

## Penerapan Algoritma *Random Forest* dalam Prediksi Kelayakan Air Minum

Khairul Abdi<sup>\*1</sup>, Angga Warjaya<sup>2</sup>, Inna Muthmainnah<sup>3</sup>, Padli Husaini Pahutar<sup>4</sup>

<sup>1,2,3,4</sup>Prodi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri  
Medan, Indonesia

Email: <sup>1</sup>khairulabdi@mhs.unimed.ac.id, <sup>2</sup>anggawarjaya@mhs.unimed.ac.id,  
<sup>3</sup>innamuthmainnah@mhs.unimed.ac.id, <sup>4</sup>padlihusaini3@gmail.com

### Abstrak

Tantangan meningkatnya kebutuhan akan air bersih di tengah ketersediaan yang terbatas mendorong penelitian ini. Pencemaran air, terutama oleh limbah rumah tangga, menyebabkan penurunan kualitas air dan kelangkaan sumber air bersih. Dampak buruknya terhadap kesehatan masyarakat, terutama diare, menunjukkan urgensi identifikasi dini kualitas air yang tidak layak konsumsi. Oleh karena itu, penelitian ini menerapkan algoritma *Random Forest* untuk klasifikasi kualitas air dan prediksi kelayakan air minum. Penggunaan teknik data mining, khususnya *Random Forest*, diharapkan dapat mengidentifikasi pola kompleks dalam data dan faktor-faktor yang mempengaruhi kualitas air. Menggunakan dataset *Water Quality* dari Kaggle, hasil penelitian menunjukkan akurasi sebesar 69%. Analisis *Feature Importance Score* memperlihatkan kontribusi relatif fitur terhadap prediksi. Kurva ROC menggambarkan optimalitas klasifikasi, sementara *Confusion Matrix* memberikan gambaran kinerja model. *Confusion Matrix* yang merinci hasil klasifikasi model *Random Forest Classifier*. Diagonal utama menunjukkan jumlah instance yang benar-benar diprediksi dengan benar untuk kategori "Potabel" dan "Tidak Potabel", masing-masing 370 dan 84. Namun, terdapat 160 instance yang salah diklasifikasikan sebagai "Potabel" dan 42 sebagai "Tidak Potabel".

**Kata kunci:** Kualitas Air, Klasifikasi, Potabilitas, *Random Forest*

### Abstract

The escalating demand for clean water amidst limited availability poses a significant challenge addressed by this research. Water pollution, primarily from household waste, leads to a decline in water quality and a scarcity of clean water sources. The adverse impact on public health, particularly diarrhea, underscores the urgency of early identification of water quality unfit for consumption. Hence, this study applies the *Random Forest* algorithm for water quality classification and prediction of drinking water suitability. The use of data mining techniques, specifically *Random Forest*, aims to identify complex patterns in data and factors influencing water quality. Utilizing the *Water Quality* dataset from Kaggle, the research yields an accuracy of 69%. Analysis of *Feature Importance Score* reveals the relative contribution of features to predictions. The ROC curve illustrates the classification's optimality, while the *Confusion Matrix* provides insights into the model's performance. The *Confusion Matrix* details the classification results of the *Random Forest Classifier* model. The main diagonal indicates the number of instances correctly predicted for the "Potable" and "Not Potable" categories, with 370 and 84 instances, respectively. However, there are 160 instances misclassified as "Potable" and 42 as "Not Potable".

**Keywords:** *Random Forest*, *Water Quality*, Potability, Classification

## 1. PENDAHULUAN

Kebutuhan akan air bersih yang terus meningkat setiap tahunnya merupakan tantangan serius di tengah ketersediaan air yang terbatas. Pertumbuhan pembangunan tanpa memperhatikan keseimbangan lingkungan, terutama di daerah perkotaan, menyebabkan terjadinya penurunan kualitas air dan sempitnya daerah resapan air [1]. Pencemaran air menjadi masalah utama di Indonesia, yang sebagian besar disebabkan oleh limbah rumah tangga dan domestik. Pencemaran ini mengakibatkan menurunnya kualitas air, sehingga menyebabkan kelangkaan sumber air bersih di berbagai daerah [2].

