

Klasifikasi Kelayakan Air Minum Berbasis Pembelajaran Mesin Menggunakan *Artificial Neural Network* dan *Random Forest*

Fega Yudistira¹, Dilla Restu Agusthiani², Eki Ahmad Zaki Hamidi³, Edi Mulyana⁴

^{1,2,3}Teknik Elektro, UIN Sunan Gunung Djati Bandung, Jl. AH. Nasution 105 Bandung, 40614, Indonesia

Korespondensi: dillarestu30@gmail.com

ARTICLE HISTORY

Received:18-06-2026

Revised:24-06-2026

Accepted:25-06-2026

Abstrak

Kualitas air minum merupakan faktor penting yang memengaruhi kesehatan masyarakat. Klasifikasi kelayakan air minum yang akurat diperlukan untuk memastikan keamanan air yang dikonsumsi. Penelitian ini bertujuan untuk membandingkan performa algoritma *Artificial Neural Network* (ANN) dan *Random Forest* dalam klasifikasi kelayakan air minum menggunakan dataset *Water Potability*. Dataset yang digunakan terdiri atas 3.276 sampel dengan sembilan parameter kualitas air, yaitu *pH*, *hardness*, *solids*, *chloramines*, *sulfate*, *conductivity*, *organic carbon*, *trihalomethanes*, dan *turbidity*. Tahap prapemrosesan data meliputi penanganan *missing value*, normalisasi menggunakan *MinMaxScaler*, serta pembagian data latih dan data uji dengan skenario 70:30 dan 80:20. Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil pengujian menunjukkan bahwa skenario pembagian data 80:20 menghasilkan performa terbaik pada kedua model. Pada skenario tersebut, ANN memperoleh nilai *accuracy* sebesar 0,66, *precision* 0,67, *recall* 0,88, dan *F1-score* 0,76, sedangkan *Random Forest* memperoleh *accuracy* 0,66, *precision* 0,66, *recall* 0,87, dan *F1-score* 0,75. Hasil penelitian menunjukkan bahwa ANN memperoleh nilai *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dibandingkan *Random Forest*. Namun demikian, perbedaan performa kedua model relatif kecil sehingga keduanya dapat dianggap memiliki kemampuan yang kompetitif dalam klasifikasi kelayakan air minum.

Kata kunci: Air Minum, Pembelajaran Mesin, *Artificial Neural Network*, *Random Forest*, Klasifikasi.

Machine Learning-Based Drinking Water Potability Classification Using *Artificial Neural Network* and *Random Forest*

Abstract

Drinking water quality is an important factor affecting public health. Accurate water potability classification is essential to ensure safe water consumption. This study aims to compare the performance of *Artificial Neural Network* (ANN) and *Random Forest* algorithms for drinking water potability classification using the *Water Potability* dataset. The dataset consists of 3,276 samples with nine water quality parameters, including *pH*, *hardness*, *solids*, *chloramines*, *sulfate*, *conductivity*, *organic carbon*, *trihalomethanes*, and *turbidity*. Data preprocessing involved missing value handling, normalization using

MinMaxScaler, and train-test splitting with 70:30 and 80:20 scenarios. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. Experimental results show that the 80:20 train-test split produced the best performance for both models. Under this scenario, ANN achieved an accuracy of 0.66, precision of 0.67, recall of 0.88, and F1-score of 0.76, while Random Forest achieved an accuracy of 0.66, precision of 0.66, recall of 0.87, and F1-score of 0.75. The results showed that ANN achieved slightly higher precision, recall, and F1-score values than Random Forest. However, the performance difference between the two models was relatively small, so both models can be considered competitive in classifying drinking water quality.

Key words: *Drinking Water, Machine Learning, Artificial Neural Network, Random Forest, Classification.*

1. Pendahuluan

Air minum merupakan kebutuhan dasar manusia yang berperan penting dalam menjaga kesehatan dan kualitas hidup masyarakat. Ketersediaan air minum yang aman juga menjadi salah satu indikator *Sustainable Development Goals* (SDGs), khususnya Tujuan 6 (*Clean Water and Sanitation*). Namun, peningkatan aktivitas industri, urbanisasi, pertumbuhan penduduk, dan perubahan lingkungan menyebabkan kualitas sumber air semakin rentan terhadap pencemaran fisik, kimia, maupun biologis. Kondisi tersebut meningkatkan risiko berbagai penyakit akibat konsumsi air yang tidak memenuhi standar kualitas, sehingga pemantauan dan evaluasi kualitas air minum menjadi aspek penting dalam menjaga kesehatan masyarakat dan mendukung pengelolaan sumber daya air yang berkelanjutan [1][2].

