

Implementasi Support Vector Machine pada Sistem Monitoring Kualitas Air Berbasis IoT

¹Juli Sulakosno, ²Rafli Ardiansyah

^{1,2}Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: ¹jsulaksono@gmail.com, ²rafardiansyahfli@gmail.com

Penulis Korespondens : Juli Sulaksono

Abstrak—Kualitas air merupakan salah satu faktor penting yang memengaruhi kesehatan manusia dan kelestarian lingkungan. Pemantauan kualitas air secara konvensional masih memerlukan pengukuran manual yang kurang efisien untuk pengawasan berkelanjutan. Penelitian ini mengembangkan sistem monitoring kualitas air berbasis Internet of Things (IoT) yang terintegrasi dengan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi kualitas air secara realtime. Sistem menggunakan sensor pH, Total Dissolved Solids (TDS), dan turbidity yang terhubung dengan mikrokontroler ESP32. Data sensor dikirim melalui jaringan WiFi menuju server dan diproses menggunakan model SVM yang telah dioptimasi menggunakan GridSearchCV. Dataset yang digunakan terdiri dari 4.259 data dengan dua kelas yaitu normal dan tercemar. Hasil evaluasi menunjukkan bahwa model SVM memperoleh akurasi sebesar 99,65% pada data pengujian. Sistem yang dikembangkan berhasil menampilkan data sensor dan hasil klasifikasi secara realtime melalui dashboard berbasis web. Hasil penelitian menunjukkan bahwa integrasi IoT dan machine learning mampu menghasilkan sistem monitoring kualitas air yang efektif, otomatis, dan mudah diakses.

Kata Kunci— GridSearchCV, Internet of Things, Kualitas Air, Support Vector Machine, Water Monitoring

Abstract— *Water quality is an important factor affecting human health and environmental sustainability. Conventional water quality monitoring generally requires manual measurements and is less efficient for continuous monitoring. This study develops an Internet of Things (IoT)-based water quality monitoring system integrated with a Support Vector Machine (SVM) algorithm to perform real-time water quality classification. The system utilizes pH, Total Dissolved Solids (TDS), and turbidity sensors connected to an ESP32 microcontroller. Sensor data are transmitted via WiFi to a server and processed using an SVM model optimized through GridSearchCV. The dataset consists of 4,259 samples with two classes, namely normal and contaminated water. Evaluation results indicate that the optimized SVM model achieved an accuracy of 99.65% on testing data. The developed system successfully displays sensor data and classification results through a web-based dashboard in real time. The integration of IoT and machine learning provides an effective, automated, and accessible solution for water quality monitoring.*

Keywords— *GridSearchCV, Internet of Things, Support Vector Machine, Water Monitoring, Water Quality*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Air adalah kebutuhan dasar yang tidak bisa lepas dari kehidupan sehari-hari, mulai dari konsumsi, sanitasi, hingga keperluan industri. Sayangnya, kualitas air di banyak wilayah terus menurun akibat aktivitas manusia dan industrialisasi yang kian pesat [1]. Kondisi ini menuntut adanya sistem pemantauan yang tidak hanya akurat, tapi juga bisa bekerja secara berkelanjutan tanpa bergantung pada pengukuran manual.

Metode pemantauan konvensional yang masih umum digunakan saat ini punya keterbatasan yang cukup jelas prosesnya lambat, biayanya relatif mahal, dan hasilnya tidak bisa langsung diperoleh [2]. Perkembangan teknologi Internet of Things (IoT) membuka peluang untuk membangun sistem monitoring yang bekerja otomatis dan memberikan data secara realtime melalui jaringan internet [1].

Namun data dari sensor saja tidak cukup. Dibutuhkan metode klasifikasi otomatis agar sistem bisa langsung memutuskan apakah kondisi air normal atau tercemar tanpa perlu diinterpretasikan secara manual. Machine Learning hadir sebagai solusi untuk kebutuhan ini [3]. Salah satu algoritma yang sudah terbukti efektif untuk klasifikasi adalah Support Vector Machine (SVM), khususnya dengan kernel RBF yang mampu menangani data non-linear dengan baik [4].