Dampak dari kualitas air yang buruk sangat serius, terutama dalam hal kesehatan masyarakat. Salah satu penyakit yang sering timbul akibat konsumsi air yang terkontaminasi adalah diare. Setiap tahun, sekitar 800 ribu orang diperkirakan meninggal karena penyakit ini [3]. Faktor penyebabnya melibatkan air yang tidak bersih, Ketidakcukupan kebersihan tangan saat mengonsumsi air dan permasalahan sanitasi dalam air minum [4].

Untuk mengatasi masalah ini, penting untuk dapat mengidentifikasi dini kualitas air yang tidak layak konsumsi. Teknik data mining, khususnya teknik klasifikasi, dapat menjadi solusi efektif untuk meramalkan terhadap kelas atau label yang sebelumnya tidak dapat diidentifikasi dan memisahkan objek-objek berdasarkan ciri-ciri spesifik [5]. *Random Forest* adalah algoritma klasifikasi yang dapat digunakan untuk tujuan ini.

Algoritma *Random Forest* merupakan algoritma yang populer dalam dunia klasifikasi. Algoritma ini menggunakan teknik *ensemble*, yang menggabungkan hasil dari beberapa model pohon keputusan untuk meningkatkan akurasi dan kinerja klasifikasi [6]. *Random Forest* dapat membantu mengidentifikasi pola kompleks dalam data [7], mengenali faktor-faktor yang mempengaruhi kualitas air, dan membuat prediksi terhadap kelayakan air untuk dikonsumsi.

Telah banyak penelitian yang membahas klasifikasi kualitas air minum, terutama dalam konteks penerapan *machine learning*. Sebagai contoh, sebuah studi yang dilakukan oleh [8] fokus pada klasifikasi kualitas air sumur dengan menggunakan algoritma *Random Forest* untuk menilai kelayakan konsumsi, menghasilkan tingkat akurasi sebesar 82%. Penelitian lain, yang dilaksanakan oleh [9], mengaplikasikan *Logistic Regression*, *SVM*, *Random Forest Classifier*, *KNN*, dan *XGBoost Classifier* untuk mengklasifikasikan Kualitas Air Minum. Temuan dari penelitian tersebut menyatakan bahwa algoritma *Random Forest Classifier* berhasil mencapai tingkat akurasi tertinggi sekitar 82%.

Dalam penelitian ini, dataset *Water Quality and Potability* dari sumber terbuka di <https://kaggle.com/> akan dimanfaatkan. Dataset ini sebelumnya telah dipergunakan dalam sebuah penelitian terdahulu [10]. Penelitian sebelumnya memanfaatkan algoritma *Decision Tree* dan *K-Nearest Neighbor (KNN)* untuk melakukan prediksi terhadap kualitas air. Hasilnya mencapai tingkat akurasi tertinggi sekitar 61%, yang dihasilkan dengan menggunakan algoritma *KNN*.

Penelitian ini akan menggunakan algoritma *Random Forest* dalam upaya Prediksi Kelayakan Air Minum melalui Klasifikasi Kualitas Air. Dengan memanfaatkan teknologi ini, diharapkan dapat dikembangkan sistem yang memberikan informasi awal kepada masyarakat dan pihak terkait mengenai kualitas air. Hal ini diharapkan dapat memungkinkan pengambilan tindakan pencegahan lebih cepat untuk menjaga kesehatan dan ketersediaan air bersih di masyarakat.

## 2. METODE PENELITIAN

Metode penelitian akan dijelaskan secara terstruktur melalui serangkaian langkah-langkah khusus yang akan digunakan untuk mengatasi permasalahan penelitian. Rincian langkah-langkah penelitian ini dapat ditemukan dalam ilustrasi pada Gambar 1.



Gambar 1. Langkah-langkah Penelitian

### 2.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini merupakan informasi terkait potabilitas air yang diambil dari sumber terbuka, yaitu <https://www.kaggle.com/>, dengan judul *Water Quality and Potability*. Dataset ini mencakup sepuluh atribut kelas, sebagaimana tergambar pada Tabel 1.

