

**KLASIFIKASI JENIS TANAMAN CABAI MENGGUNAKAN CITRA DIGITAL  
DAN REGRESI LINEAR BERGANDA**

**SKRIPSI**

**DISUSUN OLEH**

**ESA MAHENDRA**  
**NPM. 2009010013**



**UMSU**  
Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2024**

**KLASIFIKASI JENIS TANAMAN CABAI MENGGUNAKAN  
CITRA DIGITAL DAN REGRESI LINEAR BERGANDA**

**SKRIPSI**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer  
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer  
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**ESA MAHENDRA  
NIM. 2009010013**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

**MEDAN**

**2024**

LEMBAR PENGESAHAN

Judul Skripsi : KLASIFIKASI JENIS TANAMAN CABAI  
MENGUNAKAN CITRA DIGITAL DAN  
REGRESI LINEAR BERGANDA

Nama Mahasiswa : ESA MAHENDRA

NPM : 2009010013

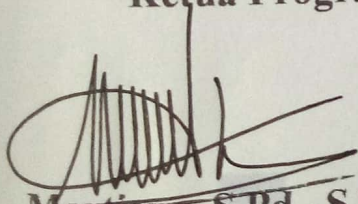
Program Studi : Sistem Informasi


Menyetujui,  
Komisi Pembanding

  
(Dr. Alkhowarizmi, S.Kom., M.Kom )  
NIDN. 0127099201

Ketua Program Studi

Dekan

  
Martiano, S.Pd., S.Kom., M.Kom  
NIDN.0128029302

  
Dr. Alkhowarizmi, S.Kom., M.Kom  
NIDN. 0127099201



PERNYATAAN ORISINILITAS

KLASIFIKASI JENIS TANAMAN CABAI MENGGUNAKAN  
CITRA DIGITAL DAN REGRESI LINEAR BERGANDA

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya saya sendiri,kecuali beberapa kutipan dan ringkasan yang masing – masing disebutkan sumbernya

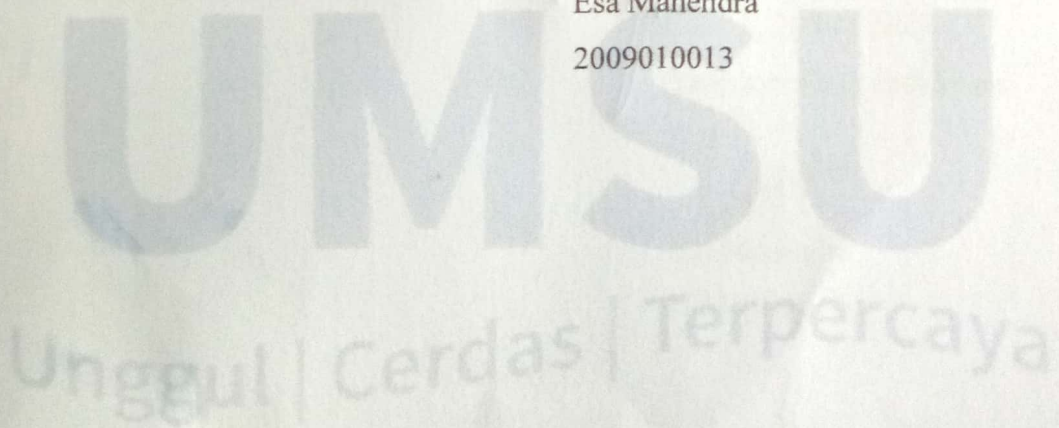
Medan, 29 Juli 2024

Yang Membuat Pernyataan



Esa Mahendra

2009010013



**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK  
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Esa Mahendra  
NPM : 2009010013  
Program Studi : Sistem Informasi  
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

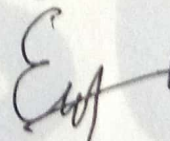
**Klasifikasi Jenis Tanaman Cabai Menggunakan Citra Digital Dan  
Regresi Linear Berganda**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 29 Juli 2024

Yang membuat pernyataan



Esa Mahendra

2009010013

## **DAFTAR RIWAYAT HIDUP**

Nama Lengkap : Esa Mahendra  
Tempat dan Tanggal Lahir : Medan, 09 November 2001  
Alamat Rumah : JL.BalaiDesa Gg. Gama. Medan Polonia  
Telepon/Faks/HP : 088261027910  
E-mail : esamahendra09@gmail.com  
Instansi Tempat Kerja : Belum Bekerja  
Alamat Kantor : -  
DATA PENDIDIKAN : SDIT ALIF TAMAT  
SMP MUHAMMADIYAH 3 MEDAN  
SMA NEGERI 1 MEDAN TAMAT



## KATA PENGANTAR



Assalamu'alaikum Wr. Wb

Dalam kesempatan ini penulis bersyukur kepada Allah SWT, karena berkat Ridho-Nya penulis mampu menampilkan skripsi yang berjudul “Klasifikasi Jenis Tanaman Cabai Menggunakan Citra Digital Dan Regresi Linear Berganda”. Skripsi ini disusun sebagai kewajiban penulis guna melengkapi tugas dan syarat untuk menyelesaikan pendidikan Strata-1 serta memperoleh gelar Sarjana Komputer Program Studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera. Dalam penulisan skripsi penulis menyadari masih banyak terdapat kekurangan dan jauh dari kata kesempurnaan ini. Oleh sebab itu penulis sangat senang menerima saran dan kritikan demi kesempurnaan penelitian ini. maka dari itu dalam kesempatan ini penulis mengucapkan rasa terimakasih yang sedalam-dalamnya kepada:

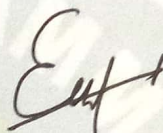
1. Bapak Prof. Dr. Agussani, M.AP selaku Rektor Universitas Muhammadiyah Sumatera Utara.
2. Bapak Dr. Alkhowarizmi, S.Kom., M.Kom selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
3. Bapak Martiono, S.Pd, S.Kom, M.Kom selaku Ketua Prodi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
4. Bapak Dr. Alkhowarizmi, S.Kom., M.Kom, selaku dosen pembimbing penulis yang telah banyak meluangkan waktunya untuk memberikan serta memberikan arahan dan bimbingan dalam penulisan skripsi ini.
5. Dosen dan staf pengajar di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara yang telah memberikan pengetahuan dan ilmu selama perkuliahan.
6. Orang tua dan keluarga tercinta yang telah membesarkan, merawat dan mendidik serta memberikan doa dan motivasi kepada penulis sampai penulis menyelesaikan pendidikan sampai menyelesaikan skripsi ini.
7. Teman-teman seperjuangan yang ada di Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara Prodi Sistem Informasi.

Penulis juga tidak lupa mengucapkan terima kasih yang sebesar-besarnya kepada sahabat dan rekan mahasiswa yang banyak memberikan balasan yang berarti bagi penulis. Penulis juga meminta maaf atas semua kekurangan dan kesalahan yang ada pada penulisan skripsi ini, semoga skripsi ini bermanfaat bagi pembaca.

Wassalamualaikum Wr. Wb.

Medan, 29 Juli 2024

Penulis



Esa Mahendra

2009010013



UMSU

Unggul | Cerdas | Terpercaya



## **ABSTRAK**

Penelitian ini berfokus pada penerapan metode klasifikasi menggunakan citra digital dan regresi linear berganda untuk mengidentifikasi jenis tanaman cabai berdasarkan fitur tekstur dan bentuk yang diekstraksi dari citra daun. Dalam prosesnya, citra digital tanaman cabai melalui tahapan pra-pemrosesan untuk meningkatkan kualitas gambar, diikuti oleh ekstraksi fitur menggunakan metode seperti Gray-Level Co-occurrence Matrix (GLCM). Penelitian ini menggunakan 100 dataset citra tanaman cabai yang diperoleh dari situs BRIN, yang kemudian dibagi menjadi data latih dan data uji untuk melatih model regresi linear berganda. Namun, hasil penelitian menunjukkan bahwa model regresi linear berganda tidak mampu menangkap kompleksitas data dengan baik, sebagaimana ditunjukkan oleh nilai R-squared negatif dan kesalahan prediksi yang besar. Oleh karena itu, disarankan untuk menerapkan teknik pengurangan dimensi dan validasi silang guna memperbaiki kinerja model dan meningkatkan akurasi dalam klasifikasi jenis tanaman cabai di masa mendatang.

Kata Kunci : Regresi Linier, Citra Digital, Tanaman Cabai

## **ABSTRACT**

*This research focuses on the application of classification methods using image processing and multiple linear regression to identify chili plant types based on texture and shape features extracted from leaf images. In the process, the digital images of chili plants undergo pre-processing stages to enhance image quality, followed by feature extraction using methods such as the Gray-Level Co-occurrence Matrix (GLCM). This study utilizes a dataset of 100 chili plant images obtained from the BRIN website, which were then split into training and testing data to train the multiple linear regression model. However, the results indicate that the multiple linear regression model was unable to capture the complexity of the data well, as demonstrated by negative R-squared values and significant prediction errors. Therefore, it is recommended to implement dimension reduction techniques and cross-validation to improve model performance and increase accuracy in the classification of chili plant types in the future.*

**Keywords:** *Linear Regression, Digital Imagery, Chili Plants*

## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
PERNYATAAN ORISINILITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	iii
DAFTAR RIWAYAT HIDUP .....	iv
KATA PENGANTAR .....	v
ABSTRAK.....	vii
<i>ABSTRACT</i> .....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL .....	xi
DAFTAR GAMBAR.....	xii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
BAB II LANDASAN TEORI.....	5
2.1 Penelitian Terdahulu .....	5
2.2 Tanaman Cabai.....	6
2.3 Klasifikasi .....	7
2.4 Citra Digital .....	8
2.5 <i>Regresi Linier Berganda</i> .....	9



2.6	Phyton .....	10
2.7	OpenCV .....	12
2.8	Pandas .....	13
2.9	NumPy .....	15
BAB III METODE PENELITIAN .....		17
3.1	Data Penelitian .....	17
3.2	Arsitektur Umum .....	19
3.3	Data Preparation.....	21
BAB IV HASIL DAN PEMBAHASAN.....		24
4.1	Pembahasan.....	24
4.1.1	Deteksi Citra .....	24
4.1.2	Teknik Pre-Processing .....	29
4.1.3	Penerapan Metode Regresi Linier Berganda.....	33
BAB V KESIMPULAN DAN SARAN .....		39
5.1	Kesimpulan .....	39
5.2	Saran .....	39
DAFTAR PUSTAKA .....		40

## **DAFTAR TABEL**

Tabel 2. 2 Penelitian Terdahulu .....	5
Tabel 3. 1 Data Penelitian .....	17
Tabel 3. 2 Pemrosesan Gambar .....	22
Tabel 3. 3 Gray-Level Co-occurrence Matrix.....	22

## DAFTAR GAMBAR

Gambar 3. 1 Arsitektur Penelitian .....	19
Gambar 4. 1 Dataset Jenis Tanaman Cabai .....	24
Gambar 4. 2 Komponen Citra YCbCr .....	25
Gambar 4. 3 Citra Binner .....	26
Gambar 4. 4 Grafik Distribusi Kemungkinan Pada Sampel Citra.....	27
Gambar 4. 5 Citra Erosi Dan Citra Dilasi.....	28
Gambar 4. 6 Grafik Dilasi Dan Grafik Erosi.....	28
Gambar 4. 7 Hasil Citra Grayscale .....	30
Gambar 4. 8 Hasil Ekstrasi .....	30
Gambar 4. 9 Proses Labeling.....	32
Gambar 4. 10 Hasil Penerapan Metode .....	33
Gambar 4. 11 Grafik Scatter Plot .....	34
Gambar 4. 12 Residuals.....	35
Gambar 4. 13 Grafik Histogram.....	36
Gambar 4. 14 Grafik Scatter Matrix .....	37



## Daftar Lampiran

Lampiran 1 Persetujuan Judul Penelitian .....	41
Lampiran 2. Berita Acara Bimbingan Proposal .....	42
Lampiran 3. Berita Acara Bimbingan Skripsi .....	43

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Tanaman cabai (*Capsicum spp.*) merupakan salah satu komoditas hortikultura yang memiliki nilai ekonomi tinggi di berbagai negara, termasuk Indonesia. Cabai sering digunakan sebagai bahan dasar bumbu masakan dan juga memiliki berbagai manfaat kesehatan. Dengan permintaan yang terus meningkat, produksi dan kualitas cabai menjadi sangat penting untuk diperhatikan. Dalam upaya meningkatkan produktivitas dan kualitas tanaman cabai, teknologi pengolahan citra digital dan analisis data menjadi alat yang semakin relevan. Penggunaan citra digital dalam klasifikasi jenis tanaman cabai merupakan langkah maju dalam bidang pertanian presisi. Citra digital membuat pengamatan yang lebih akurat dan efisien terhadap karakteristik tanaman cabai seperti bentuk daun, warna buah, dan tekstur permukaan. Teknologi ini dapat membantu petani dalam mengidentifikasi jenis cabai secara cepat dan tepat yang pada gilirannya dapat meningkatkan pengelolaan tanaman dan hasil panen (PUTRA 2022).