Penentuan kelayakan air minum secara konvensional umumnya dilakukan melalui pengujian laboratorium berdasarkan berbagai parameter kualitas air. Meskipun memiliki tingkat akurasi yang tinggi, metode ini memerlukan waktu, biaya, dan sumber daya yang relatif besar sehingga kurang efisien untuk proses evaluasi dalam skala besar [3][4].

Perkembangan teknologi machine learning memberikan peluang dalam meningkatkan efisiensi klasifikasi kualitas air. Berbagai algoritma, seperti Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbor (KNN), Random Forest, dan Artificial Neural Network (ANN), telah diterapkan untuk mengidentifikasi pola hubungan antarparameter kualitas air dan menentukan status kelayakan air minum secara otomatis. Di antara algoritma tersebut, ANN unggul dalam memodelkan hubungan nonlinier yang kompleks, sedangkan Random Forest memiliki kelebihan dalam stabilitas model, kemampuan generalisasi, dan ketahanan terhadap overfitting [5][6][7].

Berbagai penelitian terdahulu menunjukkan bahwa algoritma *machine learning* mampu menghasilkan performa yang cukup baik dalam klasifikasi kualitas air minum. Dwi Hartanti dkk [8]. membandingkan *Decision Tree, Logistic Regression, SVM, dan ANN*, serta menemukan bahwa ANN menghasilkan akurasi tertinggi sebesar 69,54%. M. Rizky Fajar dkk[9]. membandingkan *Random Forest* dan ANN, dengan *Random Forest* mencapai akurasi 67,823%. Al-Adhaileh dan Alsaade [10] mengembangkan model berbasis ANFIS, FFNN, dan KNN, di mana FFNN mencapai akurasi klasifikasi 100%,

sedangkan ANFIS memperoleh koefisien korelasi sebesar 96,17% pada prediksi *Water Quality Index*. Penelitian terbaru oleh Ishak Bintang dkk [11]. menerapkan teknik *stacking ensemble* yang menggabungkan LGBM, *Random Forest*, *Gradient Boosting*, dan *XGBoost* dengan optimasi *Bayesian Search CV* serta SMOTE untuk menangani ketidakseimbangan data. Model *stacking* yang dikembangkan mampu mencapai akurasi 85,38%, *precision* 88,02%, *recall* 85,38%, dan *F1-score* 85,23%, melampaui performa model individual. Temuan-temuan tersebut menunjukkan bahwa penggunaan metode ensemble dan teknik penanganan data yang tepat mampu meningkatkan performa klasifikasi kualitas air secara signifikan.

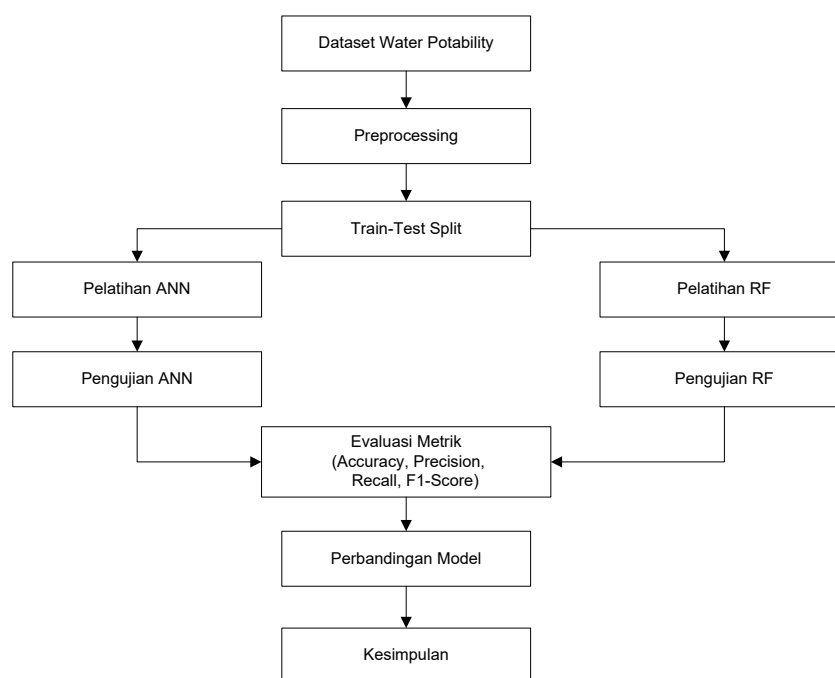
Meskipun berbagai penelitian telah berhasil meningkatkan akurasi klasifikasi, masih terdapat beberapa *research gap*. Sebagian besar penelitian berfokus pada penerapan metode *ensemble*, optimasi hiperparameter, atau teknik penyeimbangan data sehingga evaluasi terhadap performa dasar masing-masing algoritma belum banyak dilakukan. Selain itu, hasil penelitian sebelumnya menunjukkan performa yang berbeda antara ANN dan *Random Forest* meskipun menggunakan *dataset* yang serupa. Perbedaan tersebut mengindikasikan bahwa karakteristik algoritma maupun skenario pembagian data dapat memengaruhi hasil klasifikasi. Namun, penelitian yang membandingkan kedua algoritma secara langsung dengan prosedur *preprocessing* yang identik serta variasi pembagian data latih dan data uji masih relatif terbatas.