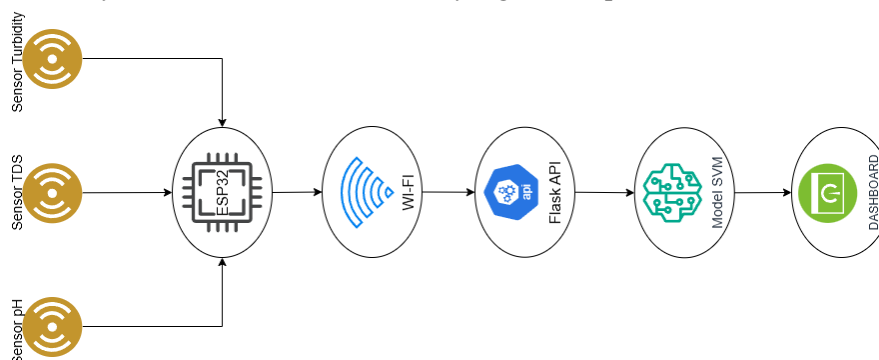
Beberapa penelitian sebelumnya sudah menggunakan SVM untuk klasifikasi kualitas air dan menunjukkan hasil yang menjanjikan [5][6]. Namun banyak juga studi yang melaporkan performa SVM kurang memuaskan dibanding algoritma lain seperti Random Forest atau XGBoost [7][8]. Hasil yang tidak konsisten ini bukan semata karena kelemahan SVM, melainkan karena kurangnya optimasi terutama dalam pemilihan kernel dan pengaturan hyperparameter yang tepat [9].

Penelitian ini mengembangkan sistem monitoring kualitas air berbasis IoT yang mengintegrasikan sensor pH, TDS, dan turbidity dengan model SVM yang dioptimasi menggunakan GridSearchCV. Berbeda dari penelitian sebelumnya yang umumnya hanya melakukan simulasi pada dataset statis, penelitian ini mengimplementasikan model langsung pada perangkat keras ESP32 untuk pengujian data sensor secara realtime, dengan hasil klasifikasi yang ditampilkan melalui dashboard berbasis web.

II. METODE

A. Arsitektur Sistem

Sistem yang dibangun terdiri dari tiga lapisan utama. Lapisan pertama adalah akuisisi data, yaitu sensor-sensor yang membaca kondisi air secara langsung. Lapisan kedua adalah pemrosesan, di mana data dikirim ke server dan diklasifikasikan oleh model SVM. Lapisan ketiga adalah visualisasi, yaitu dashboard berbasis web yang menampilkan hasil secara realtime.



Gambar 1. Arsitektur Sistem Monitoring Kualitas Air Berbasis IoT

Arsitektur sistem yang dikembangkan ditunjukkan pada Gambar 1. Data kualitas air diperoleh dari sensor pH, TDS, dan turbidity yang terhubung ke ESP32. Data kemudian dikirim melalui jaringan WiFi menuju Flask API untuk diproses menggunakan model SVM sebelum ditampilkan pada dashboard Streamlit.

B. Dataset dan Pelabelan

Dataset yang digunakan merupakan gabungan dua dataset publik dari Kaggle yang memiliki struktur fitur yang sama, menghasilkan total 4.259 sampel setelah proses pembersihan data. Dari semua parameter yang tersedia, penelitian ini hanya menggunakan tiga fitur utama yang sesuai dengan kemampuan sensor IoT, yaitu pH, TDS, dan turbidity .

Pelabelan dilakukan berdasarkan standar WHO dan Permenkes RI No. 32 Tahun 2017 [10][11]. Sampel diberi label normal jika ketiga parameter memenuhi batas standar, dan tercemar jika salah satu parameter melewati batas. Standar yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Standar Baku Air

Parameters	Unit	Quality Standard
pH	–	6,5 – 8,5 (WHO & Permenkes RI No. 32/2017)
Conductivity	μS/cm	≤ 750 μS/cm (WHO & Permenkes RI No. 32/2017)
Turbidity	NTU	≤ 5 NTU (WHO)

Distribusi kelas yang dihasilkan adalah 51% normal (2.172 sampel) dan 49% tercemar (2.087 sampel), sehingga dataset dianggap seimbang dan tidak memerlukan teknik resampling tambahan.

