

Perbandingan Algoritma Supervised Learning Terhadap Klasifikasi Citra Bunga Untuk Mengukur Akurasi Masing-Masing Algoritma

Al Danny Rian Wibisono¹, Syahrul Hidayat², Humam Maulana Tsubasanofa Ramadhan³, Ikhwanudin Ghifari⁴, Fetty Tri Anggraeny^{5*}

^{1,2,3,4} (Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur)

¹20081010010@student.upnjatim.ac.id

²20081010076@student.upnjatim.ac.id

³20081010084@student.upnjatim.ac.id

⁴20081010241@student.upnjatim.ac.id

⁵ (Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur)

*Corresponding author email: fettyanggraeny.if@upnjatim.ac.id

Abstrak— Klasifikasi citra adalah tugas penting dalam pengolahan citra yang digunakan untuk mengidentifikasi objek atau pola dalam citra. Salah satu contoh pentingnya adalah pengenalan bunga, yang bermanfaat dalam pemeliharaan taman, penelitian ekologi, dan pengawetan spesies. Namun, klasifikasi bunga secara manual rumit dan memakan waktu jika jumlah bunga yang harus diidentifikasi banyak. Untuk mengatasi tantangan tersebut, algoritma supervised learning populer dalam klasifikasi citra digunakan. Penelitian ini bertujuan membandingkan beberapa algoritma tersebut, dengan fokus pada klasifikasi citra bunga, untuk mengukur akurasi dan menentukan algoritma terbaik. Hasilnya menunjukkan bahwa algoritma Random Forest (RF) dan Logistic Regression (LR) memiliki performa terbaik, dengan akurasi RF sebesar 64.40% dan LR sebesar 62.07%. Namun, perlu dicatat bahwa hasil tersebut dapat dipengaruhi oleh faktor lain seperti kualitas dataset, metode ekstraksi fitur, dan parameter pelatihan model. Kemungkinan ada fitur lain yang lebih efektif dalam membedakan kelas bunga, yang jika digunakan dapat meningkatkan akurasi model. Oleh karena itu, perlu penyesuaian lebih lanjut untuk meningkatkan kinerja model klasifikasi.

Kata Kunci— Klasifikasi, Pengolahan Citra, Algoritma Supervised Learning, Random Forest (RF), Logistic Regression (LR).

I. PENDAHULUAN

Dalam bidang pengolahan citra, klasifikasi citra adalah salah satu tugas yang penting dan sering digunakan. Klasifikasi citra dapat membantu dalam mengidentifikasi objek atau pola tertentu dalam citra, yang memiliki aplikasi yang luas dalam berbagai bidang seperti pengenalan pola, identifikasi objek, dan analisis data visual [1].

Salah satu aplikasi penting dari klasifikasi citra adalah pengenalan bunga. Pengenalan bunga dapat memberikan manfaat dalam berbagai konteks, termasuk pemeliharaan taman, penelitian ekologi, dan pengawetan spesies. Namun, klasifikasi bunga secara manual dapat menjadi tugas yang rumit dan memakan waktu, terutama ketika jumlah bunga yang harus diidentifikasi menjadi banyak.

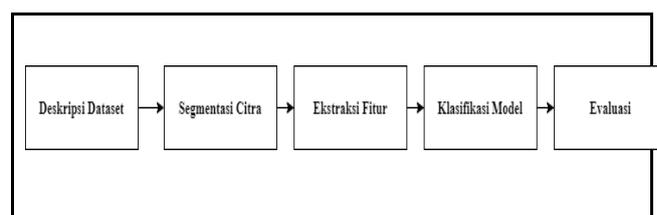
Dalam upaya untuk mengatasi tantangan tersebut, penggunaan algoritma supervised learning telah menjadi pendekatan yang populer dalam klasifikasi citra. Algoritma supervised learning menggunakan dataset yang telah diberi label untuk melatih model dan membuat prediksi yang akurat terhadap data baru.

Namun, ada berbagai algoritma supervised learning yang tersedia dan dipergunakan dalam klasifikasi citra. Setiap algoritma memiliki kelebihan dan kelemahan masing-masing dalam mengklasifikasikan citra bunga. Oleh karena itu, penting untuk melakukan perbandingan yang komprehensif untuk mengevaluasi kinerja masing-masing algoritma dalam tugas klasifikasi citra bunga.