Tabel 1. Atribut Dataset

Atribut	Keterangan
<i>pH</i>	Tingkat pH air
<i>Hardness</i>	Kekerasan air, suatu ukuran kandungan mineral.
<i>Solids</i>	Total padatan terlarut dalam air.
<i>Chloramines</i>	Konsentrasi Kloramin dalam air
<i>Sulfate</i>	Konsentrasi sulfat dalam air
<i>Conductivity</i>	Konduktivitas listrik air
<i>Organic_carbon</i>	Kandungan karbon organik dalam air
<i>Trihalomethanes</i>	Konsentrasi trihalometana dalam air
<i>Turbidity</i>	Tingkat kekeruhan, suatu ukuran kejernihan air
<i>Potability</i>	Variabel target; menunjukkan potabilitas air dengan nilai 1 (potabel) dan 0 (tidak potabel)

## 2.2. Preprocessing Data

*Preprocessing data* adalah sekumpulan teknik yang digunakan untuk membersihkan dan menormalkan data yang tidak teratur dan tidak konsisten. Hal ini dilakukan untuk menghilangkan *noise*, *missing values*, dan data yang tidak *balance* dari database [11]. Pada tahap ini dilakukan *cleaning data* pada data yang tidak lengkap, data kosong, dan berisikan *noise* dan *outlier*.

Data akan dipecah menjadi dua bagian utama, yakni data pelatihan dan data pengujian, dari keseluruhan 3276 entri dataset. Pada tahap awal ini, kami akan menyusun dataset pelatihan dan pengujian dengan perbandingan 80% untuk data pelatihan dan 20% untuk data pengujian [12]. Selanjutnya, dalam fase ini, kami akan menerapkan proses standarisasi data menggunakan persamaan (1), dengan tujuan untuk menormalkan data ke dalam rentang yang seragam [13].

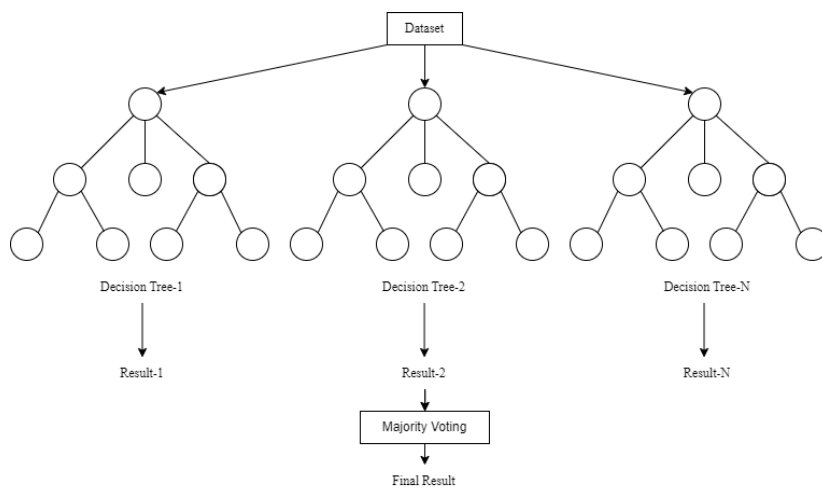
$$x' = \frac{x_i - \text{mean}(x)}{\text{std}(x)} \quad (1)$$

Keterangan:

- $x'$  = nilai setelah distandarisasi.
- $x_i$  = nilai yang akan distandarisasi.
- $\text{mean}(x)$  = nilai rata-rata kolom.
- $\text{std}(x)$  = nilai standar deviasi dari kumpulan nilai pada kolom

## 2.3. Implementasi *Random Forest*

Tahap kunci dalam penelitian ini melibatkan pengembangan model klasifikasi dengan menerapkan algoritma *Random Forest*. Metode *Random Forest* ini menggabungkan beberapa pohon keputusan untuk bekerja bersama-sama, menghasilkan suara bersama yang menentukan hasil akhir dalam mendeteksi sarkasme. Pemanfaatan data pelatihan dan variabel acak yang bersifat mandiri. dalam pendekatan ini dijadikan sebagai pendukung untuk proses klasifikasi, yang berbeda dengan fitur yang beragam. [14]. Algoritma *Random Forest* lebih tepat digunakan ketika dataset memiliki volume yang cukup besar [15]. Gambar 2 merupakan ilustrasi algoritma *random forest*.