Berdasarkan permasalahan tersebut, penelitian ini melakukan analisis komparatif antara *Artificial Neural Network* (ANN) dan *Random Forest* pada *dataset Water Potability* menggunakan tahapan *preprocessing* yang identik. Selain membandingkan performa kedua algoritma, penelitian ini juga mengevaluasi pengaruh variasi pembagian data latih dan data uji (70:30 dan 80:20) terhadap kemampuan klasifikasi model. Evaluasi dilakukan menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* yang diperoleh dari *confusion matrix*. Kontribusi utama penelitian ini meliputi: (1) menyajikan evaluasi komparatif antara ANN dan *Random Forest* menggunakan prosedur *preprocessing* yang identik pada *dataset Water Potability*; (2) menganalisis pengaruh variasi pembagian data latih dan data uji terhadap performa klasifikasi kualitas air minum; serta (3) memberikan analisis *confusion matrix* untuk mengidentifikasi karakteristik kesalahan klasifikasi pada setiap model. Hasil penelitian diharapkan memberikan pemahaman yang lebih komprehensif mengenai karakteristik dan performa ANN serta *Random Forest* dalam klasifikasi kelayakan air minum, sekaligus menjadi referensi bagi pengembangan sistem pemantauan kualitas air berbasis pembelajaran mesin.

2. Metode

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk membangun dan mengevaluasi model klasifikasi kelayakan air minum menggunakan algoritma *Artificial Neural Network* (ANN) dan *Random Forest*. Tahapan penelitian dimulai dari pengumpulan *dataset Water Potability*, dilanjutkan dengan proses prapemrosesan data yang meliputi penanganan *missing value*, normalisasi data menggunakan *MinMaxScaler*, serta pembagian data menjadi data latih dan data uji. Selanjutnya, model ANN dan *Random Forest* dilatih menggunakan data latih dan diuji menggunakan data uji untuk memperoleh hasil klasifikasi. Performa kedua model kemudian dievaluasi menggunakan

metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Alur tahapan penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Flowchart alur penelitian

2.1. Dataset

Dataset yang digunakan dalam penelitian ini adalah *Water Potability Dataset* yang diperoleh dari Kaggle. Dataset ini terdiri atas 3.276 sampel dengan sembilan parameter kualitas air, yaitu *pH*, *Hardness*, *Solids*, *Chloramines*, *Sulfate*, *Conductivity*, *Organic Carbon*, *Trihalomethanes*, dan *Turbidity*. Kesembilan parameter tersebut digunakan sebagai fitur untuk menentukan kelayakan air minum. Variabel target yang digunakan adalah **Potability**, dengan nilai **1 (Potable)** yang menunjukkan air layak dikonsumsi dan **0 (Non-potable)** yang menunjukkan air tidak layak dikonsumsi. Distribusi jumlah sampel pada masing-masing kelas ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Kelas Dataset Water Potability

