

Deteksi dan Identifikasi Jenis Bunga

disusun oleh:

- | | | | |
|-----------------------------|------------------|----------------|---------------|
| 1. M. Zidan Husaini | 211351092 | Ketua | Pagi B |
| 2. Refadli Dwi Ilham | 211351121 | Anggota | Pagi B |
| 3. Ikhsan Al Fajar | 211351062 | Anggota | Pagi B |
| 4. Desby Permata S. | 211351042 | Anggota | Pagi B |



**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI TEKNOLOGI WASTUKANCANA
PURWAKARTA
2023**

DAFTAR ISI

DAFTAR ISI	i
DAFTAR TABEL	ii
DAFTAR GAMBAR	iii
I. Business Undestanding	1
A. Latar Belakang Masalah	1
B. Tujuan	1
C. Solusi	2
D. Pengguna	3
II. Data Understanding	3
A. Sumber Data	3
B. Deskripsi Data	4
C. Explore Data	4
D. Kualitas Data	5
III. Data Preparation	7
A. Pemilihan Data	7
B. Data Preprocessing	8
C. Data Issue	8

DAFTAR TABEL

Tabel 1. 1 Pihak Pengguna dan Manfaat Proyel 3

Tabel 2. 1 Jumlah Visualisasi Data 4

DAFTAR GAMBAR

Gambar 2. 1 Sumber Dataset	4
Gambar 2. 2 Diagram Jumlah Visualisasi Data.....	5
Gambar 2. 3 Daisy.....	6
Gambar 2. 4 Dandelion	6
Gambar 2. 5 Rose	6
Gambar 2. 6 Sunflowers	7
Gambar 2. 7 Tulip.....	7
Gambar 4. 1 Pelatihan Model	10
Gambar 4. 2 Validasi Setelah Pelatihan	11
Gambar 4. 3.....	12
Gambar 4. 4.....	12
Gambar 4. 5 Confussion matrix.....	13
Gambar 4. 6 Tes model dan Save model.....	14
Gambar 4. 7 Streamlit Deploy.....	14

I. Business Understanding

A. Latar Belakang Masalah

Kemajuan dalam teknologi pengolahan citra digital telah membuka berbagai peluang baru di bidang pengenalan objek di alam dan klasifikasi tumbuhan. Salah satu penerapan teknologi ini adalah dalam proses deteksi dan identifikasi jenis bunga. Identifikasi bunga secara manual membutuhkan waktu, tenaga, dan keahlian khusus yang tidak dimiliki oleh banyak orang. Selain itu, metode identifikasi manual seringkali rentan terhadap kesalahan karena adanya variasi warna, bentuk, dan tekstur bunga yang beragam.

Dalam konteks pengolahan citra digital, deteksi dan identifikasi objek merupakan salah satu topik yang perlu dipelajari karena memiliki banyak aplikasi di berbagai teknologi. Untuk memperdalam pemahaman dalam topik ini, proyek deteksi dan identifikasi jenis bunga dipilih sebagai objek penelitian. Bunga dipilih karena kompleksitas variasi visualnya, yang menambah tantangan dalam proses pengolahan citra.

Dengan menggunakan teknik pengolahan citra digital, identifikasi bunga dapat dilakukan secara otomatis, cepat, dan lebih akurat. Teknologi ini memungkinkan pengembangan sistem yang dapat mendeteksi dan mengklasifikasikan jenis bunga berdasarkan citra digital. Proses ini mencakup metode pengolahan citra seperti pra-pengolahan gambar, ekstraksi fitur, dan penerapan algoritma klasifikasi sederhana untuk mengenali pola visual yang membedakan setiap jenis bunga. Proyek ini juga berfungsi sebagai media pembelajaran bagi mahasiswa untuk memahami teknik dasar pengolahan citra, termasuk segmentasi, pengenalan pola, dan klasifikasi.

Diharapkan bahwa hasil dari proyek ini dapat membantu mahasiswa dalam memahami konsep-konsep pengolahan citra digital serta aplikasinya dalam situasi nyata.

