



IMPLEMENTATION OF K-NEAREST NEIGHBOR FOR COLOR CLASSIFICATION

IMPLEMENTASI K-NEAREST NEIGHBOR UNTUK KLASIFIKASI WARNA

Riama Sibarani^{1*}, Muhamad Saefudin², Sukarno Bahat Nauli³

^{1*,3}Universitas Satya Negara Indonesia, Jakarta Selatan

²STMIK Jakarta STI&K, Jakarta Selatan

¹riama.sibarani@usni.ac.id, ²saefudin@staff.jak-stik.ac.id, ³sukarnobahat@usni.ac.id

INFORMASI ARTIKEL

Submitted:
20-07-2026

Accepted:
11-08-2026

Published:
14-08-2026

Keywords:
Computer Vision; Color
Detection; Euclidean
Distance; OpenCV; Python

Kata Kunci:
Computer Vision; Deteksi
Warna; Jarak Euclidean;
OpenCV; Python

ABSTRACT

The human eye's ability to identify colors has subjective limitations and is influenced by fatigue and ambient lighting conditions. This study aims to design and test a computer vision-based automatic color detection system using the Python programming language and the OpenCV library. The research method used was a laboratory experiment with a quantitative approach. The system was developed using the Euclidean distance algorithm to calculate the proximity input Red, Green, and Blue (RGB) pixel values to a reference database containing 865 standard color names. Testing was carried out using a dataset under various lighting conditions. The test results showed high effectiveness, with an average Macro Average Precision of 97.12%, Recall of 97.11%, and Total System Accuracy of 97.11%. The conclusion of this study is that the Python-based color detection system with the using the Euclidean distance algorithm is valid and effective for standardizing color identification, with the note that optimization or normalization of lighting in dim environments is necessary.

ABSTRAK

Kemampuan mata manusia dalam mengidentifikasi warna memiliki keterbatasan yang bersifat subjektif dan dipengaruhi oleh faktor kelelahan serta kondisi pencahayaan sekitar. Penelitian ini bertujuan untuk merancang dan menguji sistem deteksi warna otomatis berbasis computer vision menggunakan bahasa pemrograman Python dan pustaka OpenCV. Metode penelitian yang digunakan adalah eksperimental laboratorium dengan pendekatan kuantitatif. Sistem dikembangkan menggunakan algoritma jarak Euclidean untuk menghitung kedekatan nilai piksel input Red, Green, Blue (RGB) dengan basis data referensi yang berisi 865 nama warna standar. Pengujian dilakukan dengan Hasil pengujian menunjukkan efektivitas tinggi dengan rata-rata Macro Average Precision 97,12%, Recall 97,11%, dan Akurasi Total Sistem mencapai 97,11%. Kesimpulan dari penelitian ini adalah sistem deteksi warna berbasis Python dengan algoritma jarak Euclidean valid dan efektif digunakan untuk standarisasi identifikasi warna, dengan catatan perlu adanya optimasi atau normalisasi pencahayaan pada lingkungan yang redup.

INTRODUKSI

Warna merupakan salah satu fitur visual paling mendasar yang membawa informasi krusial mengenai karakteristik fisik, kimia, maupun biologis dari suatu objek [1]. Dalam berbagai implementasi praktis, perubahan atau perbedaan warna sering kali menjadi indikator utama dalam pengambilan keputusan kritis. Pada bidang kedokteran, misalnya, gradasi warna pada jaringan kulit menjadi parameter awal dalam mendeteksi keganasan kanker. Di sektor ilmu lingkungan, pergeseran spektrum warna pada perairan menjadi indikator instan terjadinya polusi atau ledakan alga beracun. Sementara itu, dalam industri manufaktur dan fashion, konsistensi warna menjadi standar mutlak penentu kualitas produk yang berdampak langsung pada nilai ekonomis perusahaan [2].

