

## Klasifikasi Jenis Daun Herbal Menggunakan Metode Logistic Regression dan Decision Tree Classifier Berdasarkan Fitur (Warna dan Bentuk)

Gabriela Honesty<sup>1</sup>, Mayang Sajida<sup>2</sup>, Agung Ramadhanu<sup>3</sup>

<sup>1,2,3</sup> Universitas Putra Indonesia "YPTK", Padang, Indonesia

<sup>1</sup> Gabrielahonesty12@gmail.com, <sup>2</sup>Mayangsajida024@gmail.com

### ARTICLE INFO

|          |            |           |            |
|----------|------------|-----------|------------|
| Submit   | 19-01-2024 | Review    | 22-01-2024 |
| Accepted | 30-01-2024 | Published | 05-02-2024 |

### ABSTRACT

Penelitian ini membandingkan metode klasifikasi citra daun herbal menggunakan Logistic Regression dan Decision Tree Classifier. Ekstraksi fitur dilakukan berdasarkan warna (HSV), dan bentuk (eksentrisitas dan metrik). Fitur yang digunakan untuk mengenali jenis daun meliputi bentuk, warna, dan tekstur. Tidak semua jenis fitur perlu digunakan untuk melakukan komputasi hasil ekstraksi, namun perlu diseleksi beberapa fitur yang paling berpengaruh dalam sistem temu kembali citra daun. Analisis keterkaitan korelasi antar fitur melalui seleksi fitur juga dikombinasikan dengan penggunaan kedekatan dalam menghitung similaritas pada sistem temu kembali. Dari data uji sebanyak 40 buah dan 3 jenis klasifikasi daun herbal (anggur, Ketapang, dan pisang) diperoleh keberhasilan pengujian sebanyak 37 data dengan tingkat akurasi yang layak. Hasil Berdasarkan hasil pengujian menggunakan 40 sampel citra daun, dimana masing-masing jenis daun digunakan 4 sampai 6 sampel citra, memberikan persentase keberhasilan dalam identifikasi sebesar 80,93%.

**Keyword :** Klasifikasi citra, Logistic Regression, Decision Tree Classifier

### 1. Introduction

Tanaman Herbal mempunyai banyak manfaat bagi kehidupan. Berdasarkan pengamatan manusia memiliki senyawa yang bermanfaat untuk mencegah, menyembuhkan penyakit, melakukan fungsi biologis tertentu, hingga mencegah serangan serangga dan jamur. (Noviana and Nugraha 2023)

Proses pengenalan tanaman herbal dengan cara pengambilan gambar daun herbal kemudian melakukan pengenalan pola daun dengan cara mengenali karakteristik struktural daun seperti warna, dan bentuk. Proses klasifikasi tanaman herbal dilakukan berdasarkan warna, dan bentuk daun dengan metode Decision Tree Classifier dan Logistic Regression. Proses identifikasi yang dilakukan dengan model citra masukan untuk menganalisa karakteristik dari struktur daun. Fitur warna dianalisis menggunakan Hue Saturation Value (HSV), dianalisis menggunakan Dissimilarity, Correlation, Homogeneity, Contrast, ASM dan Energy, sedangkan fitur bentuk dianalisis menggunakan metric, eccentricity. Penelitian ini menggunakan dua jenis fitur yang melibatkan bentuk dan warna. Ekstraksi fitur bentuk yang digunakan meliputi slimness, form factor/roundness, rectangularity, narrow factor, rasio antara keliling dan diameter, dan rasio antara keliling dan panjang serta lebar daun. (Arum Sari et al., n.d.)

Berdasarkan penelitian ini akan dilakukan perbandingan klasifikasi daun herbal dengan metode Decision Tree Classifier dan Logistic Regression berdasarkan ekstraksi fitur warna dengan HSV, dan ekstraksi fitur bentuk dengan eccentricity dan metric. Sebelum dilakukan tahapan pengelompokkan maka tahap pertama yang dilakukan tahap preprocessing citra dan ekstraksi fitur citra untuk mendapatkan nilai masukan yang tepat dalam tahapan klasifikasi jenis daun berdasarkan citra daun.