B. Tujuan

Tujuan dari proyek *Deteksi dan Identifikasi Jenis Bunga* ini adalah sebagai berikut:

1. Memfasilitasi Pembelajaran Teknik Pengolahan Citra

Proyek ini bertujuan untuk membantu mahasiswa memahami dan menerapkan teknik dasar dalam pengolahan citra digital, termasuk pra-pengolahan gambar, ekstraksi fitur, dan segmentasi. Dengan menjadikan bunga sebagai objek studi, mahasiswa dapat belajar secara praktis mengenai proses deteksi dan klasifikasi objek.

2. Menerapkan Algoritma Klasifikasi dalam Identifikasi Bunga

Proyek ini dirancang untuk memperkenalkan penggunaan algoritma sederhana dalam pengelompokan objek. Mahasiswa akan diajarkan cara mengkategorikan jenis bunga berdasarkan fitur visualnya, seperti warna, bentuk, dan tekstur. Tujuannya adalah untuk meningkatkan pemahaman mereka tentang cara kerja algoritma dalam mendeteksi dan mengenali pola.

3. Mengembangkan Kemampuan Analisis Visual

Dengan proyek ini, diharapkan mahasiswa dapat meningkatkan kemampuan analisis visual dan memahami tantangan yang timbul dalam pengolahan citra, terutama pada objek dengan variasi tinggi, seperti bunga. Proyek ini juga berfungsi sebagai media untuk mengeksplorasi potensi aplikasi pengolahan citra digital di berbagai bidang, khususnya dalam deteksi dan identifikasi objek.

4. Memberikan Pemahaman Konsep Pengolahan Citra Digital dalam Konteks Nyata

Hasil dari proyek ini diharapkan dapat memberikan mahasiswa pemahaman yang lebih mendalam tentang konsep-konsep pengolahan citra digital serta penerapannya dalam konteks nyata. Dengan demikian, mereka dapat memanfaatkan pengetahuan yang diperoleh dalam studi dan karir di masa depan.

C. Solusi

Untuk mengatasi kesulitan dalam identifikasi bunga secara manual yang memerlukan waktu, tenaga, dan keterampilan khusus, proyek ini akan mengembangkan sistem berbasis pengolahan citra digital. Sistem ini akan dirancang untuk melakukan deteksi dan klasifikasi jenis bunga secara otomatis. Dengan demikian, pengguna yang tidak memiliki keahlian khusus dapat dengan mudah mengenali berbagai jenis bunga.

Dengan menerapkan solusi ini, proyek diharapkan dapat mencapai beberapa tujuan, antara lain:

1. Meningkatkan Pemahaman Mahasiswa tentang teknik dasar pengolahan citra dan penerapannya, terutama dalam identifikasi bunga.
2. Memberikan Pengalaman Praktis dalam penggunaan algoritma klasifikasi dan teknik pengolahan citra, sehingga mahasiswa dapat lebih siap menghadapi tantangan di dunia nyata.
3. Membangun Sistem yang Efisien untuk deteksi dan identifikasi jenis bunga yang dapat digunakan di berbagai bidang, termasuk pendidikan, botani, dan penelitian.

D. Pengguna

Dibawah ini terdapat tabel pihak pengguna dan manfaat proyek.

Tabel 1. 1 Pihak Pengguna dan Manfaat Proyek

No	Pihak Pengguna	Manfaat yang Diperoleh
1	Mahasiswa	Memahami teknik pengolahan citra dan aplikasinya dalam studi bunga.
2	Peneliti dan Praktisi	Alat efisien untuk identifikasi bunga dalam penelitian.
3	Pecinta dan Hobi Tumbuhan	Memudahkan mengenali berbagai jenis bunga.
4	Pengembang Teknologi	Referensi untuk pengembangan aplikasi pengolahan citra.
5	Komunitas Lingkungan	Meningkatkan kesadaran tentang keanekaragaman hayati.