Gambar 2. Ilustrasi *Random Forest*

Persamaan yang diterapkan dalam algoritma *Random Forest* dapat diidentifikasi pada rumus (2).

$$f_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} n_j}{\sum_{\epsilon \text{ all nodes}} n_k} \quad (2)$$

Keterangan:

$f_i$  = fitur  $i$   
 $n_j$  = simpul  $j$

## 2.4. Evaluasi Hasil

Tahap evaluasi hasil melibatkan penggunaan *confusion matrix*, yang memberikan gambaran rinci tentang kinerja model. *Confusion matrix* adalah suatu matriks yang berisi informasi tentang keakuratan prediksi dan kesesuaian dengan keadaan sebenarnya, menunjukkan apakah prediksi tersebut benar atau salah [16]. Tabel 2 merupakan bentuk dari *confusion matrix*.

Tabel 2. *Confusion Matrix*

	<b>Positive Predicted</b>	<b>Negative Predicted</b>
<b>Positive</b>	True Positive (TP)	True Negative (TN)
<b>Negative</b>	False Positive (FP)	False Negative (FN)

*Confusion matrix* akan menghasilkan elemen-elemen seperti TP, FP, FN, dan TN. Elemen-elemen ini akan dimanfaatkan untuk mengkalkulasi performa model guna mendapatkan nilai akurasi, presisi, *recall*, dan *F1 Score*, dengan menerapkan formula (3), (4), (5), dan (6).

$$F1 \text{ Score} = \frac{Precision \times Recall}{Precision + Recall} \quad (3)$$

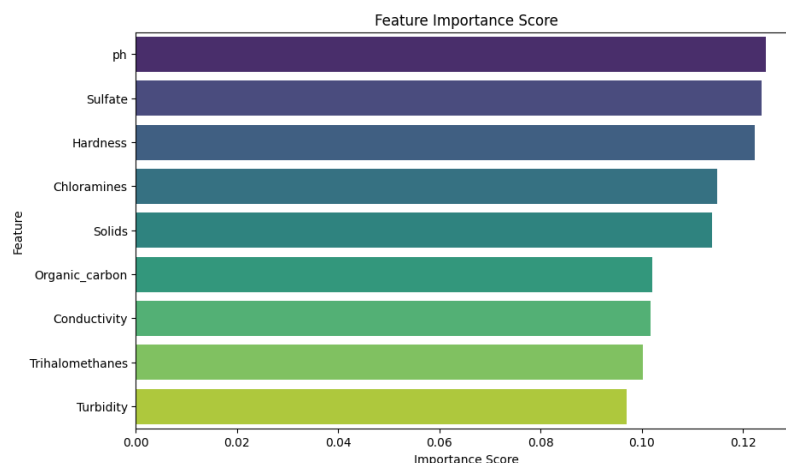
$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

### 3. HASIL DAN PEMBAHASAN

Dalam penerapan metode *random forest* melalui platform *Google Colaboratory* untuk menganalisis data kualitas air dan potabilitas, Terjadi pembagian dataset, di mana 80% dijadikan sebagai data pelatihan dan 20% sebagai data pengujian. Pada konteks ini, data pelatihan dimanfaatkan untuk melatih model, sementara data pengujian digunakan untuk mengevaluasi kinerja model.



Gambar 3. Grafik *Feature Importance Score*

Gambar 3 memperlihatkan representasi visual dari *Feature Importance Score*, yang merupakan ukuran signifikansi relatif dari setiap fitur terhadap hasil prediksi. Informasi lebih lanjut mengenai *Feature Importance Score* ini dapat ditemukan dalam Tabel 3, yang memberikan penjelasan rinci terkait pengaruh masing-masing fitur terhadap kualitas air dan potabilitas. Analisis ini berguna untuk memahami kontribusi relatif dari setiap variabel terhadap keakuratan model, memungkinkan pemahaman mendalam tentang faktor-faktor yang paling berpengaruh dalam prediksi kualitas air dan potabilitas.