Dalam konteks ini, penelitian ini bertujuan untuk melakukan perbandingan antara beberapa algoritma supervised learning yang populer dalam klasifikasi citra, dengan fokus pada klasifikasi citra bunga. Tujuannya adalah untuk mengukur akurasi masing-masing algoritma dan menentukan algoritma yang memberikan hasil terbaik dalam mengklasifikasikan citra bunga. Dengan pemahaman yang lebih baik tentang kinerja algoritma ini, dapat membantu dalam pengembangan solusi yang lebih efektif dalam klasifikasi citra bunga dan aplikasinya yang lebih luas.

II. METODOLOGI PENELITIAN

Secara garis besar tahapan penelitian ini yaitu mengidentifikasi masalah, pengumpulan data, perancangan sistem, pengkodean, dan pengambilan kesimpulan. Adapun rincian tahapan penelitian ditunjukkan dalam Gambar 1.



Gambar 1. Tahap Desain Penelitian Sistem yang dilakukan

Berdasarkan Gambar 1. tahap-tahap dari desain penelitian sistem yang dilakukan adalah sebagai berikut:

A. Tahap pertama menyiapkan dataset bunga yang belum tersegmentasi. Dataset diambil dari [kaggle](https://www.kaggle.com) yang terdiri atas 5 kelas bunga yang berbeda, yaitu Lilly, Lotus (Teratai), Sunflower (Bunga Matahari), Orchid (Anggrek) dan Tulip. Setiap kelas bunga memiliki 1000 gambar.

Tabel 1. Kelas Bunga pada Dataset

				
Lily	Lotus	Orchid	Sunflower	Tulip

B. Tahap kedua, melakukan segmentasi citra pada seluruh citra di dalam dataset bunga menggunakan metode *Thresholding*. Segmentasi citra adalah proses untuk memisahkan bagian-bagian atau objek-objek tertentu dari citra asli yang ingin dianalisis atau diproses lebih lanjut. Sebelum proses segmentasi dilakukan, citra-citra bunga dikonversi ke dalam citra *grayscale* menggunakan fungsi "cv2.cvtColor" dengan opsi "cv2.COLOR_BGR2GRAY". Kemudian, *thresholding* dilakukan pada citra *grayscale* menggunakan fungsi "cv2.threshold". *Thresholding* adalah suatu teknik untuk mengubah citra menjadi citra biner (hitam-putih) dengan cara memilih nilai batas tertentu (threshold) dan mengubah setiap piksel pada citra menjadi hitam atau putih berdasarkan apakah nilai piksel tersebut di atas atau di bawah nilai batas threshold [2]. Pada kode di atas, digunakan *thresholding* metode Otsu dengan opsi "cv2.THRESH_BINARY + cv2.THRESH_OTSU".

C. Tahap ketiga, melakukan ekstraksi fitur berdasarkan bentuk dan warna dari citra bunga yang telah diproses sebelumnya melalui segmentasi *thresholding*. Ekstraksi fitur adalah proses mengubah data input menjadi representasi yang lebih sederhana dan informatif yang mencerminkan karakteristik penting dari data tersebut. Ekstraksi fitur bentuk dilakukan dengan menggunakan metode *Hu Moments*. *Hu Moments* adalah deskriptor bentuk yang memperhitungkan momen dasar objek gambar dan memberikan fitur numerik yang dapat digunakan untuk membedakan objek satu sama lain. Kemudian pada ekstraksi fitur warna digunakan histogram HSV. Histogram HSV mengukur distribusi piksel gambar pada berbagai nilai Hue (H), Saturation (S), dan Value (V) untuk menggambarkan fitur warna dari citra bunga.

D. Tahap keempat, melakukan klasifikasi model menggunakan 6 algoritma *supervised learning* antara lain metode K-Nearest Neighbors (KNN), Logistic Regression (LR), Decision Tree (CT), Random Forest (RF), Neural Network (NN), dan Support Vector Machine (SVM).

1. K-Nearest Neighbors (KNN)

k-Nearest Neighbors (kNN) adalah algoritma yang berdasarkan pada prinsip bahwa objek dengan fitur serupa cenderung dalam lingkungan yang sama. Algoritma ini menentukan k titik terdekat dari objek yang diklasifikasikan, kemudian mengambil suara mayoritas untuk menentukan label kelas [3].

b. Logistic Regression (LR)

Logistic Regression (LR) adalah algoritma yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen dengan menghasilkan output yang berkisar antara 0 dan 1 [4].

c. Decision Tree (CT)

Classification/Decision Tree (CT) adalah algoritma yang membangun model dalam bentuk struktur pohon keputusan, di mana setiap simpul dalam pohon merepresentasikan keputusan berdasarkan fitur tertentu [5].

d. Random Forest (RF)