II. Data Understanding

Pada tahap ini adalah proses analisis data, di mana penulis fokus memahami karakteristik, struktur, dan kualitas dataset yang akan digunakan. Tahap ini, penulis mempelajari data secara mendalam untuk memastikan bahwa data tersebut relevan dan cukup baik untuk digunakan dalam analisis atau pengembangan model.

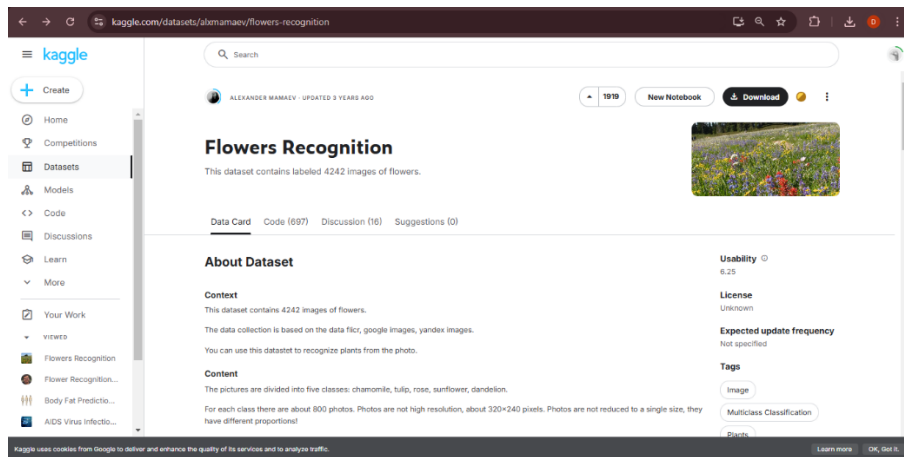
A. Sumber Data

Dataset yang digunakan berasal dari kaggle dengan judul “Flowers Recognition”, dataset ini berisi tentang 4242 gambar bunga. Pengumpulan data didasarkan pada data flicr, google images, yandex images.

Berikut link dataset dan sumberkode dari kaggle Flowers Recognition:

Link Dataset: <https://www.kaggle.com/datasets/alxmamaev/flowers-recognition>

Sumberkode: Kaggle.com, Youtube, Chat GPT, GitHub, Materi Mesin Learning.



Gambar 2. 1 Sumber Dataset

B. Deskripsi Data

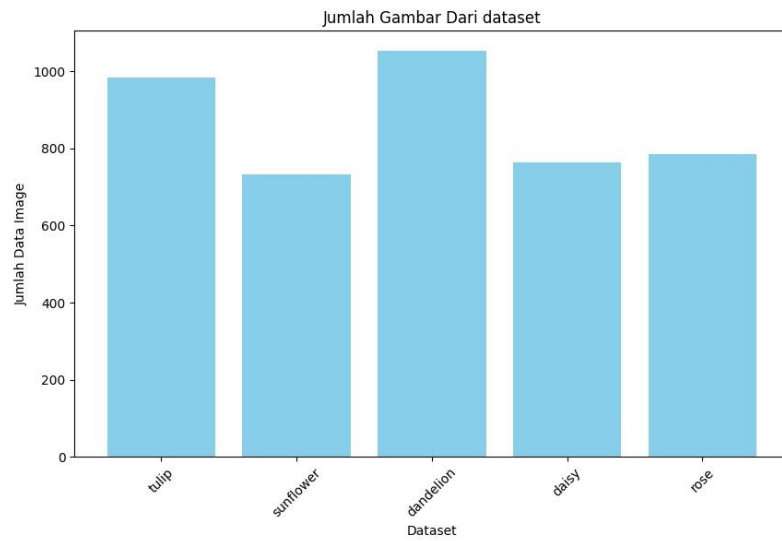
Dataset ini berasal dari kaggle dengan judul “Flowers Recognition”. Dataset ini berisi 4317 gambar bunga dengan format jpg. Pengumpulan data didasarkan pada data flicr, google images, yandex images. Dataset ini dapat digunakan untuk mengenali tanaman dari foto. Isi gambar-gambar tersebut dibagi menjadi lima kelas: kamomil, tulip, mawar, bunga matahari, dandelion. Untuk setiap kelas ada sekitar 800 foto. Foto tidak beresolusi tinggi, sekitar 320 x 240 piksel dengan ukuran tidak lebih dari 100kb. Foto tidak diperkecil menjadi satu ukuran, melainkan memiliki proporsi berbeda. Jumlah data secara keseluruhan yaitu 4317 data.