Permasalahan yang dihadapi dalam membedakan jenis tanaman cabai meliputi berbagai aspek yang kompleks. Variasi bentuk, ukuran, dan warna cabai seringkali sulit untuk dibedakan dengan pengamatan visual sederhana. Kesalahan dalam identifikasi jenis cabai dapat berdampak signifikan pada seluruh rantai pasok mulai dari penanaman, perawatan, hingga pemasaran. Identifikasi yang tidak akurat dapat mengakibatkan penggunaan teknik budidaya yang tidak sesuai dan kesalahan dalam aplikasi pestisida yang akhirnya mempengaruhi kualitas dan kuantitas hasil panen. Solusi untuk permasalahan ini adalah dengan memanfaatkan teknologi digital khususnya pengolahan citra digital. Penggunaan citra digital dalam klasifikasi jenis tanaman cabai membuat pengamatan yang lebih detail dan presisi terhadap karakteristik fisik tanaman. Teknologi ini dapat mengidentifikasi perbedaan bentuk, warna, dan tekstur cabai secara otomatis sehingga dapat membantu petani dalam menentukan jenis cabai dengan cepat dan akurat. Proses ini juga dapat dilakukan secara konsisten tanpa tergantung pada kemampuan pengamat manusia.

Metode analisis yang dapat digunakan untuk mendukung solusi ini adalah regresi linear berganda. Regresi linear berganda membuat analisis hubungan antara beberapa variabel independen yang berasal dari data citra digital dengan variabel dependen berupa jenis cabai. Dengan memasukkan berbagai fitur citra seperti intensitas warna, bentuk, dan tekstur ke dalam model regresi dapat diperoleh model prediksi yang mampu mengklasifikasikan jenis cabai secara akurat. Selain itu, model ini juga memberikan wawasan mengenai pentingnya masing-masing fitur dalam proses klasifikasi. Penggunaan regresi linear berganda dalam klasifikasi jenis tanaman cabai memberikan keuntungan dalam hal interpretabilitas dan akurasi. Model yang dihasilkan dapat menunjukkan variabel mana yang paling berpengaruh dalam klasifikasi, sehingga dapat menjadi panduan bagi petani dalam mengidentifikasi dan menangani tanaman cabai. Selain itu, validasi model dengan data yang beragam akan memastikan bahwa model ini robust dan dapat diandalkan dalam berbagai kondisi lingkungan (G. W. Saputra and Ardani 2020).

Dalam penelitian yang dilakukan oleh (Mutmainna, Achmad, and Suhardi 2017) dalam penelitiannya berhasil mengembangkan model algoritma pendugaan kadar lengas tanah menggunakan citra Landsat 8 dengan beberapa pendekatan. Pertama, menggunakan reflektansi band tunggal, model terbaik diperoleh dengan band 2 (biru), yang menunjukkan koefisien determinasi ( $R^2$ ) sebesar 0,653. Selanjutnya, pendekatan dengan indeks vegetasi menghasilkan model terbaik menggunakan Soil Adjusted Vegetation Index (SAVI), dengan koefisien determinasi ( $R^2$ ) sebesar 0,843. Ini menunjukkan bahwa SAVI memiliki korelasi tertinggi dan paling signifikan dalam menduga kadar lengas tanah dibandingkan pendekatan lainnya.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, maka penulis mencoba melakukan penelitian dengan judul “ **Klasifikasi Jenis Tanaman Cabai Menggunakan Citra Digital Dan Regresi Linear Berganda**”.



## **1.2 Rumusan Masalah**

Adapun rumusan masalah dalam penelitian ini yaitu bagaimana citra digital dapat digunakan untuk mengidentifikasi dan mengklasifikasikan jenis tanaman cabai serta Seberapa akurat model regresi linear berganda dalam mengklasifikasikan jenis tanaman cabai berdasarkan fitur citra digital.

## **1.3 Batasan Masalah**

Adapun batasan masalah dalam penelitian ini sebagai berikut :

1. Penelitian ini hanya memiliki 100 dataset jenis tanaman cabai.
2. Penelitian ini hanya menggunakan metode regresi linear berganda untuk analisis dan klasifikasi jenis cabai
3. Analisis data dan pengolahan citra digital akan dilakukan menggunakan Google Colab
4. Penelitian ini hanya akan menggunakan Python dan OpenCV untuk pengolahan citra digital.

## **1.4 Tujuan Penelitian**

Tujuan dari penelitian ini yaitu :

1. Membuat model klasifikasi jenis tanaman cabai menggunakan citra digital dan regresi linear berganda.
2. Menentukan tingkat akurasi model regresi linear berganda dalam klasifikasi jenis tanaman cabai.
3. Mengidentifikasi fitur citra digital yang paling signifikan dalam klasifikasi jenis tanaman cabai.

## **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini yaitu sebagai berikut :

1. Dengan adanya model klasifikasi yang akurat petani dapat mengidentifikasi jenis cabai secara cepat dan tepat. Hal ini akan mempermudah dalam penerapan teknik budidaya yang sesuai sehingga dapat meningkatkan produktivitas tanaman cabai.
2. Penelitian ini menambah wawasan dan pengetahuan dalam bidang teknologi pengolahan citra digital dan analisis regresi linear berganda

3. Dengan adanya klasifikasi yang akurat, kualitas cabai yang dipasarkan dapat lebih terjaga khususnya dalam aplikasi di bidang pertanian.

## BAB II

### LANDASAN TEORI

#### 2.1 Penelitian Terdahulu

Adapun penelitian terdahulu terkait penelitian yang akan dilakukan sebagai berikut :

Tabel 2. 1 Penelitian Terdahulu

Judul Penelitian	Hasil Penelitian	Tahun Penelitian
<b>Klasifikasi Tanaman Hortikultura Menggunakan Citra Digital dan Algoritma Machine Learning</b>	Penggunaan algoritma machine learning dalam klasifikasi tanaman hortikultura menunjukkan akurasi yang tinggi dengan citra digital. Dalam penelitian ini, beberapa algoritma seperti Random Forest, SVM, dan KNN digunakan untuk klasifikasi, dan hasil menunjukkan akurasi lebih dari 90%.	2019
<b>Pendugaan Lengan Tanah Inceptisol Pada Tanaman Hortikultura Menggunakan Citra Landsat 8</b>	Algoritma pendugaan menggunakan citra Landsat 8 efektif dalam memprediksi kadar lengan tanah dengan koefisien determinasi tinggi. Model terbaik diperoleh menggunakan reflektansi band tunggal, terutama band 2 (biru), dengan koefisien determinasi ( $R^2$ ) sebesar 0,653. Pendekatan dengan indeks vegetasi menggunakan SAVI menunjukkan korelasi tertinggi dengan $R^2$ sebesar 0,843.	2017
<b>Identifikasi Jenis Tanaman Cabai Menggunakan Pengolahan Citra Digital</b>	Pengolahan citra digital efektif dalam mengidentifikasi jenis cabai dengan akurasi yang lebih baik dibandingkan metode manual. Penggunaan teknik pengolahan citra seperti segmentasi dan ekstraksi fitur membantu dalam	2020



	membedakan jenis cabai berdasarkan karakteristik fisik seperti warna, bentuk, dan tekstur. Akurasi identifikasi mencapai 88%.	
<b>Penggunaan Regresi Linear Berganda dalam Klasifikasi Jenis Tanaman Pertanian</b>	Regresi linear berganda memberikan model klasifikasi yang akurat untuk berbagai jenis tanaman pertanian berdasarkan fitur citra digital. Penelitian ini menunjukkan bahwa regresi linear berganda mampu mengintegrasikan berbagai fitur citra seperti warna, tekstur, dan bentuk untuk menghasilkan model prediksi yang akurat dengan $R^2$ sebesar 0,75.	2018
<b>Analisis Citra Digital untuk Klasifikasi Jenis Cabai Menggunakan Metode Statistik</b>	Metode statistik dalam analisis citra digital menunjukkan hasil yang signifikan dalam klasifikasi jenis cabai dengan tingkat akurasi yang memadai. Penelitian ini menggunakan analisis diskriminan dan regresi logistik untuk klasifikasi dan mendapatkan akurasi klasifikasi sebesar 85%.	2021

## 2.2 Tanaman Cabai

Tanaman cabai (*Capsicum* spp.) adalah salah satu jenis tanaman hortikultura yang termasuk dalam keluarga Solanaceae. Tanaman ini dikenal dengan buahnya yang memiliki rasa pedas dan digunakan sebagai bumbu dalam berbagai masakan di seluruh dunia. Cabai memiliki beberapa spesies yang populer, seperti *Capsicum annuum*, *Capsicum frutescens*, *Capsicum chinense*, *Capsicum pubescens*, dan *Capsicum baccatum*. Setiap spesies cabai memiliki karakteristik yang berbeda, baik dari segi bentuk, warna, ukuran, maupun tingkat kepedasannya. Cabai sering dibudidayakan karena nilai ekonominya yang tinggi serta manfaat kesehatannya. Buah cabai kaya akan vitamin C, vitamin A, dan berbagai antioksidan, seperti capsaicin, yang memberikan rasa pedas sekaligus berperan sebagai anti-inflamasi dan analgesik. Tanaman cabai juga mudah tumbuh di berbagai kondisi iklim dan tanah, menjadikannya tanaman yang ideal

untuk dibudidayakan di banyak negara, terutama di daerah tropis dan subtropis (G. W. Saputra and Ardani 2020).

Tanaman cabai memiliki siklus hidup yang relatif singkat, dan dapat ditanam sepanjang tahun di daerah yang memiliki iklim hangat. Budidaya cabai melibatkan beberapa tahap mulai dari penyemaian, pemindahan bibit ke lahan tanam, pemeliharaan yang meliputi penyiraman, pemupukan, serta pengendalian hama dan penyakit, hingga tahap panen. Cabai biasanya mulai berbuah setelah 2-3 bulan dari masa tanam, dan dapat dipanen berulang kali selama beberapa bulan. Penggunaan cabai dalam kuliner sangat luas, baik dalam bentuk segar, kering, atau diolah menjadi saus dan bumbu. Selain sebagai bumbu masakan, cabai juga digunakan dalam industri farmasi dan kosmetik karena kandungan capsaicin-nya yang bermanfaat. Hal ini menjadikan tanaman cabai sebagai komoditas penting dalam perdagangan global dan lokal, serta sebagai sumber pendapatan yang signifikan bagi banyak petani (PUTRA 2022).

### **2.3 Klasifikasi**

Klasifikasi adalah proses pengelompokan objek atau fenomena ke dalam kategori atau kelas yang telah ditentukan berdasarkan karakteristik atau atribut tertentu. Dalam berbagai bidang ilmu, klasifikasi digunakan untuk mengorganisir data atau informasi sehingga dapat lebih mudah dipahami, dianalisis, dan digunakan. Klasifikasi membantu dalam mengenali pola, memahami hubungan antara berbagai elemen, dan memudahkan pengambilan keputusan. Dalam konteks ilmu biologi, klasifikasi sering merujuk pada taksonomi, yaitu pengelompokan organisme hidup ke dalam kelompok-kelompok berdasarkan kesamaan dan perbedaan dalam ciri-ciri morfologi, genetik, dan ekologis. Sistem klasifikasi ini membantu ilmuwan dalam memahami keanekaragaman hayati, evolusi, dan hubungan kekerabatan antarspesies. Taksonomi mencakup hierarki klasifikasi dari tingkat yang lebih umum ke tingkat yang lebih spesifik, seperti kingdom, phylum, class, order, family, genus, dan species (Setiawan et al. 2023).

Di bidang teknologi informasi, klasifikasi digunakan dalam pengolahan data dan kecerdasan buatan. Misalnya, dalam pembelajaran mesin (machine learning), klasifikasi adalah salah satu tugas utama di mana algoritma dilatih untuk mengelompokkan data baru ke dalam kategori yang telah ditentukan sebelumnya. Contoh penerapan klasifikasi

dalam teknologi informasi termasuk pengenalan wajah, pengelompokan email spam, dan deteksi anomali dalam keamanan siber. Dalam ilmu kesehatan, klasifikasi penyakit membantu tenaga medis dalam mendiagnosis dan merawat pasien. Sistem klasifikasi seperti International Classification of Diseases (ICD) yang disusun oleh Organisasi Kesehatan Dunia (WHO) memberikan standar global untuk mengkodekan berbagai jenis penyakit dan kondisi kesehatan. Dengan adanya sistem klasifikasi ini, data kesehatan dapat dianalisis secara konsisten di berbagai negara dan lembaga kesehatan, memungkinkan penelitian dan perencanaan kesehatan yang lebih efektif (Idris et al. 2022).

Di bidang ekonomi, klasifikasi industri atau sektor digunakan untuk mengelompokkan perusahaan atau aktivitas ekonomi berdasarkan jenis produk atau jasa yang dihasilkan. Klasifikasi ini membantu dalam analisis pasar, perencanaan ekonomi, dan kebijakan publik. Sistem klasifikasi seperti International Standard Industrial Classification (ISIC) atau North American Industry Classification System (NAICS) memberikan kerangka kerja untuk memahami struktur ekonomi dan tren industri (Idris et al. 2022).