Random Forest (RF) adalah algoritma yang membangun beberapa pohon keputusan dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan output yang lebih akurat. Algoritma ini mampu menangani data yang berdimensi tinggi dan kompleks [6].

e. Neural Network (NN)

Neural Network (NN) adalah algoritma yang terinspirasi oleh struktur dan fungsi jaringan saraf manusia. Model ini terdiri dari beberapa lapisan neuron yang saling terhubung dan dapat digunakan untuk klasifikasi atau regresi [7].

f. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma yang membagi data menjadi dua kelas dengan garis hiperplane yang memaksimalkan margin antara kelas-kelas tersebut. Algoritma ini efektif dalam menangani data yang berdimensi tinggi dan dapat menghasilkan model yang stabil dan akurat [8].

E. Tahap kelima, mengevaluasi performa klasifikasi dataset bunga pada 6 algoritma *supervised learning*. Evaluasi dilakukan dengan menggunakan metrik-metrik evaluasi seperti *confusion matrix*, *f1-score*, dan *akurasi*. Tujuan dari tahap ini adalah untuk membandingkan kinerja keenam metode klasifikasi dan mengevaluasi efektifitas setiap metode dalam mengklasifikasikan jenis bunga berdasarkan dataset yang telah diolah melewati tahapan segmentasi citra dan ekstraksi fitur.

III. HASIL DAN PEMBAHASAN

A. Ekstraksi Fitur Warna

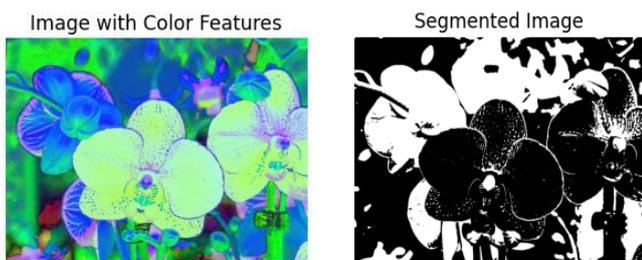
Pada proses ini warna RGB dari data citra dilakukan perubahan warna ke mode hsv atau bisa disebut dengan ekstraksi fitur dimana citra awal yang berpatokan pada pengaturan warna dasar RGB yakni Red, Green, Blue dirubah ke HSV yakni HUE (penggambaran jenis atau pewarnaan

tertentu dengan menggunakan ukuran derajat dengan kisaran 0 - 360 derajat biasanya jenis pewarnaan yang paling menonjol adalah merah, kuning, hijau, biru dan ungu), Saturation (mencerminkan kejenuhan atau kekayaan warna nilai S yang tinggi menunjukkan warna yang lebih jenuh atau kaya, sedangkan nilai S yang rendah menghasilkan warna yang lebih pucat atau abu-abu. Rentang nilai S biasanya berkisar dari 0 hingga 1 (dalam OpenCV) atau 0 hingga 255 (dalam representasi bilangan bulat), Value (Value menunjukkan tingkat kecerahan warna. Nilai V yang tinggi menunjukkan warna yang lebih cerah atau terang, sedangkan nilai V yang rendah menghasilkan warna yang lebih gelap. Rentang nilai V biasanya berkisar dari 0 hingga 1 (dalam OpenCV) atau 0 hingga 255 (dalam representasi bilangan bulat). Sehingga dari pada itu didapatkan hasil seperti Gambar 2.



Gambar 2. Sampel Bunga yang telah diekstraksi

B. Segmentasi Thresholding Otsu



Gambar 3. Sampel Bunga yang telah di segmentasi

Pada saat metode segmentasi citra menggunakan thresholding otsu disini semua dataset citra bunga yang awalnya telah diekstraksi dirubah menjadi warna hitam putih dengan metode looping agar dapat menjangkau semua data set, karena fungsi thresholding otsu adalah menentukan ambang batas (threshold) yang sesuai dengan memisahkan objek dari latar belakangnya berdasarkan tingkat keabuan atau warna piksel. Dari hal itu didapatkan hasil citra seperti pada Gambar 4.