C. Explore Data

Dalam penelitian ini penulis mengidentifikasi lima jenis bunga yang digunakan sebagai sampel untuk sistem deteksi dan identifikasi jenis bunga. Setiap jenis bunga memiliki jumlah data yang bermacam-macam seperti pada tabel 2.1.

Tabel 2. 1 Jumlah Visualisasi Data

Jenis	Jumlah
Tulip	984
Sunflower	733
Dandelion	1052
Daisy	764
Rose	784
Total	4317



Gambar 2. 2 Diagram Jumlah Visualisasi Data

D. Kualitas Data

Dataset Flowers Recognition di Kaggle yang diunggah oleh alxmamaev berisi sekitar 4.317 gambar bunga yang terbagi menjadi lima kelas, yaitu daisy, dandelion, rose, sunflower, dan tulip. Semua gambar disimpan dalam format JPEG, sehingga kompresi yang digunakan cenderung lossy (mengurangi kualitas sedikit), namun ukuran file relatif kecil dan mudah diproses. Gambar memiliki variasi warna yang baik untuk setiap kelas bunga. Warna bunga pada tiap gambar sangat representatif dan berguna untuk model klasifikasi berbasis warna atau tekstur. ebagian besar gambar diambil dalam berbagai kondisi pencahayaan dan latar belakang, sehingga cocok untuk model pembelajaran yang dapat belajar dari kondisi dunia nyata (misalnya, kondisi outdoor yang berbeda). Dataset ini cukup seimbang untuk kelima kelas, sehingga cocok untuk model klasifikasi yang tidak memerlukan teknik penyeimbangan khusus.



Gambar 2. 3 Daisy



Gambar 2. 4 Dandelion



Gambar 2. 5 Rose



Gambar 2. 6 Sunflowers



Gambar 2. 7 Tulip

III. Data Preparation

A. Pemilihan Data

Pemilihan data pada gambar untuk dataset seperti *Flowers Recognition* bisa dilakukan dengan mempertimbangkan beberapa aspek yang memastikan gambar relevan dan berkualitas untuk proses pelatihan model. Memastikan gambar memiliki resolusi yang cukup tinggi dan jelas, sehingga fitur bunga dapat dikenali oleh model. Jika terlalu rendah atau buram, pertimbangkan untuk mengecualikan gambar tersebut, gambar yang buram atau tidak fokus dapat menurunkan kinerja model. Penulis melakukan pengecekan manual atau otomatisasi menggunakan algoritma pendeteksi ketajaman untuk mengecualikan

gambar buram dengan teknik *data augmentation* seperti rotasi, zoom, perubahan pencahayaan, atau transformasi kecil pada gambar.

B. Data Preprocessing

Pada tahap ini penulis mencari terlebih dahulu apakah ada data yang akan mengganggu seperti data rusak, hilang, buram dllnya yang bisa mengganggu dalam process pemodelan dan membagi data menjadi dua yaitu data train dan data validasi untuk mengolah data agar akurasi menjadi lebih baik.

C. Data Issue

Pada dataset ini penulis menemukan beberapa gambar yang hilang, buram, dan rusak. Akhirnya penulis melakukan penghapusan data dan memasukan data kedalam daftar gambar rusak disimpan dalam 'corrupt_images.txt'.

IV. Modeling

A. Algoritma / Teknik Yang Digunakan

. Pada proyek klasifikasi batik ini, teknik yang digunakan adalah *Deep Learning* dengan memanfaatkan algoritma *Convolutional Neural Networks* (CNN). CNN merupakan jenis jaringan saraf tiruan yang dirancang khusus untuk tugas pengolahan data berbentuk gambar. Algoritma ini terkenal efektif untuk ekstraksi fitur spasial dan klasifikasi gambar.

