

# IMPLEMENTASI MACHINE LEARNING BERBASIS FITUR WARNA RGB DAN HSV UNTUK KLASIFIKASI KUALITAS AIR

Muhammad Ilham Manzis<sup>1</sup>, Ranggi Pharaningtyas Aji<sup>2</sup>

<sup>1,2</sup>Universitas Amikom Purwokerto; Jl. Letjend Pol. Soemarto No.127, Purwokerto Utara, Banyumas, Jawa Tengah 53127

## Keywords:

Machine Learning;  
Klasifikasi Kualitas Air; Fitur  
Warna Digital; RGB; HSV

## Correspondent Email:

muhamadilhammansiz259  
@gmail.com

**Abstrak.** Penurunan kualitas sumber air menjadi permasalahan penting yang memerlukan metode deteksi yang cepat, murah, dan mudah diimplementasikan. Pemantauan kualitas air secara konvensional umumnya dilakukan melalui analisis laboratorium yang membutuhkan biaya tinggi serta waktu yang relatif lama. Penelitian ini mengimplementasikan machine learning untuk mengklasifikasikan kualitas air berdasarkan analisis fitur warna digital dalam ruang warna RGB dan HSV. Dataset yang digunakan terdiri dari 647 citra air yang diklasifikasikan ke dalam dua kelas, yaitu air jernih dan air kotor. Fitur warna diekstraksi berupa nilai rata-rata R, G, B, H, S, dan V, kemudian dilakukan pelabelan otomatis berdasarkan threshold pada ruang warna HSV. Beberapa algoritma machine learning, yaitu Random Forest, Support Vector Machine, K-Nearest Neighbor, dan Decision Tree, digunakan dan dibandingkan kinerjanya. Hasil penelitian menunjukkan bahwa Decision Tree menghasilkan akurasi tertinggi sebesar 100%, diikuti oleh Random Forest sebesar 99,23%, Support Vector Machine sebesar 98,46%, dan K-Nearest Neighbor sebesar 96,92%. Analisis feature importance menunjukkan bahwa komponen Value dan Saturation pada ruang warna HSV merupakan fitur paling dominan dalam proses klasifikasi. Hasil ini menunjukkan bahwa kombinasi machine learning dan fitur warna digital efektif digunakan untuk klasifikasi kualitas air secara otomatis.



Copyright © [JITET](http://www.jitet.org) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** The declining quality of water resources requires fast, cost-effective, and easily deployable detection methods. Conventional water quality monitoring relies on laboratory analysis, which is expensive and time-consuming. This study implements machine learning to classify water quality based on digital color feature analysis in RGB and HSV color spaces. The dataset consists of 647 water images classified into clear and turbid water categories. Color features were extracted as mean values of R, G, B, H, S, and V, followed by automatic labeling using HSV-based threshold rules. Several machine learning algorithms, including Random Forest, Support Vector Machine, K-Nearest Neighbor, and Decision Tree, were evaluated and compared. The results indicate that the Decision Tree achieved the highest accuracy of 100%, followed by Random Forest at 99.23%, Support Vector Machine at 98.46%, and K-Nearest Neighbor at 96.92%. Feature importance analysis reveals that Value and Saturation components in the HSV color space are the most influential features for classification. These findings demonstrate that machine learning combined with digital color features provides an effective approach for automated water quality classification.

## 1. PENDAHULUAN

Air bersih merupakan kebutuhan dasar yang memiliki peran penting dalam menjaga

kesehatan dan keberlanjutan kehidupan manusia [1]. Namun, pada kondisi saat ini, kualitas sumber air di berbagai wilayah terus mengalami penurunan. Organisasi Kesehatan Dunia melaporkan bahwa sebagian besar populasi global masih bergantung pada sumber air yang tercemar, sehingga berpotensi menimbulkan berbagai penyakit berbasis air [2]. Penurunan kualitas air tersebut umumnya dipicu oleh aktivitas manusia, seperti pembuangan limbah industri, limbah rumah tangga, serta penggunaan bahan kimia dalam sektor pertanian [3].

