

# Klasifikasi Kualitas Air Sungai Citarum dengan Artificial Neural Network, Gaussian Naive Bayes, dan Weighted K-Nearest Neighbors

1<sup>st</sup> Daffa Asyqar Ahmad Khalisheka

Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia

khalishekahmad@student.telkomuniversity.ac.id

2<sup>nd</sup> Meta Kallista

Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia

metakallista@telkomuniversity.ac.id

3<sup>rd</sup> Ig. Prasetya Dwi Wibawa

Fakultas Teknik Elektro  
Universitas Telkom  
Bandung, Indonesia

prasdwibawa@telkomuniversity.ac.id

**Abstrak** — Sungai Citarum, sungai terpanjang di Jawa Barat dengan panjang 323 kilometer, mengalir dari Situ Cisanti di Kabupaten Bandung ke Laut Jawa. Sungai ini sangat penting bagi ekologi dan kehidupan penduduk setempat. Pada tahun 2018, Sungai Citarum dinilai sebagai sungai terkotor di dunia, yang menimbulkan masalah lingkungan dan kesehatan yang signifikan di kota-kota sekitarnya. Untuk mengetahui hasil klasifikasi kualitas air Sungai Citarum menggunakan perhitungan manual membutuhkan waktu yang lama. Oleh karena itu menggunakan *Machine Learning* untuk mengetahui hasilnya dengan cepat, tepat, dan akurat untuk menentukan kualitas air sungai apakah tidak tercemar/memenuhi baku mutu, tercemar ringan, tercemar sedang, dan tercemar berat. Jumlah kontaminasi air sungai ditentukan dengan menggunakan pendekatan *Supervised Learning* dari *Machine Learning*. Tiga model yang digunakan adalah *Artificial Neural Network (ANN)*, *Gaussian Naive Bayes*, dan *Weighted K-Nearest Neighbor with Euclidean Distance*. Setelah itu dilakukan berbagai pengujian terhadap 3 model tersebut untuk mendapatkan hasil klasifikasi kualitas air Sungai Citarum. Berdasarkan hasil pengujianya, *Weighted KNN with Euclidean Distance* mengungguli dua model lainnya untuk mengklasifikasikan kualitas air sungai, dengan akurasi sebesar 97,3%. Sedangkan *Artificial Neural Network* mendapat akurasi sebesar 94,1% dan *Gaussian Naive Bayes* mendapat akurasi sebesar 94,7%.

**Kata kunci**— Kualitas Air Sungai, Klasifikasi, ANN, GNB, W-KNN,

## I. PENDAHULUAN

Sungai Citarum merupakan sungai terpanjang di Jawa Barat, membentang sepanjang 323 kilometer dari Situ Cisanti di Kabupaten Bandung hingga Laut Jawa. Sungai ini sangat penting bagi ekologi dan kehidupan penduduk setempat karena menyediakan air untuk keperluan rumah tangga, pertanian, dan industri. Namun, dalam beberapa dekade terakhir, Sungai Citarum telah tercemar parah. Pada tahun 2018, Sungai Citarum dinobatkan sebagai sungai terkotor di dunia, yang menimbulkan risiko lingkungan dan kesehatan yang signifikan bagi daerah sekitarnya [1].

Penilaian kualitas air sungai dilakukan sesuai dengan Keputusan Menteri Lingkungan Hidup No. 115/2003 tentang

Pedoman Penentuan Status Mutu Air. Pedoman ini menjelaskan cara menentukan status mutu air dengan menggunakan pendekatan Indeks Pencemaran (IP). Kualitas air dikategorikan ke dalam empat tingkatan: tidak tercemar, tercemar ringan, tercemar sedang, dan tercemar berat. Dalam hal ini, digunakan klasifikasi mutu air kelas II sesuai dengan Peraturan Pemerintah No. 82 Tahun 2001 tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air, yang meliputi prasarana, rekreasi air, pembudidayaan ikan air tawar, peternakan, irigasi pertanian, dan peruntukan lain yang mempersyaratkan mutu air yang sama [2].

Tantangan ini menyoroti pentingnya pemantauan dan klasifikasi kualitas air sungai yang sering dilakukan untuk mengevaluasi tingkat pencemaran dan mengambil tindakan yang sesuai. Secara umum, analisis kualitas air dilakukan secara perhitungan manual, yang memakan waktu dan sumber daya. Oleh karena itu, diperlukan cara yang lebih efisien dan tepat untuk menilai kualitas air Sungai Citarum.