## **2.4 Citra Digital**

Citra digital adalah representasi visual dari sebuah objek, pemandangan, atau fenomena yang telah dikonversi ke dalam format digital, terdiri dari piksel-piksel yang masing-masing memiliki nilai numerik tertentu. Nilai-nilai ini, yang dikenal sebagai nilai intensitas atau nilai warna, membentuk gambar yang dapat dilihat dan diproses oleh komputer. Citra digital digunakan dalam berbagai bidang, termasuk fotografi, penginderaan jauh, kedokteran, dan pemrosesan video. Dalam citra digital, setiap piksel merupakan elemen terkecil dari gambar yang merepresentasikan satu titik warna atau intensitas cahaya tertentu. Piksel-piksel ini diatur dalam bentuk matriks dua dimensi, di mana setiap piksel memiliki koordinat tertentu dan nilai warna yang terdiri dari kombinasi warna dasar (merah, hijau, dan biru dalam sistem RGB) atau tingkat keabuan dalam gambar hitam putih. Resolusi citra digital, yang ditentukan oleh jumlah piksel dalam citra, sangat mempengaruhi kualitas dan ketajaman gambar (Mutaqin and Kom 2023).

Proses pembentukan citra digital melibatkan pengambilan gambar melalui sensor atau perangkat perekam, seperti kamera digital, scanner, atau satelit. Sensor ini menangkap informasi cahaya yang dipantulkan atau dipancarkan oleh objek dan mengubahnya menjadi sinyal elektrik. Sinyal ini kemudian dikonversi menjadi data digital melalui proses yang disebut digitalisasi. Data digital yang dihasilkan dapat disimpan, diproses, dan ditampilkan menggunakan perangkat elektronik seperti komputer atau layar digital (Mutaqin and Kom 2023).

Salah satu keunggulan utama citra digital adalah kemampuannya untuk diedit dan diproses dengan berbagai teknik pemrosesan citra. Pemrosesan citra digital mencakup berbagai operasi, seperti perbaikan kualitas gambar, segmentasi, pengenalan pola, dan analisis fitur. Teknologi ini memungkinkan manipulasi gambar untuk meningkatkan visualisasi, mendeteksi objek tertentu, atau mengurangi noise dan artefak yang tidak diinginkan. Contohnya, dalam fotografi digital, aplikasi pengeditan gambar memungkinkan pengguna untuk mengatur kecerahan, kontras, warna, dan detail lainnya sesuai keinginan (Mutaqin and Kom 2023).

## 2.5 *Regresi Linier Berganda*

Analisis regresi berganda adalah suatu metode untuk meramalkan nilai pengaruh dua variabel independen atau lebih terhadap satu variabel dependen. Lebih mudahnya yaitu untuk membuktikan ada tidaknya hubungan antara dua variabel atau lebih dari dua variabel independen  $X_1, X_2, X_3, \dots, X_i$  terhadap satu variabel terikat  $Y$  (Pramanda, Diara, and Sarana 2024). Persamaan umum analisis regresi yaitu sebagai berikut :

$$Y = B_x + \epsilon$$

Dimana :

- $Y$  : Variabel Dependen
- $B$  : Parameter
- $X$  : Variabel Independen
- $\epsilon$  : Error

Hubungan antara satu variabel dependen dengan satu atau lebih variabel independen dapat dinyatakan dalam regresi linier berganda. Hubungan tersebut dapat dinyatakan secara umum sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$$

Dimana :

$Y_i$	: Variabel dependen untuk pengamatan ke $i=1,2,\dots,n$
$\beta_0, \beta_1, \dots, \beta_k$	: Parameter
$X_{i1}, X_{i2}, \dots, X_{ik}$	: Variabel independen
$\varepsilon_i$	: Sisaan untuk pengamatan ke $i$

## 2.6 Python

Python adalah bahasa pemrograman tingkat tinggi yang dikenal karena kemudahan penggunaannya dan sintaks yang mudah dibaca. Dikembangkan oleh Guido van Rossum dan pertama kali dirilis pada tahun 1991, Python dirancang untuk menekankan keterbacaan kode, memungkinkan para pengembang untuk menulis logika pemrograman yang jelas untuk proyek-proyek besar dan kecil. Bahasa ini mendukung berbagai paradigma pemrograman, termasuk pemrograman berorientasi objek, pemrograman fungsional, dan pemrograman prosedural. Salah satu kekuatan utama Python adalah kemampuannya untuk berintegrasi dengan berbagai sistem dan platform, serta ketersediaan pustaka standar yang luas. Pustaka standar Python mencakup modul-modul untuk menangani berbagai tugas umum seperti manipulasi string, pengolahan file, komunikasi jaringan, dan pengujian unit. Selain itu, ada ribuan pustaka pihak ketiga yang tersedia melalui Python Package Index (PyPI) yang memperluas fungsionalitas Python untuk berbagai aplikasi, seperti analisis data, kecerdasan buatan, pengembangan web, dan banyak lagi (Nugraha, Harani, and Habibi 2020).

Python sangat populer dalam komunitas ilmu data dan kecerdasan buatan berkat pustaka-pustaka seperti NumPy, pandas, matplotlib, scikit-learn, TensorFlow, dan PyTorch. Pustaka-pustaka ini memungkinkan para peneliti dan praktisi untuk



melakukan analisis data, pembelajaran mesin, dan pemrosesan bahasa alami dengan lebih efisien. Misalnya, NumPy menyediakan dukungan untuk array multidimensi dan operasi matematika yang cepat, sementara pandas menawarkan struktur data yang fleksibel dan alat analisis data yang kuat. Di bidang pengembangan web, Python juga memiliki peran penting dengan adanya kerangka kerja seperti Django dan Flask. Django adalah kerangka kerja web tingkat tinggi yang mendorong pengembangan cepat dan desain yang bersih dan pragmatis. Flask, di sisi lain, adalah kerangka kerja mikro yang lebih ringan dan memberikan fleksibilitas lebih besar kepada pengembang untuk memilih komponen yang mereka butuhkan. Keduanya memungkinkan pembuatan aplikasi web yang skalabel dan aman dengan cepat (Hartatik et al. 2023).

Python juga digunakan secara luas dalam otomatisasi skrip dan administrasi sistem. Sintaks yang sederhana dan pustaka standar yang kuat memungkinkan administrator sistem untuk menulis skrip yang kompleks untuk mengotomatisasi tugas-tugas seperti pengelolaan server, pemrosesan batch, dan analisis log. Kemampuan Python untuk bekerja di berbagai sistem operasi, termasuk Windows, macOS, dan Linux, membuatnya menjadi alat yang berharga bagi administrator sistem dan pengembang perangkat lunak (Hartatik et al. 2023).

Dalam pendidikan, Python sering menjadi bahasa pemrograman pertama yang diajarkan kepada siswa pemula. Keterbacaan dan kesederhanaan sintaksnya membuat Python ideal untuk mengajarkan konsep-konsep dasar pemrograman tanpa membingungkan siswa. Banyak institusi pendidikan dan kursus online menggunakan Python untuk memperkenalkan dasar-dasar pemrograman, algoritma, dan struktur data. Selain itu, ada banyak sumber daya pembelajaran yang tersedia, termasuk buku, dokumentasi, tutorial, dan komunitas yang aktif, yang semuanya membantu pemula untuk belajar dan menguasai Python dengan cepat. Keberhasilan Python juga didukung oleh komunitasnya yang besar dan aktif. Komunitas Python terdiri dari pengembang, peneliti, dan pengguna dari berbagai latar belakang yang berkontribusi terhadap perkembangan bahasa ini. Mereka terlibat dalam pembuatan pustaka baru, penyempurnaan pustaka yang ada, dan pengembangan fitur-fitur baru untuk Python. Konferensi seperti PyCon juga menjadi platform di mana anggota komunitas dapat

bertukar ide, berkolaborasi, dan belajar dari satu sama lain (Nugraha, Harani, and Habibi 2020) .

## 2.7 OpenCV

OpenCV, atau Open Source Computer Vision Library, adalah pustaka perangkat lunak sumber terbuka yang dikembangkan oleh Intel pada tahun 1999 untuk mendukung aplikasi visi komputer dan pembelajaran mesin. Pustaka ini dirancang untuk menyediakan infrastruktur umum bagi aplikasi visi komputer dan mempercepat penggunaan persepsi mesin dalam produk komersial. Sejak pertama kali dirilis, OpenCV telah berkembang pesat dan kini digunakan secara luas dalam berbagai aplikasi, mulai dari penelitian akademik hingga industri teknologi tinggi. OpenCV menyediakan lebih dari 2500 algoritma yang dioptimalkan yang mencakup berbagai tugas visi komputer. Algoritma-algoritma ini meliputi deteksi dan pengenalan wajah, identifikasi objek, klasifikasi tindakan manusia dalam video, pelacakan pergerakan, segmentasi gambar, dan banyak lagi. Dengan pustaka yang kaya akan fitur ini, OpenCV memungkinkan para pengembang untuk membuat aplikasi yang kompleks dengan lebih efisien dan efektif (Putra et al. 2023).

Salah satu keunggulan utama OpenCV adalah dukungannya terhadap berbagai platform. OpenCV dapat dijalankan di sistem operasi seperti Windows, Linux, macOS, Android, dan iOS. Ini membuatnya sangat fleksibel dan dapat diintegrasikan ke dalam berbagai lingkungan pengembangan. Selain itu, OpenCV mendukung banyak bahasa pemrograman, termasuk C++, Python, Java, dan MATLAB, yang memudahkan pengguna dari berbagai latar belakang untuk mengadopsi dan menggunakan pustaka ini dalam proyek mereka. Dalam aplikasi industri, OpenCV digunakan untuk berbagai tujuan, seperti sistem pengawasan, analisis gambar medis, sistem panduan kendaraan otonom, dan realitas tertambah (AR). Misalnya, dalam sistem pengawasan, OpenCV dapat digunakan untuk mendeteksi dan melacak pergerakan objek atau orang dalam video secara real-time. Di bidang medis, OpenCV membantu dalam analisis citra medis untuk diagnosis dan penelitian. Dalam teknologi kendaraan otonom, OpenCV memainkan peran penting dalam deteksi jalur, pengenalan rambu lalu lintas, dan penghindaran hambatan (Andani et al., n.d.).

OpenCV juga sangat populer di komunitas penelitian dan pendidikan. Banyak universitas dan lembaga pendidikan menggunakan OpenCV untuk mengajarkan konsep-konsep dasar visi komputer dan pemrosesan gambar. Dengan dokumentasi yang luas dan komunitas pengguna yang aktif, OpenCV menyediakan sumber daya yang cukup bagi para pemula maupun peneliti berpengalaman untuk belajar dan mengembangkan aplikasi mereka. Tersedianya berbagai tutorial, kursus online, dan forum diskusi membantu memfasilitasi pembelajaran dan berbagi pengetahuan di antara pengguna OpenCV. Selain itu, OpenCV sering digunakan dalam proyek-proyek inovatif dan penelitian di bidang kecerdasan buatan (AI) dan pembelajaran mesin (ML). Dengan integrasi pustaka pembelajaran mesin seperti TensorFlow dan PyTorch, OpenCV memungkinkan pengembangan aplikasi AI yang canggih. Misalnya, OpenCV dapat digunakan bersama dengan model deep learning untuk deteksi objek, pengenalan wajah, dan segmentasi gambar dengan tingkat akurasi yang tinggi (Andani et al., n.d.).

## **2.8 Pandas**

Pandas adalah pustaka perangkat lunak sumber terbuka yang digunakan untuk manipulasi dan analisis data dalam bahasa pemrograman Python. Dibuat oleh Wes McKinney pada tahun 2008, Pandas menyediakan struktur data yang fleksibel dan alat analisis data yang kuat. Pustaka ini telah menjadi standar de facto dalam analisis data dengan Python, digunakan secara luas oleh ilmuwan data, analis, dan pengembang di seluruh dunia. Salah satu fitur utama Pandas adalah DataFrame, yang merupakan struktur data dua dimensi yang mirip dengan tabel dalam basis data atau lembar kerja dalam spreadsheet. DataFrame memungkinkan penyimpanan dan manipulasi data dalam bentuk baris dan kolom dengan cara yang sangat intuitif dan efisien. Pengguna dapat dengan mudah melakukan operasi seperti menggabungkan, memfilter, mengelompokkan, dan mengubah data dalam DataFrame, menjadikannya alat yang sangat berguna untuk analisis data (Kamal, n.d.).

Pandas juga mendukung berbagai format input dan output data, termasuk CSV, Excel, SQL, JSON, dan banyak lagi. Kemampuan ini memungkinkan pengguna untuk dengan mudah membaca data dari berbagai sumber, memprosesnya dalam Pandas, dan kemudian menyimpan hasilnya kembali ke format yang diperlukan. Integrasi yang

mulus dengan pustaka Python lainnya seperti NumPy, matplotlib, dan scikit-learn juga memperkuat fungsionalitas Pandas dalam ekosistem analisis data. Keunggulan lain dari Pandas adalah kemampuannya untuk menangani data yang hilang atau tidak lengkap dengan cara yang efisien. Pandas menyediakan berbagai metode untuk mendeteksi, mengisi, atau menghapus data yang hilang, yang sangat penting dalam analisis data dunia nyata di mana data seringkali tidak lengkap atau tidak sempurna. Fitur ini membantu ilmuwan data dan analis untuk membersihkan dan mempersiapkan data sebelum melakukan analisis lebih lanjut (Y. A. Saputra, Pane, and Awangga 2020).