Pemantauan kualitas air secara tradisional dilakukan melalui pengujian laboratorium dengan parameter fisik, kimia, dan biologis. Meskipun metode ini mampu menghasilkan hasil yang akurat, proses pengujian membutuhkan biaya yang relatif tinggi, peralatan khusus, serta waktu analisis yang cukup lama. Kondisi ini menjadikan metode konvensional kurang efisien untuk kebutuhan pemantauan yang bersifat cepat dan berkelanjutan, terutama pada wilayah dengan keterbatasan akses dan sumber daya [4].

Perkembangan teknologi informasi dan kecerdasan buatan membuka peluang baru dalam pemantauan kualitas air. Berbagai penelitian telah mengembangkan pendekatan berbasis pengolahan citra digital untuk menganalisis kondisi visual air [5]. Metode berbasis analisis tekstur maupun *deep learning* dilaporkan mampu memberikan performa yang baik, namun sering kali membutuhkan kompleksitas komputasi tinggi serta dataset dalam jumlah besar. Oleh karena itu, diperlukan pendekatan yang lebih sederhana namun tetap efektif.

Pemanfaatan fitur warna digital menjadi salah satu alternatif yang menjanjikan dalam analisis kualitas air. Karakteristik visual air dapat direpresentasikan melalui komponen warna pada ruang RGB dan HSV [6], [7]. Secara umum, air jernih memiliki tingkat kecerahan (*Value*) yang tinggi dengan tingkat kejenuhan (*Saturation*) yang rendah, sedangkan air kotor menunjukkan karakteristik sebaliknya. Ruang warna HSV dipilih karena mampu memisahkan informasi warna, kejenuhan, dan kecerahan secara lebih jelas serta lebih mendekati persepsi visual manusia dibandingkan ruang warna RGB.

Meskipun berbagai pendekatan telah dikembangkan, sebagian besar penelitian sebelumnya masih berfokus pada penerapan algoritma tertentu tanpa melakukan perbandingan menyeluruh terhadap beberapa algoritma *machine learning* [8]. Selain itu, analisis kontribusi masing-masing fitur warna terhadap performa klasifikasi masih terbatas, karena sebagian penelitian hanya mengevaluasi satu atau dua algoritma tanpa analisis komparatif yang mendalam [9]. Hal ini menunjukkan adanya celah penelitian yang perlu dieksplorasi lebih lanjut.

Berdasarkan latar belakang tersebut, penelitian ini mengimplementasikan metode *machine learning* berbasis fitur warna RGB dan HSV untuk klasifikasi kualitas air. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi otomatis air jernih dan air kotor, membandingkan kinerja beberapa algoritma *machine learning* seperti Decision Tree, Random Forest, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN), serta menganalisis kontribusi fitur warna terhadap akurasi klasifikasi. Hasil penelitian ini diharapkan dapat menjadi solusi alternatif yang efektif, ekonomis, dan mudah diimplementasikan dalam sistem monitoring kualitas air berbasis teknologi digital.

## 2. TINJAUAN PUSTAKA

Tinjauan pustaka disusun untuk memberikan landasan teoritis yang relevan dengan penelitian serta memperkuat pemilihan metode yang digunakan. Pembahasan difokuskan pada konsep kualitas air, pengolahan citra digital, fitur warna, dan penerapan *machine learning* dalam tugas klasifikasi.

### 2.1 Kualitas Air

Kualitas air menggambarkan kondisi fisik, kimia, dan biologis suatu sumber air yang menentukan tingkat kelayakannya untuk berbagai kebutuhan. Air dengan kualitas baik umumnya ditandai oleh tingkat kekeruhan yang rendah, warna yang jernih, serta minim kandungan zat pencemar. Penurunan kualitas air dapat terjadi akibat masuknya limbah industri, limbah rumah tangga, dan residu aktivitas pertanian, sehingga diperlukan metode pemantauan yang mampu mendeteksi

perubahan kualitas air secara cepat dan berkelanjutan.

## 2.2 Pengolahan Citra Digital

Pengolahan citra digital merupakan proses analisis gambar menggunakan bantuan komputer untuk mengekstraksi informasi tertentu dari citra. Dalam konteks pemantauan kualitas air, citra digital dapat dimanfaatkan untuk menangkap karakteristik visual air, seperti warna dan tingkat kecerahan, yang berkaitan erat dengan kondisi fisik air. Proses pengolahan citra umumnya melibatkan tahapan *preprocessing*, ekstraksi fitur, dan analisis data untuk memperoleh representasi yang sesuai dengan kebutuhan klasifikasi.