Model SVM dilatih menggunakan fitur pH, conductivity, dan turbidity karena menghasilkan distribusi data yang lebih seimbang. Pada implementasi sistem IoT, sensor yang digunakan menghasilkan nilai TDS. Oleh karena itu, nilai TDS dikonversi menjadi conductivity menggunakan persamaan $\text{conductivity} = \text{TDS} \times 2$ sebelum diproses oleh model SVM. Dengan pendekatan ini, fitur yang digunakan pada tahap implementasi tetap konsisten dengan fitur yang digunakan pada tahap pelatihan model.

C. Preprocessing

Preprocessing dilakukan dalam beberapa tahap. Pertama, kedua dataset digabung, baris dengan nilai kosong dihapus, dan fitur yang tidak dibutuhkan dibuang hingga hanya tersisa pH, TDS, dan turbidity. Kedua, data diberi label sesuai standar pada Tabel 1. Ketiga, data dinormalisasi menggunakan StandardScaler dengan rumus z-score:

$$z = (x - \mu) / s$$

di mana x adalah nilai asli, μ adalah rata-rata data latih, dan s adalah standar deviasinya. Proses fitting hanya dilakukan pada data latih agar tidak terjadi kebocoran data ke data uji [12]. Data kemudian dibagi dengan rasio 80/20 menggunakan random state 21.

D. Pelatihan Model dan Optimasi GridSearchCV

Model dibangun menggunakan SVM dengan kernel RBF yang dipilih karena kemampuannya menangani pola data non-linear dengan baik [4]. Optimasi hyperparameter dilakukan menggunakan GridSearchCV dengan 5-fold cross-validation [9]. Ruang pencarian yang digunakan adalah sebagai berikut:

1. C : {0,1 ; 1 ; 10 ; 100 ; 1000}
2. Gamma : {1 ; 0,1 ; 0,01 ; 0,001 ; 0,0001}
3. Kernel : {rbf, poly, sigmoid}

Kombinasi tersebut menghasilkan 75 konfigurasi yang diuji untuk setiap skenario. Metrik scoring yang digunakan adalah accuracy, karena pada dataset yang seimbang, accuracy terbukti memberikan sinyal optimasi yang stabil dan tidak mudah dieksploitasi secara matematis.

Model dievaluasi menggunakan confusion matrix yang menghasilkan empat nilai dasar: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Dari keempat nilai tersebut dihitung accuracy, precision, recall, dan F1-score.

E. Implementasi Sistem dan Dashboard

Model yang sudah dilatih disimpan dan diintegrasikan ke dalam server Flask yang berjalan di laptop. Setiap kali ESP32 mengirimkan data sensor, server langsung melakukan normalisasi dan klasifikasi, kemudian mengembalikan hasilnya. Dashboard dikembangkan menggunakan Streamlit dan menampilkan nilai pH, TDS, turbidity, status koneksi, serta hasil klasifikasi SVM secara realtime.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Sensor

Pengujian sensor dilakukan menggunakan empat jenis sampel air yang dipilih untuk mewakili variasi karakteristik yang berbeda-beda. Hasil pengukurannya ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pengukuran Sampel Air

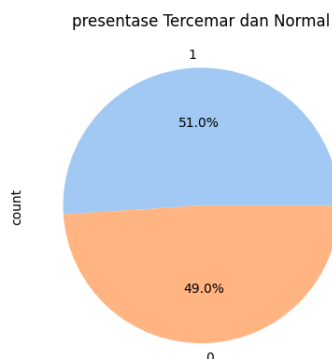
Sampel Air	pH	TDS (ppm)	Turbidity (NTU)	Prediksi
Air Bersih	7.1–8.4	183	5	Normal
Air Garam	6.1–8.4	540	5	Tercemar
Air Sabun	6.0–7.5	475	5	Tercemar
Air Kopi	6.6–8.1	276	40	Tercemar

Dari hasil pengukuran, setiap sampel menunjukkan karakteristik yang berbeda dan dapat dibedakan oleh sensor. Air garam memiliki nilai TDS yang sangat tinggi hingga 540 ppm, jauh melampaui batas normal 500 ppm. Air kopi menunjukkan nilai turbidity tertinggi yaitu 40 NTU, jauh di atas batas 5 NTU. Perbedaan karakteristik antar sampel ini membuktikan bahwa ketiga sensor bekerja dengan baik dan mampu menangkap variasi kondisi air yang sesungguhnya.

