



Klasifikasi Kualitas Air Sumur Menggunakan Metode K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) di Wilayah Jakarta Timur

Linda Nur Afifa*, Kania Gita Azzahra

Universitas Darma Persada

DOI:

<https://doi.org/10.53697/jkomitek.v5i2.3697>

*Correspondence: Linda Nur Afifa

Email: linda_nur_afifa@ft.unsada.ac.id

Received: 24-10-2025

Accepted: 25-11-2025

Published: 25-12-2025



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>)

Abstrak: Penelitian ini menerapkan algoritma *K-Nearest Neighbor* (KNN) dan Support Vector Machine (SVM) sebagai metode klasifikasi untuk menentukan apakah air termasuk dalam kategori "Layak" atau "Tidak Layak" konsumsi. Proses klasifikasi mengikuti tahapan CRISP-DM dan dilakukan melalui aplikasi berbasis Python dan Streamlit yang dirancang untuk memudahkan pengguna dalam memantau kualitas air. Selain itu, pemodelan sistem divisualisasikan menggunakan diagram UML seperti *use case*, *activity*, dan *class diagram* untuk menggambarkan struktur dan interaksi sistem secara menyeluruh. Hasil dari penelitian ini diharapkan dapat membantu masyarakat dan instansi terkait dalam mengambil keputusan berbasis data untuk menjaga kualitas air yang aman dikonsumsi.

Kata kunci: Air Sumur, Kualitas Air, *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), CRISP-DM, Streamlit, Python, UML, Klasifikasi, Jakarta Timur.

Pendahuluan

Meskipun air adalah kebutuhan pokok, identifikasi kualitas air sangat penting untuk memastikan bahwa air aman untuk digunakan. berbagai kebutuhan sehari-hari, seperti memasak, mencuci, dan minum, manusia membutuhkan air. Kualitas air dipengaruhi oleh berbagai kondisi dan komponen lingkungan yang berbeda, mengingat fungsinya yang sangat penting (Noviantoroa et al, 2022).

Peraturan Menteri Kesehatan Nomor 32 Tahun 2017 menetapkan standar untuk kualitas air. di Indonesia yang mencakup parameter fisik, kimia, dan biologi seperti warna, total *dissolved solids* (TDS), dan pH. Tujuan dari standar ini adalah untuk memastikan bahwa air yang digunakan atau dikonsumsi oleh masyarakat memenuhi standar kesehatan yang aman, sehingga dapat mencegah dampak buruk terhadap kesehatan masyarakat (Azmi et al, 2022).

Untuk mengklasifikasikan kualitas air sesuai dengan standar baku mutu, analisis berbasis data menjadi sangat penting untuk menyelesaikan masalah tersebut. Dalam proses ini, algoritma seperti *K-Nearest Neighbor* (KNN) dan Support Vector Machine (SVM) dapat digunakan. Dataset digunakan sebagai data latih dalam machine learning untuk

mengidentifikasi kualitas air yang layak konsumsi. Ini memungkinkan untuk menguji kinerja model dengan metode klasifikasi yang tepat (Dhewayani et al, 2022).

Oleh karena itu, penelitian yang berjudul “Klasifikasi Kualitas Air Sumur Menggunakan Metode *K-Nearest Neighbor* (KNN) Dan *Support Vector Machine* (SVM) Di Wilayah Jakarta Timur” dibuat untuk menerapkan dan membandingkan performa algoritma KNN dan SVM dalam mengklasifikasikan kualitas air sumur di wilayah Jakarta Timur. Hasil klasifikasi diharapkan tidak hanya memberikan informasi akurat kepada masyarakat dan pemerintah, tetapi juga menjadi dasar dalam pengambilan keputusan strategis untuk menjaga kualitas air dan kesehatan masyarakat secara berkelanjutan.

Metodologi

Metode Tahap Penelitian

Sebagai kerangka kerja utama untuk penelitian ini, pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) digunakan. Pendekatan ini terdiri dari enam tahapan:

1. Business Understanding

Mengidentifikasi permasalahan utama, yaitu klasifikasi kualitas air sumur di wilayah Jakarta Timur berdasarkan parameter fisik, kimia, dan biologi.

2. Data Understanding

Menelaah karakteristik data yang diperoleh dari pengukuran lapangan dan data sekunder resmi (Dinas Lingkungan Hidup DKI Jakarta). Parameter meliputi pH, TDS, Besi (Fe), Nitrat sebagai N, E. Coli, Total Coliform, warna, dan bau.