## 2.3 Machine Learning untuk Klasifikasi

*Machine learning* merupakan pendekatan kecerdasan buatan yang memungkinkan sistem mempelajari pola dari data dan menghasilkan keputusan atau prediksi secara otomatis. Dalam permasalahan klasifikasi, *machine learning* digunakan untuk memetakan data ke dalam kelas tertentu berdasarkan karakteristik yang dimilikinya. Beberapa algoritma yang umum digunakan dalam tugas klasifikasi antara lain Decision Tree, Random Forest, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN).

Decision Tree membentuk struktur pohon keputusan berdasarkan aturan pemisahan data, sehingga mudah dipahami dan diinterpretasikan. Random Forest merupakan metode *ensemble learning* yang mengombinasikan sejumlah pohon keputusan untuk meningkatkan akurasi dan mengurangi risiko overfitting. SVM bekerja dengan mencari hyperplane optimal yang memisahkan data antar kelas dengan margin maksimum, sedangkan K-NN mengklasifikasikan data berdasarkan kedekatan jarak dengan data latih. Keempat algoritma tersebut memiliki karakteristik dan keunggulan masing-masing dalam menyelesaikan permasalahan klasifikasi.

## 2.4 Penelitian Terkait

Berbagai penelitian sebelumnya telah mengkaji penerapan pengolahan citra dan *machine learning* dalam klasifikasi kualitas air. Pendekatan berbasis analisis tekstur citra maupun *deep learning* dilaporkan mampu menghasilkan tingkat akurasi yang tinggi,

namun sering kali memerlukan kompleksitas komputasi yang besar. Penelitian lain menunjukkan bahwa penggunaan fitur warna sederhana dalam ruang RGB dan HSV mampu memberikan performa yang kompetitif dengan kebutuhan komputasi yang lebih rendah.

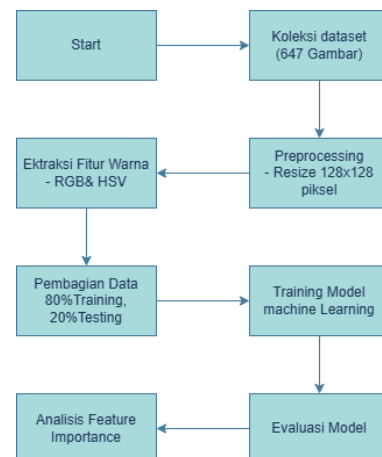
Namun demikian, sebagian besar penelitian terdahulu masih berfokus pada penerapan algoritma tertentu tanpa melakukan perbandingan performa beberapa algoritma *machine learning* secara komprehensif. Selain itu, kajian mendalam mengenai kontribusi fitur warna pada ruang RGB dan HSV terhadap hasil klasifikasi masih terbatas.

## 3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan eksperimental untuk mengembangkan dan mengevaluasi sistem klasifikasi kualitas air berbasis *machine learning* dengan memanfaatkan fitur warna digital. Metode penelitian dirancang secara sistematis agar seluruh tahapan yang dilakukan dapat direplikasi oleh peneliti lain.

### 3.1 Rancangan Penelitian

Alur penelitian terdiri dari beberapa tahap utama, meliputi pengumpulan dan persiapan dataset, *preprocessing* citra, ekstraksi fitur warna, pelabelan data, pembagian dataset, pelatihan model, serta evaluasi performa. Setiap tahap dirancang secara terintegrasi untuk mendukung tujuan penelitian. Penelitian ini menggunakan pendekatan *machine learning* berbasis fitur warna RGB dan HSV. Alur metode penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Metode Penelitian

### 3.2 Objek dan Data Penelitian

Dataset yang digunakan dalam penelitian ini berupa citra air yang dikumpulkan dari berbagai sumber terbuka. Total dataset yang digunakan berjumlah 647 citra, yang terdiri dari dua kelas, yaitu air jernih dan air kotor. Citra-citra tersebut merepresentasikan variasi kondisi visual air dengan perbedaan tingkat kecerahan dan kejenuhan warna.