Teknologi Pembelajaran Mesin, seperti *Artificial Neural Network*, *Gaussian Naive Bayes*, dan *Weighted K-Nearest Neighbor with Euclidean Distance*, memberikan solusi yang efektif untuk memantau tingkat kualitas air. Model-model ini telah terbukti berguna dalam berbagai aplikasi klasifikasi. Masalah utama yang dihadapi adalah mengklasifikasikan kualitas air Sungai Citarum dengan cepat dan akurat.

Tujuan dari penelitian ini adalah untuk menggunakan dan membandingkan tiga model *Machine Learning* untuk klasifikasi kualitas air di Sungai Citarum: *Artificial Neural Network*, *Gaussian Naive Bayes*, dan *Weighted K-Nearest Neighbor with Euclidean Distance*. Tujuan akhirnya adalah untuk menemukan model yang optimal untuk mengklasifikasikan tingkat polusi dan memungkinkan pemantauan kualitas air yang lebih efektif.

## II. KAJIAN TEORI

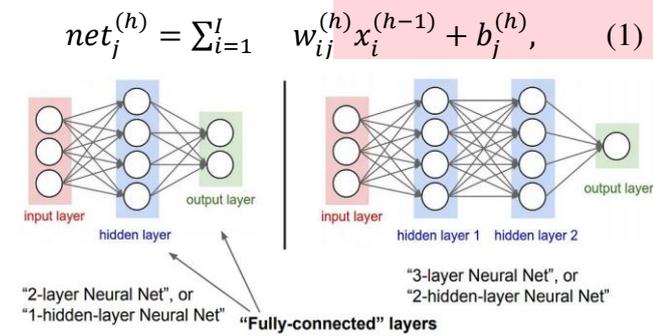
### A. *Supervised Learning* untuk Klasifikasi

*Supervised learning* adalah penggunaan data berlabel untuk melatih model untuk prediksi atau klasifikasi. Tujuannya adalah untuk mengkategorikan data berdasarkan karakteristik tertentu. Model belajar untuk mengenali pola dalam data pelatihan, memungkinkannya untuk

mengklasifikasikan data baru dengan akurasi yang tinggi. Algoritma yang digunakan antara lain ANN, GNB, dan Weighted KNN [8].

B. Artificial Neural Network (ANN)

Artificial Neural Network adalah teknologi pembelajaran mesin yang meniru cara kerja otak manusia. Jaringan saraf tiruan terdiri dari banyak lapisan neuron yang saling terhubung. Neuron-neuron ini berkolaborasi untuk memproses informasi dan membuat prediksi atau klasifikasi berdasarkan fakta-fakta yang tersedia. ANN telah menunjukkan keefektifan dalam berbagai aplikasi, termasuk pengenalan pola, klasifikasi gambar, dan pemrosesan data waktu nyata. Berikut ini persamaan untuk ANN di bawah ini [3]. Serta Bisa dilihat pada Gambar 2.1 Arsitektur Artificial Neural Network.



GAMBAR 2.1 Arsitektur Artificial Neural Network

C. Gaussian Naive Bayes

Gaussian Naive Bayes (GNB) adalah algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes. GNB mengandaikan bahwa karakteristik data yang disediakan tidak bergantung satu sama lain dan memiliki distribusi normal (*Gaussian*). Pendekatan ini terkenal dengan kemudahannya penggunaan dan efisiensinya saat bekerja dengan set data yang besar. GNB sering digunakan untuk klasifikasi teks, deteksi spam, dan analisis sentimen. Berikut persamaan dari GNB [4].