3. Data Preparation

Tahapan ini mencakup proses *cleaning*, normalisasi menggunakan Min-Max Scaling, serta transformasi data menjadi format numerik dan kategorikal (Layak dan Tidak Layak).

4. Modeling

K-Nearest Neighbor (KNN) dan *Support Vector Machine* (SVM) adalah dua algoritma yang digunakan untuk membuat model klasifikasi. Parameter seperti jumlah tetangga (k) dan fungsi kernel disesuaikan.

5. Evaluation

Untuk menilai algoritma, metrik akurasi, presisi, recall, dan skor F1 digunakan berdasarkan matrix confusion. Hasilnya digunakan untuk membandingkan efektivitas kedua algoritma.

6. Deployment

Aplikasi berbasis web yang menggunakan Streamlit menggunakan model ini, yang memungkinkan pengguna mengimpor data dan melihat hasil klasifikasi secara real-time.

Metode Pengolahan Penelitian

Pengolahan data dilakukan menggunakan Python dan berbagai *library* pendukung, seperti pandas untuk manajemen data, scikit-learn untuk implementasi algoritma *machine learning*, serta matplotlib dan seaborn untuk visualisasi.

Langkah-langkah pengolahan data sebagai berikut:

1. *Preprocessing* Data

- Menghapus duplikasi dan nilai kosong
- Melakukan encoding pada data kategorikal (warna dan bau)
- Menyusun label klasifikasi: Layak (1) dan Tidak Layak (0)

2. Normalisasi Data

Untuk meningkatkan kinerja model, gunakan metode skala Min-Max agar semua parameter berada pada skala yang sama.

3. Pelatihan Model

Model KNN dan SVM dilatih dengan data latih dan diuji dengan data uji untuk mengukur performanya.

4. Evaluasi Model

Dilakukan menggunakan confusion matrix dan classification report untuk mendapatkan nilai akurasi, precision, recall, dan F1-score.

5. Visualisasi dan Implementasi

Hasil klasifikasi ditampilkan melalui aplikasi berbasis web. Aplikasi ini menampilkan grafik, tabel klasifikasi, dan fitur cetak laporan yang berguna bagi masyarakat dan instansi terkait.

Hasil dan Pembahasan

Hasil penelitian ini menggabungkan spesifikasi perangkat lunak yang digunakan dalam proses pengembangan sistem klasifikasi kualitas air sumur, tampilan antarmuka sistem setelah proses *deployment*, struktur *database* yang dibangun, serta hasil pengujian dan analisis performa algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM).

1. Perangkat Lunak Yang Digunakan

Bahasa pemrograman utama yang digunakan dalam penelitian ini adalah Python untuk pengolahan data dan penerapan algoritma klasifikasi, dan Streamlit adalah framework yang digunakan untuk membangun aplikasi web yang interaktif dan mudah digunakan. Pustaka Python seperti pandas, numpy, scikit-learn, matplotlib, dan seaborn digunakan untuk melakukan analisis data dan pembuatan model. Selain itu, digunakan Visual Studio Code sebagai lingkungan pengembangan utama (IDE) untuk penulisan dan pengujian kode program, XAMPP sebagai server lokal untuk mendukung integrasi dengan basis data MySQL, dan Microsoft Excel untuk pengelolaan awal data. Kombinasi perangkat lunak ini dipilih untuk mendukung pengolahan data yang efisien, visualisasi hasil, dan pembuatan sistem klasifikasi yang mudah diakses dan terintegrasi.

2. Tampilan Interface Hasil Deploy

Adapun tampilan interface hasil deploy sebagai berikut:

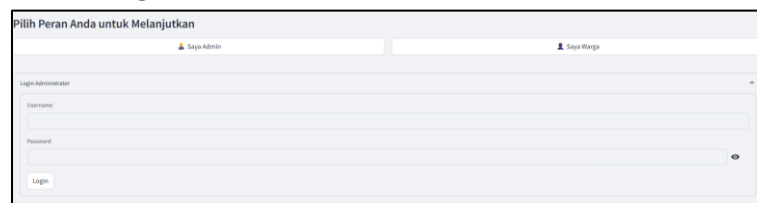
a. Tampilan beranda halaman depan



Gambar 2. Tampilan beranda halaman depan

Halaman beranda ini menyambut pengguna dengan memberikan penjelasan komprehensif tentang pentingnya kualitas air dan faktor-faktor yang memengaruhinya, serta urgensi aplikasi ini dalam membantu warga Jakarta Timur mengevaluasi kualitas air sumur. Selain itu, halaman ini juga menyediakan opsi masuk baik sebagai 'Admin' maupun 'Warga', memfasilitasi akses sesuai dengan peran pengguna.