Seluruh citra pada dataset memiliki variasi resolusi dan format, sehingga diperlukan tahap penyeragaman sebelum diproses lebih lanjut. Dataset ini digunakan sebagai dasar dalam proses ekstraksi fitur warna RGB dan HSV serta pelatihan model machine learning untuk klasifikasi kualitas air.

### 3.3 Perangkat dan Perangkat Lunak

Lingkungan pengembangan penelitian ini menggunakan bahasa pemrograman Python versi 3.8. Proses pengolahan citra dilakukan dengan memanfaatkan pustaka OpenCV, sedangkan implementasi algoritma machine learning menggunakan pustaka Scikit-learn. Pengolahan dan analisis data didukung oleh pustaka NumPy dan Pandas, sementara visualisasi data dilakukan menggunakan Matplotlib. Seluruh proses eksperimen dijalankan pada platform Google Colaboratory dengan dukungan perangkat keras berbasis GPU untuk mempercepat proses komputasi.

### 3.4 Pra-pemrosesan Citra

Tahap pra-pemrosesan dilakukan untuk menyeragamkan data citra sebelum proses ekstraksi fitur. Setiap citra diubah ukurannya menjadi  $128 \times 128$  piksel dan dikonversi ke ruang warna RGB dan HSV. Langkah ini bertujuan untuk meminimalkan pengaruh perbedaan resolusi dan mempermudah proses analisis fitur warna.

### 3.5 Ekstraksi Fitur Warna

Ekstraksi fitur warna dilakukan untuk memperoleh representasi numerik dari karakteristik visual citra air yang digunakan sebagai masukan pada proses klasifikasi. Pada penelitian ini, fitur warna diekstraksi dari dua ruang warna, yaitu **RGB** dan **HSV**, karena keduanya mampu merepresentasikan

informasi warna dan kecerahan citra secara efektif.

Pada ruang warna **RGB**, fitur yang diambil meliputi nilai rata-rata komponen **Red (R)**, **Green (G)**, dan **Blue (B)** dari setiap citra. Sementara itu, pada ruang warna **HSV**, fitur yang diekstraksi meliputi nilai rata-rata komponen **Hue (H)**, **Saturation (S)**, dan **Value (V)**. Proses konversi citra dari ruang warna RGB ke HSV dilakukan menggunakan pustaka OpenCV.

Nilai rata-rata setiap komponen warna dihitung untuk merepresentasikan karakteristik dominan citra secara keseluruhan. Kombinasi fitur RGB dan HSV ini diharapkan mampu memberikan informasi warna yang lebih komprehensif dalam membedakan kualitas air jernih dan air kotor.

### 3.6 Pelabelan Data

Pelabelan data dilakukan secara otomatis untuk mengelompokkan citra air ke dalam dua kelas, yaitu **air jernih** dan **air kotor**. Proses pelabelan didasarkan pada nilai ambang (*threshold*) tertentu pada fitur warna di ruang HSV, khususnya komponen **Value (V)** dan **Saturation (S)**. Citra dengan nilai kecerahan tinggi dan kejenuhan rendah diklasifikasikan sebagai air jernih, sedangkan citra dengan nilai kecerahan rendah dan kejenuhan tinggi diklasifikasikan sebagai air kotor. Pendekatan ini dipilih untuk mengurangi subjektivitas pelabelan manual serta meningkatkan konsistensi data.

### 3.7 Pembagian Data

Dataset yang telah melalui proses pelabelan selanjutnya dibagi menjadi dua bagian, yaitu **data latih (training data)** dan **data uji (testing data)**. Pembagian dataset dilakukan dengan rasio **80% data latih dan 20% data uji**. Data latih digunakan untuk membangun model machine learning, sedangkan data uji digunakan untuk mengevaluasi performa model terhadap data yang belum pernah dilihat sebelumnya.

### 3.8 Pelatihan dan Evaluasi Model

Pada tahap ini, dilakukan pelatihan model machine learning menggunakan empat algoritma, yaitu Decision Tree, Random Forest [10], Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN).

Algoritma Decision Tree digunakan sebagai metode klasifikasi yang membangun model prediksi berbentuk struktur pohon keputusan berdasarkan atribut data untuk memisahkan kelas secara hierarkis [11].