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \quad (2)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \quad (3)$$

D. Weighted KNN with Euclidean Distance

Weighted K-Nearest Neighbor adalah algoritma klasifikasi yang mengurutkan data berdasarkan jarak terdekat dari k tetangga. Dalam Weighted KNN, setiap tetangga diberi bobot berdasarkan jaraknya dari titik yang akan dikategorikan. Metrik Euclidean Distance digunakan untuk menghitung jarak antara dua titik dalam ruang fitur. Weighted KNN dengan Euclidean Distance sangat efektif untuk data nonlinier yang menuntut perhitungan jarak yang tepat. Perhitungan bobot dapat digunakan dengan persamaan di bawah ini [5].

$$W_{ni} = \frac{1}{d(x_{ik}, x_{jk})} \quad (4)$$

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (5)$$

E. SMOTE-ADASYN

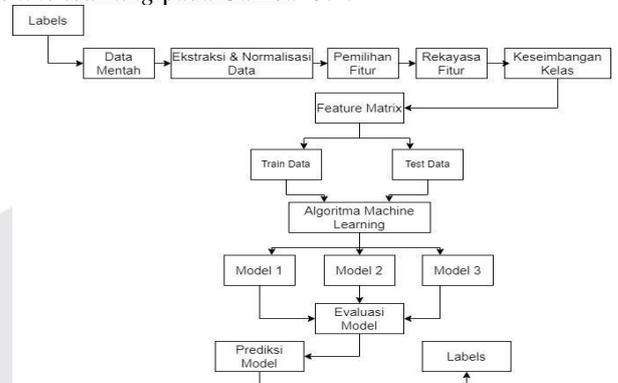
SMOTE (*Synthetic Minority Over-sampling Technique*) dan ADASYN (*Adaptive Synthetic Sampling Approach*) adalah dua metode untuk mengatasi ketidakseimbangan kelas dalam set data. SMOTE menghasilkan sampel sintesis baru dari kelas minoritas dengan menginterpolasi sampel yang sudah ada. Sebaliknya, ADASYN berkonsentrasi pada sampel yang lebih sulit diklasifikasikan dan membuat contoh sintesis baru untuk meningkatkan akurasi model secara keseluruhan. SMOTE-ADASYN digabungkan untuk meningkatkan kinerja model pembelajaran mesin dengan memberikan distribusi kelas yang lebih seimbang [6] [7].

F. Hasil Evaluasi Model

Evaluasi sangat penting untuk memastikan ketepatan dan kegunaan model pembelajaran mesin. Metrik utama meliputi: *Accuracy*, yang mengukur persentase prediksi yang benar; *Precision*, yang menilai proporsi prediksi positif yang benar; *Recall*, yang mengevaluasi proporsi positif aktual yang diidentifikasi dengan benar; *F1-score*, yang menyeimbangkan *precision* dan *recall*; dan Matriks Kebingungan, yang menampilkan distribusi positif dan negatif yang benar dan yang salah, yang menyoroti kesalahan tertentu [9].

III. METODE

Untuk pembuatan model *machine learning* klasifikasi kualitas air Sungai Citarum mengikuti alur gambar *flowchart machine learning* pada Gambar 3.1.



GAMBAR 3.1 Flowchart Model Machine Learning

A. Sumber Dataset

Data yang digunakan untuk mengembangkan model klasifikasi kualitas air Sungai Citarum dikumpulkan dari Dinas Lingkungan Hidup Provinsi Jawa Barat, yang bertanggung jawab atas pemantauan dan pengelolaan lingkungan hidup di wilayah tersebut. Data yang dikumpulkan dari Januari hingga Desember 2023 ini terdiri dari 1010 baris dan 10 kolom. pH (*Potential of Hydrogen*), TSS (Total Suspended Solids), DO (*Dissolved Oxygen*), BOD (*Biochemical Oxygen Demand*), COD (Chemical Oxygen Demand), Nitrat (NO<sub>3</sub>), Fecal Coliform, Fosfat, Indeks Pencemaran, dan Kelas kualitas air merupakan metrik yang penting.

Data tersebut kemudian diekstraksi dan distandarisasi. Selanjutnya, karakteristik penting dipilih dan direkayasa, dan penyeimbangan kelas dilakukan. Data yang telah diproses diubah menjadi matriks fitur yang selanjutnya dipisahkan menjadi data pelatihan sebanyak 80% dan pengujian sebanyak 20%. Bisa dilihat pada Tabel 3.1 *Split Dataset*.

TABEL 3.1  
*Split Dataset*

Data Pelatihan	Data Pengujian
80%	20%

**B. Training dan Atur Parameter Model**

Kemudian lanjut ke proses pembuatan dan training 3 model *machine learning*. Serta penyetelan *hypertuning parameter* agar mendapatkan hasil yang terbaik. Bisa dilihat pada Tabel 3.2 Model *Machine Learning* dan *Parameter*.

