

***CLUSTERING CITRA TANAMAN OBAT MENGGUNAKAN
METODE K-MEANS***

(Skripsi)

Oleh

**VIRA VERINA
1917051037**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

ABSTRAK

CLUSTERING CITRA TANAMAN OBAT MENGGUNAKAN METODE K-MEANS

Oleh

VIRA VERINA

Tanaman obat memiliki manfaat besar dalam menyembuhkan berbagai penyakit. Banyak masyarakat yang menggunakan tanaman obat sebagai sumber tanaman obat keluarga (TOGA). Bagian tanaman yang dapat dimanfaatkan sebagai obat adalah daun. Daun sulit diidentifikasi karena banyak daun yang memiliki bentuk dan warna yang mirip. Oleh karena itu, pengolahan citra dengan *K-Means Clustering* diterapkan untuk membantu mengidentifikasi daun tanaman obat. Citra akan *resizing* menjadi 600 x 800 piksel. Ciri tekstur pada citra akan dihitung menggunakan *Gray Level Co-occurrence Matrix* (GLCM). GLCM menghasilkan nilai numerik yang kemudian digunakan untuk proses *K-Means Clustering*. *K-Means Clustering* akan mengelompokkan citra berdasarkan kesamaan nilai fitur dan *centroid* terdekatnya. Jumlah *cluster* optimal yaitu $k=3$ berdasarkan perhitungan nilai SSE. Hasil penelitian 900 dataset citra daun terbagi menjadi 3 *cluster* dengan akurasi tertinggi 51,54% pada *cluster* 2 yang memprediksi daun binahong. Proses *K-Means Clustering* kurang baik dikarenakan penentuan nilai *centroid* berpengaruh terhadap hasil *cluster*. *K-Means Clustering* diimplementasikan ke dalam sistem web menggunakan *flask*.

Kata Kunci: *K-Means Clustering*, *Gray Level Co-occurrence Matrix* (GLCM), Tanaman Obat, *Flask*

ABSTRACT

CLUSTERING OF MEDICINAL PLANT IMAGES USING K-MEANS METHOD

Oleh

VIRA VERINA

Medicinal plants have significant benefits in treating various diseases. Many communities utilize medicinal plants as a source of family herbal medicine (TOGA). The part of the plant that is often used for medicinal purposes is the leaves. Identifying leaves can be challenging as many of them have similar shapes and colors. Therefore, image processing with K-Means Clustering is applied to aid in the identification of medicinal plant leaves. The leaf images are scaled to 600 x 800 pixels. The texture features of the images are computed using the Gray Level Co-occurrence Matrix (GLCM) method. GLCM generates numerical values that are later used in the K-Means Clustering process. K-Means Clustering method groups the images based on the similarity of their feature values and their closest centroids. The optimal number of clusters is determined to be $k=3$ based on the calculation of the Sum of Squared Errors (SSE). The results show that out of 900 leaf image datasets, the K-Means Clustering divides them into 3 clusters, with the highest accuracy of 51.54% achieved in cluster 2, which predicts the binahong leaves. K-Means Clustering process is less effective due to the influence of centroid initialization on the clustering results. K-Means Clustering is implemented into a web-based system using Flask.

Keywords: K-Means Clustering, Gray Level Co-occurrence Matrix (GLCM), Medicinal Plant, Flask

**CLUSTERING CITRA TANAMAN OBAT MENGGUNAKAN
METODE K-MEANS**

Oleh

VIRA VERINA

Skripsi

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA KOMPUTER**

Pada

Jurusan Ilmu Komputer

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023**

Judul Skripsi : **CLUSTERING CITRA TANAMAN OBAT
MENGUNAKAN METODE K-MEANS**

Nama Mahasiswa : **Vira Verina**

Nomor Pokok Mahasiswa : 1917051037

Program Studi : **S1 Ilmu Komputer**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. **Komisi Pembimbing**



Rizky Prabowo, M.Kom.
NIP. 198808072019031011



Ridho Sholehurrohman, M. Mat.
NIP. 232111970128101

2. **Ketua Jurusan Ilmu Komputer**

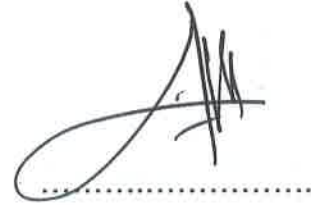


Didik Kurniawan, S.Si., M.T.
NIP. 198004192005011004

MENGESAHKAN

1. Tim Penguji

Ketua : Rizky Prabowo, M.Kom.



Sekretaris : Ridho Sholehurrohman, M. Mat.