Support Vector Machine (SVM) merupakan algoritma klasifikasi yang bekerja dengan membangun hyperplane optimal untuk memisahkan data ke dalam kelas yang berbeda [12].

Algoritma K-Nearest Neighbor (K-NN) digunakan sebagai salah satu metode klasifikasi berbasis jarak yang sederhana dan efektif dalam berbagai studi komparatif algoritma machine learning [13].

Proses pelatihan dilakukan menggunakan data latih dengan parameter standar dari pustaka Scikit-learn.

Evaluasi performa model dilakukan menggunakan data uji dengan beberapa metrik evaluasi, antara lain akurasi (accuracy), presisi (precision), recall, dan F1-score [14]. Selain itu, confusion matrix digunakan untuk menganalisis kemampuan model dalam mengklasifikasikan masing-masing kelas secara lebih rinci [15] [16].

Untuk mengukur konsistensi dan stabilitas performa model, dilakukan validasi silang menggunakan metode 5-fold cross-validation [17]. Cross-validation adalah teknik evaluasi model yang membagi dataset menjadi beberapa subset agar performa model dapat diestimasi secara stabil dari berbagai partisi data. Hasil evaluasi ini digunakan sebagai dasar perbandingan performa antar algoritma machine learning yang digunakan.

### 3.9 Analisis Feature Importance

Analisis *feature importance* dilakukan untuk mengidentifikasi tingkat kontribusi masing-masing fitur warna terhadap hasil klasifikasi. Analisis ini bertujuan untuk mengetahui fitur yang paling berpengaruh dalam membedakan kualitas air jernih dan air kotor [18].

### 3.10 Etika Penelitian

Seluruh data citra yang digunakan dalam penelitian ini diperoleh dari sumber publik dan tidak melanggar hak cipta. Penelitian dilakukan dengan menjunjung prinsip transparansi dan reproduktibilitas melalui

penyajian metode dan proses penelitian secara jelas.

## 4. HASIL DAN PEMBAHASAN

### 4.1 Karakteristik Dataset dan Hasil Ekstraksi Fitur

Berdasarkan analisis terhadap 647 citra air yang digunakan dalam penelitian ini, diperoleh perbedaan karakteristik fitur warna yang cukup signifikan antara kelas air jernih dan air kotor. Nilai rata-rata komponen *Value* (V) pada ruang warna HSV menunjukkan kecenderungan lebih tinggi pada citra air jernih dibandingkan citra air kotor. Sebaliknya, nilai *Saturation* (S) pada citra air kotor cenderung lebih tinggi dibandingkan citra air jernih. Pola ini mencerminkan karakteristik visual air, di mana air jernih memiliki tingkat kecerahan yang lebih tinggi dan warna yang lebih netral.

Hasil ekstraksi fitur warna RGB juga menunjukkan perbedaan distribusi nilai antar kelas, meskipun pemisahan kelas tidak sejelas pada ruang warna HSV. Temuan ini mengindikasikan bahwa fitur warna pada ruang HSV lebih representatif dalam membedakan kualitas air dibandingkan fitur RGB.

### 4.2 Performa Klasifikasi Model Machine Learning

Evaluasi performa klasifikasi dilakukan menggunakan empat algoritma *machine learning*, yaitu Decision Tree, Random Forest, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN). Hasil pengujian menunjukkan bahwa seluruh model mampu mengklasifikasikan kualitas air dengan tingkat akurasi yang tinggi. Decision Tree menghasilkan akurasi tertinggi sebesar 100%, diikuti oleh Random Forest sebesar 99,23%, SVM sebesar 98,46%, dan K-NN sebesar 96,92%.

Untuk menguji konsistensi performa model, dilakukan validasi silang menggunakan metode *5-fold cross-validation*. Hasil validasi menunjukkan bahwa Decision Tree dan Random Forest memiliki stabilitas performa yang tinggi dengan variasi nilai akurasi yang relatif kecil. Hal ini menunjukkan bahwa model tidak hanya unggul pada data uji, tetapi juga memiliki kemampuan generalisasi yang baik terhadap data yang tidak dilihat sebelumnya.

Tabel 1 menyajikan perbandingan performa keempat algoritma *machine learning*

berdasarkan metrik akurasi, presisi, recall, dan F1-score.