Meskipun warna memegang peran yang sangat vital, proses pengenalan dan klasifikasi warna sejauh ini masih sering mengandalkan persepsi visual manusia. Pendekatan konvensional ini memiliki kelemahan mendasar karena sifatnya yang sangat subjektif dan inkonsisten. Persepsi warna pada manusia dipengaruhi oleh berbagai faktor eksternal dan internal, seperti intensitas pencahayaan sekitar (metamerism), kelelahan mata, usia, hingga keterbatasan biologis seperti defisiensi penglihatan warna (buta warna parsial). Kesalahan interpretasi warna akibat faktor-faktor tersebut dapat berakibat fatal, mulai dari salah menegakkan diagnosis medis, keterlambatan mitigasi bencana lingkungan, hingga kegagalan produksi massal di industri. Oleh karena itu, diperlukan suatu sistem pengenalan warna otomatis yang objektif, konsisten, dan memiliki tingkat akurasi tinggi [3].

Untuk mengatasi keterbatasan tersebut, pemanfaatan teknologi computer vision yang dikombinasikan dengan algoritma machine learning menjadi solusi yang paling relevan. Salah satu algoritma klasifikasi yang terbukti efektif dan efisien dalam menangani data spasial warna adalah K-Nearest Neighbors (K-NN). Algoritma K-NN bekerja dengan cara mentransformasikan representasi visual warna ke dalam koordinat numerik yang pasti, baik melalui ruang warna RGB, HSV, maupun CIE Lab*. Melalui perhitungan jarak matematis (seperti Euclidean Distance) terhadap data latih yang telah terlabeli. K-NN mampu melakukan klasifikasi objek baru berdasarkan prinsip kedekatan karakteristik terkecil melalui mekanisme voting mayoritas tetangga terdekat [4].

Keunggulan utama K-NN terletak pada kesederhanaan komputasinya namun tetap mampu menghasilkan keputusan yang presisi untuk data yang bersifat kontinu seperti gradasi warna. Dengan meminimalkan intervensi subjektivitas manusia, penerapan K-NN dalam pengenalan warna diharapkan dapat mengeliminasi risiko human error, mempercepat waktu pemrosesan, dan memastikan setiap keputusan klinis maupun industri diambil berdasarkan fondasi data kuantitatif yang valid. Berdasarkan urgensi tersebut, penelitian ini difokuskan pada pengembangan dan analisis sistem klasifikasi warna menggunakan metode K-Nearest Neighbors (K-NN) untuk memastikan akurasi tinggi dalam proses penentuan keputusan.

Implementasi K-NN secara langsung pada dataset warna memberikan keandalan tinggi karena proses klasifikasi didasarkan pada perbandingan riil terhadap data sampel yang sudah terlabeli dengan baik. Dengan meminimalkan intervensi subjektivitas manusia dan mengandalkan nilai mutlak komponen penyusun warna dasar ini. Berdasarkan urgensi tersebut, penelitian ini difokuskan pada pengembangan dan analisis sistem klasifikasi warna menggunakan metode K-Nearest Neighbors (K-NN) berbasis ruang warna RGB untuk memastikan akurasi tinggi dalam proses penentuan keputusan [5].

METODE PENELITIAN

Penelitian ini menggunakan desain eksperimen dengan pendekatan kuantitatif. Penelitian berfokus pada implementasi dan pengujian sebuah perangkat lunak untuk mendeteksi warna secara otomatis. Eksperimen dilakukan dengan memanipulasi input kondisi pencahayaan dan objek gambar untuk mengukur Tingkat akurasi system berdasarkan parameter warna matematika.

Alur penelitian dilakukan secara bertahap, melalui siklus implementasi dan pengujian sebagai berikut:
[Mulai] → [Pengumpulan Dataset] → [Pra-pemrosesan Gambar] → [Ekstraksi Nilai RGB/HSV] → [Perhitungan Jarak Jauh (Euclidean)] → [Pencocokan Nama Warna] → [Uji Akurasi] → [Selesai]

Perangkat lunak yang digunakan adalah Windows/Linux/macOS, bahasa pemrograman Python 3.x, serta pustaka pendukung yaitu OpenCV (*Open Source Computer Vision Library*) untuk pemrosesan gambar, Pandas untuk

manajemen basis data warna, dan NumPy untuk komputasi matriks. Data diolah dalam penelitian ini adalah Data Sekunder (Basis Data Warna): Dataset referensi berupa file CSV (*Comma-Separated Values*) yang berisi daftar nama warna standar beserta nilai parameter RGB masing-masing.