Penguji
Utama : Rico Andrian, S.Si., M.Kom.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: 03 Agustus 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Vira Verina

NPM : 1917051037

menyatakan bahwa skripsi saya yang berjudul “**Clustering Citra Tanaman Obat Menggunakan Metode *K-Means***” merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 03 Agustus 2023



Vira Verina

NPM. 1917051037

RIWAYAT HIDUP



Penulis bernama Vira Verina yang lahir di Tangerang pada tanggal 29 Februari 2000, sebagai anak pertama dari dua bersaudara. Penulis menyelesaikan pendidikan formal di SD Negeri 6 Metro Pusat dan selesai pada tahun 2012. Kemudian melanjutkan pendidikan menengah pertama di SMP Negeri 6 Metro yang diselesaikan pada tahun 2015, lalu melanjutkan pendidikan menengah atas di SMA Negeri 1 Metro yang diselesaikan pada tahun 2018.

Pada tahun 2019 penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan antara lain:

1. Mengikuti kegiatan Pengenalan Kehidupan Kampus bagi Mahasiswa Baru (PKKMB) Universitas Lampung, PKKMB FMIPA, dan Pengenalan Jurusan Ilmu Komputer pada tahun 2019.
2. Menjadi anggota Adapter Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2019/2020.
3. Menjadi anggota divisi 3D (Desain, Dokumentasi, dan Dekorasi) pada kegiatan Gebyar Anggota Muda Ilmu Komputer (ADAPTER) Himpunan Mahasiswa Jurusan Ilmu Komputer pada periode 2019/2020.
4. Mengikuti kegiatan Karya Wisata Ilmiah (KWI) XXX FMIPA di Desa Tambah Dadi, Kecamatan Purbolinggo, Kabupaten Lampung Timur pada tahun 2019.
5. Menjadi anggota bidang Eksternal Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2019/2020.

6. Menjadi Sekretaris Bidang Eksternal Himpunan Mahasiswa Jurusan Ilmu Komputer 2020/2021.
7. Menjadi anggota divisi HPD (Humas, Publikasi, Dokumentasi) pada Pekan Raya Jurusan Ilmu Komputer (PRJ) IX Himpunan Mahasiswa Jurusan Ilmu Komputer tahun 2021.
8. Mengikuti ujian sertifikasi *Junior Web Developer* oleh Badan Nasional Sertifikasi Profesi (BNSP) pada tahun 2022.
9. Menjadi Asisten Dosen mata kuliah Keamanan Sistem Informasi (KSI) dan Teknologi & Aplikasi Mobile (TAM) pada tahun 2022.
10. Melaksanakan Kerja Praktik di UPT Badan Pelindungan Pekerja Migran Indonesia (BP2MI) Lampung pada bulan Februari tahun 2022.
11. Melaksanakan KKN di Desa Sinar Banten, Kecamatan Talang Padang, Kabupaten Tanggamus, Lampung pada bulan Juni tahun 2022.

PERSEMBAHAN

Alhamdulillahirobbilalamin

Puji dan syukur tercurahkan kepada Allah Subhanahu Wa Ta'ala atas segala Rahmat dan Karunia-Nya sehingga saya dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Nabi Muhammad Shallallahu Alaihi Wasallam.

Kupersembahkan karya ini kepada:

Kedua Orang Tuaku Tercinta

Yang senantiasa memberikan yang terbaik, dan melantunkan do'a yang selalu menyertaiku. Kuucapkan pula terima kasih sebesar-besarnya karena telah mendidik dan membesarkanku dengan cara yang dipenuhi kasih sayang, dukungan, dan pengorbanan yang belum bisa terbalaskan.

Adikku serta Keluarga Besar

Yang selalu memberikan doa, dukungan dan semangat.

Sahabat dan Teman-temanku

Yang telah menemani, mendukung, dan memberikanku kebahagiaan.

Seluruh Keluarga Besar Ilmu Komputer 2019

Yang selalu memberikan semangat dan dukungan.

Almamater Tercinta, Universitas Lampung dan Jurusan Ilmu Komputer

Tempat bernaung mengemban semua ilmu untuk menjadi bekal hidup.

MOTTO

“There can be miracles, when you believe.”

(Mariah Carey & Whitney Houston)

“As you sow, so shall you reap.”

(Anonim)

“Sesungguhnya sesudah kesulitan, itu ada kemudahan.”

(QS. Al Insyirah ayat 5)

“Syukuri dan Hargai yang kamu miliki”

(Anonim)

“When you love what you have, you have everything you need.”