TABEL 3.2  
*Model Machine Learning dan Parameter*

Model ML	Parameter
<i>Artificial Neural Network</i>	<ul style="list-style-type: none"> <li>- Input dimension</li> <li>- Dense layers (units: 128, 64, 32, 16, output: y_train.shape[1])</li> <li>- Activation functions (ReLU, softmax)</li> <li>- Batch normalization</li> <li>- Dropout rate (0.3)</li> <li>- Optimizer (Adam, learning rate: 0.01)</li> <li>- Loss function (categorical crossentropy)</li> <li>- Metrics (accuracy)</li> </ul>
<i>Gaussian Naive Bayes</i>	<ul style="list-style-type: none"> <li>- var_smoothing: 1e-3</li> </ul>
Weighted KNN	<ul style="list-style-type: none"> <li>- n_neighbors: 20</li> <li>- metric: euclidean</li> </ul>

Setelah itu didapatkan hasil klasifikasi kualitas air sungai berdasarkan kategori kelas kualitas air sungai yang ada pada Tabel 3.3 Kategori Kualitas Air Sungai sesuai Kelas.

Tabel 3.3  
Kategori Kualitas Air Sungai sesuai Kelas

Kualitas Air Sungai	Kelas
Tidak Tercemar	1
Tercemar Ringan	2
Tercemar Sedang	3

Tercemar Berat	4
----------------	---

**C. Evaluasi Model *Machine Learning***

Metrik evaluasi berguna untuk menganalisis kinerja model pembelajaran mesin. Setiap indikator memberikan wawasan unik tentang kinerja model, membantu dalam pemilihan model yang optimal untuk pekerjaan tertentu. Berikut persamaan dari evaluasi menggunakan berbagai metrik yaitu *confusion matrix*, *accuracy*, *precision*, *recall*, dan *F1-score* [10] [11]. Bisa dilihat pada Tabel 3.4 *Confusion Matrix*.

TABEL 3.4  
*Confusion Matrix*

		Actual Value	
		True	False
Predicted Value	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

Setelah mengetahui hasil pengelompokan kelas-kelas pada kualitas air sungai, kemudian dilakukan penghitungan *accuracy*, *precision*, *recall*, dan *F1-score* dengan persamaan sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

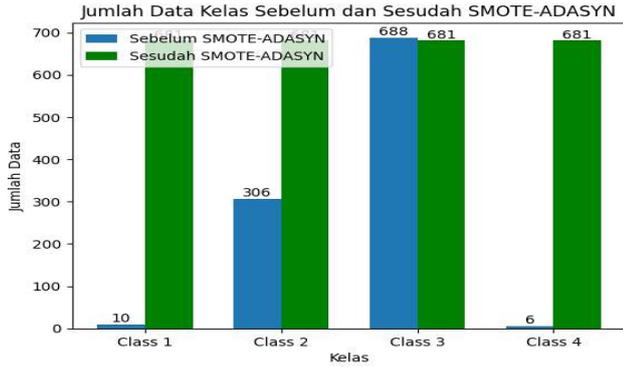
$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

*Accuracy* itu menghitung proporsi prediksi yang benar dari prediksi model secara menyeluruh. *Precision* itu menghitung rasio prediksi positif yang benar untuk setiap prediksi positif yang dibuat oleh model. *Recall* menghitung persen dari semua kasus positif yang ditemukan oleh model. Dan *F1-score* itu skor rata-rata harmonis dari *precision* dan *recall* yang memastikan bahwa keduanya seimbang [11].

**IV. HASIL DAN PEMBAHASAN**

Semua pekerjaan pembuatan dan pengujian model machine learning dilakukan di *Google Colab code editor online* dengan *Python*. Untuk pengujian hasil 3 model machine learning sudah dilakukan dengan dataset kualitas air sungai yang tidak seimbang. Data-data yang tidak seimbang itu data yang ada di kolom kelas kualitas air sungai yaitu kelas 1 sebanyak 10 data, kelas 2 sebanyak 306 data, kelas 3 sebanyak 688 data, dan kelas 4 sebanyak 6 total ada 1010 baris data. Hasil akurasi nya beragam tetapi rendah jauh dari yang diharapkan dan dilihat hasil *confusion matrix* nya model tidak dapat mengklasifikasikan kelas dengan benar. Sehingga solusi untuk mengatasi masalah ini dengan menggunakan

SMOTE-ADASYN untuk menyeimbangkan data-data kelas tersebut.