Kelas	Jumlah Sampel
Non-potable (0)	1998
Potable (1)	1278
Total	3276

Berdasarkan Tabel 1, terlihat bahwa distribusi kelas pada dataset tidak seimbang (*class imbalance*). Kelas *Non-potable* memiliki proporsi sebesar 60,98%, sedangkan kelas *Potable* memiliki proporsi sebesar 39,02%. Ketidakeimbangan kelas ini berpotensi menyebabkan model lebih cenderung mengenali kelas mayoritas dibandingkan kelas minoritas. Oleh karena itu, penelitian ini tidak hanya menggunakan metrik *accuracy*, tetapi juga *precision*, *recall*, dan *F1-score* untuk memberikan evaluasi yang lebih representatif terhadap performa model. Selain itu, hasil eksplorasi data menunjukkan adanya nilai yang hilang (*missing value*) pada beberapa atribut, yaitu *pH* sebanyak 491 data, *Sulfate* sebanyak 781 data, dan *Trihalomethanes* sebanyak 162 data. Keberadaan

missing value tersebut ditangani pada tahap *preprocessing* sebelum data digunakan dalam proses pelatihan dan pengujian model. Dengan memanfaatkan dataset tersebut, penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Artificial Neural Network* (ANN) dan *Random Forest* dalam melakukan klasifikasi kelayakan air minum secara otomatis.

2.2.Data Preprocessing

Tahap *data preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model *machine learning*. Proses ini bertujuan untuk menghasilkan data yang bersih, konsisten, dan siap dianalisis sehingga dapat meningkatkan performa model klasifikasi. Tahapan pertama adalah penanganan *missing value* pada atribut yang memiliki nilai kosong. Hasil eksplorasi menunjukkan bahwa atribut **pH**, **Sulfate**, dan **Trihalomethanes** masing-masing memiliki 491, 781, dan 162 data yang hilang. Apabila tidak ditangani, keberadaan *missing value* dapat memengaruhi proses pembelajaran model dan menurunkan akurasi klasifikasi. Penelitian ini menggunakan *median imputation* untuk menggantikan nilai yang hilang karena lebih robust terhadap *outlier* dibandingkan *mean imputation* serta mampu mempertahankan distribusi data. Selanjutnya, dilakukan normalisasi menggunakan *MinMaxScaler* untuk mengubah rentang nilai setiap fitur menjadi 0–1 sehingga skala antarfitur seragam dan proses pembelajaran model menjadi lebih optimal.

2.3.Artificial Neural Network (ANN)

Artificial Neural Network (ANN) merupakan algoritma *machine learning* yang mampu mempelajari hubungan nonlinier antarvariabel. Pada penelitian ini, ANN digunakan untuk mengklasifikasikan kelayakan air minum berdasarkan sembilan parameter kualitas air pada dataset *Water Potability*. Arsitektur jaringan terdiri atas satu input layer dengan 9 neuron, dua hidden layer masing-masing berisi 32 dan 16 neuron, serta satu output layer dengan satu neuron untuk menghasilkan klasifikasi biner antara air layak minum (*potable*) dan tidak layak minum (*non-potable*) [12][13]. Model menggunakan fungsi aktivasi ReLU pada *hidden layer*, fungsi aktivasi Sigmoid pada *output layer*, optimizer Adam, dan *loss function* Binary Crossentropy. Proses pelatihan dilakukan selama 100 *epoch* untuk memperoleh performa klasifikasi yang optimal[14].

2.4.Random Forest

Random Forest merupakan algoritma *machine learning* berbasis *ensemble* yang menggabungkan sejumlah *decision tree* melalui mekanisme *majority voting* untuk menghasilkan klasifikasi yang lebih akurat dan stabil. Pada penelitian ini, *Random Forest* digunakan untuk mengklasifikasikan kelayakan air minum berdasarkan sembilan parameter kualitas air pada dataset *Water Potability*. Model dikonfigurasi dengan *n_estimators = 100*, *max_depth = None*, dan *random_state = 42* untuk memperoleh proses pelatihan yang optimal, konsisten, dan dapat direproduksi. Konfigurasi tersebut diharapkan mampu menghasilkan performa klasifikasi yang akurat dan andal [15][16][17].