SANWACANA

Puji syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala atas rahmat dan hidayah-Nya, serta petunjuk dan pedoman dari Rasulullah Nabi Muhammad Shollallahu Alaihi Wasallam penulis dapat menyelesaikan skripsi yang berjudul **“Clustering Citra Tanaman Obat Menggunakan Metode K-Means”** dengan baik dan lancar.

Keberhasilan skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Allah Subhanahu Wa Ta'ala yang telah memberikan rahmat dan karunia-Nya selama menjalani penelitian hingga dapat menyelesaikan skripsi dengan baik.
2. Kedua Orangtua yang selalu memberikan dukungan moral, material, motivasi, kasih sayang yang tak terhingga, dan senantiasa mendoakan selama ini. Semoga Allah SWT selalu memberikan keberkahan dan kebahagiaan untuk kedua orang tuaku di dunia dan akhirat.
3. Adikku Devina Yunanda dan keluarga yang selalu memberikan doa, kasih sayang, dan dukungannya sehingga penelitian ini terselesaikan.
4. Bapak Didik Kurniawan, S.Si., M.T., selaku Ketua Jurusan Ilmu Komputer Universitas Lampung.
5. Ibu Anie Rose Irawati, ST., M.Cs., selaku Sekretaris Jurusan Ilmu Komputer Universitas Lampung dan koordinator skripsi.
6. Ibu Yohana Tri Utami, M.Kom., selaku dosen pembimbing akademik yang telah memberikan arahan dan bimbingan selama setiap semester.
7. Bapak Rizky Prabowo, M.Kom., selaku dosen pembimbing utama dalam penelitian ini, yang selalu memberikan bimbingan, ilmu, dan saran.
8. Bapak Ridho Sholehurrohman, M. Mat., selaku dosen pembimbing kedua dalam penelitian ini, yang selalu memberikan bimbingan, ilmu, dan saran.

9. Bapak Rico Andrian, S.Si., M.Kom., selaku dosen pembahas, yang telah memberikan ilmu, kritik, dan saran dalam penelitian ini.
10. Seluruh Bapak dan Ibu Dosen Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah memberikan ilmu dan pengalamannya selama masa perkuliahan.
11. Seluruh karyawan FMIPA Universitas Lampung, Ibu Ade Nora, Bang Zainuddin, Mas Syam, dan Mas Nofal yang telah membantu segala urusan administrasi.
12. Hani Cita Lestari dan Finka Marisa Geananda Sufie sebagai teman seperjuangan (partner) selama perkuliahan hingga penelitian dan penyusunan skripsi ini.
13. Teman seperjuangan dikala suka dan duka selama perkuliahan Sendy Hani Pramita, Olivia Desti Riana, Salsabilla Julia Farhana, Devi Ramadhia Fitri, dan Zahara Liza Mulyani yang selalu memberikan bantuan dan semangat.
14. Agnes Pramudani, Suci Annisa Putri, Irsyifa Mayzela Afnan, Lidwina Ratih Tri Wulandari, R. Nata Trisna Hardini, Retno Isni Hasanah, Suci Hasanah Bertha, dan Jonathan Michael yang selalu memberi dukungan serta mengapresiasi setiap langkah dalam menyelesaikan skripsi ini.
15. Teman-teman seangkatan Jurusan Ilmu Komputer 2019 di Universitas Lampung.

Penulis menyadari dalam penulisan ini masih jauh dari kata sempurna disebabkan terbatasnya kemampuan, pengetahuan, dan pengalaman. Semoga skripsi ini memberikan manfaat bagi civitas Universitas Lampung dan pihak yang membaca.

Bandar Lampung, 03 Agustus 2023

Vira Verina

NPM. 1917051037

DAFTAR ISI

| | Halaman |
|--|--------------|
| DAFTAR ISI | xiii |
| DAFTAR TABEL | xv |
| DAFTAR GAMBAR | xvi |
| DAFTAR KODE PROGRAM | xviii |
| I. PENDAHULUAN | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah..... | 4 |
| 1.4 Tujuan Penelitian..... | 4 |
| 1.5 Manfaat Penelitian..... | 4 |
| II. TINJAUAN PUSTAKA | 5 |
| 2.1 Penelitian Terdahulu..... | 5 |
| 2.2 Landasan Teori | 9 |
| 2.2.1 Tanaman Obat | 9 |
| 2.2.2 Daun Sirih Hijau | 9 |
| 2.2.3 Daun Binahong | 10 |
| 2.2.4 Daun Keji Beling | 11 |
| 2.2.5 <i>Clustering</i> | 11 |
| 2.2.6 <i>K-Means Clustering</i> | 12 |
| 2.2.7 Metode <i>Elbow</i> | 14 |
| 2.2.8 <i>Unsupervised Learning</i> | 14 |
| 2.2.9 Pengolahan Citra Digital..... | 14 |
| 2.2.10 <i>Gray Level Co-occurrence Matrix (GLCM)</i> | 16 |
| 2.2.11 <i>Python</i> | 17 |
| 2.2.12 Metode <i>Waterfall</i> | 18 |
| 2.2.13 <i>Data Flow Diagram (DFD)</i> | 19 |
| III. METODOLOGI PENELITIAN | 21 |