B. Proses Pembagian Dataset

Pada proyek klasifikasi dan deteksi bungan ini, dataset dibagi menjadi dua subset utama yaitu training set dan testing set. Proses pembagian dilakukan menggunakan fungsi `train_test_split` dari *library sklearn*, dengan mempertimbangkan proporsi data yang seimbang antar kelas.

Tahapan Pembagian Dataset:

1. Pembacaan Dataset

Dataset berupa gambar batik dibaca dari folder yang terorganisir berdasarkan kelas. Gambar-gambar ini diubah menjadi format array menggunakan library *Pillow* dan diubah ukurannya menjadi dimensi yang seragam, yaitu 320x240 piksel.

2. Proses Pembagian Dataset

Dataset dibagi menjadi 80% data pelatihan dan 20% data pengujian. Pembagian dilakukan dengan parameter `stratify` untuk menjaga distribusi proporsi kelas yang sama pada kedua subset

C. Proses Pelatihan Model

```
✓ 0d model = Sequential([
    data_augmentation,
    layers.Rescaling(1./255),
    Conv2D(16, 3, padding='same', activation='relu'),
    MaxPooling2D(),
    Conv2D(32, 3, padding='same', activation='relu'),
    MaxPooling2D(),
    Conv2D(64, 3, padding='same', activation='relu'),
    MaxPooling2D(),
    Dropout(0.2),
    Flatten(),
    Dense(128, activation='relu'),
    Dense(5)
])

✓ 0d [16] model.compile(optimizer='adam',
    loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),
    metrics=['accuracy'])
```

Gambar 4. 1 Pelatihan Model

Arsitektur model yang digunakan dalam aplikasi klasifikasi bunga ini adalah Convolutional Neural Network (CNN), yang terdiri dari beberapa lapisan konvolusional dan pooling, diikuti oleh lapisan dense untuk klasifikasi. Model dimulai dengan data augmentation yang melibatkan operasi acak seperti pembalikan horizontal, rotasi, dan zoom untuk meningkatkan variasi data pelatihan dan mengurangi risiko overfitting. Gambar input yang berukuran 180x180 piksel dengan 3 saluran warna (RGB) kemudian diproses melalui tiga lapisan Conv2D yang masing-masing memiliki 16, 32, dan 64 filter.

Lapisan konvolusional ini menggunakan fungsi aktivasi ReLU dan diikuti oleh lapisan MaxPooling2D untuk mengurangi dimensi spasial gambar, sambil mempertahankan fitur penting. Setelah lapisan konvolusional dan pooling, terdapat lapisan Dropout dengan tingkat 0.2 untuk mengurangi overfitting dengan menonaktifkan 20% neuron secara acak selama pelatihan. Data kemudian diratakan menggunakan lapisan Flatten, yang mengubah data 2D menjadi vektor 1D, yang selanjutnya diteruskan ke lapisan Dense dengan 128 neuron dan fungsi aktivasi ReLU. Pada lapisan output, model menggunakan lapisan Dense dengan 5 neuron yang mewakili lima kelas bunga yang berbeda (*Daisy*, *Dandelion*, *Rose*, *Sunflower*, dan *Tulip*), tanpa fungsi aktivasi, karena model menggunakan *SparseCategoricalCrossentropy* dengan *logits* untuk menghitung loss dan melakukan klasifikasi. Dengan arsitektur ini, model dapat mempelajari fitur penting dari gambar bunga dan mengklasifikasikannya ke dalam kategori yang tepat berdasarkan pola visual yang terdeteksi.