Hasil tersebut menunjukkan bahwa kombinasi sensor pH, TDS, dan turbidity mampu membedakan karakteristik masing-masing sampel air sehingga layak digunakan sebagai fitur masukan pada proses klasifikasi.

B. Hasil Pengolahan Dataset

Sebelum proses pelatihan model dilakukan, dataset dianalisis untuk mengetahui distribusi kelas yang digunakan pada penelitian. Dataset yang telah melalui proses pembersihan dan pelabelan terdiri atas dua kelas, yaitu kelas normal dan tercemar.



Gambar 2. Distribusi Data Kelas Normal dan Tercemar

Berdasarkan Gambar 2, distribusi data menunjukkan komposisi yang relatif seimbang antara kedua kelas. Kelas normal memiliki proporsi sebesar 51%, sedangkan kelas tercemar memiliki proporsi sebesar 49%. Distribusi yang seimbang ini mengurangi risiko terjadinya bias model terhadap salah satu kelas dan mendukung proses pelatihan SVM yang lebih optimal.

C. Hasil Gridsearch dan Evaluasi Model SVM

Setelah proses optimasi menggunakan GridSearchCV dengan 5-fold cross-validation, kombinasi hyperparameter terbaik yang ditemukan adalah $C = 1000$, $\gamma = 0,1$, dengan kernel RBF. Nilai cross-validation score terbaik yang diperoleh adalah 0,986, menunjukkan bahwa model sudah dioptimasi dengan baik sebelum diuji pada data uji.

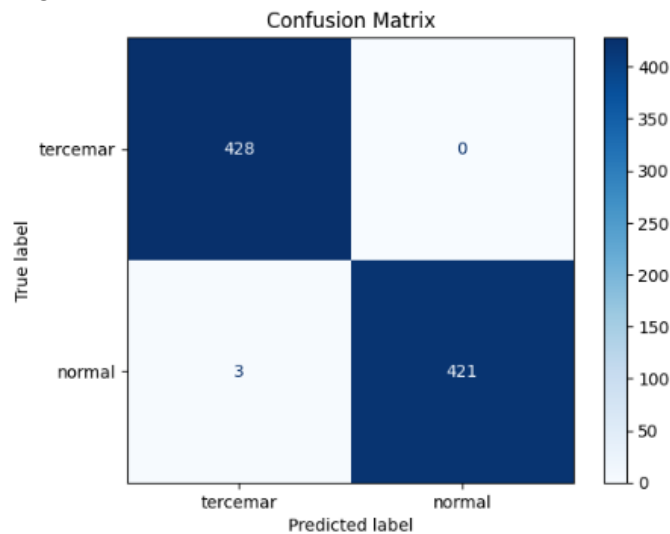
Hasil evaluasi lengkap pada data uji ditunjukkan pada Tabel 3.

Tabel 3. Hasil Evaluasi Model

Scoring	Acc.	Prec.	Rec.	F1	TN/FN/FP/TP
Accuracy	0,996	1,00	1,00	1,00	428/3/0/421
Precision	0,996	1,00	1,00	1,00	428/3/0/421
Recall	0,497	0,25	0,50	0,33	0/0/428/424

Model dengan scoring accuracy dan precision menghasilkan performa yang sangat baik dengan akurasi 99,65% dan F1-score sempurna 1,00. Keduanya menghasilkan confusion matrix yang identik, yang menunjukkan bahwa pada dataset yang seimbang, kedua metrik ini memberikan sinyal optimasi yang sama ke GridSearchCV.