| | | |
|------------|---|-----------|
| 3.1 | Tempat dan Waktu Penelitian | 21 |
| 3.2 | Sumber Data | 21 |
| 3.3 | Alat Pendukung | 21 |
| 3.3.1 | Perangkat Keras (<i>Hardware</i>) | 21 |
| 3.3.2 | Perangkat Lunak (<i>Software</i>)..... | 22 |
| 3.4 | Tahapan Penelitian | 22 |
| 3.4.1 | Studi Literatur | 23 |
| 3.4.2 | Pengumpulan Data Citra Daun..... | 23 |
| 3.4.3 | <i>Pre-Processing</i> | 24 |
| 3.4.4 | Ekstrasi Fitur | 24 |
| 3.4.5 | <i>K-Means Clustering</i> | 25 |
| 3.4.6 | Metode <i>Elbow</i> | 25 |
| 3.4.7 | Pengembangan Sistem | 25 |
| IV. | HASIL DAN PEMBAHASAN | 30 |
| 4.1 | <i>Pre-Processing</i> Citra | 30 |
| 4.2 | Ekstrasi Fitur (GLCM) | 31 |
| 4.3 | <i>K-Means Clustering</i> | 33 |
| 4.3.1 | Metode <i>Elbow</i> | 38 |
| 4.4 | Pengembangan Sistem..... | 42 |
| V. | KESIMPULAN DAN SARAN | 45 |
| 5.1 | Kesimpulan..... | 45 |
| 5.2 | Saran..... | 45 |
| | DAFTAR PUSTAKA | 46 |
| | LAMPIRAN..... | 49 |

DAFTAR TABEL

| | Halaman |
|--|---------|
| Tabel 1. Penelitian Terdahulu | 5 |
| Tabel 2. Simbol DFD | 19 |
| Tabel 3. Hasil Perhitungan Daun Binahong..... | 32 |
| Tabel 4. Hasil Perhitungan Daun Sirih Hijau | 32 |
| Tabel 5. Hasil Perhitungan Daun Keji Beling..... | 33 |
| Tabel 6. Rata-rata Nilai Fitur dari 10 Data | 34 |
| Tabel 7. <i>Centroid</i> Awal..... | 35 |
| Tabel 8. Hasil Perhitungan Jarak Data ke <i>Centroid</i> Awal | 36 |
| Tabel 9. Anggota <i>Cluster</i> | 36 |
| Tabel 10. <i>Centroid</i> Baru..... | 37 |
| Tabel 11. Hasil Perhitungan Jarak Data ke <i>Centroid</i> Baru | 37 |
| Tabel 12. Hasil Clustering | 40 |

DAFTAR GAMBAR

| | Halaman |
|--|---------|
| Gambar 1. Daun Sirih Hijau | 10 |
| Gambar 2. Daun Binahong..... | 10 |
| Gambar 3. Daun Keji Beling..... | 11 |
| Gambar 4. <i>K-Means Clustering</i> | 12 |
| Gambar 5. Ilustrasi proses <i>K-Means Clustering</i> | 13 |
| Gambar 6. Citra RGB..... | 15 |
| Gambar 7. Citra <i>Grayscale</i> | 15 |
| Gambar 8. Arah sudut GLCM..... | 16 |
| Gambar 9. Tahapan <i>waterfall</i> | 19 |
| Gambar 10. Tahapan Penelitian | 22 |
| Gambar 11. Data Daun Binahong | 23 |
| Gambar 12. Data Daun Sirih Hijau | 23 |
| Gambar 13. Data Daun Keji Beling | 24 |
| Gambar 14. GLCM | 24 |
| Gambar 15. <i>Flowchart</i> Sistem | 26 |
| Gambar 16. <i>Context Diagram</i> | 27 |
| Gambar 17. DFD Level 0..... | 28 |
| Gambar 18. <i>Design Interface</i> | 28 |
| Gambar 19. <i>Scaling</i> | 30 |
| Gambar 20. Plot Metode <i>Elbow</i> | 39 |
| Gambar 21. Plot Hasil <i>K-Means Clustering</i> | 40 |