Pandas juga mendukung operasi waktu dan tanggal yang kompleks, memungkinkan analisis data berbasis waktu seperti deret waktu. Fitur ini termasuk konversi antara format tanggal dan waktu yang berbeda, resampling, pergeseran, dan perhitungan perbedaan waktu. Kemampuan ini sangat berguna dalam berbagai aplikasi, termasuk analisis keuangan, pelacakan kinerja bisnis, dan analisis tren jangka panjang. Dalam pendidikan dan penelitian, Pandas telah menjadi alat yang sangat populer untuk mengajarkan dan mempelajari analisis data. Banyak kursus dan buku tentang ilmu data menggunakan Pandas sebagai bagian penting dari kurikulum mereka. Dokumentasi yang luas, komunitas yang aktif, dan ketersediaan berbagai tutorial dan sumber daya online membuat Pandas mudah diakses oleh pemula maupun pengguna berpengalaman. Secara keseluruhan, Pandas adalah pustaka yang sangat kuat dan serbaguna untuk manipulasi dan analisis data dalam Python. Dengan struktur data yang intuitif, dukungan untuk berbagai format data, kemampuan untuk menangani data yang hilang, dan fitur-fitur analisis waktu, Pandas telah menjadi alat yang sangat penting dalam toolkit ilmuwan data dan analis. Keberlanjutan dan perkembangan Pandas di masa depan akan terus mendorong inovasi dalam analisis data dan ilmu data secara keseluruhan (Rifky et al. 2024).

## 2.9 NumPy

NumPy, singkatan dari Numerical Python, adalah pustaka sumber terbuka yang digunakan untuk komputasi ilmiah dalam bahasa pemrograman Python. Dibuat oleh Travis Oliphant pada tahun 2005, NumPy menyediakan dukungan untuk array dan matriks multidimensi serta berbagai fungsi matematika yang efisien untuk operasi pada array ini. Pustaka ini menjadi fondasi bagi banyak pustaka dan alat lain dalam ekosistem ilmu data dan komputasi ilmiah di Python. Salah satu fitur utama NumPy adalah objek array n-dimensinya, yang disebut `ndarray`. Array ini memungkinkan penyimpanan dan manipulasi data dalam bentuk yang sangat efisien, baik dari segi memori maupun komputasi. Operasi matematika pada `ndarray` sangat cepat karena NumPy menggunakan implementasi yang dioptimalkan, sering kali memanfaatkan pustaka seperti BLAS dan LAPACK yang ditulis dalam bahasa pemrograman C dan Fortran (Nugraha, Harani, and Habibi 2020).

NumPy juga menyediakan berbagai fungsi matematika dan statistik yang kuat, termasuk operasi aritmatika dasar, fungsi trigonometri, aljabar linear, transformasi Fourier, dan banyak lagi. Fungsi-fungsi ini memungkinkan pengguna untuk melakukan komputasi ilmiah yang kompleks dengan mudah dan efisien. Kemampuan ini sangat penting dalam berbagai bidang, seperti analisis data, pembelajaran mesin, simulasi ilmiah, dan pengolahan sinyal. Selain itu, NumPy memiliki kemampuan integrasi yang baik dengan pustaka lain dalam ekosistem Python. Pustaka ini sering digunakan bersama dengan `pandas` untuk manipulasi data, `matplotlib` untuk visualisasi data, dan `scikit-learn` untuk pembelajaran mesin. Integrasi yang mulus ini memungkinkan ilmuwan data dan peneliti untuk membangun pipeline analisis data yang kompleks dengan lebih mudah (Nugraha, Harani, and Habibi 2020).

NumPy juga mendukung operasi vektorisasi, yang memungkinkan eksekusi operasi pada seluruh array tanpa perlu menggunakan loop Python. Vektorisasi ini tidak hanya membuat kode lebih bersih dan lebih mudah dibaca, tetapi juga secara signifikan meningkatkan kinerja komputasi. Operasi yang diimplementasikan dengan vektorisasi biasanya jauh lebih cepat dibandingkan dengan implementasi loop tradisional di Python. Dalam pendidikan dan penelitian, NumPy adalah alat yang sangat penting untuk mengajarkan dan mempelajari komputasi ilmiah dan analisis data. Banyak



kursus dan buku teks tentang ilmu data, fisika, dan matematika menggunakan NumPy sebagai bagian dari kurikulum mereka. Dokumentasi yang luas, komunitas yang aktif, dan berbagai tutorial dan sumber daya online membuat NumPy mudah diakses oleh pemula maupun pengguna berpengalaman. Secara keseluruhan, NumPy adalah pustaka yang sangat kuat dan serbaguna untuk komputasi ilmiah dan analisis data dalam Python. Dengan dukungan untuk array multidimensi, berbagai fungsi matematika yang efisien, dan integrasi yang baik dengan pustaka lain, NumPy telah menjadi fondasi penting dalam ekosistem ilmu data Python. Keberlanjutan dan perkembangan NumPy di masa depan akan terus mendorong inovasi dalam komputasi ilmiah dan analisis data (Batubara and Awangga 2020)

### BAB III METODE PENELITIAN






#### 3.1 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data pengelompokan jenis tanaman cabai. Data pengelompokan jenis tanaman cabai diperoleh dari situs [brin.go.id](http://brin.go.id). Dalam sistem ini, para petani yang terdaftar mengisi data terkait jenis tanaman cabai yang mereka budidayakan, termasuk karakteristik fisik dan kondisi pertumbuhannya. Data ini kemudian digunakan untuk mengidentifikasi dan mengklasifikasikan jenis tanaman cabai yang ditanam. Total data yang digunakan pada penelitian ini adalah 100 baris data yang berhasil diambil dari situs [brin.go.id](http://brin.go.id). Data ini mencakup informasi seperti gambar tanaman cabai yang sudah di tanam. Adapun datanya sebagai berikut :

Tabel 3. 1 Data Penelitian

No.	Gambar Tanaman
1.	
2.	

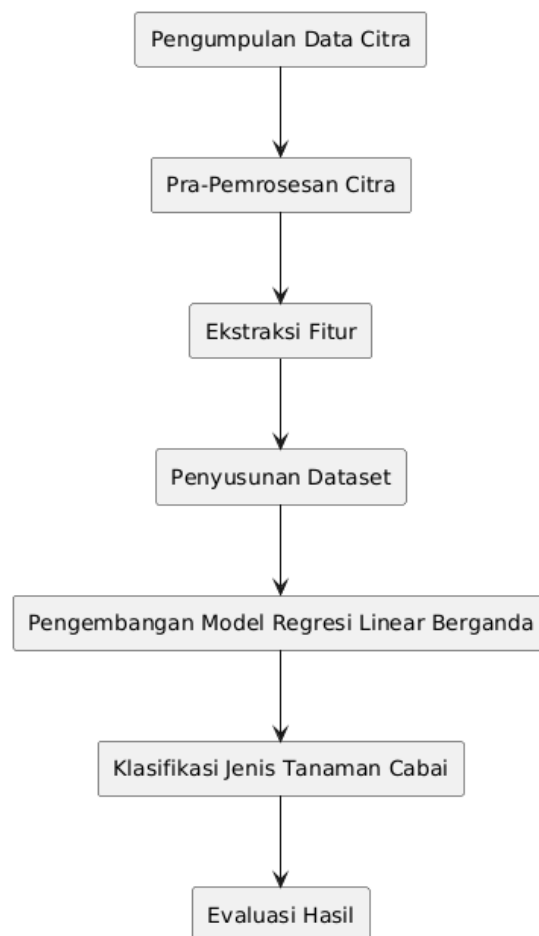


3.	
4.	
...	
97	
98	
99	



### 3.2 Arsitektur Umum

Penelitian ini menggunakan arsitektur umum yang terdiri dari beberapa tahap untuk mengklasifikasikan jenis tanaman cabai menggunakan pengelolaan citra digital dan regresi linear berganda. Berikut adalah tahap-tahap utama dari arsitektur umum yang digunakan:



Gambar 3. 1 Arsitektur Penelitian

Adapun keterangan gambar 3.1 yaitu sebagai berikut :

#### 1. Pengumpulan Data Citra

Data citra tanaman cabai dikumpulkan dari situs [brin.go.id](http://brin.go.id). Citra digital mencakup gambar berbagai jenis tanaman cabai dengan karakteristik fisik yang berbeda seperti bentuk daun, warna buah, dan tekstur permukaan.

#### 2. Pra-Pemrosesan Citra

Pada tahap ini, citra digital yang telah dikumpulkan melalui beberapa langkah pra-pemrosesan untuk meningkatkan kualitas gambar dan mempersiapkan data untuk analisis lebih lanjut. Langkah-langkah ini mencakup pengubahan Ukuran yaitu menyelaraskan ukuran semua citra ke dimensi yang seragam. Pengurangan Noise yaitu menggunakan teknik filtering untuk mengurangi noise dan artefak yang tidak diinginkan dalam gambar. Serta melakukan normalisasi yaitu menormalisasi intensitas piksel untuk memastikan konsistensi dalam analisis.

#### 3. Ekstraksi Fitur

Fitur-fitur penting dari citra digital diekstraksi untuk digunakan dalam model regresi linear berganda. Fitur-fitur ini meliputi bentuk Daun menggunakan teknik segmentasi untuk mengidentifikasi dan mengukur bentuk daun.

#### 4. Penyusunan Dataset

Data fitur yang telah diekstraksi disusun menjadi sebuah dataset yang siap digunakan untuk analisis lebih lanjut. Setiap baris dalam dataset mewakili satu citra tanaman cabai dan mencakup fitur-fitur yang diekstraksi serta label jenis cabai.

#### 5. Pengembangan Model Regresi Linear Berganda

Model regresi linear berganda dikembangkan menggunakan dataset yang telah disusun. Model ini bertujuan untuk mengidentifikasi hubungan antara fitur-fitur citra digital (variabel independen) dan jenis tanaman cabai (variabel dependen). Tahap ini meliputi melatih model regresi menggunakan sebagian dari dataset, memvalidasi model menggunakan data yang tidak digunakan dalam pelatihan untuk memastikan keakuratan dan generalisasi model. Dan model regresi linear berganda yang telah dilatih digunakan untuk mengklasifikasikan jenis tanaman cabai baru berdasarkan fitur-fitur citra yang diekstraksi. Proses ini melibatkan:

#### 6. Prediksi Jenis Tanaman

Menggunakan model untuk memprediksi jenis cabai dari citra baru.

#### 7. Evaluasi Hasil

Mengevaluasi keakuratan prediksi dengan membandingkan hasil klasifikasi dengan label yang diketahui (jika tersedia).

### 3.3 Data Preparation

Untuk mempersiapkan dataset agar dapat dianalisis dengan baik, diperlukan beberapa tahapan yang dikerjakan secara semi-otomatis untuk data tanaman cabai, yaitu:

#### 1. Pengunduhan Data

Data pengelompokan jenis tanaman cabai diunduh dari situs [brin.go.id](http://brin.go.id). Data ini berupa gambar dalam format JPG yang menggambarkan bentuk daun cabai serta karakteristik lainnya

#### 2. Konversi Format

Gambar-gambar yang diunduh dalam format JPG dikonversi ke format tabel CSV untuk menyimpan metadata dan hasil ekstraksi fitur. Langkah ini membuat data untuk lebih mudah diproses dan dianalisis menggunakan alat analisis data seperti Pandas di Python.

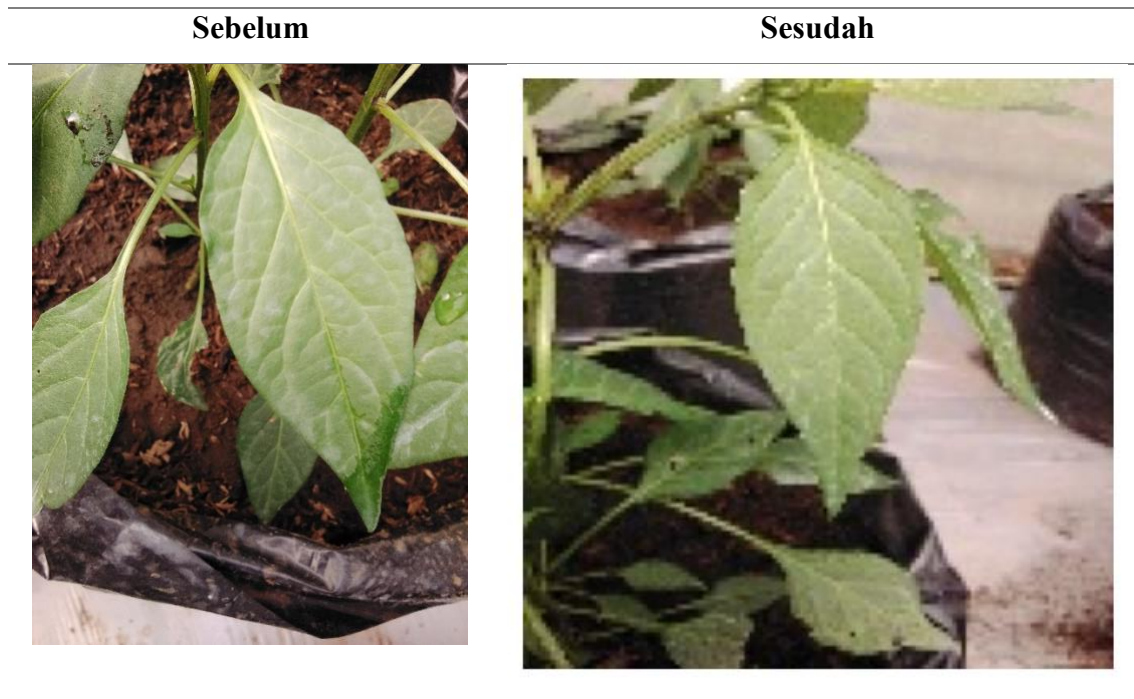
#### 3. Pembersihan Data

Mengidentifikasi dan menghapus gambar yang duplikat untuk memastikan setiap entri uni Mengisi atau menghapus gambar yang hilang atau rusak dalam dataset. Misalnya, gambar yang tidak lengkap atau buram dapat dihapus jika tidak terlalu banyak.