Tabel 1. Perbandingan performa model klasifikasi kualitas air

Model	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
Decision Tree	100.00	100.00	100.00	100.00
Random Forest	99.23	99.25	99.22	99.23
SVM	98.46	98.48	98.45	98.46
K-NN	96.92	96.95	96.90	96.92

Hasil ini menunjukkan bahwa seluruh model memiliki performa yang sangat baik, dengan Decision Tree dan Random Forest menunjukkan keunggulan yang lebih konsisten dibandingkan algoritma lainnya.

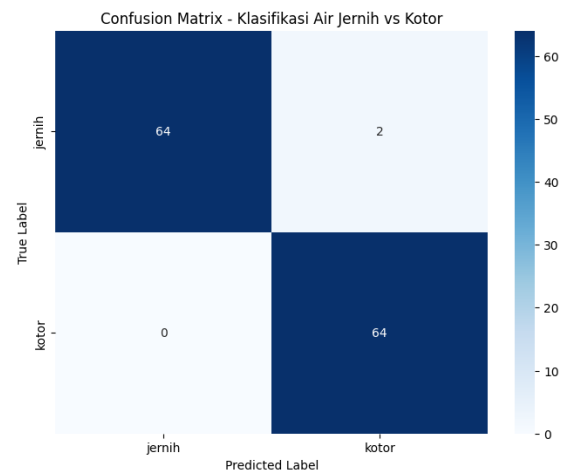
Selain evaluasi menggunakan data uji, konsistensi performa model juga dianalisis melalui validasi silang menggunakan metode *5-fold cross-validation*. Hasil validasi silang menunjukkan bahwa algoritma Decision Tree memiliki rata-rata akurasi sebesar **99,69%** dengan deviasi standar **0,38%**, sedangkan Random Forest mencapai rata-rata akurasi sebesar **99,53% ± 0,62%**. Hasil ini menunjukkan bahwa kedua model memiliki stabilitas performa yang baik serta kemampuan generalisasi yang konsisten terhadap data yang tidak dilihat sebelumnya.

#### 4.3 Evaluasi Detail Klasifikasi

Analisis lebih lanjut terhadap hasil klasifikasi dilakukan menggunakan confusion matrix untuk melihat distribusi prediksi benar dan salah pada masing-masing kelas. Confusion matrix menunjukkan bahwa model Random Forest mampu mengklasifikasikan sebagian besar citra air jernih dan air kotor secara tepat, dengan jumlah kesalahan klasifikasi yang sangat minimal.

Visualisasi confusion matrix pada Gambar 2 memperlihatkan pemisahan kelas yang jelas antara air jernih dan air kotor. Hal ini menunjukkan bahwa kombinasi fitur warna digital dan algoritma machine learning yang digunakan mampu merepresentasikan

perbedaan karakteristik kedua kelas secara efektif.



Gambar 2. Confusion matrix hasil klasifikasi kualitas air menggunakan Random Forest

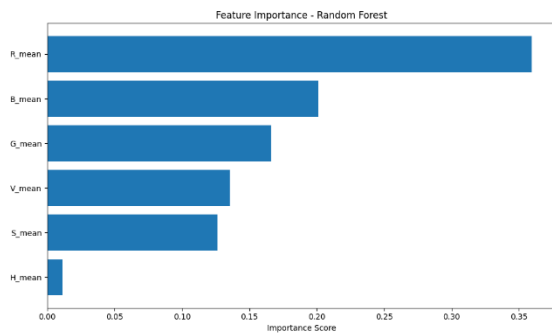
#### 4.4 Analisis Feature Importance

Analisis feature importance dilakukan untuk mengetahui kontribusi relatif masing-masing fitur warna terhadap hasil klasifikasi. Hasil analisis menunjukkan bahwa komponen Value (V) pada ruang warna HSV merupakan fitur yang paling dominan dalam menentukan kualitas air, diikuti oleh Saturation (S) dan Hue (H). Kontribusi fitur warna pada ruang RGB relatif lebih kecil dibandingkan fitur HSV.

Dominannya komponen Value dan Saturation menunjukkan bahwa tingkat kecerahan dan kejenuhan warna memiliki peran yang sangat penting dalam membedakan air jernih dan air kotor. Temuan ini memperkuat asumsi bahwa ruang warna HSV lebih informatif dibandingkan RGB dalam analisis kualitas air berbasis citra.