| | |
|--|----|
| Gambar 22. Tampilan Awal Sistem | 43 |
| Gambar 23. Tampilan <i>Preview</i> Citra | 43 |
| Gambar 24. Tampilan Hasil <i>Clustering</i> | 44 |
| Gambar 25. Hasil <i>Cluster</i> 1 | 50 |
| Gambar 26. Hasil <i>Cluster</i> 1 (lanjutan) | 50 |
| Gambar 27. Hasil <i>Cluster</i> 1 (lanjutan 2) | 50 |
| Gambar 28. Hasil <i>Cluster</i> 1 (lanjutan 3) | 50 |
| Gambar 29. Hasil <i>Cluster</i> 1 (lanjutan 4) | 50 |
| Gambar 30. Hasil <i>Cluster</i> 2 | 50 |
| Gambar 31. Hasil <i>Cluster</i> 2 (lanjutan) | 50 |
| Gambar 32. Hasil <i>Cluster</i> 2 (lanjutan 2) | 50 |
| Gambar 33. Hasil <i>Cluster</i> 2 (lanjutan 3) | 50 |
| Gambar 34. Hasil <i>Cluster</i> 3 | 50 |
| Gambar 35. Hasil <i>Cluster</i> 3(lanjutan) | 50 |

DAFTAR KODE PROGRAM

| | Halaman |
|--|---------|
| Kode Program 1. Perhitungan GLCM | 31 |
| Kode Program 2. <i>Export</i> Perhitungan GLCM ke <i>Excel</i> | 31 |
| Kode Program 3. <i>K-Means Clustering</i> pada <i>Python</i> | 33 |
| Kode Program 4. Membaca Data Excel..... | 34 |
| Kode Program 5. <i>Euclidean Distance</i> | 35 |
| Kode Program 6. Prediksi <i>Cluster</i> | 38 |
| Kode Program 7. Perhitungan SSE | 39 |
| Kode Program 8. <i>Export</i> model ke format <i>pkl</i> | 42 |

I. PENDAHULUAN

1.1 Latar Belakang

Tanaman obat merupakan tumbuhan yang memiliki manfaat untuk mencegah dan menyembuhkan penyakit. Tanaman obat biasanya dapat dijadikan obat tradisional maupun bahan baku pembuatan obat kimiawi. Salah satu bagian tanaman yang dapat digunakan sebagai obat adalah daun. Sering ditemui daun yang satu dengan lainnya memiliki warna dan bentuk yang serupa. Sehingga, dibutuhkan Pengolahan Citra untuk mempermudah manusia mengidentifikasi daun tanaman obat.

Pengolahan Citra merupakan proses mengolah dan menginterpretasi citra dengan bantuan komputer. Tahap awal dari pengolahan citra adalah *preprocessing*. *Preprocessing* merupakan persiapan dan penyesuaian citra seperti menyeragamkan ukuran citra dan konversi warna. Selanjutnya, dilakukan ekstraksi fitur untuk mengambil ciri dari citra agar dapat dikenali/dibedakan dengan citra lainnya. *Gray Level Co-occurrence Matrix* (GLCM) merupakan salah satu ekstraksi fitur untuk menganalisis tekstur. GLCM merepresentasikan hubungan spasial *pixel* pada citra yang memiliki intensitas warna keabuan (Widodo *et al.*, 2018).

Hasil pengolahan citra akan *dicluster* berdasarkan kemiripan karakteristik data. *Cluster* adalah sekelompok data yang serupa dalam satu *cluster* dan berbeda dengan *cluster* lainnya. *Clustering* merupakan salah satu metode *unsupervised learning* yang membagi data menjadi beberapa *cluster*.

Clustering dibagi menjadi dua jenis, yaitu *Hierarchical Clustering* dan *Non-Hierarchical (Partitional) Clustering*. *Hierarchical Clustering* adalah metode yang digunakan untuk data yang tidak terlalu banyak dan jumlah *cluster* belum diketahui (Astuti *et al.*, 2018). Metode ini dilakukan secara bertahap dan disajikan dalam bentuk dendogram. Metode hirarki dibagi menjadi dua, yaitu *Agglomerative* (Penggabungan) dan *Divisive* (Pembagian). Sedangkan, *Non-Hierarchical (Partitional) Clustering* adalah metode yang ditentukan terlebih dahulu jumlah *clusternya* dan efisien untuk data yang banyak. Metode ini dilakukan dengan mempartisi data dan secara iteratif. Metode non hirarki yang banyak digunakan yaitu *K-Means* dan *K-Medoids*.