D. Validasi

```
[18] history = model.fit(train_ds, epochs=15, validation_data=val_ds)

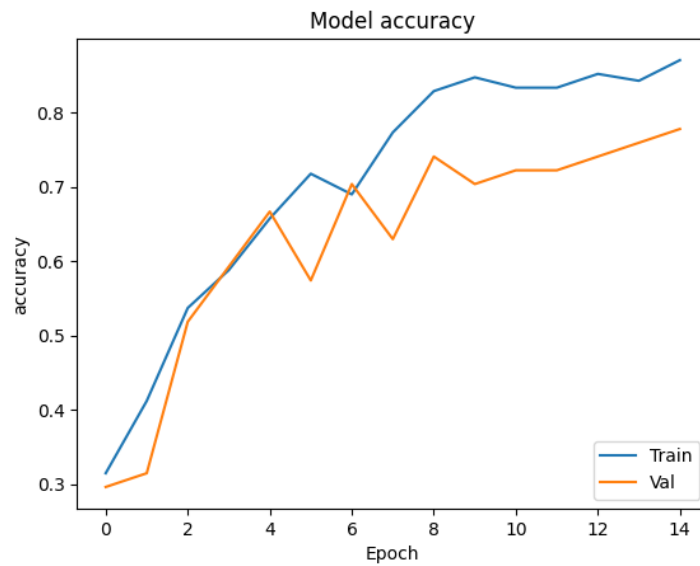
Epoch 1/15
7/7 ----- 22s 3s/step - accuracy: 0.2692 - loss: 1.6452 - val_accuracy: 0.2963 - val_loss: 1.4736
Epoch 2/15
7/7 ----- 10s 1s/step - accuracy: 0.4466 - loss: 1.3667 - val_accuracy: 0.3148 - val_loss: 1.2555
Epoch 3/15
7/7 ----- 9s 1s/step - accuracy: 0.5033 - loss: 1.1322 - val_accuracy: 0.5185 - val_loss: 1.0073
Epoch 4/15
7/7 ----- 10s 1s/step - accuracy: 0.5898 - loss: 1.0225 - val_accuracy: 0.5926 - val_loss: 0.9006
Epoch 5/15
7/7 ----- 8s 1s/step - accuracy: 0.6380 - loss: 0.8265 - val_accuracy: 0.6667 - val_loss: 0.8844
Epoch 6/15
7/7 ----- 10s 1s/step - accuracy: 0.7237 - loss: 0.7001 - val_accuracy: 0.5741 - val_loss: 0.9407
Epoch 7/15
7/7 ----- 12s 1s/step - accuracy: 0.7266 - loss: 0.6967 - val_accuracy: 0.7037 - val_loss: 0.7322
Epoch 8/15
7/7 ----- 9s 1s/step - accuracy: 0.7753 - loss: 0.6133 - val_accuracy: 0.6296 - val_loss: 0.8883
Epoch 9/15
7/7 ----- 8s 1s/step - accuracy: 0.8350 - loss: 0.5959 - val_accuracy: 0.7407 - val_loss: 0.7951
Epoch 10/15
7/7 ----- 12s 1s/step - accuracy: 0.8552 - loss: 0.4998 - val_accuracy: 0.7037 - val_loss: 0.7643
Epoch 11/15
7/7 ----- 10s 1s/step - accuracy: 0.8209 - loss: 0.4586 - val_accuracy: 0.7222 - val_loss: 0.7992
Epoch 12/15
7/7 ----- 8s 1s/step - accuracy: 0.8297 - loss: 0.4110 - val_accuracy: 0.7222 - val_loss: 0.8068
Epoch 13/15
7/7 ----- 12s 1s/step - accuracy: 0.8647 - loss: 0.3801 - val_accuracy: 0.7407 - val_loss: 0.7631
Epoch 14/15
7/7 ----- 9s 1s/step - accuracy: 0.8529 - loss: 0.4147 - val_accuracy: 0.7593 - val_loss: 0.6641
Epoch 15/15
7/7 ----- 8s 1s/step - accuracy: 0.8632 - loss: 0.3792 - val_accuracy: 0.7778 - val_loss: 0.5627
```

Gambar 4. 2 Validasi Setelah Pelatihan

Selanjutnya melakukan pelatihan model dengan data pelatihan (*train_data*) sebanyak 5 kali dengan kode “*epoch*” dan mengecek performanya pada data validasi (*val_data*) setelah setiap kali selesai melatih. Hasil dari proses ini disimpan dalam variabel “*history*”, yang berisi informasi seperti keakuratan model dan tingkat kesalahan (*loss*) selama pelatihan dan validasi. Hasil epoch menunjukkan ***accuracy 0,8632 loss 0,3792 val_accuracy 0,7778 val_loss 0,56***.