2.5.Matrik Evaluasi

Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung berdasarkan *confusion matrix* yang terdiri atas *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Penggunaan keempat

metrik tersebut memberikan evaluasi yang komprehensif terhadap kemampuan model dalam melakukan klasifikasi [18][11]. *Accuracy* mengukur proporsi prediksi yang benar terhadap seluruh data, *precision* menunjukkan ketepatan prediksi kelas positif, *recall* mengukur kemampuan model dalam mendeteksi data positif, sedangkan *F1-score* merupakan rata-rata harmonik antara *precision* dan *recall*. Persamaan perhitungan masing-masing metrik ditunjukkan pada Persamaan (1)–(4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

3. Hasil dan Pembahasan

Penelitian ini membandingkan performa algoritma *Artificial Neural Network* (ANN) dan *Random Forest* dalam mengklasifikasikan kelayakan air minum menggunakan *dataset* *Water Potability*. Pengujian dilakukan dengan dua skenario pembagian data, yaitu 70:30 dan 80:20, untuk menganalisis pengaruh proporsi data latih dan data uji terhadap performa model. Evaluasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* yang dihitung berdasarkan *confusion matrix*.

3.1. Hasil Evaluasi Model

Hasil pengujian model ANN dan *Random Forest* pada kedua skenario pembagian data disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model

Model	Variasi Dataset	Akurasi	Precision	Recall	F1-Score
<i>Random Forest</i>	70:30	0.63	0.60	0.45	0.51
ANN	70:30	0.65	0.63	0.48	0.55
<i>Random Forest</i>	80:20	0.66	0.66	0.87	0.75
ANN	80:20	0.66	0.67	0.88	0.76

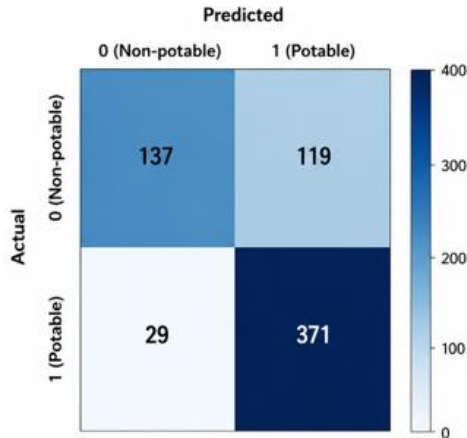
Berdasarkan Tabel 2, kedua algoritma menunjukkan peningkatan performa pada pembagian data 80:20 dibandingkan 70:30. Pada skenario 70:30, ANN memperoleh *accuracy* sebesar 0,65, lebih tinggi dibandingkan *Random Forest* sebesar 0,63. ANN juga menghasilkan nilai *precision*, *recall*, dan *F1-score* yang lebih baik, yaitu masing-masing 0,63, 0,48, dan 0,55, sedangkan *Random Forest* memperoleh 0,60, 0,45, dan 0,51.

Pada skenario 80:20, kedua model memperoleh *accuracy* yang sama, yaitu 0,66. Namun, ANN tetap menunjukkan performa yang sedikit lebih baik dengan *precision* 0,67, *recall* 0,88, dan *F1-score* 0,76, sedangkan *Random Forest* memperoleh *precision* 0,66, *recall* 0,87, dan *F1-score* 0,75. Tingginya nilai *recall* pada kedua model menunjukkan

kemampuan yang baik dalam mengidentifikasi air layak minum sehingga kesalahan klasifikasi pada kelas positif dapat diminimalkan.

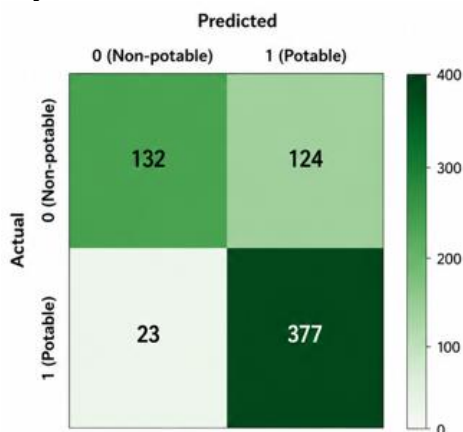
3.2. Confusion Matrix Model

Untuk menganalisis performa klasifikasi secara lebih rinci, digunakan *confusion matrix* yang menunjukkan jumlah prediksi benar dan salah pada setiap kelas. Gambar 2 menunjukkan bahwa model Random Forest mampu mengidentifikasi sebagian besar sampel air layak minum (*potable*) dengan benar, yang ditunjukkan oleh jumlah *True Positive* (TP) lebih besar daripada *False Negative* (FN). Hasil ini sejalan dengan nilai *recall* sebesar 0,87 pada Tabel 2.