Penelitian ini menggunakan metode *K-Means Clustering*. *K-Means Clustering* mengelompokkan data secara iteratif dan menentukan jumlah *cluster* di awal. *K-Means Clustering* mengelompokkan data berdasarkan jarak minimum ke *centroid*. Metode ini dapat meminimalkan kesamaan karakteristik data antar *cluster* dan mengoptimalkan kesamaan karakteristik data dalam satu *cluster*.

Penelitian (Desmanto & Angreni, 2014) menggunakan metode *K-Means Clustering* dalam pengelompokkan citra digital menghasilkan akurasi sebesar 83%. Perhitungan akurasi berdasarkan jumlah citra digital yang sesuai tingkat kemiripan warna setiap *cluster*. Penentuan nilai *centroid* awal sangat berpengaruh terhadap hasil *cluster*. Terbukti pengujian 30 citra dengan *centroid* awal data citra 1, citra 2, dan citra 3 menghasilkan 3 *cluster*. *Cluster* 1 ada 5 citra, *cluster* 2 ada 15 citra, dan *cluster* 3 ada 10 citra. Lalu, pengujian 30 citra dengan *centroid* awal data citra 4, citra 5, dan citra 6 menghasilkan 3 *cluster*. *Cluster* 1 ada 13 citra, *cluster* 2 ada 10 citra, dan *cluster* 3 ada 7 citra.

Menurut (Widodo *et al.*, 2018) ekstraksi ciri *Gray Level Co-occurrence Matrix* (GLCM) citra jeruk keprok untuk klasifikasi mutu menghasilkan akurasi tertinggi sebesar 82,5% dengan jumlah data latih sebesar 20 dan 30 data, pada jarak ketetanggaan piksel (*distance*) sebesar 2 dan pada arah GLCM 45° . Arah diagonal ($45^\circ/135^\circ$) nilai homogenitas kecil sehingga variasi intensitas dalam citra tinggi. Ciri tersebut merepresentasikan nilai tekstur pada citra jeruk. GLCM dapat mengetahui perbedaan nilai antar piksel pada citra. Jika nilai antar piksel homogen maka nilai kontras kecil. Jika nilai antar piksel tidak homogen maka nilai kontras besar.

(Atina, 2017) melakukan penelitian segmentasi citra paru menggunakan metode *K-Means Clustering* mendapatkan hasil terbaik pada jumlah *cluster* 8. Citra yang dihasilkan pada $k = 8$ memvisualisasikan pembentukan warna yang lebih nyata dan jelas. Menurutnya, citra ini dapat dijadikan data *Computerized Aid Diagnose* (CAD) untuk mendeteksi jenis penyakit pada paru secara komputerisasi.

Penelitian terdahulu menunjukkan bahwa pengolahan citra dapat membantu mengidentifikasi suatu citra dengan baik. Penulis akan melakukan penelitian yang berjudul "*Clustering Citra Tanaman Obat Menggunakan Metode K-Means*". Penelitian ini untuk mengelompokkan tanaman obat yang berjenis daun berdasarkan tekstur.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana mengembangkan model *Clustering* tanaman obat menggunakan metode *K-Means*?
2. Bagaimana cara implementasi terhadap model *Clustering* tanaman obat menggunakan metode *K-Means*?

1.3 Batasan Masalah

Agar penelitian tidak menyimpang dari tujuan yang diharapkan, maka ditetapkan batasan masalah dalam penelitian ini yaitu:

1. Metode yang digunakan dalam penelitian ini adalah *K-Means Clustering* menggunakan bahasa *Python*.
2. Data dalam penelitian ini berjumlah 900 *dataset* citra tanaman obat dari 3 jenis tanaman yaitu daun sirih hijau, daun binahong, dan daun keji beling yang diperoleh dari penelitian sebelumnya.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini yaitu:

1. Membuat model *Clustering* tanaman obat menggunakan metode *K-Means*.
2. Melakukan implementasi terhadap model *Clustering* tanaman obat menggunakan metode *K-Means*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dalam penelitian ini yaitu:

1. Menambah pengetahuan terkait Pengolahan Citra menggunakan metode *K-Means Clustering*.
2. Mengetahui cara membuat model dan implementasi terhadap model *Clustering* tanaman obat menggunakan metode *K-Means*.