#### 4. Pra-Pemrosesan Gambar

Menyelaraskan ukuran semua gambar ke dimensi yang seragam untuk memastikan konsistensi dalam analisis.

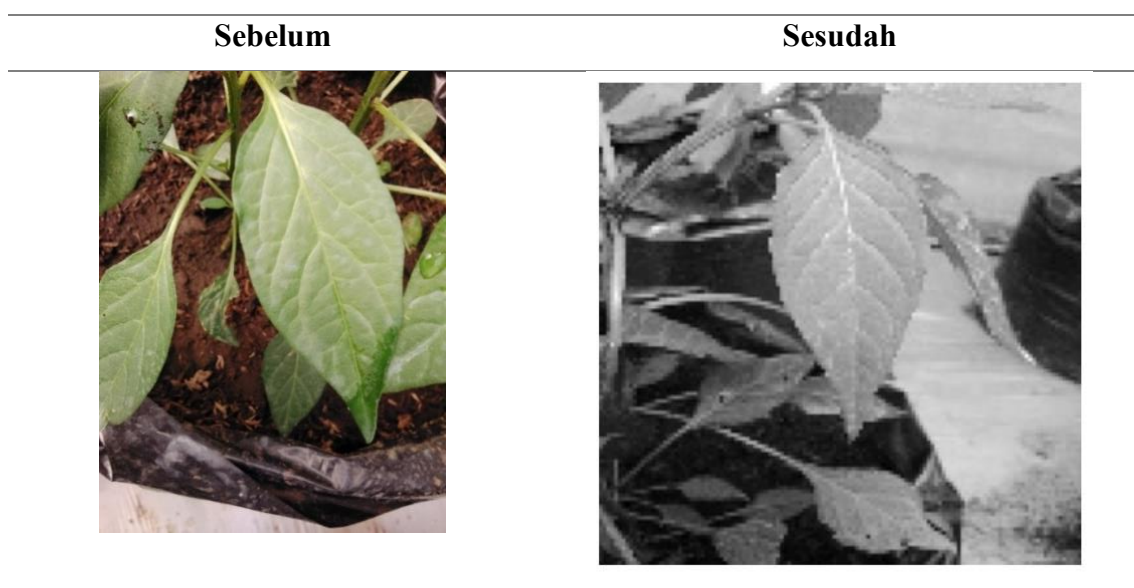
Tabel 3. 2 Pemrosesan Gambar



#### 5. Analisis tekstur

menggunakan metode seperti GLCM (Gray-Level Co-occurrence Matrix) untuk mengidentifikasi pola tekstur pada permukaan daun.

Tabel 3. 3 Gray-Level Co-occurrence Matrix



0	5f3ec52b-bd7a-41b7-aec7-ce388b5cd9d5.jpeg	Image	Contrast	Dissimilarity	\
			2295.04154	31.886687	
		Homogeneity	Energy	Correlation	ASM
0		0.052813	0.009391	0.708124	0.000088

Gambar 3.2 Hasil Ekstraksi

Gambar 4.1 menunjukkan hasil tabel yang dihasilkan dari ekstraksi fitur GLCM (Gray-Level Co-occurrence Matrix) untuk gambar daun cabai. Tabel ini mencantumkan nama gambar dan nilai-nilai fitur GLCM yang diekstraksi dari gambar tersebut

#### 6. Penyusunan Dataset

Data fitur yang telah diekstraksi disusun menjadi sebuah dataset yang siap digunakan untuk analisis lebih lanjut. Setiap baris dalam dataset mewakili satu gambar daun cabai dan mencakup fitur-fitur yang diekstraksi serta label jenis cabai.

## BAB IV

### HASIL DAN PEMBAHASAN

#### 4.1 Pembahasan

Pada penelitian ini, dibuat suatu sistem untuk mengenali jenis tanaman cabai menggunakan metode Regresi Linear Berganda. Sedangkan untuk mendeteksi dan mengolah citra tanaman cabai digunakan metode pengolahan citra digital dengan menggunakan library OpenCV dan teknik-teknik seperti segmentasi gambar dan ekstraksi fitur GLCM (Gray-Level Co-occurrence Matrix). Citra tanaman cabai sebanyak 100 image dari berbagai jenis cabai yang diperoleh dari database BRIN disimpan di dalam database yang dibagi menjadi dua bagian yaitu data latih dan data uji. Masing-masing jenis cabai mempunyai beberapa image yang terdiri dari berbagai variasi bentuk, warna, dan tekstur.



Gambar 4. 1 Dataset Jenis Tanaman Cabai

##### 4.1.1 Deteksi Citra

Pada penelitian ini untuk mendeteksi citra tanaman cabai digunakan metode pengolahan citra digital. Pengolahan citra digital adalah proses pemisahan antara objek utama (foreground) yaitu tanaman cabai dengan latar belakang (background). Metode ini akan memisahkan nilai RGB menjadi nilai-nilai intensitas dan karakteristik warna. Warna asli pada citra tanaman cabai seringkali masih mengandung efek pencahayaan yang dapat mengubah karakteristik visual seperti warna, bentuk, dan tekstur, sehingga perlu dilakukan konversi ke dalam bentuk representasi fitur yang lebih stabil terhadap perubahan pencahayaan. Untuk mengurangi efek pencahayaan tersebut digunakan model seperti GLCM (Gray-Level Co-occurrence Matrix) yang membantu dalam ekstraksi fitur tekstur seperti kontras, homogenitas, dan korelasi, yang lebih robust terhadap perubahan cahaya.





Gambar 4. 2 Komponen Citra YCbCr

Gambar ini menunjukkan hasil dari proses konversi citra tanaman cabai dari ruang warna RGB ke ruang warna YCbCr, serta pemisahan komponen-komponen Y, Cb, dan Cr yang dihasilkan dari proses ini. Citra RGB adalah representasi asli dari daun tanaman cabai dalam format warna Red, Green, Blue, yang merupakan format umum untuk penyimpanan dan tampilan citra digital. Format RGB ini memanfaatkan kombinasi dari tiga warna dasar untuk menampilkan warna penuh pada setiap piksel, yang sangat sesuai dengan cara mata manusia merespons cahaya. Setelah citra RGB dikonversi ke ruang warna YCbCr, citra ini dibagi menjadi tiga komponen: Y (luminance), Cb (chrominance-blue), dan Cr (chrominance-red). Komponen Y (luminance) menyimpan informasi mengenai intensitas cahaya atau kecerahan setiap piksel tanpa memperhatikan warna. Ini ditampilkan sebagai citra dalam skala abu-abu yang hanya mengandung informasi mengenai terang atau gelapnya setiap area dalam gambar. Komponen ini sangat penting karena mata manusia lebih sensitif terhadap perubahan kecerahan daripada perubahan warna.

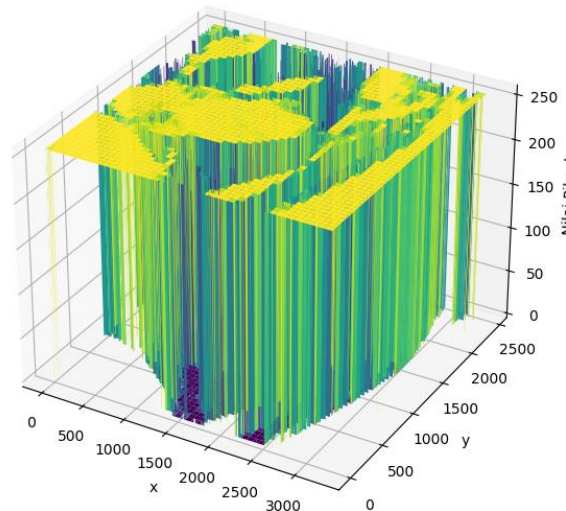
Selanjutnya, komponen Cb (chrominance-blue) dan Cr (chrominance-red) masing-masing merepresentasikan perbedaan antara warna biru dan merah pada citra RGB dengan nilai kecerahan (luminance). Komponen Cb mengindikasikan seberapa "biru" piksel dibandingkan dengan kecerahannya, sedangkan komponen Cr menunjukkan seberapa "merah" piksel tersebut. Informasi ini berguna dalam berbagai aplikasi pengolahan citra, terutama karena informasi warna dapat dikompres lebih banyak dibandingkan informasi luminance, tanpa terlalu mengurangi kualitas visual yang dirasakan.



Gambar 4. 3 Citra Binner

Gambar ini menampilkan hasil dari proses konversi citra tanaman cabai menjadi citra biner. Citra biner ini merupakan representasi dari gambar asli yang telah diubah menjadi bentuk dengan dua nilai intensitas, yaitu hitam dan putih. Proses ini biasanya dimulai dengan mengubah gambar menjadi grayscale, di mana gambar hanya mengandung berbagai tingkatan abu-abu tanpa informasi warna. Selanjutnya, teknik thresholding diterapkan untuk membagi piksel menjadi dua kategori: piksel yang berada di atas nilai ambang tertentu ditampilkan sebagai putih, sementara piksel yang berada di bawah ambang tersebut ditampilkan sebagai hitam. Pada citra biner ini, bagian-bagian tanaman cabai yang berwarna lebih terang, seperti daun, ditampilkan sebagai area putih, sedangkan area yang lebih gelap, seperti latar belakang atau bayangan, ditampilkan sebagai area hitam. Penggunaan citra biner ini memudahkan dalam proses analisis lebih lanjut, misalnya dalam deteksi tepi, segmentasi objek, atau penghitungan area tertentu dalam gambar. Dengan hanya dua tingkat intensitas, citra biner memberikan representasi yang sangat kontras, yang dapat mengurangi kompleksitas pemrosesan data visual dan mempermudah identifikasi fitur-fitur penting dalam gambar, seperti bentuk dan kontur dari daun cabai

Grafik Distribusi Kemungkinan pada Sampel Citra



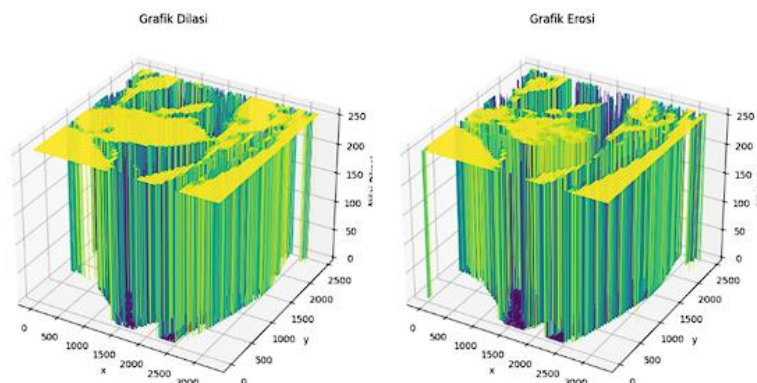
Gambar 4. 4 Grafik Distribusi Kemungkinan Pada Sampel Citra

Gambar ini menampilkan grafik distribusi 3D dari nilai piksel pada sebuah sampel citra. Grafik ini menggambarkan hubungan antara posisi piksel dalam citra (diwakili oleh sumbu x dan y) dengan nilai intensitas piksel (diwakili oleh sumbu z) dalam citra tersebut. Setiap batang dalam grafik ini mewakili nilai intensitas dari piksel pada posisi tertentu dalam citra, di mana warna dan ketinggian batang menunjukkan perbedaan dalam intensitas atau nilai piksel. Secara visual, grafik ini memberikan gambaran yang jelas mengenai distribusi kecerahan atau kegelapan di seluruh citra. Area dengan nilai intensitas tinggi, yang kemungkinan besar merupakan bagian dari objek terang atau bercahaya dalam citra, ditampilkan sebagai batang yang tinggi dan berwarna cerah. Sebaliknya, area yang lebih gelap dalam citra ditampilkan dengan batang yang lebih rendah dan berwarna lebih gelap. Dengan menganalisis grafik ini, kita bisa memahami bagaimana intensitas cahaya tersebar di seluruh citra dan mengidentifikasi pola atau fitur penting seperti tepi, tekstur, atau bentuk yang muncul dalam citra.



