minum
R6	-	-	Tidak Diterima	Tidak Layak Minum

Tabel 5. 2 Aturan kondisi cukup layak minum (multi-parameter evaluative rule)

Rule	pH	TDS	Turbidity	Status
R7	Sedikit Asam	Cukup	Cukup	Cukup Layak Minum
R8	Sedikit Basa	Cukup	Cukup	Cukup Layak Minum
R9	Netral	Cukup	Baik	Cukup Layak Minum

R10	Netral	Baik	Cukup	Cukup Layak Minum
R11	Sedikit Asam	Baik	Baik	Cukup Layak Minum
R12	Sedikit Basa	Baik	Baik	Cukup Layak Minum
R13	Sedikit Asam	Baik	Sempurna	Cukup Layak Minum
R14	Sedikit Basa	Baik	Sempurna	Cukup Layak Minum
R15	Sedikit Asam	Sempurna	Baik	Cukup Layak Minum
R16	Sedikit Basa	Sempurna	Baik	Cukup Layak Minum

Tabel 5. 3 Aturan kondisi layak minum (strict quality rule)

Rule	pH	TDS	Turbidity	Status
R17	Sedikit Asam	Sempurna	Sempurna	Layak Minum
R18	Sedikit Basa	Sempurna	Sempurna	Layak Minum
R19	Netral	Sempurna	Sempurna	Layak Minum
R20	Netral	Baik	Baik	Layak Minum
R21	Netral	Sempurna	Baik	Layak Minum
R22	Netral	Baik	Sempurna	Layak Minum

BAB VI

IMPLEMENTASI LOGIKA FUZZY

6.1 Fungsi Keanggotaan

Fungsi keanggotaan digunakan untuk memetakan nilai input tegas (crisp) menjadi derajat keanggotaan fuzzy dengan rentang nilai antara 0 hingga 1. Derajat keanggotaan ini menunjukkan tingkat keterwakilan suatu nilai input terhadap himpunan linguistik tertentu. Proses ini merupakan tahap awal dalam sistem inferensi fuzzy sebelum dilakukan penalaran berdasarkan aturan IF–THEN.

Pada penelitian ini, variabel input yang digunakan terdiri dari pH, Total Dissolved Solids (TDS), dan Turbidity. Masing-masing variabel dibagi ke dalam beberapa himpunan linguistik.

6.1.1 Fungsi Keanggotaan pH

Variabel pH dibagi menjadi lima himpunan linguistik, yaitu Asam, Sedikit Asam, Netral, Sedikit Basa, dan Basa. Fungsi keanggotaan pH direpresentasikan menggunakan fungsi segitiga dan trapesium. Berikut merupakan tabel himpunan Fuzzy dari pH:

Tabel 6. 1 Himpunan Fuzzy dari pH

Himpunan	Parameter (a,b,c,d)
Asam	(0, 0, 6.0, 6.5)
Sedikit Asam	(6.0, 6.5, 6.5, 7.0)
Netral	(6.5, 7.0, 7.0, 7.5)
Sedikit Basa	(7.0, 7.5, 8.5, 8.5)
Basa	(8.5, 9.0, 14, 14)

6.1.2 Fungsi Keanggotaan TDS

Variabel TDS dibagi menjadi lima himpunan linguistik, yaitu Excellent, Good, Fair, Poor, dan Unacceptable. Pembagian ini didasarkan pada rentang nilai TDS dalam satuan ppm. Setiap rentang nilai dipetakan ke dalam fungsi keanggotaan fuzzy agar dapat diproses dalam mekanisme inferensi. Berikut merupakan tabel himpunan Fuzzy dari TDS:

Tabel 6. 2 Himpunan Fuzzy dari TDS

Himpunan	Parameter (a,b,c,d)
Sempurna	(0, 0, 200, 300)
Baik	(200, 300, 500, 600)
Cukup	(500, 600, 750, 850)
Buruk	(750, 850, 950, 1000)
Tidak Diterima	(950, 1000, 1000, 1000)