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini dibutuhkan penelitian terdahulu sebagai berikut:

Tabel 1. Penelitian Terdahulu

| No | Peneliti | Judul | Data | Metode | Hasil |
|----|--------------------------------|--|---|--|---|
| 1. | (Darnita <i>et al.</i> , 2020) | Penerapan <i>K-Means Clustering</i> Pada Aplikasi Android Tanaman Herbal | Metode Data tanaman dibagi 3 bagian yaitu tanaman berbunga, tanaman berbuah, dan rempah-rempah. | Metode pengembangan sistem yaitu <i>Waterfall</i> dan <i>K-Means Clustering</i> untuk pengclusteran. | Penggunaan <i>k random</i> tidak ada jaminan untuk menemukan kumpulan <i>cluster</i> yang optimal, namun apabila nilai yang diperoleh acak untuk penginisialisasi kurang baik maka pengelompokkan yang didapat tidak optimal. |
| 2. | (Astuti <i>et</i> | Perbandingan | Data jumlah | <i>K-Means</i> dan | Hasil <i>cluster</i> yang terbentuk dengan metode |

| | | | | |
|--------------------|--|---|-------------------------------|---|
| <i>al.</i> , 2018) | Algoritma <i>K-Means</i> dan Algoritma <i>K-Medoids</i> dalam Pengelompokkan Komoditas Tanaman Biofarmaka di Provinsi Jawa Tengah Tahun 2018 | produksi tanaman biofarmaka tahun 2018 di Jawa Tengah | <i>K-Medoids</i> | Cluster K-means dan metode Cluster K-medoids adalah sama. Diperoleh hasil <i>clusternya</i> yaitu <i>cluster</i> 1 ada 2 kabupaten, <i>cluster</i> 2 ada 32 kabupaten/kota dan <i>cluster</i> 3 ada 1 kabupaten. Dari <i>cluster</i> yang terbentuk diperoleh urutan kelompok kabupaten/kota dengan jumlah produksi yang paling banyak hingga kelompok kabupaten/kota yang jumlah produksi paling sedikit berturut-turut adalah <i>cluster</i> 3, <i>cluster</i> 1, <i>cluster</i> 2. Cluster 3 adalah kategori <i>cluster</i> tinggi yang artinya kelompok kabupaten yang berada di <i>cluster</i> 3 memproduksi semua tanaman biofarmaka dengan hasil rata-rata jumlah produksinya yang tinggi dengan nilai rata-ratanya yaitu 2.476.562,9. Daerah yang berada di <i>cluster</i> 3 adalah Kabupaten Wonogiri. |
| 3. | (Marhaena Penentuan Sangrai <i>nto et al.</i> , | Lama 30 kg biji kopi arabika dan 60 Kopi | Analisis regresi dan korelasi | Hubungan variabel citra (RGB) sampel kopi arabika dan robusta dengan variasi |

| | | | | |
|-----------------------------------|---|--|--|--|
| 2015) | Berdasarkan Variasi Derajat Sangrai Menggunakan Model Warna RGB pada Pengolahan Citra Digital (<i>Digital Citra Processing</i>) | biji kopi linier. Korelasi standar derajat kuat dengan nilai koefisien | kopi yang digunakan untuk mengetahui hubungan antara data menggunakan parameter citra dengan lama sangrai. | standar derajat kuat dengan nilai koefisien korelasi berkisar pada interval 0,8-1,0. |
| 4. (Jatmika & Purnamasa ri, 2014) | Rancang Bangun Alat Pendeteksi Kematangan Apel Menggunakan Metode <i>Citra Processing</i> Berdasarkan Komposisi Warna | Buah apel yang berjenis <i>rome beauty</i> . | <i>Citra Processing</i> berdasarkan komposisi warna dan Bahasa pemrograman <i>Delphi</i> . | Pengujian dilakukan 6 kali berdasarkan pencahayaan dan jarak apel dengan <i>webcam</i> . Berdasarkan hasil pengujian, keakuratan data terjadi pada kondisi cahaya lampu 14 watt dengan jarak 4 meter dan jarak <i>webcam</i> 13 cm. Pada kondisi tersebut tingkat eror 0%. Perhitungan nilai jarak antar histogram dapat menentukan kematangan apel. Dimana jika selisih jarak sama dengan nol maka citra memiliki |