Regresi Logistik adalah salah satu teknik dalam statistik yang umumnya digunakan untuk pemodelan prediktif pada data dengan output berupa variabel biner, seperti klasifikasi dalam image processing di mana gambar diklasifikasikan ke dalam kategori tertentu. Dalam konteks pengolahan citra (image processing), Regresi Logistik digunakan untuk klasifikasi citra berdasarkan fitur-fitur yang diambil dari citra tersebut. Misalnya, dalam jurnal yang disebutkan sebelumnya, fitur-fitur seperti warna, dan bentuk dari citra daun herbal dapat diekstraksi, kemudian digunakan sebagai input untuk model Regresi Logistik. (Noviana and Nugraha 2023)

Langkah-langkah umum dalam menggunakan Regresi Logistik untuk klasifikasi citra dalam image processing termasuk:

Pertama; Ekstraksi Fitur: Fitur-fitur penting dari citra diekstraksi. Fitur-fitur ini bisa berupa warna, tekstur, bentuk, atau fitur lainnya yang relevan dengan jenis klasifikasi yang dilakukan. Kedua; Pembentukan Model: Fitur-fitur yang diekstraksi dari citra digunakan sebagai input untuk model Regresi Logistik. Model ini akan belajar untuk memahami pola atau karakteristik fitur-fitur yang berkaitan dengan kategori atau klasifikasi yang diinginkan. Ketiga; Pelatihan Model: Model Regresi Logistik dilatih dengan menggunakan data yang telah diklasifikasikan sebelumnya. Dalam konteks ini, citra daun herbal yang sudah diklasifikasikan akan digunakan sebagai data pelatihan untuk mengajarkan model tentang hubungan antara fitur-fitur yang diekstraksi dan kategori klasifikasi yang diinginkan. Keempat; Evaluasi Model: Setelah pelatihan, model dievaluasi menggunakan data yang tidak digunakan sebelumnya (data uji) untuk mengukur seberapa baik model dapat mengklasifikasikan citra yang belum pernah dilihat sebelumnya. Kelima; Prediksi: Setelah model dinilai baik, itu dapat digunakan untuk mengklasifikasikan citra baru berdasarkan fitur-fitur yang diekstraksi.

Regresi Logistik sering digunakan karena kelebihanannya dalam mengatasi masalah klasifikasi biner pada data yang memadai dalam menangani masalah overfitting dengan jumlah fitur yang relatif besar, serta relatif mudah untuk diinterpretasi. Namun, untuk beberapa kasus, model klasifikasi lainnya seperti Decision Tree, Random Forest, atau algoritma Deep Learning mungkin memberikan kinerja yang lebih baik tergantung pada kompleksitas masalah dan struktur data.

Metode Decision Tree Classifier adalah teknik klasifikasi yang menghasilkan aturan keputusan berbentuk struktur pohon. Dalam konteks pengolahan citra (image processing), Decision Tree Classifier digunakan untuk mengklasifikasikan citra berdasarkan fitur-fitur yang diekstraksi dari citra tersebut. (Noviana and Nugraha 2023). Langkah-langkah umum dalam menggunakan Decision Tree Classifier untuk klasifikasi citra dalam image processing termasuk:

Pertama; Ekstraksi Fitur: Fitur-fitur penting diekstraksi dari citra, seperti warna, tekstur, bentuk, atau fitur lainnya yang relevan dengan klasifikasi yang dilakukan. Kedua; Pembentukan Pohon Keputusan: Fitur-fitur yang diekstraksi dari citra digunakan sebagai input untuk membangun struktur pohon keputusan. Decision Tree Classifier memilih fitur-fitur yang paling penting untuk memisahkan kelas-kelas citra. Ketiga; Pembelajaran Model: Model Decision Tree dibentuk dengan menggunakan data yang telah diklasifikasikan sebelumnya. Citra-citra yang sudah diberi label akan digunakan sebagai data pelatihan untuk mengajarkan model tentang aturan-aturan keputusan yang dibentuk dari fitur-fitur yang diekstraksi. Keempat; Evaluasi Model: Setelah pelatihan, model dievaluasi menggunakan data yang tidak digunakan sebelumnya

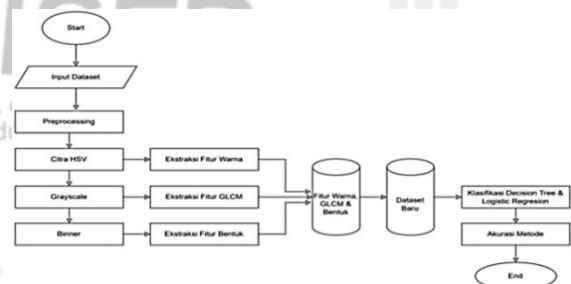
(data uji) untuk mengukur seberapa baik model dapat mengklasifikasikan citra yang belum pernah dilihat sebelumnya. Kelima; Prediksi: Setelah model dinilai baik, ia dapat digunakan untuk mengklasifikasikan citra baru berdasarkan fitur-fitur yang diekstraksi.