GAMBAR 4.1 Bar Chart Data Tidak Seimbang dan Seimbang

Pada Gambar 4.1 Bar Chart Data Tidak Seimbang dan Seimbang menunjukkan bahwa data-data sebanyak 1010 baris data yang tidak seimbang tadi, setelah menggunakan SMOTE-ADASYN menjadi seimbang dibagi rata setiap kelas. Kelas 1 ada 681 data, kelas 2 ada 681 data, kelas 3 ada 681 data, dan kelas 4 ada 681. Setelah itu dilakukan pengujian-pengujian 3 model *machine learning* beserta hasil evaluasi modelnya menggunakan tabel hasil akurasi dan gambar grafik *confusion matrix*. Hasil *training* model ANN mendapat 95%, GNB mendapat 91%, dan *Weighted KNN with Euclidean Distance* mendapat 95%.

Berikut ini adalah menampilkan hasilnya dari pengujian-pengujian 3 model *machine learning* yang sudah dilakukan dengan *testing on new data* sebanyak 150 baris data yang belum dilihat oleh model-model tersebut sebagai bentuk validasi model berhasil mengklasifikasikan kualitas air sungai.

1. Artificial Neural Network

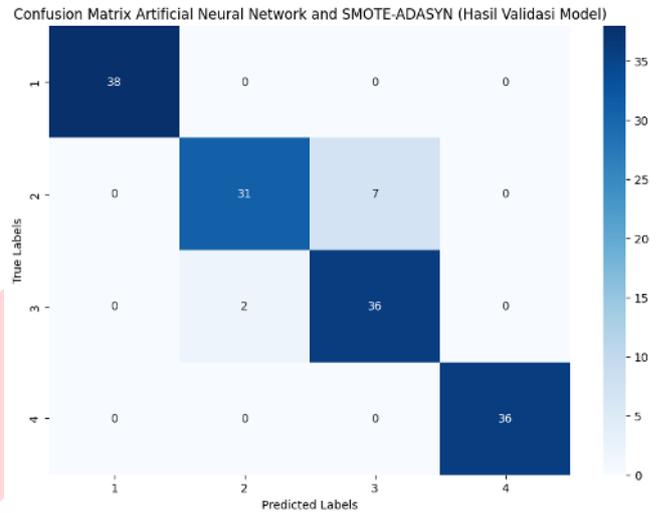
Berikut ini hasil evaluasi model dari ANN dengan tabel yang berisi *accuracy*, *precision*, *recall*, dan *f1-score*. Serta gambar visualisasi grafik *confusion matrix* dari model klasifikasinya.

TABEL 4.1 Classification Report ANN

Artificial Neural Network (Accuracy: 94.1%)			
Kelas	Precision	Recall	F1-Score
Kelas 1	100%	100%	100%
Kelas 2	94%	82%	87%
Kelas 3	84%	95%	89%
Kelas 4	100%	100%	100%

Laporan klasifikasi untuk *Artificial Neural Network* (ANN) bisa dilihat pada Tabel 4.1 *Classification Report ANN* dengan akurasi 94.1% menunjukkan kinerja yang bervariasi di seluruh kelas. Kelas 1 dan 4 bekerja dengan

sangat baik, dengan skor F1 masing-masing 100% dan 100%, sementara Kelas 2 dan 3 memiliki skor F1 yang lebih rendah, yaitu 87% dan 89%. Secara keseluruhan, meskipun memiliki akurasi yang tinggi tetapi model belum mengklasifikasikan dengan baik di kelas 2 bagian *recall* dan juga *F1-score* nya dan kelas 3 di bagian *F1-score* saja.



GAMBAR 4.2 Confusion Matrix ANN

Hasil visualisasi grafik *confusion matrix* untuk model ANN bisa dilihat pada Gambar 4.2 *Confusion Matrix ANN* menunjukkan bahwa klasifikasi yang baik untuk Kelas 1 dan Kelas 4, dengan masing-masing 38 dan 36 prediksi yang akurat. Namun, terdapat kesalahan klasifikasi yang signifikan pada Kelas 2 dan Kelas 3, dengan jumlah 7 prediksi di bagian Kelas 2 yang salah diklasifikasikan sebagai Kelas 3, serta 2 prediksi di bagian Kelas 3 yang salah diklasifikasikan sebagai Kelas 2. Hal ini mencerminkan kinerja yang kuat secara keseluruhan, tetapi menggarisbawahi kekurangan model dalam membedakan antara Kelas 2 dan 3.