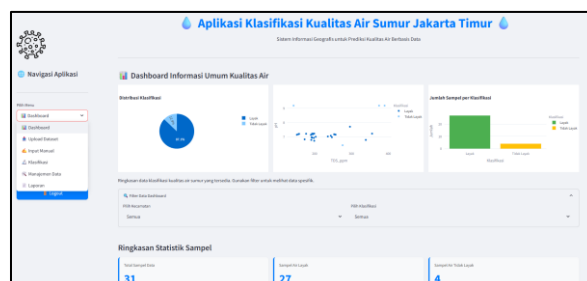
b. Tampilan Halaman Login Admin



Gambar 3. Tampilan Halaman Login Admin

Gambar ini menunjukkan halaman pemilihan peran untuk melanjutkan masuk ke dalam sistem. Pengguna dapat memilih untuk masuk sebagai "Admin" atau "Warga". Di bawah pilihan peran tersebut, terdapat formulir "Login Administrator" yang meminta input Username dan Password untuk proses login admin.

c. Tampilan Dashboard Admin



Gambar 4. Tampilan Dashboard Admin

Dashboard "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur" adalah sistem informasi geografis berbasis data yang memungkinkan pengguna (admin atau warga) untuk memantau, menganalisis, dan memfilter kualitas air sumur, menampilkan distribusi kelayakan air, sebaran TDS, jumlah sampel per klasifikasi, dan ringkasan statistik dari total

31 sampel, di mana 27 sampel (87.1%) diklasifikasikan layak dan 4 sampel (12.9%) tidak layak.

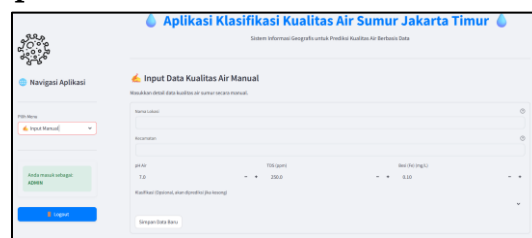
d. Tampilan Halaman Upload Dataset Admin



Gambar 5. Halaman Upload Dataset Admin

Gambar ini menunjukkan halaman "Unggah Dataset Baru (Excel)" dari "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur", di mana pengguna dengan peran "ADMIN" dapat mengunggah file Excel (.xlsx) baru untuk mengganti data yang ada di database, dengan peringatan untuk membuat cadangan jika diperlukan, dan memiliki batas ukuran file 200MB.

e. Tampilan Halaman Input Manual Admin



Gambar 6. Halaman Input Manual Admin

Gambar ini menampilkan halaman "Input Data Kualitas Air Manual" dari "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur", tempat pengguna dengan peran "ADMIN" dapat memasukkan detail data kualitas air sumur secara manual, mencakup kolom untuk Nama Lokasi, Kecamatan, pH Air, TDS (ppm), Besi (mg/L), dan Klasifikasi (opsional, akan diprediksi jika dikosongkan), sebelum menyimpan data baru.

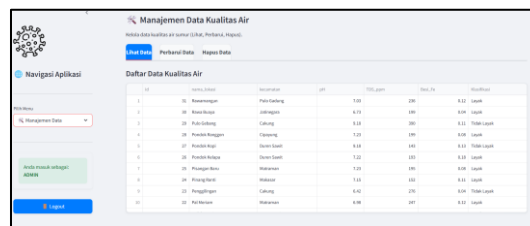
f. Tampilan Halaman Klasifikasi Admin



Gambar 7. Halaman Klasifikasi Admin

Gambar ini menampilkan halaman "Prediksi Kualitas Air Berdasarkan Input Baru" dari "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur", di mana pengguna dengan peran "ADMIN" dapat memasukkan parameter air (pH, TDS, kandungan Besi (Fe), Nitrat sebagai N, E. Coli, Total Coliform, warna, dan bau) untuk mendapatkan prediksi klasifikasi kualitas air (Layak/Tidak Layak) menggunakan model KNN dan SVM yang sudah dilatih.