Temuan ini sejalan dengan karakteristik visual air yang telah dijelaskan pada Subbab 4.1, di mana air jernih cenderung memiliki tingkat kecerahan yang lebih tinggi dan kejenuhan warna yang lebih rendah dibandingkan air kotor.

Kontribusi relatif masing-masing fitur warna terhadap hasil klasifikasi ditunjukkan pada Gambar 3.

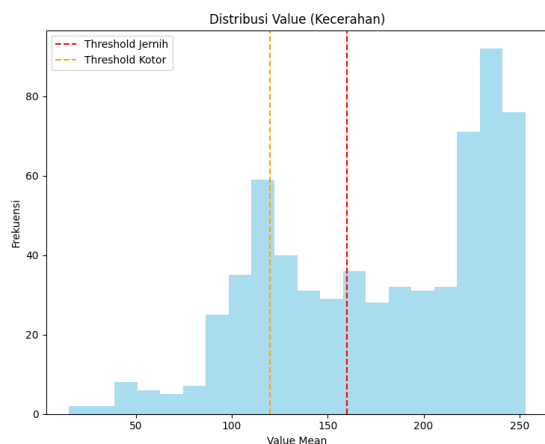


Gambar 3. Feature importance model Random Forest untuk klasifikasi kualitas air

#### 4.5 Distribusi Nilai Fitur sebagai Dasar Klasifikasi

Distribusi nilai rata-rata *Value* (V) pada seluruh dataset menunjukkan adanya pemisahan yang cukup jelas antara kelas air jernih dan air kotor. Sebagian besar citra air jernih berada pada rentang nilai *Value* yang tinggi, sedangkan citra air kotor terkonsentrasi pada nilai yang lebih rendah. Penggunaan threshold tertentu pada komponen *Value* dan *Saturation* terbukti efektif dalam mendukung proses pelabelan otomatis dan klasifikasi kualitas air.

Visualisasi distribusi nilai *Value* yang ditunjukkan pada Gambar 4 menegaskan bahwa fitur ini dapat dijadikan dasar yang kuat dalam membedakan kualitas air secara visual.



Gambar 4. Distribusi nilai Value (V) dengan threshold klasifikasi kualitas air

#### 4.6 Pembahasan Hasil

Tingginya akurasi yang dicapai oleh algoritma Decision Tree dapat dijelaskan oleh kemampuannya dalam menangkap pola pemisahan data berbasis aturan threshold yang secara alami terdapat pada fitur warna.

Karakteristik ini menjadikan Decision Tree sangat efektif dalam menangani data dengan hubungan non-linear sederhana, seperti yang ditemukan pada fitur warna RGB dan HSV.

Random Forest juga menunjukkan performa yang sangat baik dengan tingkat stabilitas yang tinggi pada proses validasi silang. Hal ini mengindikasikan bahwa metode *ensemble learning* mampu mengurangi risiko overfitting dan meningkatkan kemampuan generalisasi model. Sementara itu, SVM dan K-NN tetap memberikan performa yang kompetitif, meskipun cenderung lebih sensitif terhadap variasi data dan noise pada citra.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa pemanfaatan fitur warna digital sederhana yang dikombinasikan dengan algoritma *machine learning* mampu menghasilkan sistem klasifikasi kualitas air yang akurat dan efisien. Pendekatan ini memiliki keunggulan dari sisi kebutuhan komputasi yang relatif rendah, sehingga berpotensi diterapkan pada sistem monitoring kualitas air secara real-time dan berbasis perangkat dengan sumber daya terbatas.

Dengan demikian, pendekatan yang diusulkan mampu menjawab celah penelitian yang telah diidentifikasi pada bagian pendahuluan.

## 5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan sistem klasifikasi kualitas air berbasis machine learning dengan memanfaatkan fitur warna RGB dan HSV. Hasil pengujian menunjukkan bahwa seluruh algoritma yang digunakan mampu memberikan performa klasifikasi yang tinggi, dengan Decision Tree dan Random Forest menunjukkan akurasi serta stabilitas performa yang paling unggul.