Analisis Data Algoritma Sistem

Proses komputasi data warna di dalam kode Python mengikuti tahapan matematis berikut:

- Proses komputasi data warna di dalam kode Python mengikuti tahapan matematis berikut:

Pixel input = (R₁, G₁, B₁)

- Perhitungan Jarak Euclidean

Sistem membandingkan nilai warna input (R₁, G₁, B₁) dengan seluruh baris data warna referensi (R₂, G₂, B₂) yang ada di basis data CSV. Kedekatan warna dihitung menggunakan rumus Jarak Euclidean:

$$d = \sqrt{(R_2 - R_1)^2 + (G_2 - G_1)^2 + (B_2 - B_1)^2}$$

- Penentuan Keputusan (Klasifikasi)

Program mencari nilai jarak terkecil atau nilai minimum dari seluh hasil perhitungan. Warna terpilih = min (d).

Warna dengan nilai d paling minimum akan diambil label dan muncul di layer.

Pengujian dan Validasi

Sistem diuji untuk mengetahui tingkat keandalannya. Hasil deteksi otomatis oleh Python akan divalidasi silang dengan pengamatan visual manusia (sebagai Ground Truth). Tingkat keberhasilan dihitung menggunakan rumus akurasi:

$$\text{Akurasi} = \frac{\text{jumlah gambar terdeteksi benar}}{\text{Total seluruh sampel gambar}} \times 100\%$$

System dinyatakan berhasil dan valid jika persentasi akurasi total mencapai batas minimal $\geq 85\%$

Alur Pengolahan Data

Proses klasifikasi warna menggunakan algoritma K-Nearest Neighbors (K-NN) pada penelitian ini dilakukan melalui beberapa tahapan terstruktur. Secara umum, data numerik RGB dari dataset yang berjumlah 865 baris akan melewati tahap prapemrosesan, pemisahan data (*data splitting*), pemodelan K-NN, hingga tahap evaluasi akhir [6].

Dataset yang digunakan merupakan data numerik hasil ekstraksi ruang warna RGB (Red, Green, Blue) dengan ukuran dataset berskala menengah, yaitu berjumlah 865 baris data sampel. Pengelompokkan objek dibatasi pada 3 (tiga) label kelas warna target yang spesifik, yaitu Turquoise, Magenta, dan Terracotta, guna memastikan fokus pengujian pada akurasi algoritma K-NN dalam mengenali variasi gradasi dari ketiga kelas tersebut. Evaluasi performa algoritma K-Nearest Neighbors (K-NN) diukur berdasarkan parameter akurasi (*accuracy*), presisi (*precision*), dan matriks evaluasi standar lainnya (*confusion matrix*).

Struktur Dataset dan Label Kelas Warna

Dataset yang digunakan dalam penelitian ini memuat 865 sampel data warna. Setiap baris data memiliki atribut berupa nilai numerik komponen R, G, dan B yang berkisar antara 0 hingga 255 sebagai fitur masukan (*input features*). Data tersebut didistribusikan ke dalam 3 (tiga) label kelas warna target, yaitu Turquoise, Magenta, dan Terracotta. Penentuan ketiga label kelas warna ini berfungsi sebagai acuan bagi algoritma untuk melakukan proses pembelajaran (*learning*) dan pengujian (*testing*), sehingga komputer dapat mengenali batas-batas numerik yang memisahkan antara spektrum warna hijau-biru cerah (Turquoise), keunguan (Magenta), dan merah bata kecokelatan (Terracotta).