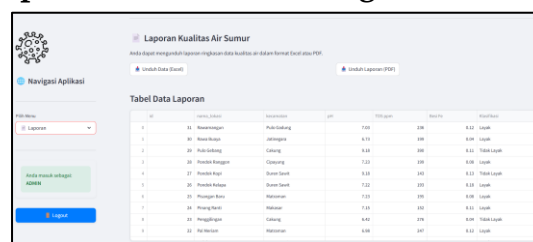
g. Tampilan Halaman Manajemen Data Admin



Gambar 8. Halaman Manajemen Data Admin

Gambar ini menunjukkan halaman "Manajemen Data Kualitas Air" dari "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur", di mana pengguna dengan peran "ADMIN" dapat melihat, memperbarui, dan menghapus data kualitas air sumur yang tercantum dalam tabel "Daftar Data Kualitas Air". Tabel tersebut menampilkan informasi seperti ID, nama lokasi, kecamatan, pH, TDS_ppm, Besi_Fe, dan klasifikasi (Layak/Tidak Layak).

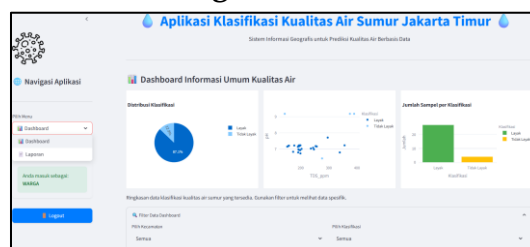
h. Tampilan Halaman Laporan (Admin dan Warga)



Gambar 9. Halaman Laporan

Gambar ini menampilkan halaman "Laporan Kualitas Air Sumur" dari "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur", di mana pengguna dengan peran "ADMIN" dapat mengunduh ringkasan data kualitas air dalam format Excel atau PDF, serta melihat "Tabel Data Laporan" yang berisi ID, nama lokasi, kecamatan, pH, TDS ppm, Besi Fe, dan klasifikasi dari setiap sampel air.

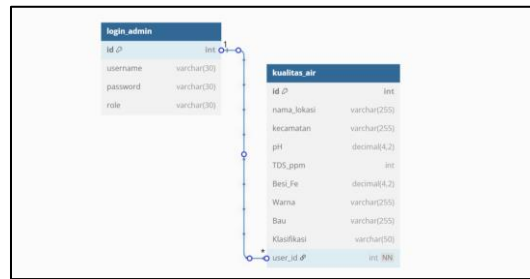
i. Tampilan Halaman Dashboard Warga



Gambar 10. Halaman Dashboard Warga

Halaman ini Dashboard "Aplikasi Klasifikasi Kualitas Air Sumur Jakarta Timur" yang diakses oleh pengguna "WARGA" menampilkan ringkasan informasi kualitas air melalui distribusi klasifikasi (87.1% layak, 12.9% tidak layak), sebaran data pH vs. TDS, jumlah sampel per klasifikasi, serta fitur filter berdasarkan kecamatan dan klasifikasi.

j. Struktur Database



Gambar 11. Struktur Database

Database ini dirancang untuk mengelola data pengguna (admin) dan informasi terkait kualitas air sumur yang diukur berdasarkan berbagai parameter. Terdapat dua tabel utama: tabel *login_admin* yang menyimpan data akun pengguna seperti username, password, dan peran (*role*), serta tabel *kualitas_air* yang menyimpan data hasil pengukuran kualitas air dari berbagai lokasi, termasuk parameter seperti pH, TDS, kandungan Besi (Fe), Nitrat sebagai N, E. Coli, Total Coliform, warna, dan bau, dan klasifikasi kelayakan air. Relasi *one-to-many* menghubungkan tabel *login_admin* dengan tabel *kualitas_air*, di mana satu admin dapat menginput banyak data kualitas air. Atribut *user_id* pada tabel *kualitas_air* berfungsi sebagai *foreign key* yang merujuk ke *id* pada tabel *login_admin*. Dengan desain ini, sistem memungkinkan setiap admin untuk merekam dan mengelola data kualitas air secara independen berdasarkan lokasi masing-masing.