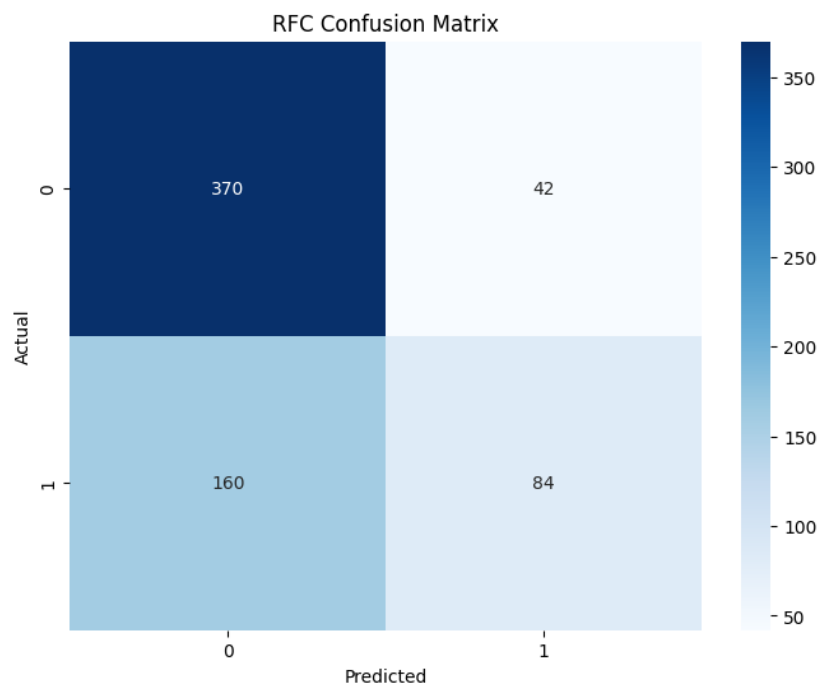
Fitur	Nilai
Ph	0,124
Sulfate	0,123
Hardness	0,122
Chloramines	0,114
Solids	0,113
Organic Carbon	0,101
Conductivity	0,101
Trihalomethanes	0,100
Turbidity	0,096

Pada Gambar 4, terlihat kurva *Receiver Operating Characteristic (ROC)*, ditampilkan visualisasi yang menggambarkan hubungan antara Tingkat Positif Sejati (TPR) dan Tingkat Positif Palsu (FPR). Kehadiran kurva mendekati sudut kiri atas menunjukkan peningkatan performa klasifikasi. Melalui pengujian menggunakan metode *Random Forest*, diperoleh hasil akurasi sebesar 69%, mencerminkan seberapa baik model dapat memberikan prediksi yang sesuai dengan data uji. Evaluasi ini menggambarkan pemahaman lebih rinci tentang kemampuan model dalam mengidentifikasi keaslian dan ketidakkeaslian kualitas air berdasarkan fitur-fitur yang diberikan.

Tabel 4. *Classification Report*

Metric	Potabel	Tidak Potabel	AVG
Akurasi	0,69	0,69	0,69
Presisi	0,67	0,70	0,68
Recall	0,34	0,90	0,62
F1 Score	0,45	0,79	0,62

Evaluasi kinerja model *Random Forest Classifier* (RFC) untuk memprediksi kualitas air potabel atau tidak potabel ditampilkan dalam Tabel 4 *Classification Report*. Dengan tingkat akurasi sebesar 69%, model ini menunjukkan kemampuan yang memadai dalam memberikan prediksi yang sesuai dengan data uji. Presisi, yang mengukur ketepatan model dalam mengidentifikasi kelas tertentu, menunjukkan nilai sebesar 67% untuk kelas "Potabel" dan 70% untuk kelas "Tidak Potabel." Sementara itu, recall, yang mencerminkan kemampuan model dalam mendeteksi instance yang sebenarnya positif, variatif antara kelas "Potabel" dengan nilai 34% dan kelas "Tidak Potabel" dengan nilai 90%. *F1 Score*, sebagai *harmonic mean* antara presisi dan *recall*, memberikan nilai holistik sebesar 0.45 untuk kelas "Potabel" dan 0.79 untuk kelas "Tidak Potabel".