Keunggulan Decision Tree Classifier termasuk kemampuan untuk memberikan aturan keputusan yang mudah diinterpretasi dan dipahami. Pohon keputusan dapat diperiksa untuk melihat logika di balik setiap keputusan klasifikasi. Namun, dalam beberapa kasus, Decision Tree mungkin cenderung overfitting jika tidak diatur dengan baik atau jika terlalu kompleks.

Untuk masalah klasifikasi citra yang lebih kompleks atau dengan fitur-fitur yang sangat banyak, terkadang digunakan variasi dari Decision Tree seperti Random Forest (yang merupakan ensemble dari banyak Decision Tree) untuk meningkatkan kinerja dan mencegah overfitting. Setiap metode klasifikasi, termasuk Decision Tree Classifier, memiliki kelebihan dan kelemahan masing-masing dan penerapannya bisa tergantung pada kompleksitas masalah, jenis fitur, dan ukuran dataset yang digunakan.

## 2. Research Methods

Tahap penelitian yang pertama menginput dataset sebanyak 11 yaitu: Base, Cermai, Dapdap, JS, Kayutoktok, MD, Menuh, Piduh, Pucuk, Pule dan TB.



Gambar 1. Alur kerja metodologi penelitian

### 2.1 Preprocessing

Preprocessing merupakan suatu proses untuk menghilangkan bagian-bagian yang tidak diperlukan pada gambar input untuk proses selanjutnya

### 2.2 Klasifikasi

Klasifikasi merupakan proses menemukan sekumpulan model atau fungsi yang menggambarkan dan membedakan konsep atau kelas-kelas data, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek atau data yang label kelasnya tidak diketahui. Klasifikasi terdiri atas dua tahap, yaitu tahap training (pelatihan) dan prediksi (klasifikasi). Pada tahap pelatihan dibentuk sebuah model domain permasalahan dari setiap kasus atau instance yang ada. Setiap teknik klasifikasi menggunakan algoritma pembelajaran untuk mengklasifikasi model yang memberikan hubungan yang paling sesuai antara himpunan atribut dan label kelas dari data input. Pendekatan umum yang digunakan

dalam masalah klasifikasi adalah pertama, training set berisi record yang mempunyai label kelas yang diketahui haruslah tersedia. Training set digunakan untuk membangun model klasifikasi, yang kemudian diaplikasikan ke test set, yang berisi record-record dengan label kelas yang tidak diketahui. (Mentari Awanda, dkk 2018.)

### 2.3 Ekstraksi Fitur Warna

Ciri warna digunakan apabila objek-objek yang akan dikenali mempunyai warna yang berbeda. Misalnya, untuk membedakan citra buah apel Amerika berwarna merah dengan apel Malang berwarna hijau digunakan ciri warna sebagai parameternya. Parameter-parameter warna didapat dengan cara normalisasi setiap komponen warna RGB (Red Green Blue) pada citra menggunakan persamaan 1 dibawah ini : (Andono, dkk., 2017).

$$r = r / (r+g+b) \dots\dots\dots(1)$$

$$g = g / (r+g+b) \dots\dots\dots(2)$$

$$b = b / (r+g+b) \dots\dots\dots(3)$$

R = nilai persentasi merah

G = nilai persentasi hijau

B = nilai persentasi biru

r = jumlah nilai merah

g = jumlah nilai hijau

b = jumlah nilai biru

Dengan memanfaatkan nilai r, g, dan b yang telah dinormalisasi, rumus transformasi RGB ke HSV

$$V = \max(r, g, b) \dots\dots\dots(4)$$

$$S = \begin{cases} 0 & \text{jika } V = 0 \\ V - \frac{\min(r, g, b)}{V} & \text{jika } V > 0 \end{cases} \dots\dots\dots(5)$$

$$H = \begin{cases} 0 & \text{jika } S = 0 \\ \frac{60 \times (g-b)}{S \times V} & \text{jika } V = r \\ 60 \times \left[ 2 + \frac{(b-r)}{S \times V} \right] & \text{jika } V = g \\ 60 \times \left[ 4 + \frac{(r-g)}{S \times V} \right] & \text{jika } V = b \end{cases} \dots\dots\dots(6)$$

$$H = H + 360 \text{ jika } H < 0 \dots\dots\dots(7)$$

Keterangan:

H = Hue.