k. Hasil Evaluasi

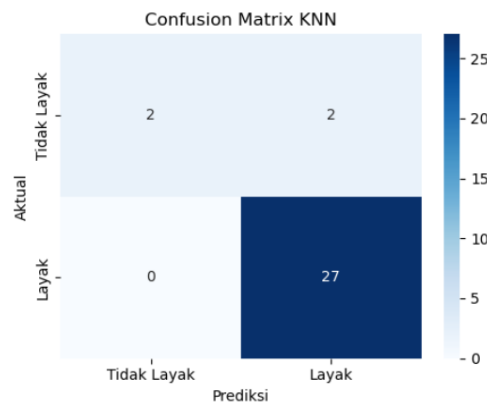
Berdasarkan hasil evaluasi, algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) menunjukkan performa yang baik dalam mengklasifikasikan kualitas air sumur menjadi kategori Layak dan Tidak Layak. Evaluasi dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score yang diperoleh dari confusion matrix. Hasil evaluasi menunjukkan bahwa algoritma SVM memiliki nilai recall yang lebih tinggi dibandingkan KNN dalam mendeteksi data air yang Layak, sementara KNN memiliki keunggulan pada nilai presisi. Dengan demikian, kedua algoritma dapat digunakan secara efektif, namun pemilihan metode akhir dapat disesuaikan dengan kebutuhan spesifik sistem, apakah lebih mengutamakan ketepatan atau kelengkapan deteksi.

A. Confusion Matrix Evaluasi KNN

Untuk menilai kinerja algoritma *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan kualitas air sumur, digunakan Confusion Matrix yang membandingkan hasil prediksi dengan label aktual dari data uji.

Tabel 1. Confusion Matix Evaluasi KNN

Aktual	Prediksi	
	Layak	Tidak Layak
Layak	27 (TP)	0 (FN)
Tidak Layak	2 (FP)	2 (TN)



Gambar 12. Confusion Mstrix KNN

Tabel dan Gambar di atas merupakan visualisasi dari Confusion Matrix hasil klasifikasi kualitas air sumur menggunakan algoritma *K-Nearest Neighbor* (KNN). Matriks ini menunjukkan jumlah prediksi yang benar dan salah berdasarkan label aktual dari data uji.

Hasil berikut didapat dari Confusion Matrix:

- a. TP (True Positive) = 27
- c. FP (False Positive) = 2
- c. FN (False Negative) = 0
- d. TN (True Negative) = 2

Berikut adalah perhitungan untuk mencari matriks evaluasi:

a) *Accuracy* (Akurasi)

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

$$Accuracy = \frac{27 + 2}{27 + 2 + 2 + 0}$$

$$Accuracy = \frac{29}{31}$$

$$= 93.55\%$$

Berdasarkan hasil perhitungan, dapat disimpulkan bahwa model KNN memberikan hasil prediksi yang benar untuk 29 dari 31 data uji.

b) *Precision* (Presisi)

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{27}{27 + 2}$$

$$= \frac{27}{29}$$

$$= 93.10\%$$

Dengan mempertimbangkan hasil perhitungan sebelumnya, dapat disimpulkan bahwa dari semua data yang diprediksi Layak, terdapat 93.10% yang benar-benar Layak.

c) *Recall* (Sensitivitas / True Positive Rate)

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ \text{Recall} &= \frac{27}{27 + 0} \\ &= \frac{27}{27} \\ &= 100\% \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa model berhasil menemukan seluruh data yang benar-benar Layak (tidak ada yang terlewat).

d) F1-Score

$$\begin{aligned} F1 - \text{Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ F1 - \text{Score} &= 2 \times \frac{0.9310 \times 1.00}{0.9310 + 1.00} \\ &= 0.963 \\ &= 96.30\% \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa F1-Score memberikan nilai seimbang antara presisi dan kemampuan model menemukan data yang benar-benar layak

Tabel 2. Hasil Akurasi KNN

KNN	
	<i>Percentage Split</i>
<i>Accuracy</i>	93.55%
<i>Precision</i>	93.10%
<i>Recall</i>	100%
F1-Score	96.30%