Gambar 4. *Confusion Matrix*

Gambar 4 menampilkan *Confusion Matrix* yang merinci hasil klasifikasi model *Random Forest Classifier*. Diagonal utama menunjukkan jumlah instance yang benar-benar diprediksi dengan benar untuk kategori "Potabel" dan "Tidak Potabel", masing-masing 370 dan 84. Namun, terdapat 160 instance yang salah diklasifikasikan sebagai "Potabel" dan 42 sebagai "Tidak Potabel".

#### 4. KESIMPULAN

Penelitian ini menggunakan algoritma *Random Forest* untuk memprediksi kelayakan air minum. Dengan menerapkan teknik data mining pada dataset *Water Quality* dari Kaggle, hasil penelitian menunjukkan bahwa model *Random Forest* mencapai akurasi sebesar 69%. Analisis *Feature Importance Score* mengidentifikasi kontribusi relatif setiap fitur terhadap prediksi, sementara *ROC curve* memberikan gambaran optimalitas klasifikasi. *Confusion matrix* merinci hasil klasifikasi, di mana 370 instance benar-benar diprediksi untuk kategori "Potabel" dan 84 untuk "Tidak Potabel".

Meskipun terdapat 160 *instance* yang salah diklasifikasikan sebagai "Potabel" dan 42 sebagai "Tidak Potabel", penelitian ini memberikan wawasan penting terkait aplikabilitas algoritma *Random Forest* dalam konteks prediksi kualitas air.

Dengan akurasi sebesar 69%, model *Random Forest* dalam penelitian ini menunjukkan kemampuan yang memadai dalam memberikan prediksi yang sesuai dengan data uji. Meskipun terdapat ruang untuk peningkatan akurasi, hasil ini memberikan dasar yang kuat untuk mempertimbangkan penerapan algoritma *Random Forest* sebagai alat prediktif dalam penilaian kualitas air.