Gambar 4. 5 Citra Erosi Dan Citra Dilasi

Gambar ini menunjukkan hasil dari dua operasi morfologi yang umum digunakan dalam pemrosesan citra, yaitu erosi dan dilasi, yang diterapkan pada gambar tanaman cabai. Pada citra pertama, hasil dari proses erosi ditampilkan. Erosi adalah teknik yang menghilangkan piksel-piksel pada tepi objek di dalam gambar, membuat area terang pada gambar menjadi lebih kecil, sehingga detail halus atau objek kecil mungkin hilang. Pada gambar ini, tepi daun dan elemen-elemen kecil lainnya telah dikurangi atau bahkan dihilangkan, membuat objek utama tampak lebih tipis dan lebih terfokus. Sebaliknya, citra kedua menampilkan hasil dari proses dilasi, yang memperluas area terang pada citra. Dalam dilasi, objek dalam citra tampak lebih besar dan menyebar, mengisi celah-celah atau lubang kecil yang mungkin muncul setelah erosi atau karena noise dalam citra asli. Pada gambar ini, area putih yang menunjukkan daun menjadi lebih besar dan lebih penuh, dengan tepi yang lebih tebal dan objek yang lebih menonjol dibandingkan dengan hasil erosi.



Gambar 4. 6 Grafik Dilasi Dan Grafik Erosi

Gambar ini menampilkan dua grafik 3D yang memperlihatkan hasil dari proses morfologi pada citra, yaitu dilasi dan erosi. Grafik di sebelah kiri menunjukkan hasil dari proses dilasi, sementara grafik di sebelah kanan menunjukkan hasil dari proses erosi. Kedua grafik ini menggambarkan distribusi nilai piksel setelah diterapkannya operasi morfologi pada citra biner. Pada grafik dilasi, dapat melihat bagaimana objek dalam citra tampak lebih besar dan menyebar. Ini disebabkan oleh operasi dilasi yang menambahkan piksel ke tepi objek, memperluas area terang dan mengisi celah-celah kecil dalam gambar. Efek dilasi ini membuat struktur utama dalam citra, seperti bentuk daun, menjadi lebih menonjol dengan permukaan yang lebih luas dan nilai piksel yang lebih tinggi di berbagai area. Sebaliknya, grafik erosi menunjukkan hasil yang berbeda. Erosi menghilangkan piksel di tepi objek, yang menyebabkan area terang dalam gambar menjadi lebih kecil. Hal ini terlihat dari grafik erosi, di mana struktur objek tampak lebih kecil dan area dengan nilai piksel tinggi lebih sedikit dibandingkan dengan grafik dilasi. Efek erosi ini menyederhanakan citra dengan menghilangkan detail kecil dan memperkecil objek utama dalam gambar.

#### **4.1.2 Teknik Pre-Processing**

Pre-processing adalah langkah awal yang sangat penting dalam proses pengolahan citra, di mana data mentah diubah menjadi format yang lebih sesuai untuk analisis lebih lanjut. Langkah ini bertujuan untuk meningkatkan kualitas citra dan mengurangi kompleksitas dengan menghilangkan noise, menyelaraskan ukuran, serta menormalisasi intensitas piksel. Dengan melakukan pre-processing, data citra dapat dipersiapkan sedemikian rupa sehingga fitur-fitur penting lebih mudah diekstraksi dan dianalisis pada tahap berikutnya, memastikan hasil yang lebih akurat dan efisien dalam berbagai aplikasi pengolahan citra.

##### **1. Konversi RGB menjadi grayscale**

Konversi RGB menjadi grayscale ini banyak digunakan dalam sistem pengolahan citra digital karena dapat menyederhanakan proses analisis dengan mengurangi informasi warna menjadi satu tingkat kecerahan. Ini sangat penting dalam aplikasi seperti klasifikasi jenis tanaman cabai, di mana informasi bentuk dan tekstur lebih relevan dibandingkan dengan informasi warna. Proses ini juga mengurangi kebutuhan memori

dan mempercepat perhitungan, sehingga mempermudah dan mempercepat langkah-langkah analisis selanjutnya. Gambar 4.6 adalah beberapa contoh hasil dari konversi citra RGB menjadi grayscale dalam konteks pengolahan citra tanaman cabai.



Gambar 4. 7 Hasil Citra Grayscale

Gambar ini menampilkan dua hasil visualisasi dari citra tanaman cabai dalam hal ini citra asli dalam format RGB dan citra yang telah dikonversi menjadi grayscale. Pada citra RGB asli, daun tanaman cabai ditampilkan dengan semua informasi warna yang direkam oleh kamera, yang mencakup spektrum warna merah, hijau, dan biru. Informasi warna ini penting untuk visualisasi secara natural, namun seringkali kurang relevan untuk analisis bentuk dan tekstur dalam pengolahan citra digital. Sebaliknya, citra grayscale yang ditampilkan di sebelah kanan merupakan hasil dari konversi citra RGB. Dalam citra grayscale, setiap piksel hanya memiliki satu nilai intensitas yang menggambarkan kecerahan, dari hitam (nilai terendah) hingga putih (nilai tertinggi). Dengan menghilangkan informasi warna, citra grayscale menyederhanakan data visual, membuatnya lebih mudah untuk menganalisis fitur-fitur penting seperti tekstur, tepi, dan bentuk dari daun cabai. Proses konversi ini sangat berguna dalam aplikasi pengolahan citra, seperti klasifikasi tanaman, di mana warna tidak selalu menjadi faktor utama dan fokus lebih diarahkan pada struktur dan pola yang ada dalam gambar.

	Contrast	Correlation	Energy	Homogeneity
0	0.000285	184.041622	0.971970	3284.990604
1	0.000199	181.724724	0.978169	4161.784851
2	0.000264	122.080939	0.979250	2945.219606
3	0.000248	173.212799	0.977758	3894.892279
4	0.000264	122.080939	0.979250	2945.219606

Gambar 4. 8 Hasil Ekstraksi

Hasil ekstraksi fitur GLCM (Gray-Level Co-occurrence Matrix) dari gambar yang ditampilkan menunjukkan beberapa karakteristik tekstur penting. GLCM digunakan untuk mengukur seberapa sering pasangan piksel dengan nilai intensitas tertentu muncul

dalam hubungan spasial tertentu pada gambar. Dari hasil ekstraksi ini, kita dapat melihat beberapa fitur utama yaitu Contrast, Correlation, Energy, dan Homogeneity. Contrast mengukur intensitas kontras atau variasi antara piksel dan tetangganya pada gambar, dan dalam hasil ini, nilai contrast sangat rendah (sekitar 0.0002), yang menunjukkan bahwa gambar memiliki kontras rendah atau sedikit variasi intensitas antara piksel yang bersebelahan. Correlation, yang mengukur seberapa berkorelasi satu piksel dengan tetangganya, memiliki nilai yang sangat tinggi (di atas 100), menunjukkan adanya korelasi linear yang kuat antara intensitas piksel di seluruh gambar. Energy, yang merupakan ukuran keseragaman dalam gambar, menunjukkan nilai yang mendekati 1 (sekitar 0.97 hingga 0.98), mengindikasikan bahwa gambar sangat homogen dengan distribusi intensitas yang seragam. Homogeneity, yang mengukur kedekatan distribusi elemen dalam GLCM terhadap diagonal, juga memiliki nilai yang sangat tinggi (ribuan), menunjukkan bahwa distribusi intensitas piksel dalam gambar sangat dekat satu sama lain, mencerminkan tingkat keseragaman yang tinggi.

## 2. Labeling Data

Proses labeling data dalam penelitian ini merupakan langkah krusial yang bertujuan untuk mengkategorikan setiap sampel gambar tanaman cabai berdasarkan jenisnya. Dengan menambahkan label pada setiap gambar, kita dapat mengelompokkan data sesuai dengan jenis cabai, seperti cabai merah, cabai hijau, atau cabai rawit. Label ini akan menjadi dasar untuk melatih model regresi linear berganda, yang kemudian digunakan untuk mengklasifikasikan jenis tanaman cabai berdasarkan fitur tekstur yang telah diekstraksi dari citra digital. Proses labeling ini memastikan bahwa model memiliki data yang terstruktur dengan baik, memungkinkan prediksi yang lebih akurat dalam mengidentifikasi jenis cabai dari gambar.



```

Dataset awal:
  Contrast  Correlation  Energy  Homogeneity  Label
0  0.000285  184.041622  0.971970  3284.990604  0
1  0.000199  181.724724  0.978169  4161.784851  0
2  0.000264  122.080939  0.979250  2945.219606  0
3  0.000248  173.212799  0.977758  3894.892279  0
4  0.000264  122.080939  0.979250  2945.219606  0

```

Jumlah baris dalam dataset: 108

```

Dataset setelah penambahan label:
  Contrast  Correlation  Energy  Homogeneity  Label
0  0.000285  184.041622  0.971970  3284.990604  0
1  0.000199  181.724724  0.978169  4161.784851  0
2  0.000264  122.080939  0.979250  2945.219606  0
3  0.000248  173.212799  0.977758  3894.892279  0
4  0.000264  122.080939  0.979250  2945.219606  0

```

Gambar 4. 9 Proses Labeling

Awalnya, dataset memiliki empat kolom utama: Contrast, Correlation, Energy, dan Homogeneity, yang mewakili fitur tekstur yang telah diekstraksi dari gambar. Dataset ini terdiri dari 108 baris, yang masing-masing merepresentasikan satu sampel gambar. Sebelum penambahan label, dataset hanya berisi nilai-nilai dari fitur-fitur tersebut tanpa informasi tentang kategori atau jenis dari gambar yang dianalisis. Setelah menambahkan kolom Label, setiap baris dalam dataset sekarang juga memiliki informasi tentang kategori dari gambar yang diwakilinya. Dalam contoh ini, semua label pada baris yang ditampilkan adalah 0, yang mungkin mengindikasikan bahwa baris-baris ini berasal dari satu kategori yang sama, misalnya "Cabai Merah". Label ini penting karena akan digunakan dalam model klasifikasi untuk memprediksi jenis cabai berdasarkan fitur-fitur yang diekstraksi. Penambahan label memungkinkan dataset ini untuk digunakan dalam algoritma machine learning, di mana model akan dilatih untuk mengenali pola dalam fitur-fitur tekstur yang berhubungan dengan jenis cabai tertentu. Proses pelabelan ini merupakan langkah penting dalam membangun model klasifikasi, karena label inilah yang nantinya menjadi target yang diprediksi oleh model. Jika ada lebih dari satu jenis cabai dalam dataset, kolom Label akan menunjukkan variasi yang sesuai, dengan angka yang berbeda untuk setiap jenis cabai. Proses ini memastikan bahwa model memiliki data yang terstruktur dengan baik untuk pelatihan dan pengujian, sehingga dapat melakukan klasifikasi jenis cabai secara efektif berdasarkan citra digitalnya.

### 3. Pembagian Dataset

Dalam penelitian ini, langkah berikutnya adalah membagi dataset menjadi dua bagian utama: training set dan test set. Pembagian ini dilakukan untuk memastikan bahwa model yang akan dilatih dapat diuji secara akurat pada data yang tidak digunakan selama pelatihan. Training set digunakan untuk melatih model regresi linear berganda, sementara test set digunakan untuk menguji kinerja model tersebut. Pembagian dataset biasanya dilakukan dengan rasio 80:20 di mana sebagian besar data dialokasikan untuk pelatihan model, dan sisanya digunakan untuk pengujian. Hal ini penting untuk memastikan bahwa model tidak hanya bekerja dengan baik pada data pelatihan tetapi juga mampu menggeneralisasi dengan baik pada data baru yang belum pernah dilihat sebelumnya. Untuk melakukan pembagian dataset ini secara acak dan memastikan distribusi data yang serupa antara training set dan test set, menggunakan fungsi `train_test_split` dari library `scikit-learn`. Pendekatan ini membantu dalam meminimalkan bias dan memberikan evaluasi yang lebih akurat terhadap performa model dalam mengklasifikasikan jenis tanaman cabai berdasarkan citra digital.

#### 4.1.3 Penerapan Metode Regresi Linier Berganda

Penerapan metode Regresi Linier Berganda dalam penelitian ini bertujuan untuk memprediksi jenis tanaman cabai berdasarkan fitur-fitur tekstur yang diekstraksi dari citra digital. Model ini dilatih menggunakan dataset yang telah dilabeli, dengan harapan dapat mengklasifikasikan jenis cabai secara akurat berdasarkan hubungan linear antara fitur-fitur yang ada.