2. Gaussian Naive Bayes

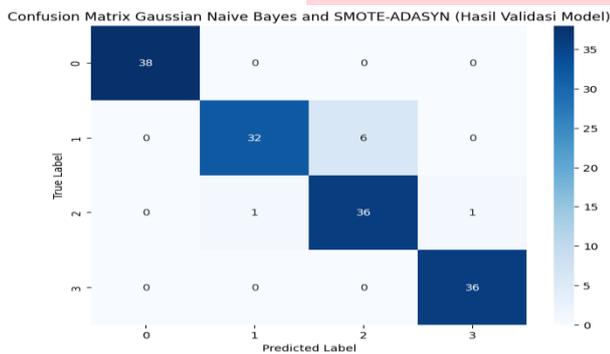
Berikut ini hasil evaluasi model dari GNB pada Tabel 4.4 *Classification Report GNB* yang berisi *accuracy*, *precision*, *recall*, dan *f1-score*. Serta gambar visualisasi grafik *confusion matrix* dari model klasifikasinya.

TABEL 4.2 Classification Report GNB

Gaussian Naive Bayes (Accuracy: 94.7%)			
Kelas	Precision	Recall	F1-Score
Kelas 1	100%	100%	100%
Kelas 2	97%	84%	90%
Kelas 3	86%	95%	90%

Kelas 4	97%	100%	99%
---------	-----	------	-----

Laporan klasifikasi pada Tabel 4.2 *Classification Report* GNB menunjukkan kinerja yang luar biasa di sebagian besar kelas, dengan akurasi 94.7%. Kelas 1 memiliki presisi, *recall*, dan juga skor F1 yang sempurna yaitu dengan hasil 100%. Kelas 4 juga bekerja dengan sangat baik, dengan skor F1 sebesar 99%, *recall* 100%, dan *precision* 97%. Namun, Kelas 2 memiliki kinerja yang agak kurang baik, dengan skor F1 sebesar 90%; meskipun memiliki presisi yang tinggi (97%), *recall* hanya 84%. Kelas 3 memiliki skor F1 90%, dengan presisi 86% dan *recall* 95%. Secara keseluruhan, meskipun GNB sangat akurat, ada kekurangan yang signifikan dalam hasil *recall* di Kelas 2 dan Kelas 3 di bagian *precision*.



GAMBAR 4.3  
Confusion Matrix GNB

Hasil visualisasi grafik *confusion matrix* pada Gambar 4.3 *Confusion Matrix* GNB menunjukkan klasifikasi yang baik untuk Kelas 1 dan 4, masing-masing dengan 38 prediksi yang benar di kelas 1 dan 36 di kelas 4. Namun, beberapa kesalahan klasifikasi terjadi pada Kelas 2 dan Kelas 3, dengan 6 prediksi di Kelas 2 yang salah diklasifikasikan sebagai Kelas 3, 1 prediksi di Kelas 3 yang salah diklasifikasikan sebagai Kelas 2, dan 1 prediksi di Kelas 3 yang salah diklasifikasikan sebagai Kelas 4. Hal ini menunjukkan kinerja yang baik secara keseluruhan, tetapi menyoroti keterbatasan model yang signifikan dalam membedakan antara Kelas 2 dan Kelas 3.

3. *Weighted KNN with Euclidean Distance*

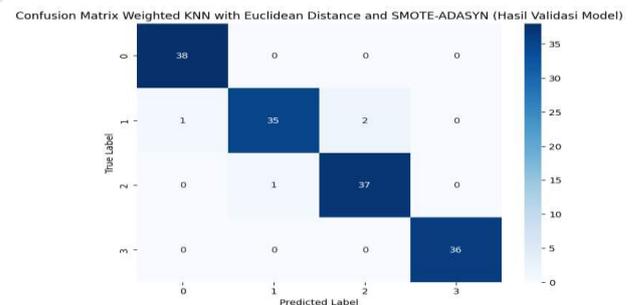
Berikut ini hasil evaluasi model pada Tabel 4.3 *Classification Report* GNB yang berisi *accuracy*, *precision*, *recall*, dan *f1-score*. Serta gambar visualisasi grafik *confusion matrix* dari model klasifikasinya.