## DAFTAR PUSTAKA

- [1] M. Sofiana, A. Kadarsah, dan D. Sofarini, "Kualitas Air Terdampak Limbah Sebagai Indikator Pembangunan Berkelanjutan Di Sub DAS Martapura Kabupaten Banjar," *Jukung (Jurnal Tek. Lingkungan)*, vol. 8, no. 1, hal. 18–31, 2022, doi: <http://dx.doi.org/10.20527/jukung.v8i1.12966>.
- [2] D. Mulyanti, "Kearifan Lokal Masyarakat Terhadap Sumber Mata Air Sebagai Upaya Konservasi Dan Pengelolaan Sumber Daya Lingkungan," *Bina Huk. Lingkung.*, vol. 6, no. 3, hal. 410–424, 2022, doi: <https://doi.org/10.24970/bhl.v6i3.286>.
- [3] H. Said, N. Matondang, dan H. N. Irmanda, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Techno.Com*, vol. 21, no. 2, hal. 256–267, 2022, doi: <https://doi.org/10.33633/tc.v21i2.5901>.
- [4] M. A. Fadilah, H. D. L. Damanik, dan Yulianto, "Kejadian Diare Pada Balita Berdasarkan Perilaku Cuci Tangan Pakai Sabun dan Sarana Air Minum di Wilayah Kerja Puskesmas Kabupaten Ogan Ilir," *J. Sanitasi Lingkung.*, vol. 2, no. 1, hal. 60–65, 2022, doi: <https://doi.org/10.36086/jsl.v2i1.878>.
- [5] R. F. Putra *et al.*, *Data Mining : Algoritma dan Penerapannya*. Jambi: PT. Sonpedia Publishing Indonesia, 2023. [Daring]. Tersedia pada: [https://books.google.co.id/books?hl=en&lr=&id=zLHGEEAAQBAJ&oi=fnd&pg=PA22&dq=P+utra,+R.+F.,+Zebua,+R.+S.+Y.,+Budiman,+B.,+Rahayu,+P.+W.,+Bangsa,+M.+T.+A.,+Zulfadhilah,+M.,+...+%26+Andiyan,+A.+\(2023\).+DATA+MINING:+Algoritma+dan+Penerapannya.+P+T.+Sonpedia+Publishing+Indonesia.&ots=tzLeatVPK5&sig=JL\\_Vj7a1s0IG02rgKWcCmlXuCJM&redir\\_esc=y#v=onepage&q&f=false](https://books.google.co.id/books?hl=en&lr=&id=zLHGEEAAQBAJ&oi=fnd&pg=PA22&dq=P+utra,+R.+F.,+Zebua,+R.+S.+Y.,+Budiman,+B.,+Rahayu,+P.+W.,+Bangsa,+M.+T.+A.,+Zulfadhilah,+M.,+...+%26+Andiyan,+A.+(2023).+DATA+MINING:+Algoritma+dan+Penerapannya.+P+T.+Sonpedia+Publishing+Indonesia.&ots=tzLeatVPK5&sig=JL_Vj7a1s0IG02rgKWcCmlXuCJM&redir_esc=y#v=onepage&q&f=false)
- [6] A. F. Nugraha, R. F. A. Aziza, dan Y. Pristyanto, "Penerapan metode Stacking dan Random Forest untuk Meningkatkan Kinerja Klasifikasi pada Proses Deteksi Web Phishing," *J. Infomedia Tek. Inform. Multimed. Jar.*, vol. 7, no. 1, hal. 39–44, 2022, doi: <https://dx.doi.org/10.30811/jim.v7i1.2959>.
- [7] I. Werdiningsih *et al.*, "Identifikasi Penipuan Kartu Kredit Pada Transaksi Ilegal Menggunakan Algoritma Random Forest dan Decision Tree," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, hal. 477–484, 2023, doi: <https://doi.org/10.32736/sisfokom.v12i3.1730>.
- [8] M. M. Mutoffar, M. Naseer, dan A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 4, no. 2, hal. 138–146, 2022, doi: <https://doi.org/10.53580/naratif.v4i2.160>.
- [9] L. Savitri dan R. Nursalim, "Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning," *Diophantine J. Math. Its Appl.*, vol. 2, no. 01, hal. 30–36, 2023, doi: <https://doi.org/10.33369/diophantine.v2i01.28260>.
- [10] Y. Christian, Jacky, P. A. Winata, Ricky, dan N. Jeonanto, "Prediksi Kualitas Air Menggunakan Algoritma Naïve Bayes Dan Random Forest," *Komputek*, vol. 6, no. 2, hal. 42–48, 2022, doi: [10.24269/jkt.v6i2.1313](https://doi.org/10.24269/jkt.v6i2.1313).
- [11] O. W. Yuda, D. Tuti, L. S. Yee, dan Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, hal. 122–131, 2022, doi: <https://doi.org/10.33372/stn.v8i2.885>.

- [12] A. Putri *et al.*, “Comparison of K-NN, Naive Bayes and SVM Algorithms for Final-Year Student Graduation Prediction,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, hal. 20–26, 2023, doi: <https://doi.org/10.57152/malcom.v3i1.610>.
- [13] Athifaturrofifah, R. Goejantoro, dan D. Yuniarti, “Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018),” *J. EKSPONENSIAL*, vol. 10, no. 2, hal. 143–152, 2019.
- [14] D. Alita dan A. Rahman, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *J. Komputasi*, vol. 8, no. 2, hal. 50–58, 2020, doi: <http://dx.doi.org/10.23960%2Fkomputasi.v8i2.2615>.
- [15] L. J. Muhammad, A. A. Haruna, I. A. Mohammed, M. Abubakar, B. G. Badamasi, dan J. M. Amshi, “Performance Evaluation of Classification Data Mining Algorithms on Coronary Artery Disease Dataset,” in *International eConference on Computer and Knowledge Engineering (ICCKE)*, 2019, hal. 1–5. doi: 10.1109/ICCKE48569.2019.8964703.
- [16] Yuyun, N. Hidayah, dan S. Sahibu, “Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap Penanganan Covid-19 Menggunakan Data Twitter,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 4, hal. 820–826, 2021, doi: <https://doi.org/10.29207/resti.v5i4.3146>.