Mean Squared Error (MSE): 0.6583044725730125

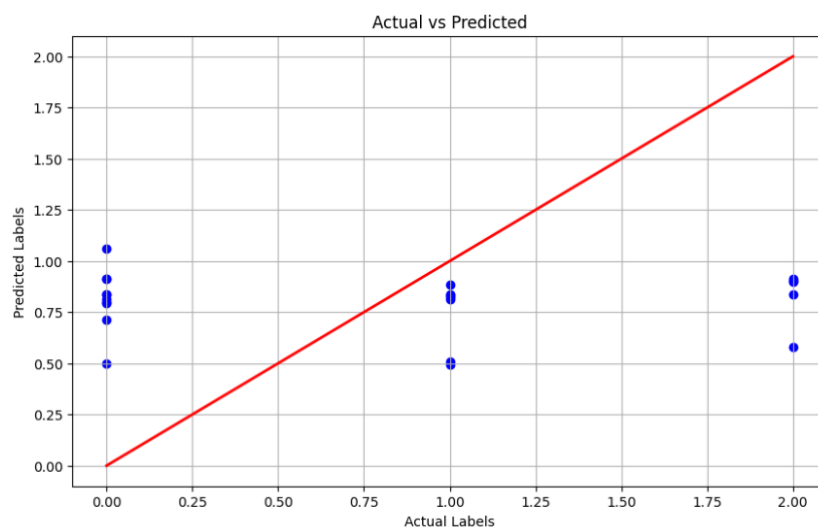
R-squared ( $R^2$ ): -0.04465365483717365

Perbandingan Nilai Asli dan Prediksi:

	Actual	Predicted
0	1	0.813997
1	0	0.796617
2	0	0.914845
3	2	0.898951
4	1	0.492673

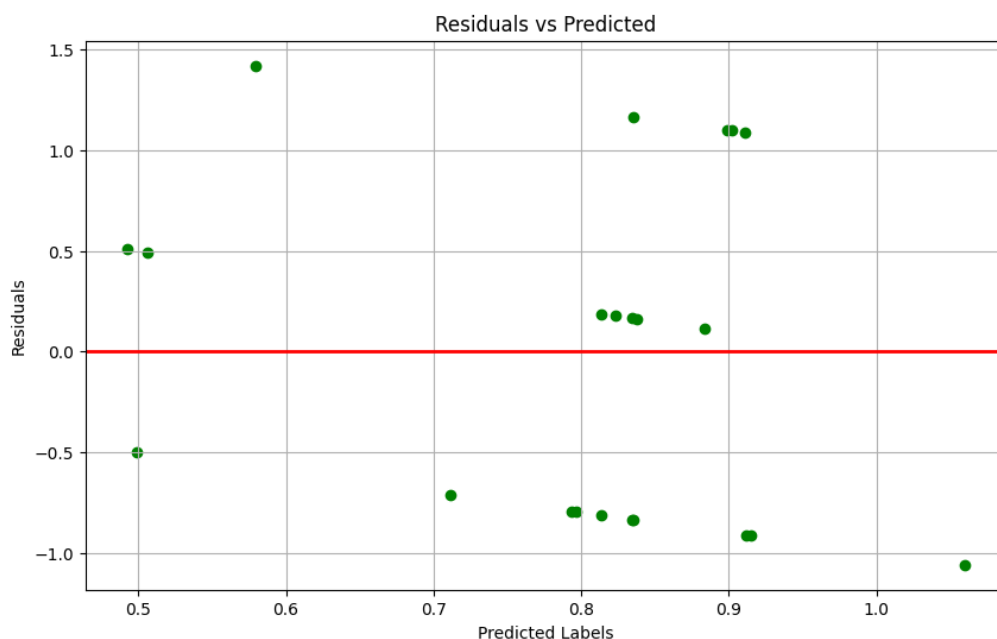
Gambar 4. 10 Hasil Penerapan Metode

Dari hasil yang didapat, penulis dapat melihat beberapa indikator kinerja dari model regresi linear berganda yang diterapkan pada dataset. Mean Squared Error (MSE) adalah metrik yang digunakan untuk mengukur rata-rata dari kuadrat kesalahan antara nilai aktual dan prediksi. Nilai MSE yang lebih kecil menunjukkan bahwa model memiliki kesalahan yang lebih sedikit dalam prediksi. Dalam hal ini, MSE yang didapat adalah sekitar 0.6583, yang menunjukkan bahwa ada kesalahan yang cukup signifikan antara prediksi model dan nilai aktual. Selain itu, R-squared ( $R^2$ ) adalah metrik yang menunjukkan seberapa baik model menjelaskan variabilitas data. Nilai  $R^2$  berkisar antara 0 hingga 1, di mana nilai 1 menunjukkan model yang sangat baik dalam memprediksi data. Namun, dalam kasus ini, nilai  $R^2$  yang diperoleh adalah negatif (-0.0447), yang mengindikasikan bahwa model tidak dapat menjelaskan variabilitas data dengan baik. Nilai negatif ini menunjukkan bahwa model performanya lebih buruk daripada model baseline yang hanya memprediksi rata-rata dari nilai-nilai target. Perbandingan antara nilai aktual dan prediksi juga menunjukkan bahwa model sering kali salah dalam memprediksi kategori. Sebagai contoh, data pertama dan keempat memiliki nilai aktual 0 dan 1, namun model memprediksikan nilai yang cukup jauh dari target tersebut. Ini lebih lanjut mengindikasikan bahwa model regresi linear berganda mungkin tidak cocok untuk tugas klasifikasi ini atau ada masalah pada data atau fitur yang digunakan.



Gambar 4. 11 Grafik Scatter Plot

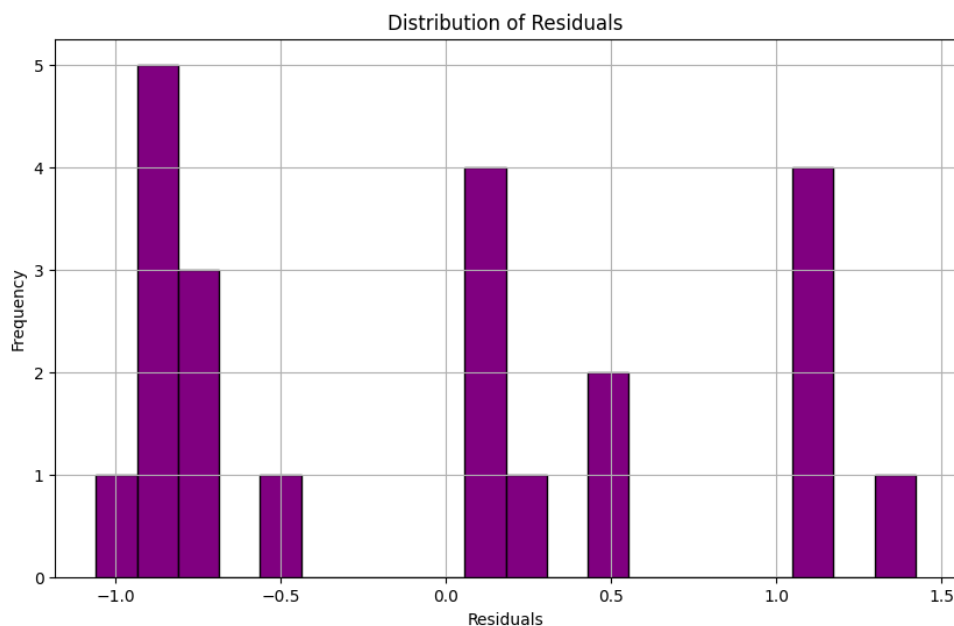
Grafik scatter plot yang ditampilkan memperlihatkan hubungan antara nilai aktual dan prediksi dari model regresi linear berganda yang diterapkan pada dataset tanaman cabai. Garis merah diagonal dalam grafik ini mewakili garis identitas, di mana nilai prediksi seharusnya berada jika model melakukan prediksi yang sempurna. Namun, hasilnya menunjukkan bahwa sebagian besar titik data tidak berada di garis identitas ini, yang mengindikasikan adanya perbedaan signifikan antara nilai prediksi dan nilai aktual. Banyak titik data yang menyimpang jauh dari garis merah, menandakan kesalahan prediksi yang cukup besar oleh model. Selain itu, cluster nilai yang muncul pada sumbu x, yang mewakili label aktual 0, 1, dan 2, tidak berdekatan dengan nilai prediksi yang diharapkan. Misalnya, banyak prediksi untuk label 0 berada di kisaran yang lebih tinggi daripada yang seharusnya, dan prediksi untuk label lain juga tidak konsisten dengan nilai yang diharapkan. Kesalahan prediksi ini menunjukkan bahwa model regresi linear berganda tidak mampu menangkap pola yang ada dalam data dengan baik, sehingga tidak berhasil dalam mengklasifikasikan jenis tanaman cabai secara akurat.



Gambar 4. 12 Residuals

Grafik residuals vs predicted yang ditampilkan menggambarkan hubungan antara nilai prediksi dari model regresi linear berganda dengan residuals, yaitu selisih antara nilai aktual dan nilai prediksi. Garis merah horizontal pada grafik ini menunjukkan garis nol, yang merupakan titik di mana prediksi model tepat sama dengan nilai aktual, sehingga

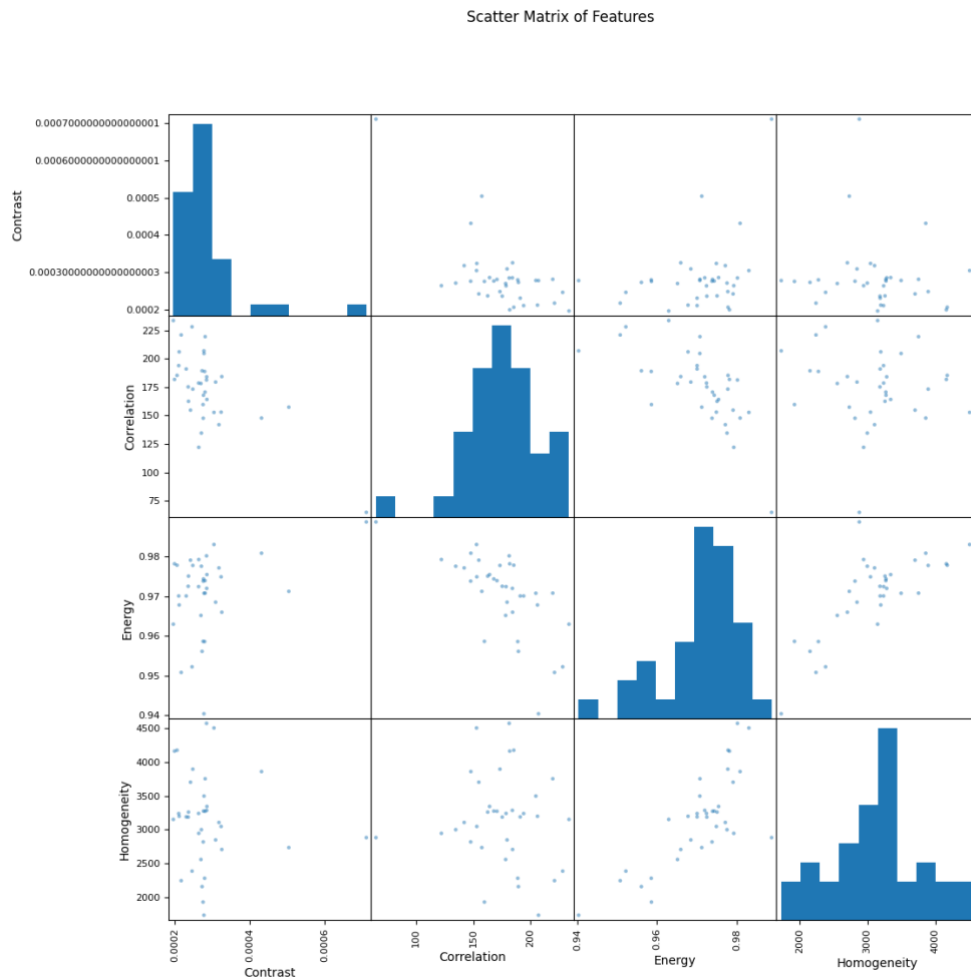
residualnya adalah nol. Dalam grafik ini, dapat dilihat bahwa sebagian besar titik data tidak berada di garis nol, yang menunjukkan bahwa model menghasilkan residuals atau kesalahan prediksi yang signifikan. Beberapa titik bahkan menunjukkan residuals positif yang tinggi, yang berarti model memprediksi nilai yang jauh lebih rendah daripada nilai aktual. Sebaliknya, terdapat juga beberapa titik dengan residuals negatif yang besar, menunjukkan bahwa model memprediksi nilai yang lebih tinggi daripada nilai aktual. Sebaran residuals yang tidak terdistribusi secara acak di sekitar garis nol menunjukkan bahwa model mungkin tidak menangkap pola dalam data dengan baik. Idealnya, jika model bekerja dengan baik, residuals akan tersebar secara acak di sekitar garis nol tanpa pola yang jelas, menandakan bahwa kesalahan prediksi terjadi secara konsisten di seluruh rentang prediksi. Namun, dalam kasus ini, residuals cenderung mengelompok pada nilai-nilai tertentu, yang menunjukkan adanya pola kesalahan yang bisa disebabkan oleh ketidakcocokan model regresi linear berganda untuk data ini.