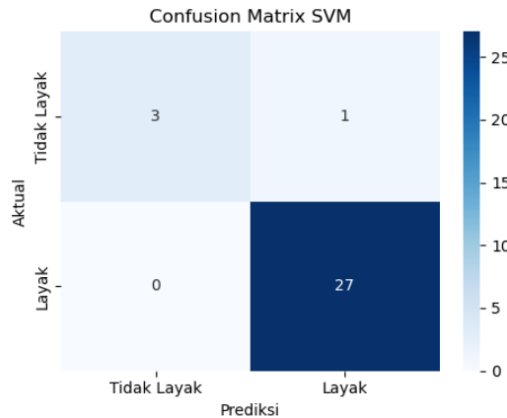
Berdasarkan Tabel 2, algoritma *K-Nearest Neighbor* (KNN) menunjukkan kemampuan yang sangat baik untuk klasifikasi dengan nilai akurasi sebesar 93,55%, yang berarti sebagian besar prediksi yang dibuat oleh model berdasarkan label asli. Nilai precision sebesar 93,10% menunjukkan bahwa prediksi kelas Layak oleh model sebagian besar benar. Sementara itu, recall mencapai 100%, yang berarti seluruh data Layak berhasil dideteksi dengan benar oleh model. Kombinasi presisi dan recall yang tinggi menghasilkan F1-score sebesar 96,30%, menandakan keseimbangan yang kuat antara ketepatan dan kelengkapan dalam klasifikasi. Hasil ini menunjukkan bahwa KNN sangat efektif dalam mengidentifikasi kualitas air sumur dengan akurasi yang tinggi.

B. Confusion Matrix Evaluasi SVM

Untuk mengevaluasi performa algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan kualitas air sumur, digunakan Confusion Matrix yang membandingkan hasil prediksi dengan label aktual dari data uj

Tabel 3. Confusion Matix Evaluasi SVM

		Prediksi	
		Tidak Layak	Layak
Aktual	Tidak Layak	3 (TN)	1 (FP)
	Layak	0 (FN)	27 (TP)



Gambar 12. Confusion Mstrix SVM

Tabel dan Gambar di atas merupakan visualisasi dari Confusion Matrix hasil klasifikasi kualitas air sumur menggunakan algoritma Support Vector Machine (SVM). Matriks ini menunjukkan jumlah prediksi yang benar dan salah berdasarkan label aktual dari data yang diuji. Hasil dari Confusion Matrix tersebut adalah sebagai berikut:

- a. TP (True Positive) = 27
- b. FP (False Positive) = 1
- c. FN (False Negative) = 0
- d. TN (True Negative) = 3

Berikut adalah perhitungan untuk mencari matriks evaluasi:

a) *Accuracy* (Akurasi)

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

$$Accuracy = \frac{27 + 3}{27 + 3 + 1 + 0}$$

$$Accuracy = \frac{30}{31}$$

$$= 96.77\%$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa akurasi menunjukkan proporsi jumlah prediksi yang benar dari seluruh data uji. Dalam kasus ini, model mampu memprediksi 30 dari 31 data dengan benar.

b) *Precision* (Presisi)

$$\begin{aligned}
 Precision &= \frac{TP}{TP + FP} \\
 Precision &= \frac{27}{27 + 1} \\
 &= \frac{27}{28} \\
 &= 96.43\%
 \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa dari 31 prediksi "Layak" yang dilakukan oleh model, ternyata 30 di antaranya benar-benar "Layak". Ini menunjukkan bahwa model jarang salah saat mengatakan suatu data termasuk kategori Layak, meskipun masih terdapat satu kesalahan klasifikasi positif palsu

c) *Recall* (Sensitivitas / *True Positive Rate*)

$$\begin{aligned}
 Recall &= \frac{TP}{TP + FN} \\
 Recall &= \frac{27}{27 + 0} \\
 &= \frac{27}{27} \\
 &= 1.00 \\
 &= 100\%
 \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa karena semua data yang benar-benar "Layak" berhasil dikenali oleh model tanpa ada yang terlewat, recall mencapai nilai sempurna 100%.

d) F1-Score

$$\begin{aligned}
 F1 - Score &= 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 F1 - Score &= 2 \times \frac{0.9643 \times 1.00}{0.9643 + 1.00} \\
 &= \frac{1.9286}{1.9643} \\
 &= 0.9818 \\
 &= 98.18\%
 \end{aligned}$$

Berdasarkan hasil perhitungan di atas, dapat disimpulkan bahwa F1-Score adalah rata-rata harmonik antara presisi dan recall. Nilai ini berguna ketika kita ingin menyeimbangkan antara kemampuan model untuk menemukan kelas positif secara tepat (presisi) dan mengidentifikasi semua kelas positif yang ada. Dengan nilai 98.18%, dapat dikatakan bahwa model memiliki keseimbangan yang sangat baik antara kedua aspek tersebut.