S = Saturation.

V = Value.

r = Nilai R yang dinormalisasi.

g = Nilai G yang dinormalisasi.

b = Nilai B yang dinormalisasi.

### 2.3 Ekstraksi Fitur Bentuk

Ciri bentuk merupakan karakter dari suatu objek konfigurasi oleh garis dan kontur. Fitur bentuk bergantung pada Teknik yang digunakan. Kategori tersebut berdasarkan batas (boundry-based) dan

berdasarkan daerah (region-based). Untuk membedakan bentuk objek satu dengan objek lainnya dapat menggunakan parameter eccentricity dan metric. (Noviana and Nugraha 2023)

Eccentricity merupakan nilai perbandingan antara jarak foci ellips minor dengan foci ellips mayor suatu objek. Eccentricity mempunyai rentang nilai antara 0 hingga 1. Objek berbentuk memanjang atau mendekati bentuk garis lurus maka mendekati angka 1, sedangkan objek yang berbentuk bulat atau lingkaran dengan mendekati angka 0. Metric merupakan nilai perbandingan antara luas dan keliling objek. Metric memiliki rentang nilai antara 0 hingga 1. Objek berbentuk memanjang atau mendekati bentuk garis lurus maka nilai metric mendekati 0 sedangkan objek berbentuk bulat atau lingkaran maka nilai metric mendekati angka 1. Parameter eccentricity dan metric merupakan teknik ekstraksi fitur yang bertujuan untuk mengambil atau mengekstraksi nilai eccentricity dan metric.

#### 1) Contrast

Contrast merepresentasikan variasi antar derajat keabuan suatu daerah. Jika terletak jauh dari pusat diagonal utama, maka nilai kontras akan besar. Rumus menghitung kontras ditunjukkan pada Persamaan 4 (Kusuma, dkk., 2017).

$$\text{var} = \sum_i \sum_j (i-j)^2 p(i, j) \dots\dots\dots(4)$$

#### 2) Correlation

Nilai korelasi merepresentasikan tingkat abu-abu ketergantungan linier antara piksel pada posisi tertentu terhadap piksel lain. Rumus menghitung korelasi ditunjukkan pada Persamaan 5 (Kusuma, dkk., 2017).

$$Cr = \sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y \sigma_x \sigma_y \dots\dots\dots(5)$$

#### 3) Energy

Energy menyatakan tingkat keseragaman piksel-piksel suatu citra. Semakin tinggi nilai Energy, maka semakin seragam teksturnya. Rumus untuk menghitung Energy dapat dilihat pada persamaan 6 (Kusuma, dkk., 2017).

$$\text{Energy} = \sum_i \sum_j p(i, j)^2 \dots\dots\dots(6)$$

#### 4) Homogeneity

Homogeneity menyatakan ukuran kedekatan setiap elemen dari cooccurrence matrix. Rumus untuk menghitung Homogeneity dapat dilihat pada persamaan 7 (Kusuma, dkk., 2017).

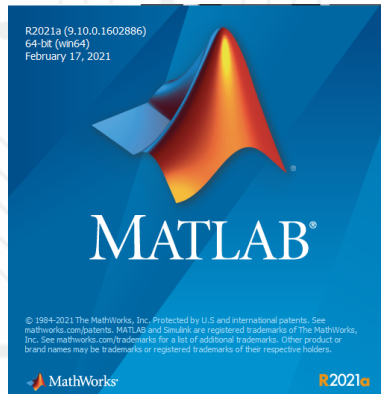
$$\text{Homogeneity} = \sum_i \sum_j p(i, j) (1 + |i-j|) \dots\dots\dots(7)$$

## 3. Results and Discussions

Berdasarkan hasil pengujian yang dilakukan dengan menggunakan 2 (dua) metode yaitu Logistic Regression dan Decision Tree Classifier dengan dataset daun 275 maka didapatkan hasil akurasi yang paling tinggi adalah Decision Tree Classifier dengan hasil akurasi train 100% dan akurasi test 78,31 sedangkan Logistic Regression dengan hasil akurasi train 72,9% dan akurasi test 60,24.