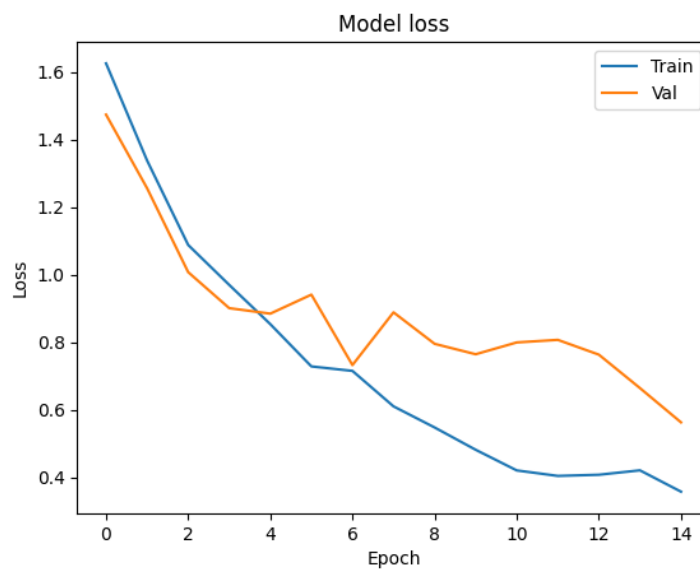
V. Evaluation

Pada tahap evaluasi dilakukan langkah untuk menilai kinerja algoritma atau model yang telah dibuat. Evaluasi ini membantu memastikan apakah solusi yang diterapkan bekerja dengan baik dalam pendeteksian jenis bunga.



Gambar 4. 3 Data Accuracy

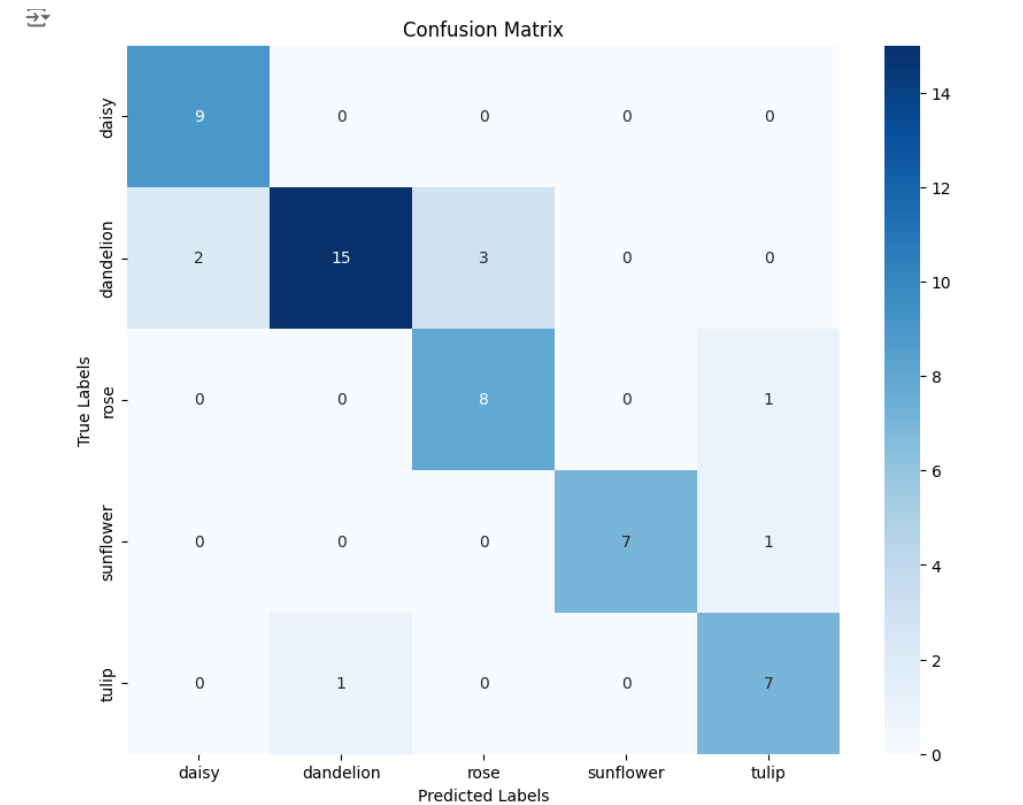
Dalam gambar 4.3 terlihat akurasi berkembang seiring bertambahnya epoch