Proses Pemisahan Dataset (*Data Splitting*)

Untuk menguji keandalan model K-NN secara objektif, seluruh data sampel yang tersedia di dalam dataset tidak boleh digunakan seutuhnya untuk proses pelatihan. Oleh karena itu, penelitian ini menerapkan proses pemisahan dataset (*data splitting*) dengan proporsi pembagian 80% untuk data pelatihan dan 20% untuk data pengujian secara acak (*random sampling*), dengan rincian sebagai berikut:

1. Data Pelatihan (*Training Data*):

Data training dialokasikan sebesar 80% dari total dataset, yaitu sebanyak 692 baris data. Bagian data ini berfungsi sebagai basis pengetahuan atau titik-titik koordinat referensi di dalam ruang tiga dimensi RGB untuk ketiga kelas warna (Turquoise, Magenta, Terracotta). Ketika algoritma K-NN menerima data baru, titik-titik data training inilah yang akan dihitung jarak geometrisnya menggunakan rumus Euclidean Distance.

2. Data Pengujian (*Testing Data*):

Data testing dialokasikan sebesar 20% dari total dataset, yaitu sebanyak 173 baris data. Data ini sama sekali tidak ikut serta dalam proses pembentukan model dan disimpan sebagai data pengujian eksternal. Fungsi utama data testing adalah untuk menguji kemampuan generalisasi model K-NN. Nilai RGB dari data testing akan dimasukkan ke dalam sistem, lalu label prediksi yang dihasilkan oleh K-NN akan dicocokkan dengan label asli (*ground truth*) untuk menghitung persentase akurasi sistem.

Perhitungan Jarak dengan Euclidean Distance

Inti dari algoritma K-Nearest Neighbors (K-NN) dalam melakukan klasifikasi warna adalah mengukur kedekatan atau tingkat kemiripan antar objek berdasarkan jarak geometrisnya. Pada penelitian ini, metode yang digunakan untuk menghitung jarak antara data pengujian (*testing*) dengan seluruh data pelatihan (*training*) di dalam ruang warna RGB (*Red, Green, Blue*) adalah rumus Euclidean Distance.

Euclidean Distance mengukur jarak garis lurus terkecil antara dua buah titik di dalam ruang multi-dimensi. Karena fitur warna yang digunakan terdiri dari 3 dimensi atribut (komponen R, G dan B), maka rumus Euclidean Distance dapat dinyatakan sebagai berikut:

$$d(x, y) = \sqrt{(Rx - Ry)^2 + (Gx - Gy)^2 + (Bx - By)^2}$$

Keterangan :

$d(x, y)$ = jarak geometris (Euclidean Distance) antara titik data x dan titik data y

Rx, Gx, Bx = Nilai intensitas komponen warna *Red, Green*, dan *Blue* pada data yang diuji (*testing*)

Ry, Gy, By = Nilai intensitas komponen warna *Red, Green*, dan *Blue* pada data *training*

Proses komputasi ini akan dilakukan secara berulang terhadap seluruh 692 data pelatihan untuk setiap satu data pengujian yang masuk. Semakin kecil nilai jarak d yang dihasilkan, maka semakin tinggi tingkat kemiripan warna antara kedua titik tersebut. Nilai-nilai jarak ini kemudian diurutkan dari yang terkecil hingga terbesar untuk diambil sebanyak K tetangga terdekat sebagai dasar penentuan voting kelas warna (Turquoise, Magenta, atau Terracotta).

Implementasi kode program menggunakan Python dan pustaka Scikit-Learn untuk mengklasifikasikan warna (Turquoise, Magenta, Terracotta) menggunakan algoritma K-NN dengan pembagian dataset 865 baris data, 80:20 split [7].

HASIL DAN PEMBAHASAN

Visualisasi Spasial Data Warna pada Ruang RGB

Sebelum dilakukan pengujian performa klasifikasi menggunakan algoritma K-Nearest Neighbors (K-NN), dilakukan analisis sebaran data visual terlebih dahulu untuk memahami karakteristik geometris dari 865 baris dataset yang digunakan. Berdasarkan hasil ekstraksi fitur komponen Red, Green, dan Blue (RGB), sebaran posisi spasial dari ketiga kelas warna (Turquoise, Magenta, dan Terracotta) dipetakan ke dalam koordinat tiga dimensi (3D *Scatter Plot*).