### 3. Pengujian

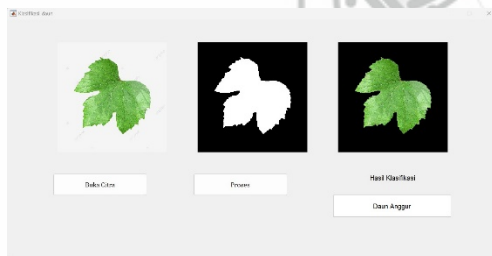
Pengujian dilakukan dengan menggunakan aplikasi Matlab. Aplikasi Matlab merupakan salah satu aplikasi yang dapat melakukan analisis dan olah data citra dan penerapannya pada *artificial intelligence*. Berikut adalah tampilan awal aplikasi Matlab.



Gambar 2. Tampilan awal aplikasi matlab

#### a. Pengujian dengan jenis daun anggur

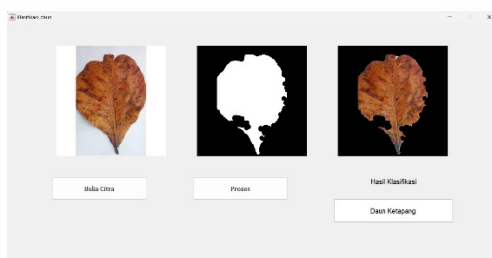
Pengujian pertama dengan melakukan ujicoba terhadap data daun dengan klasifikasi jenis daun anggur. Dari 7 data latih dan data uji jenis daun anggur diperoleh tingkat keberhasilan sebesar 71,4% yakni 5 data berhasil diidentifikasi dengan baik sebagai daun anggur.



Gambar 3. Tampilan proses ujicoba jenis daun anggur

#### b. Pengujian dengan jenis daun ketapang

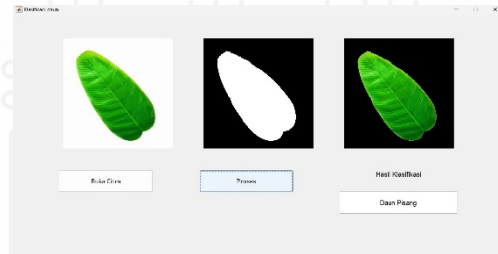
Pengujian kedua dengan melakukan ujicoba terhadap data daun dengan klasifikasi jenis daun ketapang. Dari 6 data latih dan data uji jenis daun anggur diperoleh tingkat keberhasilan sebesar 66,7% yakni 4 data berhasil diidentifikasi dengan baik sebagai daun ketapang.



Gambar 4. Tampilan proses ujicoba jenis daun ketapang

#### c. Pengujian dengan jenis daun pisang

Pengujian ketiga dengan melakukan ujicoba terhadap data daun dengan klasifikasi jenis daun pisang. Dari 9 data latih dan data uji jenis daun anggur diperoleh tingkat keberhasilan sebesar 77,7% yakni 7 data berhasil diidentifikasi dengan baik sebagai daun pisang.



Gambar 5. Tampilan proses ujicoba jenis daun pisang

### 4. Conclusion

Dari penelitian yang dilakukan hasil uji coba klasifikasi citra daun herbal maka penggunaan metode Logistic Regression dengan fitur warna dan fitur bentuk didapatkan nilai akurasi dari data uji sebanyak 40 buah dan 3 jenis klasifikasi daun herbal (anggur, Ketapang, dan pisang) diperoleh keberhasilan pengujian sebanyak 37 data dengan tingkat akurasi yang layak. Hasil Berdasarkan hasil pengujian menggunakan 40 sampel citra daun, dimana masing-masing jenis daun digunakan 4 sampai 6 sampel citra, memberikan persentase keberhasilan dalam identifikasi sebesar 80,93%.

### Reference

- Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Beneroso, D., & Robinson, J. (2021). A tool for assessing and providing personalised formative feedback at scale within a second in engineering courses. *Education for Chemical Engineers*, 36, 38–45.
- Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Ibrahim, S. U., Mustafa, N., Supa'at, I., Zainal, N., & Ahmad, M. Y. (2021). The effects of adaptive feedback on student's learning gains. *International Journal of Advanced Computer Science and Applications*, 12(7), 68–80.
- Bimba, A. T., Idris, N., Al-Hunaiyyan, A., Mahmud, R. B., & Mohd Shuib, N. L. (2017). Adaptive feedback in computer-based.
- Andono, P. N., Sutojo, T., & Muljono. (2017). *Pengolahan Citra Digital* Yogyakarta: Andi.



**JISED**  
Journal of Information System  
and Education Development