Sebaliknya, scoring recall menghasilkan performa yang sangat buruk dengan akurasi hanya sekitar 50%, setara dengan tebakan acak. Yang menarik, nilai TN selalu 0 artinya model sama sekali tidak bisa mendeteksi air tercemar karena seluruh sampel diprediksi sebagai normal. Ini terjadi karena GridSearchCV mengeksplorasi definisi matematis recall, yaitu $TP/(TP+FN)$, dengan cara memilih parameter yang memprediksi semua sampel sebagai kelas positif sehingga $FN = 0$ dan nilai recall menjadi 1,0 secara otomatis. Temuan ini menegaskan bahwa recall tidak cocok digunakan sebagai satu-satunya metrik optimasi pada kasus klasifikasi biner dengan dataset seimbang. Evaluasi model juga dilakukan menggunakan confusion matrix untuk mengetahui distribusi prediksi yang dihasilkan.



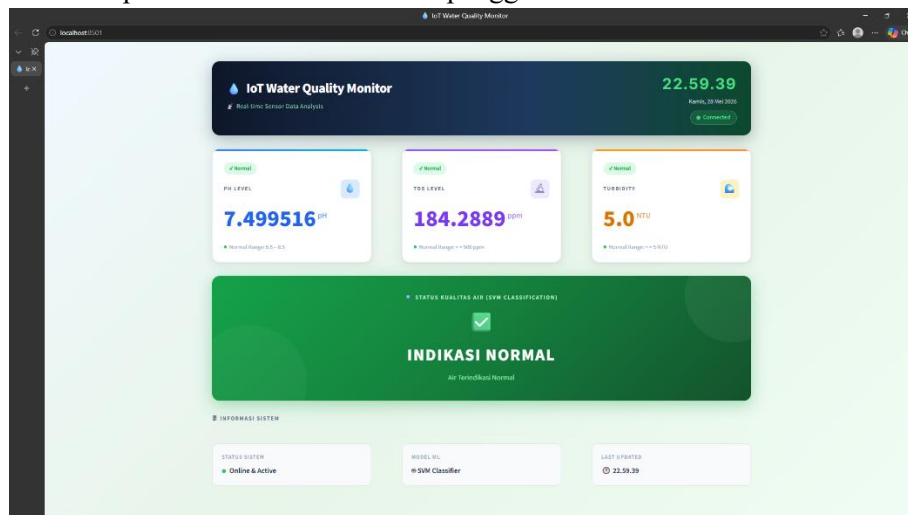
Gambar 3. Confusion Matrix

Berdasarkan confusion matrix, model berhasil mengklasifikasikan 428 data tercemar dan 421 data normal dengan benar. Hanya terdapat tiga kesalahan klasifikasi pada data pengujian sehingga menghasilkan akurasi sebesar 99,65%.

D. Implementasi Sistem Realtime

Implementasi sistem realtime dilakukan dengan mengintegrasikan sensor pH, TDS, dan turbidity dengan mikrokontroler ESP32 sebagai perangkat akuisisi data. Data hasil pembacaan sensor dikirim secara nirkabel melalui jaringan WiFi menuju server menggunakan Flask API. Pada sisi server, data yang diterima diproses menggunakan model Support Vector Machine (SVM) yang telah dilatih dan dioptimasi menggunakan GridSearchCV.

Sebelum dilakukan proses klasifikasi, data sensor terlebih dahulu melalui tahap normalisasi menggunakan StandardScaler yang sama dengan proses pelatihan model. Hasil normalisasi kemudian digunakan sebagai masukan bagi model SVM untuk menghasilkan prediksi kualitas air secara otomatis. Dengan mekanisme tersebut, sistem mampu melakukan klasifikasi kualitas air tanpa memerlukan proses analisis manual dari pengguna.



Gambar 4. Tampilan Dashboard Sistem

Seperti yang terlihat pada Gambar 4, Hasil klasifikasi ditampilkan secara realtime melalui dashboard berbasis Streamlit yang menampilkan informasi nilai pH, TDS, turbidity, status koneksi sistem, serta hasil prediksi kualitas air. Dashboard memungkinkan pengguna memantau kondisi kualitas air secara langsung melalui antarmuka web yang mudah diakses.