Tabel 3. Hasil Akurasi SVM

SVM	
<i>Percentage Split</i>	
<i>Accuracy</i>	96.77%
<i>Precision</i>	96.43%
<i>Recall</i>	100%
F1-Score	98.18%

Tabel 3 menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki kemampuan klasifikasi yang sangat baik dalam menentukan kualitas air sumur.. Nilai akurasi sebesar 96,77% menandakan bahwa sebagian besar prediksi SVM sesuai dengan label sebenarnya. Dengan *precision* sebesar 96,43%, model memiliki tingkat ketepatan yang sangat baik dalam mengklasifikasikan air Layak. Recall mencapai 100%, menunjukkan bahwa seluruh data Layak berhasil dideteksi secara sempurna. Kombinasi presisi dan recall tersebut menghasilkan F1-score sebesar 98,18%, yang mencerminkan keseimbangan optimal antara ketepatan dan sensitivitas model. Dengan demikian, SVM terbukti lebih unggul dibandingkan KNN dalam klasifikasi kualitas air pada penelitian ini.

Simpulan

Hasil menunjukkan bahwa penelitian ini berhasil membuat sistem klasifikasi kualitas air sumur di wilayah Jakarta Timur dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM). Berdasarkan hasil evaluasi, kedua algoritma menunjukkan performa yang tinggi, namun SVM memberikan hasil terbaik dengan nilai akurasi sebesar 96,77%, *precision* 96,43%, recall 100%, dan F1-score 98,18%, mengungguli KNN yang memperoleh akurasi 93,55%. Sistem berbasis web Streamlit memungkinkan pengguna memasukkan parameter air dan melihat hasil klasifikasi secara langsung dan interaktif. Hasil penelitian ini diharapkan dapat berfungsi sebagai pedoman bagi masyarakat. dan instansi terkait dalam memantau serta meningkatkan kualitas air sumur secara praktis dan ilmiah.

Untuk penelitian selanjutnya, disarankan agar pengembangan sistem dilakukan dengan memperluas dataset dari berbagai wilayah, menambahkan parameter lingkungan seperti pH tanah dan curah hujan, serta mengintegrasikan teknologi Internet of Things (IoT) guna melakukan pemantauan kualitas air secara real-time. Selain itu, pengujian dengan model pembelajaran mesin lainnya, seperti Random Forest atau Neural Network, dapat dilakukan untuk membandingkan performa dan meningkatkan akurasi klasifikasi pada skala yang lebih luas.