6.1.3 Fungsi Keanggotaan Turbidity

Variabel Turbidity (NTU) juga dibagi menjadi lima himpunan linguistik, yaitu Excellent, Good, Fair, Poor, dan Unacceptable. Nilai Turbidity yang diperoleh dari sensor dikonversi menjadi derajat keanggotaan fuzzy berdasarkan rentang nilai yang telah ditentukan pada sumber referensi. Berikut merupakan tabel himpunan Fuzzy dari Turbidity:

Tabel 6. 3 Himpunan Fuzzy dari Turbidity

Himpunan	Parameter (a,b,c,d)
Sempurna	(0, 0, 1, 1)
Baik	(1, 1, 5, 5)
Cukup	(5, 5, 25, 25)
Buruk	(25, 25, 100, 100)
Tidak Diterima	(100, 100, 300, 300)

6.2 Mekanisme Inferensi dan Defuzzifikasi

6.2.1 Mekanisme Inferensi

Sistem yang dikembangkan menggunakan metode inferensi fuzzy Mamdani dengan evaluasi aturan secara paralel (forward chaining). Pada mekanisme ini, seluruh aturan IF–THEN dievaluasi secara bersamaan berdasarkan nilai keanggotaan hasil fuzzifikasi.

Operator logika fuzzy yang digunakan adalah:

- AND direpresentasikan dengan operator MIN
- OR direpresentasikan dengan operator MAX

Aturan pada sistem dibagi menjadi tiga kelompok, yaitu:

1. Aturan kondisi tidak layak minum (single-parameter dominant rule), di mana satu parameter ekstrem sudah cukup untuk menyatakan air tidak layak minum.
2. Aturan kondisi cukup layak minum (multi-parameter evaluative rule), yang membutuhkan kombinasi kondisi pH, TDS, dan Turbidity secara bersamaan.
3. Aturan kondisi layak minum (strict quality rule), yang hanya aktif jika seluruh parameter berada pada kondisi baik atau sangat baik.

6.2.2 Mekanisme Defuzzifikasi

Hasil inferensi berupa himpunan fuzzy output, yaitu Tidak Layak Minum, Cukup Layak Minum, dan Layak Minum. Untuk memperoleh keputusan akhir yang bersifat tegas (crisp), dilakukan proses defuzzifikasi.

Nilai hasil defuzzifikasi kemudian dibandingkan dengan batas nilai yang telah ditentukan, yaitu:

1. Nilai < 50 menunjukkan Tidak Layak Minum.
2. Nilai antara 50 hingga < 80 menunjukkan Cukup Layak Minum.
3. Nilai ≥ 80 menunjukkan Layak Minum.

Proses defuzzifikasi ini memungkinkan sistem memberikan keputusan akhir yang dapat digunakan secara praktis.

6.3 Mekanisme Menyajikan Kesimpulan/Hasil Inferensi

Mekanisme penyajian hasil inferensi dilakukan melalui beberapa tahapan, yaitu:

1. Pengambilan nilai input pH, TDS, dan Turbidity.
2. Proses fuzzifikasi untuk menentukan derajat keanggotaan setiap variabel input.
3. Evaluasi seluruh aturan IF–THEN secara paralel menggunakan mekanisme inferensi fuzzy Mamdani.
4. Penggabungan hasil inferensi dan proses defuzzifikasi untuk memperoleh nilai output tegas.
5. Penentuan status kelayakan air minum berdasarkan nilai hasil defuzzifikasi.

Hasil akhir sistem disajikan dalam bentuk status kelayakan air minum, yaitu Tidak Layak Minum, Cukup Layak Minum, atau Layak Minum. Dengan mekanisme ini, keputusan yang dihasilkan tidak hanya bersifat kategoris, tetapi juga didukung oleh proses inferensi yang sistematis dan terukur.