Berdasarkan hasil implementasi, seluruh komponen sistem berhasil terintegrasi dengan baik mulai dari akuisisi data sensor, pengiriman data melalui jaringan WiFi, pemrosesan menggunakan model machine learning, hingga visualisasi hasil pada dashboard. Integrasi tersebut menunjukkan bahwa sistem mampu melakukan monitoring dan klasifikasi kualitas air secara realtime dalam satu alur kerja yang utuh (end-to-end system).

E. Perbandingan Dengan Penelitian Sebelumnya

Tabel 4. Perbandingan dengan penelitian sebelumnya

Penelitian	Model/Metode	Akurasi
GL Pritalia (2022) [5]	6 algoritma (termasuk SVM)	Random Forest tertinggi (85%)
S. Ramya dkk. (2024) [9]	14 algoritma (SVM)	Gradient Boosting tertinggi 99,64%, SVM (99,54%)

M. K. Nallakaruppan dkk. (2024) [7]	RF, DT, Naive Bayes, LR, SVM	Random Forest nyaris (99,9%), sedangkan performa SVM paling rendah (63%)
Akash B. Koli dkk. (2025) [8]	SVM, XGBoost	XGBoost (94%) lebih unggul dari SVM (64%)
Jauhari Maulani & Mayang Sari (2023) [6]	KNN, SVM	SVM (69%) lebih akurat dari KNN (66%)
Metode yang diusulkan	GridsearchCV-SVM	Optimasi mencapai akurasi 99,6% (split 80/20, RS 21), melampaui performa SVM standar

Tabel 4 menunjukkan perbandingan performa metode yang diusulkan dengan beberapa penelitian sebelumnya pada klasifikasi kualitas air. Sebagian besar penelitian terdahulu berfokus pada perbandingan berbagai algoritma machine learning untuk memperoleh tingkat akurasi terbaik. Pritalia melaporkan bahwa Random Forest memperoleh akurasi tertinggi sebesar 85%, sedangkan S. Ramya dkk. memperoleh akurasi hingga 99,64% menggunakan algoritma Gradient Boosting dan 99,54% menggunakan SVM. Penelitian Koli dkk. menunjukkan bahwa XGBoost menghasilkan akurasi sebesar 94%, sedangkan SVM hanya mencapai 64%. Sementara itu, Maulani dan Sari melaporkan bahwa SVM memperoleh akurasi sebesar 69%.

Pada penelitian ini, model SVM yang dioptimasi menggunakan GridSearchCV berhasil mencapai akurasi sebesar 99,65%. Hasil tersebut menunjukkan bahwa performa algoritma SVM sangat dipengaruhi oleh proses optimasi hyperparameter yang digunakan. Selain itu, penggunaan dataset yang seimbang serta proses preprocessing yang sesuai turut berkontribusi terhadap peningkatan performa model.

Berbeda dengan sebagian besar penelitian sebelumnya yang berfokus pada evaluasi model klasifikasi, penelitian ini tidak hanya menghasilkan model dengan akurasi tinggi tetapi juga berhasil mengimplementasikan model tersebut pada sistem monitoring kualitas air berbasis IoT yang mampu melakukan klasifikasi secara realtime.

IV. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem pemantauan kualitas air berbasis Internet of Things (IoT) dengan menggunakan sensor pH, TDS, dan kekeruhan yang diintegrasikan dengan algoritma Support Vector Machine (SVM). Hasil pengujian menunjukkan bahwa ketiga sensor tersebut mampu membedakan karakteristik berbagai sampel air yaitu air bersih, air laut, air sabun, dan air kopi berdasarkan nilai pH, TDS, dan kekeruhan yang dihasilkan.