TABEL 4.3  
*Classification Report* W-KNN with Euclidean Distance

W-KNN with Euclidean Distance (Accuracy: 97.3%)			
Kelas	Precision	Recall	F1-Score
Kelas 1	97%	100%	99%

Kelas 2	97%	92%	95%
Kelas 3	95%	97%	96%
Kelas 4	100%	100%	100%

Laporan klasifikasi untuk pada Tabel 4.3 *Classification Report* W-KNN with Euclidean Distance menunjukkan kinerja yang baik di sebagian besar kelas, dengan akurasi 97.3%. Kelas 1 memiliki presisi dan *recall* yang baik, dengan *F1-score* sebesar 97% dan *recall* sebesar 100%. Kelas 4 juga bekerja dengan sangat baik, dengan nilai F1 sebesar 100%, *recall* dan *precision* sama-sama mendapat 100%. Di Kelas 2 memiliki kinerja yang baik pula, dengan skor F1 95%, untuk presisi sebesar 97%, dan *recall* sebesar 92%. Kelas 3 memiliki skor F1 96%, dengan presisi 95% dan *recall* 97%. Secara keseluruhan, W-KNN sangat akurat untuk mengklasifikasikan tiap kelas-kelas kualitas air sungai dibandingkan 2 model lainnya yaitu ANN dan GNB.



GAMBAR 4.4  
Confusion Matrix W-KNN with Euclidean Distance

Hasil visualisasi grafik pada Gambar 4.4 *Confusion Matrix* W-KNN with Euclidean Distance menunjukkan klasifikasi yang baik untuk Kelas 1 dan 4, masing-masing dengan 38 prediksi di kelas 1 dan 36 prediksi di kelas 4 yang benar. Namun, terdapat kesalahan klasifikasi yang sedikit kurang baik pada Kelas 2 dan 3. 1 prediksi di bagian Kelas 2 salah diklasifikasikan sebagai Kelas 1, 2 prediksi di bagian Kelas 2 juga salah diklasifikasikan sebagai Kelas 3, serta ada 1 prediksi di bagian Kelas 3 yang salah juga diklasifikasikan sebagai Kelas 2. Hal ini menunjukkan kinerja yang baik secara keseluruhan, namun menunjukkan ada sedikit kekurangan saja dalam membedakan antara Kelas 1, 2, dan 3.

V. KESIMPULAN

Berdasarkan percobaan dan pengujian 3 model *machine learning* yang sudah dilakukan dilihat hasilnya dari masing-masing *classification report* dan visualisasi grafik *confusion matrix* tiap model untuk melakukan klasifikasi kualitas air Sungai Citarum menggunakan dataset dari Dinas Lingkungan Hidup Provinsi Jawa Barat. Didapatkan hasil akurasi yang baik dari *Weighted KNN with Euclidean Distance* mengungguli dua model lainnya untuk mengklasifikasikan kualitas air sungai, dengan akurasi sebesar 97,3%. Sedangkan *Artificial Neural Network* mendapat akurasi sebesar 94,1%

dan *Gaussian Naive Bayes* mendapat akurasi sebesar 94,7%. Sehingga hal ini menunjukkan model-model tersebut dapat mengklasifikasikan kualitas air sungai sesuai kelasnya yaitu kelas 1 tidak tercemar atau memenuhi baku mutu, kelas 2 tercemar ringan, kelas 3 tercemar sedang, dan kelas 4 tercemar berat.