BAB VII

ANALISIS, PENGUJIAN, DAN KESIMPULAN

7.1 Pengujian Fitur Machine Learning dan Sistem Pakar

- Machine Learning

Berikut hasil pengujian model menggunakan data baru:

Tabel 7. 1 Hasil Uji Coba Model Terhadap Data Baru

No	pH	Solids (TDS)	Turbidity	Potability (prediksi)
1	7.5	200.0	0.5	1
2	4.5	100.0	0.8	1
3	1.0	230.0	10.0	1
4	3.0	323.233	2.1	0
5	10.0	150.22	1.6	0

- Sistem Pakar

Berikut hasil pengujian sistem pakar terhadap data baru:

Tabel 7. 2 Hasil Uji Coba Sistem Pakar Terhadap Data Baru

No	pH	Solids (TDS)	Turbidity	Rule	Keputusan
1	7.5	200.0	0.5	R18	Layak Minum
2	4.5	100.0	0.8	R1	Tidak Layak Minum
3	1.0	230.0	10.0	R1	Tidak Layak Minum
4	3.0	323.233	2.1	R1	Tidak Layak Minum
5	10.0	150.22	1.6	R2	Tidak Layak Minum

7.2 Status Integrasi

Dalam sistem yang dikembangkan, Machine Learning dan Sistem Pakar diintegrasikan untuk memberikan penilaian kualitas air secara bersamaan, di mana masing-masing memiliki bobot kontribusi yang berbeda dalam pengambilan keputusan akhir. Machine Learning diberikan porsi sebesar 25%, sedangkan Sistem Pakar memiliki porsi lebih dominan yaitu 75%. Hasil dari kedua metode tersebut kemudian digabungkan melalui proses perhitungan tertentu untuk menentukan klasifikasi akhir kualitas air, yaitu *layak minum*, *cukup layak*, atau *tidak layak untuk diminum*.

Penerapan skema ini dilatarbelakangi oleh fakta bahwa kualitas air minum sangat berkaitan erat dengan kesehatan manusia, sehingga keputusan yang dihasilkan bersifat sensitif dan tidak dapat sepenuhnya bergantung pada satu pendekatan saja. Oleh karena itu, sistem tidak hanya mengandalkan prediksi berbasis data dari Machine Learning, tetapi juga didukung oleh Sistem Pakar yang menggunakan aturan dan batasan nilai parameter berdasarkan studi literatur serta sumber referensi terpercaya.

Penetapan bobot Machine Learning sebesar 25% dilakukan karena tingkat akurasi model yang diperoleh masih berada pada nilai 77,5%, yang dinilai belum cukup tinggi untuk dijadikan satu-satunya dasar pengambilan keputusan. Dengan tingkat akurasi tersebut, prediksi Machine Learning masih berpotensi menghasilkan kesalahan klasifikasi. Oleh sebab itu, Sistem Pakar diberikan bobot yang lebih besar karena mampu melakukan penilaian berbasis pengetahuan domain dan logika aturan terhadap nilai sensor, sehingga dapat meningkatkan keandalan dan kehati-hatian dalam menentukan kelayakan air untuk dikonsumsi.

7.3 Kesimpulan dan Saran

Berdasarkan hasil perancangan dan pengujian yang telah dilakukan, dapat disimpulkan bahwa Sistem Monitoring Kualitas Air Berbasis IoT dengan Prediksi Machine Learning dan Sistem Pakar mampu berjalan dengan baik sesuai dengan tujuan penelitian. Sistem ini memanfaatkan tiga parameter utama, yaitu pH, jumlah zat padat terlarut (TDS), dan kekeruhan, yang diolah melalui model machine learning serta sistem pakar dengan total 24 aturan untuk menghasilkan kesimpulan tingkat kelayakan air minum. Hasil pengujian menunjukkan bahwa sistem mampu memberikan klasifikasi kondisi air secara logis dan konsisten, serta menyajikan rekomendasi yang sesuai dengan kondisi parameter yang terukur. Integrasi antara pemantauan berbasis IoT, prediksi machine learning, dan sistem pakar menjadikan sistem lebih informatif dan mendukung pengambilan keputusan terkait kualitas air.