Analisis feature importance mengungkapkan bahwa komponen Value (V) dan Saturation (S) pada ruang warna HSV merupakan fitur yang paling berkontribusi dalam membedakan air jernih dan air kotor. Temuan ini menegaskan bahwa pemanfaatan fitur warna HSV lebih efektif dibandingkan RGB dalam analisis kualitas air berbasis citra.

Secara keseluruhan, pendekatan yang diusulkan mampu menjawab celah penelitian yang telah diidentifikasi, yaitu perbandingan performa beberapa algoritma machine learning



serta analisis kontribusi fitur warna, dengan kebutuhan komputasi yang relatif rendah dan potensi implementasi yang luas.

Penelitian selanjutnya disarankan untuk menggunakan dataset dengan jumlah dan variasi citra yang lebih besar, serta mempertimbangkan kondisi lingkungan yang lebih beragam, seperti pencahayaan dan refleksi permukaan air. Selain itu, pengembangan metode pelabelan yang lebih adaptif dan eksplorasi fitur tambahan, seperti tekstur atau deep feature, dapat dilakukan untuk meningkatkan performa dan robustitas sistem klasifikasi.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Teknologi Informasi Universitas Amikom Purwokerto atas dukungan fasilitas dan lingkungan akademik yang mendukung terlaksananya penelitian ini. Penulis juga menyampaikan terima kasih kepada seluruh pihak yang telah memberikan bantuan dan masukan selama proses penelitian.

### DAFTAR PUSTAKA

- [1] I. Essamlali, H. Nhaila, and M. El Khaili, "Heliyon Advances in machine learning and IoT for water quality monitoring: A comprehensive review," *Heliyon*, vol. 10, no. 6, p. e27920, 2024, doi: 10.1016/j.heliyon.2024.e27920.
- [2] F. Edition, I. The, and F. Addendum, *Guidelines for Drinking-water Quality*. 2017.
- [3] P. B. McIntyre *et al.*, "Global threats to human water security and river biodiversity," 2010, doi: 10.1038/nature09440.
- [4] M. Zhu *et al.*, "Eco-Environment & Health A review of the application of machine learning in water quality evaluation," *Eco-Environment Heal.*, vol. 1, no. 2, pp. 107–116, 2022, doi: 10.1016/j.eehl.2022.06.001.
- [5] I. L. Soto and Y. Concha-sánchez, "An Image-Based Water Turbidity Classification Scheme Using a Convolutional Neural Network," pp. 1–23, 2025.
- [6] K. Basha, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. Mary, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 41–48, 2015, doi: 10.1016/j.procs.2015.07.362.
- [7] H. A. Rabbani, M. A. Rahman, and B. Rahayudi, "Perbandingan Ruang Warna RGB dan HSV dalam Klasifikasi Kematangan Biji Kopi," vol. 5, no. 6, pp. 2243–2248, 2021.
- [8] A. Lokman, W. Zakiah, and W. Ismail, "A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis," pp. 1–31, 2025.
- [9] A. N. Ahmed, F. B. Othman, H. A. Afan, and A. Elsha, "Machine learning methods for better water quality prediction," vol. 578, no. May, 2019, doi: 10.1016/j.jhydrol.2019.124084.
- [10] L. E. O. Breiman, "Random Forests," pp. 5–32, 2001.
- [11] J. R. Quinlan, "Induction of Decision Trees," pp. 81–106, 2007.
- [12] L. Saitta, "Support-Vector Networks," vol. 297, pp. 273–297, 1995.
- [13] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k - nearest neighbour classifier performance for heterogeneous data sets," *SN Appl. Sci.*, vol. 1, no. 12, pp. 1–15, 2019, doi: 10.1007/s42452-019-1356-9.
- [14] P. M. B. Ong and E. R. Punzalan, "Comparative Analysis of RGB and HSV Color Models in Extracting Color Features of Green Dye Solutions," pp. 1–7, 2014.
- [15] D. M. . Powers, "EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION," pp. 37–63.
- [16] P. A. Effendi and T. Ernawati, "ANALISIS SENTIMEN ULASAN APLIKASI GAME ALGORITMA," vol. 13, no. 3, pp. 1–8, 2024.
- [17] R. Kohavi and S. Edu, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," pp. 1137–1143, 1993.
- [18] A. Fisher and C. Rudin, "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously," vol. 20, pp. 1–81, 2019.