Berdasarkan visualisasi grafik 3D yang dihasilkan melalui pustaka Matplotlib, terlihat bahwa ketiga kelas warna tersebut membentuk kelompok atau kluster (*cluster*) data yang terpisah secara signifikan di dalam ruang metrik RGB.

- Kluster Turquoise terkonsentrasi pada area dengan nilai komponen *Green* dan *Blue* yang tinggi namun memiliki nilai *Red* yang rendah.
- Kluster Magenta menempati area koordinat dengan intensitas *Red* dan *Blue* yang dominan serta nilai *Green* yang sangat minim.
- Kluster Terracotta mengelompok pada wilayah koordinat dengan nilai *Red* menengah ke atas, dengan proporsi campuran *Green* dan *Blue* yang cenderung rendah dan konstan.

Pemisahan spasial (*spatial separation*) yang jelas antar kluster ini memberikan fondasi yang sangat kuat bagi algoritma K-NN. Karena data tidak saling tumpang tindih (*overlapping*) secara masif, maka perhitungan jarak geometris menggunakan Euclidean Distance diproyeksikan mampu menentukan batasan keputusan (*decision boundary*) yang sangat tegas dan akurat. Karakteristik sebaran data ini meminimalkan ambiguitas sistem, sehingga memperkecil peluang terjadinya kesalahan pengenalan warna saat dilakukan proses pengujian (*testing*).

Analisis Matriks Evaluasi (*Confusion Matrix*)

Untuk mengukur performa pengenalan warna menggunakan algoritma K-Nearest Neighbors (K-NN) secara objektif, hasil prediksi model terhadap 173 data pengujian (*testing data*) dipetakan ke dalam sebuah matriks evaluasi standar yang disebut Confusion Matrix. Matriks ini menyajikan perbandingan antara label warna asli (*ground truth*) dengan label warna hasil prediksi sistem untuk ketiga kelas warna target, yaitu Turquoise, Magenta, dan Terracotta.

Berdasarkan hasil pengujian program Python (dengan parameter (K=3)), diperoleh data sebaran matriks evaluasi sebagaimana disajikan pada Tabel 1 berikut:

	Prediksi: Turquoise	Prediksi: Magenta	Prediksi: Terracotta	Total Aktual
Aktual: Turquoise	56	2	0	58
Aktual: Magenta	1	55	1	57
Aktual: Terracotta	0	1	1	57
Prediksi	57	58	58	173

Sumber : Olahan Peneliti

Berdasarkan data pada Tabel 4.1, performa klasifikasi untuk masing-masing kelas warna dapat dijabarkan sebagai berikut:

1. Kelas Turquoise: Dari 58 data aktual, sistem berhasil memprediksi dengan benar (*True Positive*) sebanyak 56 data, sedangkan 2 data lainnya salah diklasifikasikan (*False Negative*) sebagai warna Magenta.
2. Kelas Magenta: Dari 57 data aktual, sistem memprediksi dengan benar sebanyak 55 data. Terjadi kesalahan klasifikasi sebanyak 2 data, di mana 1 data terdeteksi sebagai Turquoise dan 1 data terdeteksi sebagai Terracotta.
3. Kelas Terracotta: Dari 58 data aktual, sistem berhasil mengenali 57 data dengan sangat presisi, dan hanya terdapat 1 data yang meleset ke dalam kategori Magenta.

Perhitungan Metrik Akurasi, Presisi, dan Recall

Guna memperkuat analisis data di atas, dilakukan perhitungan nilai metrik evaluasi turunan yang meliputi *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Ringkasan hasil performa keseluruhan model K-NN disajikan pada Tabel 2 di bawah ini:

Tabel 2. Ringkasan Performa Klasifikasi Warna K-NN

Kelas Warna	<i>Precision (%)</i>	<i>Recall (%)</i>	<i>F1-Score (%)</i>
Turquoise	98,24%	96,55%	97,39%
Magenta	94,83%	96,49%	95,65%
Terracotta	98,28%	98,28%	98,28%
Rata-rata (<i>Macro Avg</i>)	97,12%	97,11%	97,11%
Akurasi Total Sistem			97,11%

Sumber : Olahan Peneliti

Akurasi Global (97.11%): Secara keseluruhan, algoritma K-NN berbasis ruang warna RGB mampu mengklasifikasikan 168 data dengan benar dari total 173 data uji. Tingkat akurasi yang tinggi ini membuktikan bahwa metode Euclidean Distance sangat efektif dalam membedakan batas-batas nilai numerik RGB untuk ketiga kelas warna ini.