Sebagai saran untuk pengembangan di masa mendatang, sistem ini masih dapat ditingkatkan dengan menambahkan parameter pendukung lain agar hasil analisis menjadi lebih akurat dan komprehensif. Salah satu parameter yang direkomendasikan adalah kadar oksigen terlarut (Dissolved Oxygen/DO), karena berpengaruh terhadap kualitas dan kelayakan air. Selain itu, penambahan parameter lain seperti suhu, kandungan logam berat, atau mikrobiologi air juga dapat dipertimbangkan. Pengembangan lebih lanjut pada model machine learning, perluasan basis aturan sistem pakar, serta peningkatan visualisasi dan akurasi data IoT diharapkan mampu menghasilkan sistem monitoring kualitas air yang lebih adaptif, andal, dan aplikatif untuk penggunaan di lingkungan nyata.

LAMPIRAN

Link repositori git

<https://github.com/raflyfurqon/Sistem-Monitoring-Kualitas-Air-untuk-Kelayakan-Minum->

Link video

https://drive.google.com/drive/folders/1FocAUR2oHP45Z2IQ-DXW_0o2txAPanTB?usp=sharing

Kontribusi Setiap Tim

No.	Nama	Kontribusi
1.	Ardhan Anggana Prasetya	<ul style="list-style-type: none">- Membuat file python penghubung nilai sensor dengan model machine learning
2.	Ardita Aura Syawaliani	<ul style="list-style-type: none">- Merangkai implementasi wiring- Menyusun dan merapikan komponen ke dalam wadah- Membuat poster
3.	Edgard Altamerano Ferdinand	<ul style="list-style-type: none">- Wiring hardware dan implmentasi wairing
4.	Excel Thrive Valerian Haryanto	<ul style="list-style-type: none">- Merangkai implmentasi wiring- Membuat laporan
5.	Faiz Yudiana Fathul Hakim	<ul style="list-style-type: none">- Merangkai implmentasi wairing- Laporan- Membuat Deskripsi Project
6.	Indra Nurdiansyah	<ul style="list-style-type: none">- Merangkai alur sistem- Membuat diagram flowchart- Merancang UI dashboard web
7.	Labib Musyari Azis	<ul style="list-style-type: none">- Membuat arsitektur sistem- Validasi data pakar

8.	Muhammad Hikam Baidawih	<ul style="list-style-type: none"> - Membuat notebook dan model machine learning - Mencari dataset - Mengkalibrasi sensor
9.	Muhammad Iqbal Ridho	<ul style="list-style-type: none"> - Merangkai komponen sensor dan wiring, serta mendesain konsep hardware dan implementasi hardware - Membuat kode program Esp32 untuk sensor turbidity, pH, dan TDS
10.	Muhammad Raihan Ramadhan	<ul style="list-style-type: none"> - Membuat UI dashboard web - Mengisi laporan bab 2 dan bab 4 - Merangkai wiring alat
11.	Muthmainnah Indriani Mulyana	<ul style="list-style-type: none"> - Merangkai implementasi wiring - Menyusun dan merapihkan komponen ke dalam wadah
12.	Rafly Furqon Musharaf	<ul style="list-style-type: none"> - Merancang sistem pakar - Kalibrasi Sensor - Membuat laporan - Budgeting dan membeli komponen
13.	Riyadhil Haqqy Arifin	<ul style="list-style-type: none"> - Menyusun dan merapihkan semua komponen ke dalam wadah - Troubleshooting komponen - Mengkalibrasi sensor-sensor

Screenshot Activity Git