Gambar 4. 4 Data Loss

Dalam gambar 4.4 tervisualisasikan bahwa model loss turun yang menunjukkan tren menurun dan stabil angka rendah, menjadikan model berhasil dengan baik dari data train. Meskipun val loss fluktuatif, rata – rata lossnya tetap rendah, menunjukan bahwa model memiliki kemampuan generalisasi yang baik untuk data validasi.

Selanjutnya melakukan perhitungan dan memvisualisasikan confusion Matrix untuk mengevaluasi kinerja model pada klasifikasi penyakit pada daun kentang. Kode ini memberikan cara untuk mengavulasi performma model secara detail. Dengan adanya visual dari confusion matrix mempermudah dalam memahami dimana model bekerja dengan baik, dapat mengidentifikasi kelemahan seperti kelas yang sering salah klasifikasi, dan dapat memanfaatkan informasi untuk meningkatkan model.



Gambar 4. 5 *Confussion matrix*

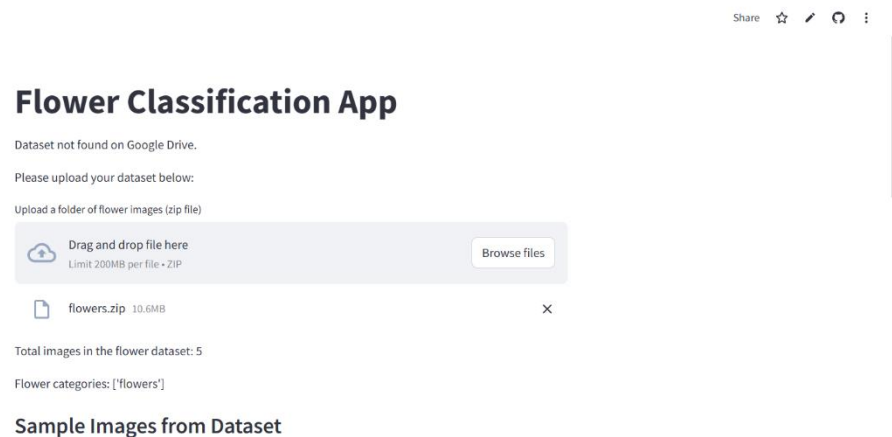
Selanjutnya melakukan tes akurasi model, dan didapatkan akurasi 97,542291879653393% menunjukkan kesalahan prediksi model sangat kecil, yang menandakan model mampu mempelajari pola data latih dengan sangat baik. Setelah selesai melakukan proses permodelan, gambar ini menjelaskan proses penyimpanan model klasifikasi penyakit pada daun kentang yang sudah di latih ke dalam file di komputer.

```
classifiy_images('/content/drive/MyDrive/flowers/daisy/107592979_aaa9cdf78_m.jpg')  
1/1 ————— 0s 74ms/step  
'The Image belongs to daisy with a score of 97.54229187965393'  
  
[ ] from joblib import dump  
  
model_filename = "svm_modelflower.joblib"  
dump(model, model_filename)  
print(f"Model saved as {model_filename}")  
Model saved as svm_modelflower.joblib
```

Gambar 4. 6 Test model dan Save model

VI. Deployment

A. Unggah Gambar Bunga



Gambar 4. 7 Streamlit Deploy

Share ☆ ✎ ↺ ⋮



Uploaded Image

Classifying...

The image belongs to flowers with a score of 100.00%

< Manage app