Gambar 2. Confusion Matrix Random Forest

Selain itu, nilai *precision* sebesar 0,66 menunjukkan bahwa sebagian besar prediksi air layak minum yang dihasilkan model sesuai dengan kondisi sebenarnya. Dengan *accuracy* sebesar 0,66 dan *F1-score* sebesar 0,75, Random Forest menunjukkan performa klasifikasi yang cukup baik dalam membedakan air layak minum dan tidak layak minum pada *dataset* Water Potability.



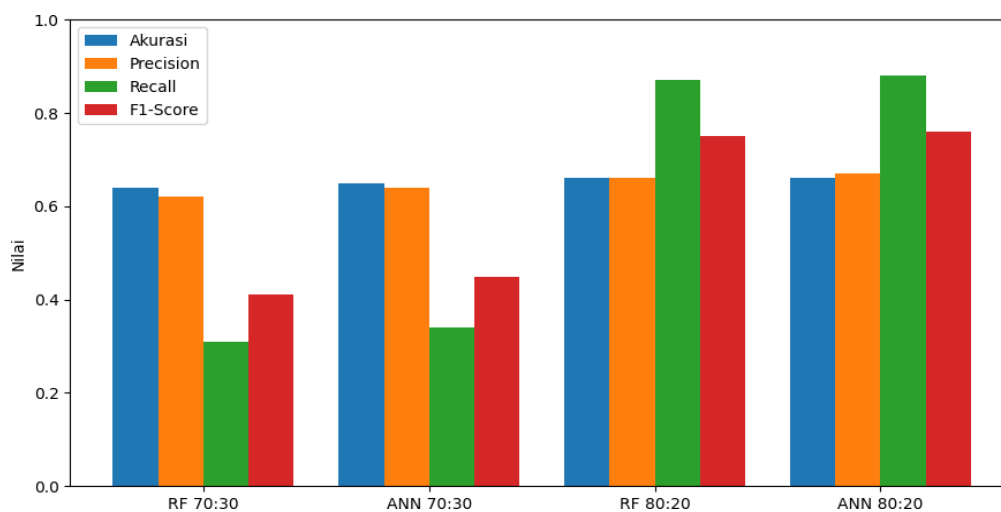
Gambar 3. Confusion Matrix Artificial Neural Network (ANN)

Gambar 3 menunjukkan *confusion matrix* model *Artificial Neural Network* (ANN). Hasil pengujian menunjukkan bahwa ANN mampu mengklasifikasikan sebagian besar sampel air layak minum (*potable*) dengan benar, yang ditunjukkan oleh jumlah *True Positive* (TP) yang lebih tinggi dibandingkan *False Negative* (FN). Temuan ini sejalan dengan nilai *recall* sebesar 0,88, *precision* sebesar 0,67, dan *F1-score* sebesar 0,76 yang diperoleh model. Berdasarkan *confusion matrix* dan metrik evaluasi pada Tabel 2, kedua model menunjukkan performa yang baik dalam klasifikasi kelayakan air minum. Namun, ANN

memperoleh nilai *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dibandingkan Random Forest. Meskipun demikian, perbedaan performa kedua model relatif kecil sehingga keduanya dapat dianggap memiliki kemampuan klasifikasi yang kompetitif pada *dataset* Water Potability.

3.3. Analisis Visualisasi Kinerja Model

Untuk memperjelas perbandingan performa model, hasil evaluasi divisualisasikan dalam bentuk diagram batang sebagaimana ditunjukkan pada Gambar 4.



Gambar 4. Perbandingan Kinerja Model Random Forest dan ANN

Gambar 4 menunjukkan perbandingan nilai *accuracy*, *precision*, *recall*, dan *F1-score* antara Random Forest dan ANN pada dua skenario pembagian *dataset*. Secara umum, pembagian data 80:20 menghasilkan performa yang lebih baik dibandingkan 70:30 pada kedua algoritma. Hal ini menunjukkan bahwa proporsi data latih yang lebih besar memungkinkan model mempelajari pola data dengan lebih baik sehingga meningkatkan kemampuan generalisasi. Peningkatan performa paling terlihat pada nilai *recall*, yang meningkat dari 0,45 menjadi 0,87 pada Random Forest dan dari 0,48 menjadi 0,88 pada ANN. Selain itu, nilai *F1-score* juga meningkat dari 0,51 menjadi 0,75 pada Random Forest dan dari 0,55 menjadi 0,76 pada ANN. Hasil tersebut menunjukkan bahwa kedua model menjadi lebih efektif dalam mendeteksi kelas positif sekaligus menghasilkan keseimbangan yang lebih baik antara *precision* dan *recall*.