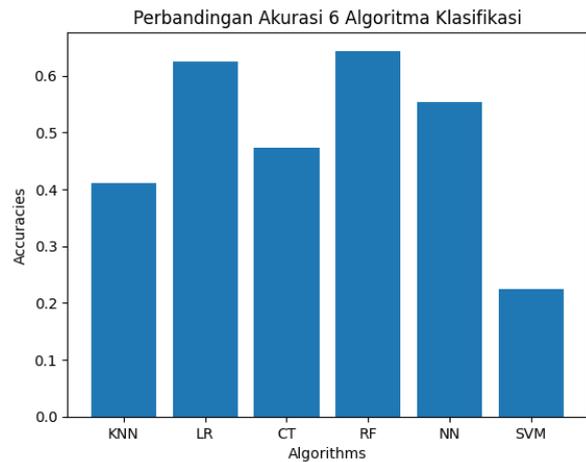
C. Ekstaksi Fitur Bentuk

Citra hasil segmentasi threshold kemudian diolah dengan menggunakan fungsi cv2.moments untuk menghitung momen-momen citra. Kemudian, momen-momen ini digunakan untuk menghitung Hu moments (cv2.HuMoments) yang merepresentasikan fitur bentuk dari citra. Hu moments ini kemudian diratakan (np.ravel) menjadi satu dimensi dan disimpan dalam variabel shape_features.

Setelah ekstraksi fitur bentuk dilakukan, kemudian dilakukan combine antara fitur-fitur bentuk (shape_features) dan warna (color_features) digabungkan menggunakan np.concatenate menjadi satu array yang disimpan dalam variabel combined_features. Variabel ini mencetak fitur-fitur gabungan dari sebuah data. Berikut adalah beberapa data awal yang ditampilkan:

```
[ 9.15700000e+03  0.00000000e+00  0.00000000e+00 ... -
8.38334783e-24
 9.60133939e-16  6.40397891e-23]
[1.30000000e+02  0.00000000e+00  0.00000000e+00 ...
1.54405412e-25
 1.77365284e-16  2.61128826e-26]
[ 2.00000000e+00  0.00000000e+00  0.00000000e+00 ... -
5.97288749e-28
 1.75959059e-17 -3.71297630e-26]
...
```

D. Perhitungan Akurasi



Gambar 4. Grafik Perbandingan Akurasi

Dari hasil ekstraksi lalu segmentasi fitur didapatkan hasil dari pada berikut :

1. Accurcy
 - K-Nearest Neighbors (KNN) : 41.20%
 - Logistic Regression (LR) : 62.07%
 - Decision Tree (CT) : 48.13%
 - Random Forest (RF) : 64.40%
 - Neural Network : 49.27%
 - Support Vector Machine (SVM) : 22.53%

Tabel 3. Evaluasi Metrix Performa

accuracy			0.62	1500
macro avg	0.62	0.62	0.62	1500
weighted avg	0.62	0.62	0.62	1500

Akurasi merupakan persentase prediksi yang benar dari total data yang dievaluasi. Dalam kasus ini, akurasi tertinggi diperoleh oleh algoritma Random Forest (RF) dengan akurasi sebesar 64.40%, diikuti oleh Logistic Regression (LR) dengan akurasi sebesar 62.07%. Algoritma K-Nearest Neighbors (KNN), Decision Tree (CT), Neural Network (NN), dan Support Vector Machine (SVM) memiliki akurasi yang lebih rendah.

2. Confusion Matrix LR:

$$\begin{bmatrix}
 [166 & 35 & 47 & 16 & 52] \\
 [44 & 196 & 41 & 5 & 23] \\
 [52 & 31 & 166 & 10 & 42] \\
 [20 & 8 & 15 & 229 & 18] \\
 [37 & 29 & 25 & 19 & 174]
 \end{bmatrix}$$

Confusion matrix memberikan gambaran tentang performa klasifikasi untuk setiap kelas [9]. Matriks ini menunjukkan jumlah prediksi yang benar (diagonal utama) dan jumlah prediksi yang salah (selain diagonal utama). Dalam kasus ini, confusion matrix untuk Logistic Regression (LR) menunjukkan jumlah prediksi yang benar dan salah untuk setiap kombinasi kelas. Misalnya, terdapat 166 prediksi yang benar untuk kelas Lilly, 35 prediksi yang salah untuk kelas Lilly, dan seterusnya.

3. Classification Report LR:

Tabel 2. Laporan Klasifikasi

Type	Precision	Recall	f1-score	support
Lilly	0.52	0.53	0.52	316
Lotus	0.66	0.63	0.64	309
Orchid	0.56	0.55	0.56	301
Sunflower	0.82	0.79	0.80	290
Tulip	0.56	0.61	0.59	284

Classification report memberikan metrik presisi, recall, dan f1-score untuk setiap kelas [10]. Presisi mengukur sejauh mana prediksi positif yang benar, recall mengukur sejauh mana kelas yang positif terdeteksi dengan benar, dan f1-score adalah ukuran harmonis antara presisi dan recall. Dalam kasus

ini, classification report untuk Logistic Regression (LR) menunjukkan metrik presisi, recall, dan f1-score untuk setiap kelas, serta rata-rata presisi, recall, dan f1-score secara keseluruhan.