Model SVM, yang dioptimalkan menggunakan GridSearchCV, menghasilkan kombinasi hyperparameter terbaik dengan akurasi 99,65% pada data uji. Hasil ini menunjukkan bahwa proses optimasi hyperparameter memainkan peran krusial dalam meningkatkan kinerja klasifikasi kualitas air dan menghasilkan model yang mampu membedakan antara kondisi air normal dan terkontaminasi dengan akurasi tinggi.

Model yang telah dilatih kemudian berhasil diimplementasikan pada sistem IoT berbasis ESP32, Flask API, dan dasbor Streamlit. Sistem ini mampu mengumpulkan data sensor, mentransmisikan data melalui jaringan WiFi, melakukan klasifikasi otomatis menggunakan machine learning, dan menampilkan hasil pemantauan real-time melalui antarmuka web. Dengan demikian, penelitian ini menunjukkan bahwa integrasi IoT dan machine learning dapat berfungsi sebagai solusi yang efektif, otomatis, dan mudah diakses untuk pemantauan kualitas air.

DAFTAR PUSTAKA

- [1] I. Essamlali, H. Nhaila, and M. El Khaili, “Advances in machine learning and IoT for water quality monitoring: A comprehensive review,” Mar. 30, 2024, *Elsevier Ltd.* doi: 10.1016/j.heliyon.2024.e27920.
- [2] P. Jayaraman, K. K. Nagarajan, P. Partheeban, and V. Krishnamurthy, “Critical review on water quality analysis using IoT and machine learning models,” *International Journal of Information Management Data Insights*, vol. 4, no. 1, Apr. 2024, doi: 10.1016/j.jjime.2023.100210.
- [3] Stacyana Jesika, Suci Ramadhani, and Yohanna Permata Putri, “Implementasi Model Machine Learning dalam Mengklasifikasi Kualitas Air,” *Jurnal Ilmiah Dan Karya Mahasiswa*, vol. 1, no. 6, pp. 382–396, Nov. 2023, doi: 10.54066/jikma.v1i6.1162.
- [4] S. Sheila Maulina Putri, M. Arhami, and J. B. Tekniknologi Informasi dan Komputer Politeknik Negeri Lhokseumawe Jln, “JAISE : Journal of Artificial Intelligence and Software Engineering Penerapan Metode SVM pada Klasifikasi Kualitas Air,” 2023. [Online]. Available: <https://www.kaggle.com/datasets/adityakadiwal/water->
- [5] G. L. Pritalia, “Analisis Komparatif Algoritme Machine Learning pada Klasifikasi Kualitas Air Layak Minum,” 2022.
- [6] J. Maulani and M. Sari, “Komparasi Metode K-Nearest Neighbor (Knn) Dengan Support Vector Machine (Svm) Terhadap Tingkat Akurasi Klasifikasi Kualitas Air,” 2023.
- [7] M. K. Nallakaruppan, E. Gangadevi, M. L. Shri, B. Balusamy, S. Bhattacharya, and S. Selvarajan, “Reliable water quality prediction and parametric analysis using explainable AI models,” *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56775-y.
- [8] A. B. Koli, B. Faijan, S. Akil, B. Kantilal, D. Madhukar, and P. R. Sanjay, “A Hybrid Approach to Water Quality Classification Using SVM and Xgboost Method,” doi: 10.51244/IJRSI.
- [9] S. Ramya, S. Srinath, and P. Tuppad, “Comprehensive analysis of multiple classifiers for enhanced river water quality monitoring with explainable AI,” *Case Studies in Chemical and Environmental Engineering*, vol. 10, Dec. 2024, doi: 10.1016/j.cscee.2024.100822.
- [10] World Health Organization, *Guidelines for Drinking-water Quality*, 4th edition. Geneva: WHO press, 2011.
- [11] Kementerian Kesehatan Republik Indonesia, “Peraturan Menteri Kesehatan Republik Indonesia Nomor 32 Tahun 2017 tentang Standar Baku Mutu Kesehatan Lingkungan,” Jakarta, 2017.
- [12] N. P. E. M. Anggarini and A. Muliantara, “Memprediksi Kelulusan Mahasiswa Graduate dan Dropout dengan Support Vector Machine dan GridSearchCV,” *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 2, no. 3, pp. 475–480, 2024.