3.4. Perbandingan ANN dan Random Forest

Hasil penelitian menunjukkan bahwa ANN memperoleh nilai *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dibandingkan Random Forest. Namun, perbedaan performa kedua model relatif kecil sehingga keduanya memiliki kemampuan yang kompetitif dalam mengklasifikasikan kelayakan air minum. Temuan ini sejalan dengan penelitian Dwi Hartanti dkk. yang melaporkan performa ANN yang baik pada klasifikasi kualitas air, serta menunjukkan bahwa kinerja model *machine learning* dipengaruhi oleh karakteristik *dataset*, tahapan *preprocessing*, dan konfigurasi parameter model.

3.5. Pembahasan Hasil Penelitian

Berdasarkan hasil pengujian, pembagian data 80:20 menghasilkan performa yang lebih baik dibandingkan 70:30, menunjukkan bahwa proporsi data latih yang lebih besar

mampu meningkatkan kemampuan model dalam mempelajari pola data. Selain itu, ANN memperoleh nilai *precision*, *recall*, dan *F1-score* yang sedikit lebih tinggi dibandingkan Random Forest. Meskipun demikian, selisih performa kedua algoritma relatif kecil sehingga keduanya memiliki kemampuan klasifikasi yang kompetitif pada *dataset* Water Potability. Oleh karena itu, ANN maupun Random Forest dapat dipertimbangkan sebagai alternatif yang efektif untuk diterapkan pada sistem klasifikasi kelayakan air minum.

4. Kesimpulan

Penelitian ini bertujuan membandingkan performa algoritma *Artificial Neural Network* (ANN) dan Random Forest dalam klasifikasi kelayakan air minum menggunakan *dataset* Water Potability. Tahapan penelitian meliputi *data preprocessing* berupa penanganan *missing value*, normalisasi menggunakan *MinMaxScaler*, serta pembagian data menjadi data latih dan data uji dengan skenario 70:30 dan 80:20. Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa pembagian data 80:20 menghasilkan performa terbaik pada kedua algoritma. Pada skenario tersebut, kedua model memperoleh *accuracy* sebesar 0,66, sedangkan ANN menghasilkan *precision* 0,67, *recall* 0,88, dan *F1-score* 0,76, sedikit lebih tinggi dibandingkan Random Forest yang memperoleh *precision* 0,66, *recall* 0,87, dan *F1-score* 0,75. Meskipun demikian, selisih performa kedua model relatif kecil sehingga ANN dan Random Forest sama-sama menunjukkan kemampuan klasifikasi yang kompetitif. Oleh karena itu, kedua algoritma layak diterapkan untuk klasifikasi kelayakan air minum, dengan pemilihan model dapat disesuaikan dengan kebutuhan implementasi, kompleksitas sistem, dan sumber daya komputasi yang tersedia.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Dekan Fakultas Sains dan Teknologi serta Ketua Jurusan Teknik Elektro UIN Sunan Gunung Djati Bandung atas dukungan, arahan, dan motivasi yang diberikan sehingga penelitian ini dapat terlaksana dengan baik.

Daftar Pustaka

- [1] A. K. Verma and S. Prakash, "WATER RESOURCES IN A CHANGING WORLD : SCIENCE , MANAGEMENT AND POLICY Water and Its Impact on Human Health," pp. 61–63, 2025, doi: 10.5281/zenodo.15183465.
- [2] D. R. Tripathi, "Safe Drinking Water-Threats of Pollution to Potable Water For The Community," *J. Surv. Fish. Sci.*, vol. 10, no. 2, pp. 1854–1858, 2023, doi: 10.53555/g86wes45.
- [3] E. R. G. Ramakanth Reddy Tetali, Salomi K, BNV Sai Durga G, Sharon Pushpa P, "Analysis of Water Quality Parameters Across Diverse Sources: Research Article," *J. Pharma Insights Res.*, vol. 2, no. 3 SE-Articles, pp. 210–216, doi: 10.69613/3jxm7e23.
- [4] M. H. Pawar, "Advancements and Challenges in Water Analysis : A Comprehensive Review of Methods and Applications," 2025.