#### REFERENSI

- [1] A. W. Utami, "Kualitas air sungai Citarum," 2019
- [2] A. Danades, D. Pratama, D. Anggraini, dan D. Anggriani, "Comparison of accuracy level K-Nearest Neighbor algorithm and Support Vector Machine algorithm in classification water quality status," 2017. doi: 10.1109/icsengt.2016.7849638.
- [3] L. A. Putri dan Suwanda, "Implementasi Metode Artificial Neural Network (ANN) Algoritma Backpropagation untuk Klasifikasi Kualitas Udara di Provinsi DKI Jakarta Tahun 2021," Bandung Conference Series: Statistics, vol. 3, no. 2, 2023, doi: 10.29313/bcss.v3i2.7826.
- [4] Q. Hasanah, H. Oktavianto, dan Y. D. Rahayu, "Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung," Jurnal Smart Teknologi, vol. 3, no. 4, 2022.
- [5] A. Mustofa, O. Okfalisa, E. P. Cynthia, Y. Yelfi, dan S. K. Gusti, "Klasifikasi Penerima Bantuan Covid-19 Menggunakan Metode Weighted K-Nearest Neighbour," Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI), vol. 5, no. 3, 2022, doi: 10.32672/jnkti.v5i3.4399.
- [6] N. V. Chawla, K. W. Bowyer, L. O. Hall, dan W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, 2002, doi: 10.1613/jair.953.
- [7] H. He, Y. Bai, E. A. Garcia, dan S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," dalam Proceedings of the International Joint Conference on Neural Networks, 2008. doi: 10.1109/IJCNN.2008.4633969.
- [8] E. Retnoningsih dan R. Pramudita, "Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python," BINA INSANI ICT JOURNAL, vol. 7, no. 2, 2020, doi: 10.51211/biict.v7i2.1422.
- [9] I. M. Esa Darmayasa Adi Putra, I. Fauzi, K. S. Krishna Prasad, K. S. Krishna Prasad, I. M. Putra Arya Winata, dan I. W. Widhiada, "Evaluasi Model Machine Learning Klasifikasi Gerak Tangan Untuk Sistem Kontrol Prototipe Protesis Tangan," Majalah Ilmiah Teknologi Elektro, vol. 22, no. 1, 2023, doi: 10.24843/mite.2023.v22i01.p18.
- [10] Ž. Vujović, "Classification Model Evaluation Metrics," International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.
- [11] A. Sharma, "Confusion Matrix in Machine Learning," [www.Geeksforgeeks.org](http://www.Geeksforgeeks.org), 2018.
- [12] R. K. Dinata, H. Akbar, dan N. Hasdyna, "Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus," ILKOM Jurnal Ilmiah, vol. 12, no. 2, 2020, doi: 10.33096/ilkom.v12i2.539.104-111.
- [13] D. Irawan Saputra dan D. L. Hakim, "Implementasi Algoritma Gaussian Naive Bayes Classifier Untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler," EPSILON: Journal of Electrical Engineering and Information Technology, vol. 20, no. 2, 2022, doi: 10.55893/epsilon.v20i2.94.
- [14] H. Hananti dan K. Sari, "Perbandingan Metode Support Vector Machine (SVM) dan Artificial Neural Network (ANN) pada Klasifikasi Gizi Balita," Seminar Nasional Official Statistics, vol. 2021, no. 1, 2021, doi: 10.34123/semnasoffstat.v2021i1.1014.
- [15] A. Kaur, M. Khurana, P. Kaur, dan M. Kaur, "Classification and Analysis of Water Quality Using Machine Learning Algorithms," dalam Lecture Notes in Electrical Engineering, 2021. doi: 10.1007/978-981-33-4866-0\_48.
- [16] S. Y. Abuzir dan Y. S. Abuzir, "Machine learning for water quality classification," Water Quality Research Journal, vol. 57, no. 3, 2022, doi: 10.2166/wqrj.2022.004.
- [17] E. K. Sari dan O. E. Wijaya, "Penentuan Status Mutu Air Dengan Metode Indeks Pencemaran Dan Strategi Pengendalian Pencemaran Sungai Ogan Kabupaten Ogan Komering Ulu," Jurnal Ilmu Lingkungan, vol. 17, no. 3, 2019, doi: 10.14710/jil.17.3.486-491.
- [18] Kementerian Lingkungan Hidup Republik Indonesia, "Keputusan Menteri Lingkungan Hidup No. 115 Tahun 2003 tentang Pedoman Penentuan Status Mutu Air," Jakarta, Indonesia: KLHK, 2003.
- [19] Peraturan Pemerintah Republik Indonesia, "Peraturan Pemerintah No. 82 Tahun 2001 tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air," Jakarta, Indonesia: 2001.
- [20] K. L. Hidup and K. R. Indonesia, "Indeks Kualitas Lingkungan Hidup Tahun 2013," 2014. [Daring]. Tersedia: <https://ppkl.menlhk.go.id/website/filebox/1149/230626135350IKLH%202013.pdf>