Gambar 4. 13 Grafik Histogram

Grafik histogram yang ditampilkan menggambarkan distribusi residuals, yaitu perbedaan antara nilai aktual dan nilai prediksi dari model regresi linear berganda yang digunakan dalam penelitian ini. Residuals ini mencerminkan kesalahan prediksi yang dibuat oleh model, dan distribusi dari residuals dapat memberikan wawasan tentang kinerja model tersebut. Dalam grafik ini, residuals tersebar di beberapa interval, dengan

beberapa kelompok residuals yang terlihat jelas. Kita dapat melihat bahwa residuals tersebar dari sekitar -1,0 hingga 1,5. Idealnya, jika model bekerja dengan baik, kita mengharapkan distribusi residuals berbentuk lonceng simetris (distribusi normal) yang berpusat di sekitar nol, yang menandakan bahwa kesalahan prediksi tersebar secara merata tanpa bias. Namun, pada grafik ini terlihat bahwa ada tiga puncak yang jelas dalam distribusi residuals yaitu satu kelompok di sekitar -1, satu lagi di sekitar 0,5, dan yang lainnya di sekitar 1,0. Distribusi ini menunjukkan bahwa model cenderung membuat kesalahan dalam beberapa kelompok yang berbeda, bukan secara acak. Residuals yang besar dan tersebar ini menunjukkan bahwa model tidak dapat menangkap hubungan yang baik antara fitur dan label, menyebabkan prediksi yang kurang akurat.



Gambar 4. 14 Grafik Scatter Matrix

Grafik scatter matrix yang ditampilkan memberikan visualisasi hubungan antara empat fitur utama yang diekstraksi dari gambar dalam penelitian ini, yaitu Contrast, Correlation, Energy, dan Homogeneity. Setiap sel dalam matriks ini menunjukkan scatter plot dari pasangan fitur yang berbeda, sementara histogram pada diagonal menunjukkan distribusi dari masing-masing fitur secara individu. Histogram untuk Contrast menunjukkan bahwa sebagian besar nilainya berkonsentrasi pada kisaran yang sangat rendah, sementara histogram untuk Correlation memperlihatkan distribusi yang lebih merata namun tetap memiliki puncak di sekitar 150 hingga 200. Energy dan Homogeneity juga menunjukkan distribusi yang cenderung terfokus pada rentang nilai tertentu, yang menunjukkan bahwa sebagian besar sampel memiliki karakteristik tekstur yang seragam dalam hal energy dan homogeneity. Scatter plot antar fitur dalam matriks ini tidak menunjukkan adanya korelasi linear yang kuat. Misalnya, hubungan antara Contrast dan Correlation tampak tersebar secara acak tanpa pola linear yang jelas, dan hal serupa terlihat dalam hubungan antara Energy dan fitur-fitur lainnya. Ketiadaan korelasi yang kuat ini mengindikasikan bahwa fitur-fitur yang digunakan mungkin memiliki hubungan yang kompleks atau tidak linear, yang bisa menjadi tantangan bagi model regresi linear berganda dalam menangkap pola-pola yang ada.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berikut merupakan kesimpulan dari penelitian ini yaitu sebagai berikut :

1. Hasil dari model regresi linear berganda menunjukkan bahwa model ini tidak memberikan performa prediksi yang optimal dalam mengklasifikasikan jenis tanaman cabai berdasarkan fitur-fitur tekstur yang diekstraksi dari citra digital. Hal ini terlihat dari nilai R-squared ( $R^2$ ) yang negatif dan distribusi residuals yang tidak simetris serta menyebar jauh dari nol.
2. Analisis scatter matrix menunjukkan bahwa tidak ada korelasi linear yang signifikan antara fitur-fitur yang digunakan (Contrast, Correlation, Energy, Homogeneity). Ini menjadi salah satu alasan mengapa model regresi linear berganda tidak mampu melakukan prediksi dengan baik, karena model ini mengandalkan adanya hubungan linear antara fitur-fitur tersebut.
3. Distribusi residuals yang terlihat dari histogram menunjukkan adanya beberapa kelompok kesalahan yang signifikan, baik positif maupun negatif. Distribusi ini menunjukkan bahwa model sering kali membuat kesalahan prediksi yang besar, mengindikasikan ketidakmampuan model untuk menangkap kompleksitas dalam data

#### **5.2 Saran**

Berikut merupakan saran dari penelitian ini yaitu sebagai berikut :

1. Menerapkan teknik reduksi dimensi seperti Principal Component Analysis (PCA) dapat membantu mengurangi kompleksitas data dengan menjaga fitur-fitur yang paling signifikan. Hal ini bisa membantu model dalam fokus pada informasi yang paling relevan, yang berpotensi meningkatkan akurasi prediksi.
2. Disarankan untuk menggunakan teknik cross-validation untuk mengevaluasi kinerja model secara lebih mendalam. Dengan membagi data ke dalam beberapa subset dan melakukan pelatihan serta pengujian berulang kali.



## DAFTAR PUSTAKA

- Andani, Ir, M T Dr Eng Muhammad Niswar, Andini Dani Achmad, and Eng Ir Wardi. n.d. "APLIKASI IOT DAN JARINGAN SENSOR NIRKABEL UNTUK MENURUNKAN RISIKO PENULARAN COVID-19."
- Batubara, Nur Arkhamia, and Rolly Maulana Awangga. 2020. *Tutorial Object Detection Plate Number With Convolution Neural Network (CNN)*. Vol. 1. Kreatif.
- Hartatik, Hartatik, Ghefra Rizkan Gaffara, Habibi Azka Nasution, Ardiansyah Ardiansyah, I Nyoman Alit Arsana, Urnika Mudhifatul Jannah, and S T Iwan Adhicandra. 2023. *PENGENALAN PEMROGRAMAN DASAR DUNIA KODING*. PT. Sonpedia Publishing Indonesia.
- Idris, Mohamad, Riza Ibnu Adam, Yulrio Brianorman, Rinaldi Munir, and Dimitri Mahayana. 2022. "Kebenaran Dalam Perspektif Filsafat Ilmu Pengetahuan Dan Implementasi Dalam Data Science Dan Machine Learning." *Jurnal Filsafat Indonesia* 5 (2): 173–81.
- Kamal, Silmi Afifah. n.d. "Implementasi Hidden Markov Model Dengan Algoritma Forward-Backward Untuk Prediksi Buku Peminjamandi Perpustakaan Studi Kasus: Perpustakaan SPS UIN Jakarta." Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta.
- Mutaqin, Ghani, and S Kom. 2023. *Teknik Penghapusan Kabut Pada Citra Digital*. Nas Media Pustaka.
- Mutmainna, Nurilmi Dwi, Mahmud Achmad, and Suhardi Suhardi. 2017. "Pendugaan Lengan Tanah Inceptisol Pada Tanaman Hortikultura Menggunakan Citra Landsat 8." *Jurnal Agritechno*, 135–51.
- Nugraha, Fikri Aldi, Nisa Hanum Harani, and Roni Habibi. 2020. *Analisis Sentimen Terhadap Pembatasan Sosial Menggunakan Deep Learning*. Kreatif.
- Pramanda, Heru, Wahyu Diara, and David Sarana. 2024. "Analisis Pengaruh Pengalaman Dan Karakter Sumber Daya Manusia Terhadap Kualitaskerjaan Proyek Konstruksi: Studi Kasus: Kecamatan Kuta Alam Dan Syiah Kuala." *Tameh* 13 (1): 55–66.
- PUTRA, I KOMANG WIDARMA. 2022. "UJI POTENSI ATRAKTAN DAUN CENGKEH DAN DAUN KEMANGI TERHADAP HAMA LALAT BUAH (*Bactrocera* Spp) PADA CABAI (*Capsicum Annuum* L.)." Universitas

Maharaswati Denpasar.

- Putra, I Nyoman Tri Anindia, Ketut Sepdyana Kartini, Yoga Kristian Suyitno, I Made Sugiarta, and Ni Kadek Era Puspita. 2023. "Penerapan Library Tensorflow, Cvzone, Dan Numpy Pada Sistem Deteksi Bahasa Isyarat Secara Real Time." *Jurnal Krisnadana* 2 (3): 412–23.
- Rifky, Sehan, Lalu Puji Indra Kharisma, H Achmad Ruslan Afendi, Segar Napitupulu, Mustika Ulina, Wulan Sri Lestari, I Made Dendi Maysanjaya, Kelvin Kelvin, Frans Mikael Sinaga, and Mutmainnah Muchtar. 2024. *Artificial Intelligence: Teori Dan Penerapan AI Di Berbagai Bidang*. PT. Sonpedia Publishing Indonesia.
- Saputra, Gede Wisnu, and I Gusti Agung Ketut Sri Ardani. 2020. "Pengaruh Digital Marketing, Word of Mouth, Dan Kualitas Pelayanan Terhadap Keputusan Pembelian." Udayana University.
- Saputra, Yogi Aditya, Syafrial Fachri Pane, and Rolly Maulana Awangga. 2020. *Big Data: Implementasi Hadoop Mapreduce Pada Pemetaan Sekolah Menggunakan Python*. Kreatif.
- Setiawan, Zunan, Muhammad Fajar, Arif Mudi Priyatno, Anggi Yhurinda Perdana Putri, Mediana Aryuni, Siti Yuliyanti, Harya Widiputra, Budanis Dwi Meilani, Rohmat Nur Ibrahim, and Rezania Agramanisti Azdy. 2023. *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia.

# LAMPIRAN

## Lampiran 1 Persetujuan judul Penelitian



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**


UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019  
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003  
<https://fikti.umsu.ac.id> [fikti@umsu.ac.id](mailto:fikti@umsu.ac.id) [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.linkedin.com/company/umsumedan) [umsumedan](https://www.youtube.com/channel/UC...)

### PERSETUJUAN TOPIK/JUDUL PENELITIAN


Nomor Agenda :  
Nama : Esa Mahendra  
NPM : 2009010013  
Tanggal Persetujuan : 22 Juli 2024  
Topik Yang Disetujui Program Studi : Segmentasi Citra digital Pada Tanaman Cabai dengan metode Yolo V7 dalam mengenali jenis Cabai  
Nama Dosen Pembimbing : Dr. Al Khowarizmi, S.Kom., M.Kom  
Judul Yang Disetujui Dosen Pembimbing : klasifikasi jenis tanaman Cabai menggunakan citra digital dan regresi linear berganda

Medan, 22 Juli 2024

Disahkan oleh  
Ketua Program Studi  
Sistem Informasi



Persetujuan  
Dosen Pembimbing



(Dr. Al Khowarizmi, S.Kom., M.Kom)



## Lampiran 2 Berita Acara Bimbingan Proposal



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003  
<https://fiki.umsu.ac.id> [fiki@umsu.ac.id](mailto:fiki@umsu.ac.id) [umsumedan](https://www.facebook.com/umsumedan) [umsumedan](https://www.instagram.com/umsumedan) [umsumedan](https://www.linkedin.com/company/umsumedan) [umsumedan](https://www.youtube.com/channel/UC...)

### Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Esa Mahendra Program Studi : Sistem Informasi  
 NPM : 2009010013 Konsentrasi :  
 Nama Dosen Pembimbing : Dr. Al-Khowarizmi, S.Kom., M.Kom. Judul Penelitian : Klasifikasi jenis tanaman cabai menggunakan Citra digital dan Regresi linear berganda

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
8/7 2024	Pergantian Judul topik Penelitian	al
17/7 2024	Penambahan Penelitian terdahulu	al
20/7 2024	Data Penelitian yang ditambah menjadi 100 Data	al
24/7 2024	Pembuatan dan Menampilkan hasil dari Ekstraksi Fitur C-LCM.	al
29/7	- ACC Semu -	al

Medan, Senin, 29 Juli 2024

Diketahui oleh :

Ketua Program Studi  
 Sistem Informasi

Martiano, S.Pd., S.Kom., M.Kom.

Disetujui oleh :

Dosen Pembimbing

Dr. Al-Khowarizmi, S.Kom., M.Kom.



## Lampiran 2 Berita Acara Bimbingan Skripsi



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH  
**UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**  
**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI**

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 83/SK/BAN-PT/Akred/PT/III/2019  
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003  
<https://fiki.umsu.ac.id>    [fiki@umsu.ac.id](mailto:fiki@umsu.ac.id)    [fumsumedan](#)    [umsumedan](#)    [umsumedan](#)    [umsumedan](#)

### Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Esa Mahendra                      Program Studi : Sistem Informasi  
 NPM : 2009010013    Konsentrasi :  
 Nama Dosen Pembimbing : Dr. Al-Khowarizmi, S.Kom., M.Kom.                      Judul Penelitian : Klasifikasi jenis tanaman cabai menggunakan Citra digital dan Regresi linear berganda

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
8/7 2024	Pergantian Judul topik Penelitian	al
17/7 2024	Penambahan Penelitian terdahulu	al
20/7 2024	Data Penelitian yang ditambah menjadi 100 Data	al
24/7 2024	Pembuatan dan Menampilkan hasil dari Ekstraksi Fitur CLCM.	al
29/7 2024	- ACC Semu -	al

Medan, 29 Juli 2024

Diketahui oleh :  
 Ketua Program Studi  
 Sistem Informasi  
  
 Martiano, S.Pd., S.Kom., M.Kom.

Disetujui oleh :  
 Dosen Pembimbing  
  
 Dr. Al-Khowarizmi, S.Kom., M.Kom.