4. F1-Score LR: 0.6234644275875117

F1-score adalah ukuran yang menggabungkan presisi dan recall menjadi satu angka. Ini berguna ketika terdapat ketidakseimbangan antara kelas-kelas yang diamati. Dalam kasus ini, f1-score untuk Logistic Regression (LR) adalah 0.623, yang menunjukkan keseimbangan antara presisi dan recall dalam melakukan klasifikasi.

KESIMPULAN

Berdasarkan hasil pengujian dan pembahasan di atas, dapat ditarik kesimpulan bahwa algoritma Random Forest (RF) dan Logistic Regression (LR) memiliki performa yang lebih baik daripada algoritma lain dalam melakukan klasifikasi citra bunga berdasarkan dataset yang digunakan. Algoritma Random Forest mencapai akurasi tertinggi sebesar 64.40%, diikuti oleh Logistic Regression dengan akurasi 62.07%. Namun, penting untuk dicatat bahwa hasil tersebut dapat dipengaruhi oleh faktor-faktor lain, seperti kualitas dataset, metode ekstraksi fitur, dan parameter yang digunakan dalam pelatihan model. Ada kemungkinan bahwa penggunaan fitur warna dan bentuk dalam ekstraksi fitur memiliki keterbatasan tertentu. Mungkin terdapat fitur-fitur lain yang lebih efektif dalam membedakan kelas-kelas bunga, dan jika diekstraksi, dapat meningkatkan akurasi model. Oleh karena itu, hasil yang diperoleh perlu ditafsirkan dengan hati-hati, dan mungkin diperlukan penyesuaian lebih lanjut untuk meningkatkan kinerja model klasifikasi.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebanyak-banyaknya kepada Dosen Pembimbing mata kuliah Visi Komputer, Ibu Fetty Tri Anggraeny, S.Kom, M.Kom. yang telah memberikan berbagai saran dan masukannya dalam penyelesaian artikel ini. Serta Terima kasih disampaikan kepada Tim SANTIKA yang telah meluangkan waktu untuk membuat *template* ini.

REFERENSI

[1] Nuraini, Rini, dkk., 2023. Building of Informatics, Technology and Science (BITS). *Klasifikasi Jenis Tanaman Fast Growing Species Menggunakan Algoritma Radial Basis Function Berdasarkan Citra Daun*, 4 (4), pp. 2005-2014.

[2] Samsu, LM, dkk., 2020. Infotek : Jurnal Informatika dan Teknologi. *Komparasi Algoritma Denoising Dan Binarization Dengan Adaptive Thresholding Dan Morfologi Untuk Meningkatkan Kualitas Keterbacaan Citra Naskah Lontar (Takepan) Sasak*, 3 (2), pp. 121-127.

[3] Ramadona, Suci, dkk., 2021. JITEL (Jurnal Ilmiah Telekomunikasi, Elektronika, dan Listrik

- Tenaga). *Indoor location tracking pegawai berbasis Android menggunakan algoritma k-nearest neighbor*, 1 (1), pp.51-58.
- [4] Phiophuead, T., Kunsuwan, N., 2019. ENGINEERING JOURNAL. *Logistic Regression Analysis of Factors Affecting Travel Mode Choice for Disaster Evacuation*, 23 (6), pp.400-417.
- [5] Esananda, S.C, dkk., 2021. Jurnal Informatika dan Sistem Informasi (JIFoSI). *Penerapan Algoritma Decision Tree Dalam Menentukan Prestasi Akademik Siswa*, 2 (2), pp.413-424.
- [6] Agustiani, Sarifah, dkk., 2022. Jurnal Komputasi. *Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram*, 10 (1), pp.65-74.
- [7] Al-Khowarizmi, 2020. JITE (Journal of Informatics and Telecommunication Engineering). *Model Classification of Nominal Value and The Original of IDR Money By Applying Evolutionary Neural Network*, 3 (2), pp.259-265.
- [8] Nalatissifa, Hiya, dkk., 2020. Jurnal Informatika Universitas Pamulang. *Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja*, 5 (4), pp.578-584.
- [9] Ridwan, A., 2020. Jurnal Sistem Komputer dan Kecerdasan Buatan. *Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus*, 4 (1), pp.15-21.
- [10] Maghfiroh, Alfiah, dkk., 2023. Building of Informatics, Technology and Science (BITS). *Klasifikasi Penipuan pada Rekening Bank menggunakan Pendekatan Ensemble Learning*, 4 (4), pp. 1883-1891.