- [5] J. Pandey and S. Verma, "Machine Learning Model for Water Quality Analytics BT - Data Science and Big Data Analytics," D. Mishra, X. S. Yang, A. Unal, and D. S. Jat, Eds., Singapore: Springer Nature Singapore, 2024, pp. 719–729.
- [6] M. V. Subbarao, "Enhancing Water Quality Monitoring : Machine Learning-Based Estimation of Water Quality Index (WQI)," vol. 10, no. 4, pp. 189–195, 2025.
- [7] D. Chandu and D. B. Bidikar, "Random Forest Based ML Model for Potability of Water," *Int. J. Sci. Res. Eng. Manag.*, vol. 09, no. 07, pp. 1–9, 2025, doi: 10.55041/ijresrem51665.
- [8] D. Hartanti and A. I. Pradana, "Komparasi Algoritma Machine Learning dalam Identifikasi Kualitas Air," *SMARTICS J.*, vol. 9, no. 1, pp. 1–6, 2023.
- [9] I. Binanto, M. R. F. Mali, B. A. D. N, and A. Jaya, "Perbandingan Algoritma Random Forest Dan Artificial Neural Network Untuk Dataset Water Potability," *Jtriste*, vol. 11, no. 1, pp. 53–59, 2024, doi: 10.55645/jtriste.v11i1.510.
- [10] M. H. Al-Adhaileh and F. W. Alsaade, "Modelling and prediction of water quality by using artificial intelligence," *Sustain.*, vol. 13, no. 8, pp. 1–18, 2021, doi: 10.3390/su13084259.
- [11] I. B. D, P. N. Andono, R. A. Pramunendar, and A. Winarno, "Pendekatan Machine Learning dengan Teknik Stacking untuk Memprediksi Kualitas Air Minum," vol. 6, no. 4, pp. 2546–2558, 2025, doi: 10.47065/bits.v6i4.7014.
- [12] S. Y. Kurniawan, S. Sanjaya, Y. Vitriani, and I. Afrianty, "Klasifikasi Kelayakan Air Minum dengan Backpropagation Neural Network Berbasis Penanganan Missing Value dan Normalisasi," *J. Inf. Syst. Res.*, vol. 6, no. 1, pp. 76–84, 2024, doi: 10.47065/josh.v6i1.5871.
- [13] I. Darmawan, Muhammad Fatchan, and Andri Firmansyah, "Classification of Drinking Water Potability With Artificial Neural Network Algorithm," *Int. J. Integr. Sci. Technol.*, vol. 2, no. 5, pp. 506–515, 2024, doi: 10.59890/ijist.v2i5.1874.
- [14] E. Systems, A. Al-ataby, B. N. Getu, and H. Attia, "Machine Learning-Based Water Quality Prediction," pp. 1–20, 2026.
- [15] F. Bouchraki, S. Hamchaoui, L. L. Ayad, Y. Fetouh, C. Mezhoud, and A. Berreksi, "Water Quality Assessment Using the Random Forest Classification Model," *Water Environ. Res.*, vol. 97, no. 10, p. e70197, Oct. 2025, doi: <https://doi.org/10.1002/wer.70197>.
- [16] S. M. Alomani, N. I. Alhawiti, and A. Alhakamy, "Prediction of Quality of Water According to a Random Forest Classifier," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 892–899, 2022, doi: 10.14569/IJACSA.2022.01306105.
- [17] C. T. Tran, "Ensemble Learning Approaches for Classification With High-Dimensional Data," *J. Sci. Tech.*, vol. 12, no. 01, pp. 83–96, 2023, doi: 10.56651/lqdtu.jst.v12.n1.659.ict.
- [18] Y. Febriani, M. L. M. Akhlak, M. Apriniyadi, and A. R. Saleh, "Evaluation of Deep Learning Model Performance for Water Quality Classification: Comparison of CNN, LSTM, and Hybrid CNN-LSTM," pp. 1–26, 2025, [Online]. Available: https://www.researchgate.net/publication/395878608_Evaluation_of_Deep_Learning_Model_Performance_for_Water_Quality_Classification_Comparison_of_CNN_LSTM_and_Hybrid_CNN-LSTM