Daftar Pustaka

- Azmi, B. N., Hermawan, A., & Avianto, D. (2022). Jurnal Sistem dan Teknologi Informasi Analisis Pengaruh PCA Pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression. 7(2). <http://jurnal.unmuhjember.ac.id/index.php/IUSTINDO>
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., Jajuli, M., & Waluyo, R. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. Jurnal Teknologi Dan Informasi. <https://doi.org/10.34010/jati.v12i1>
- Fitrianti, I., Voutama, A., & Umaidah, Y. (2023). Clustering Film Populer Pada Aplikasi Netflix Dengan Menggunakan Algoritma K-Means Dan Metode CRISP-DM Clustering Popular Movies on Netflix App Using K-Means Algorithm and CRISP-DM Method. In JTISI. 4 (2).
- Hafeez, A., & Sial, A. H. (2021). Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python. International Journal of Advanced Trends in Computer Science and Engineering, 10(1), 277–281. <https://doi.org/10.30534/ijatcse/2021/391012021>
- Irfan, M., Siregar, H., & Handoko, J. T. (2023). Pengembangan Dan Integrasi Aplikasi Prediksi Jumlah Gagal Produksi PC Menggunakan Metode Triple Exponential Smoothing Pada Sistem Aplikasi Produksi Di PT Tera Data Indonusa,Tbk.
- Junaidi, S., Devegi, M., & Kurniawan, H. (2023). Pelatihan Pengolahan dan Visualisasi Data Penduduk menggunakan Python. *ADMA: Jurnal Pengabdian Dan Pemberdayaan Masyarakat*, 4(1), 151–162. <https://doi.org/10.30812/adma.v4i1.2963>
- Kumala Sari, P., & Randy Suryono, R. (2024). Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse (Vol. 7, Issue 1).
- Lavanya, A., Gaurav, L., Sindhuja, S., Seam, H., Joydeep, M., Uppalapati, V., Ali, W., & S.D, V. (2023). Assessing the Performance of Python Data Visualization Libraries: A Review. *International Journal of Computer Engineering in Research Trends*, 10(1), 28–39. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104>
- Malik Namus Akbar, F. (2024). Metode KNN (K-Nearest Neighbor) untuk Menentukan Kualitas Air. Jurnal TEKNO KOMPAK, 18(1).
- Maulani, J., & Sari, M. (2023). Komparasi Metode K-Nearest Neighbor (Knn) Dengan Support Vector Machine (Svm) Terhadap Tingkat Akurasi Klasifikasi Kualitas Air.
- Mutoffar, M. M., & Fadillah, A. (2022a). Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, 4(2), 138–146. <https://doi.org/10.53580/naratif.v4i2.160>
- Mutoffar, M. M., & Fadillah, A. (2022b). Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, 4(2), 138–146. <https://doi.org/10.53580/naratif.v4i2.160>
- Mutaqqin et al. (2023). Pengenalan Data Mining. 1–2.
- Noviantoroa, A., Silviana, A. B., Fitriani, R. R. & Permatasari, H. P. (2022). Rancangan Dan Implementasi Aplikasi Sewa Lapangan Badminton Wilayahdepok Berbasis Web. *Jurnal Teknik Dan Science*. <https://doi.org/10.56127/jts.v1i2.108>

- Permenkes No. 32. (2017).
- Sarosa, M. (2022). *Pemrograman Python Dalam Contoh dan Penerapan (Vol. 1)*. Media Nusa Creative (MNC Publishing).
- Rahman, F. Y., Purnomo, I. I., & Hijriana, N. (2022). Penerapan Algoritma Data Mining Untuk Klasifikasi Kualitas Air. *Technologia: Jurnal Ilmiah*, 13(3), 228. <https://doi.org/10.31602/tji.v13i3.7070>
- Rahman, F. D., Zulfa, M. I. & Taryana, A. (2024). Clustering Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest. *Jurnal Sinta: Sistem Informasi Dan Teknologi Komputasi*, 1(2), 90–97. <https://doi.org/10.61124/sinta.v1i2.15>
- Ramadhani, B., & Suryono, R. R. (2024). Komparasi Algoritma Naïve Bayes dan Logistic Regression Untuk Analisis Sentimen Metaverse. *Jurnal Media Informatika Budidarma*, 8(2), 714. <https://doi.org/10.30865/mib.v8i2.7458>
- Salah, E., & Din, U. (2020). Popular Python Libraries And Their Application Domains. *International Journal of Advance Engineering and Research Development*. 270–276.
- Saputra, M., Sidabuke, J. P., Sinulingga, R. P. & Tamba, R. B. (2023). Analisis Metode Algoritma K-Nearest Neighbor (KNN) Dan Naive Bayes Untuk Klasifikasi Diabetes Mellitus. *Jurnal TEKINKOM*, 6(2). <https://doi.org/10.37600/tekinkom.v6i2.942>
- Suendri. (2018). Implementasi Diagram UML (Unified Modelling Language) Pada Perancangan Sistem Informasi Remunerasi Dosen Dengan Database Oracle (Studi Kasus: UIN Sumatera Utara Medan). *ALGORITMA: Jurnal Ilmu Komputer Dan Informatika*. <http://www.omg.org>
- Tangkelayuk, A. (2022). The Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes, dan Decision Tree. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1109–1119. <https://doi.org/10.35957/jatisi.v9i2.2048>
- Ulum, S., Fahmi, A. R., Rizkika, P., & Rozikin, C. (2023). Perbandingan Performa Algoritma KNN dan SVM dalam Klasifikasi Kelayakan Air Minum. *In Generation Journal*. 7 (2).
- Waskom, M. (2021). Seaborn: Statistical Data Visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>.
- Yuniar, N. M. (2023). Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve baiyes. *Jurnal Sains Dan Teknologi*, 5(1), 243–246. <https://doi.org/10.55338/saintek.v5i1.1383>