Analisis Presisi Tinggi pada Terracotta (98.28%): Kelas Terracotta memiliki tingkat presisi dan recall yang seimbang dan paling tinggi. Hal ini dikarenakan karakteristik komponen warna Terracotta (merah bata) secara matematis berada pada kuadran sumbu (R) tinggi dengan sumbu (G) dan (B) rendah, membuat posisinya sangat terisolasi dari kelompok warna Turquoise yang dominan hijau-biru.

Ambiguitas Minor pada Magenta (94.83%): Nilai presisi terendah berada pada kelas Magenta. Hal ini mengindikasikan adanya sedikit *overlapping* atau tumpang tindih nilai numerik pada area perbatasan gradasi warna. Beberapa piksel warna Turquoise dan Terracotta yang memiliki intensitas pencahayaan tertentu cenderung bergeser mendekati koordinat Magenta, sehingga memicu kesalahan voting oleh algoritma K-NN

KESIMPULAN

Secara umum, hasil evaluasi membuktikan bahwa implementasi K-NN pada dataset berukuran 865 baris ini mampu memberikan hasil klasifikasi warna yang sangat objektif dan konsisten, sehingga memperkecil celah kesalahan manusia (*human error*) dalam pengambilan keputusan berbasis warna digital. Pegujian dilakukan dengan hasil pengujian menunjukkan efektivitas tinggi dengan rata-rata Macro Average Precision 97,12%, Recall 97,11%, dan Akurasi Total Sistem mencapai 97,11%. Sistem deteksi warna berbasis Python dengan algoritma jarak Euclidean valid dan efektif digunakan untuk standarisasi identifikasi warna, dengan catatan perlu adanya optimasi atau normalisasi pencahayaan pada lingkungan yang redup.

DAFTAR PUSTAKA

- [1] D. N. Fajar Paksi, "Warna Dalam Dunia Visual," *IMAJI Film Fotogr. Telev. Media Baru*, vol. 12, no. 2, pp. 90–97, Jul. 2021, doi: 10.52290/i.v12i2.49.
- [2] R. Arista, "POTENSI BLOOMING ALGA PADA AIR SUNGAI DI DESA JUMENENG KECAMATAN MOJOANYAR KABUPATEN MOJOKERTO," vol. 1, no. 1, 2025.
- [3] L. Farokhah, "Implementasi K-Nearest Neighbor untuk Klasifikasi Bunga Dengan Ekstraksi Fitur Warna RGB," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 7, no. 6, pp. 1129–1136, Dec. 2020, doi: 10.25126/jtiik.2020722608.

- [4] D. Irfansyah, N. Ruseno, and A. Yani, "PENERAPAN COMPUTER VISION DALAM SISTEM KEAMANAN BERBASIS ARTIFICIAL INTELLIGENCE (AI), ROBOTIK, DAN OTOMATISASI," vol. 1, no. 1, 2025.
- [5] N. Puspitasari, A. Septiarini, and A. R. Aliudin, "METODE K-NEAREST NEIGHBOR DAN FITUR WARNA UNTUK KLASIFIKASI DAUN SIRIH BERDASARKAN CITRA DIGITAL," *PROSISKO J. Pengemb. Ris. Dan Obs. Sist. Komput.*, vol. 10, no. 2, pp. 165–172, Aug. 2023, doi: 10.30656/prosisko.v10i2.6924.
- [6] R. Helilintar, "K-Nears Neighbours".
- [7] W. McKinney, *Python for data analysis: data wrangling with pandas, NumPy, and Jupyter*, Third edition. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly, 2022.