| | | | | | | |
|----|--------------------------|--|--|---------------------------|-------------------------------|--|
| | | | | | | kemiripan distribusi warna. Range yang didapat 0%-11% (matang), 12%-17% (mengkal), dan 18%-100% (mentah). |
| 5. | (Apriyanti et al., 2016) | Algoritma <i>K-Means</i> Clustering dalam Pengolahan Citra Digital Landsat | Citra sumberdaya bumi (Studi kasus: Banjarbaru, Kalimantan Selatan). | satelit (Studi Vektor RGB | <i>K-Means</i> Clustering dan | Algoritma <i>K-Means</i> menghasilkan 12 kelas yang ditentukan berdasarkan warna dan nilai piksel. 12 kelas pembagian wilayah Banjarbaru tersebut yaitu, padang rumput, bandara udara, pertambangan, lahan terbuka, hutan tanaman, rawa, semak, belukar, pemukiman, perkebunan, pertanian lahan kering, dan lahan bervegetasi. Berdasarkan hasil kecocokan dengan lapangan yaitu 94,4% sesuai dan 5,6% tidak sesuai. |

2.2 Landasan Teori

2.2.1 Tanaman Obat

Tanaman obat merupakan tumbuhan yang memiliki khasiat untuk mencegah dan menyembuhkan berbagai penyakit. Tanaman obat dapat dijadikan obat tradisional dengan mengolah bagian tumbuhan secara alami. Tanaman obat juga dapat dijadikan sebagai bahan baku obat kimiawi. Tanaman obat dapat tumbuh secara liar maupun ditanam pada pekarangan rumah. Banyak masyarakat yang secara sengaja menanam berbagai macam tanaman obat tersebut agar dapat dikonsumsi dengan mudah. Tanaman obat ini dikenal sebagai sumber tanaman obat keluarga (TOGA) (Sarno, 2019).

2.2.2 Daun Sirih Hijau

Daun sirih hijau (*Piper betle Linn*) merupakan tanaman obat yang memiliki kandungan antibakteri untuk mengobati jerawat, antioksidan bagi tubuh, dan antiseptik yang dapat menyembuhkan luka (Noventi & Carolia, 2016). Daun sirih hijau tumbuh merambat dan menjalar hingga ketinggian 15 m dan memiliki karakteristik daun berbentuk jantung, bertulang daun melengkung, akar tunggang, bau yang khas, dan tekstur permukaan bawah agak kasar.



Gambar 1. Daun Sirih Hijau

2.2.3 Daun Binahong

Daun binahong (*Anredera cordifolia*) mirip dengan daun sirih hijau namun daun binahong betulang daun menyirip dan daunnya agak tebal. Daun binahong juga memiliki manfaat menyembuhkan berbagai penyakit seperti mencegah diabetes, mempercepat penyembuhan luka bakar, menurunkan risiko penyakit jantung, dan mengobati sariawan.



Gambar 2. Daun Binahong

2.2.4 Daun Keji Beling

Daun keji beling (*Strobilanthes crispus*) dikenal sebagai obat tradisional untuk mengatasi diabetes mellitus karena ekstrak daun keji beling menunjukkan dapat menurunkan kadar glukosa darah (Nurhidayah *et al.*, 2015). Selain itu, daun keji beling juga dapat mengobati batu ginjal, sembelit, dan baung air kecil yang kurang lancar. Daun keji beling memiliki bentuk daun elips-lonjong, tepi daun bergigi, permukaan daun berbulu halus, pangkal dan ujung daun meruncing, bertulang daun menyirip, dan daun berpasangan dengan susunan berlawanan.



Gambar 3. Daun Keji Beling

2.2.5 Clustering

Clustering merupakan proses pengelompokkan data berdasarkan kemiripan karakteristik (Lestari, 2018). Dalam *clustering* data tidak perlu diberi label. Terdapat dua jenis metode *clustering* yaitu:

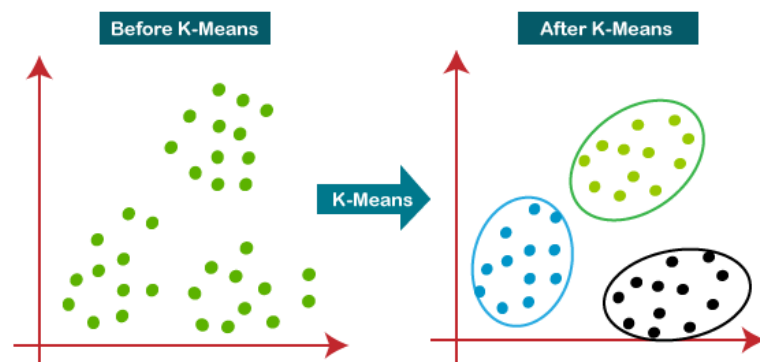
1. *Hierarchical Clustering* adalah metode yang membangun tingkatan *cluster* dan dilakukan secara

terstruktur. Biasanya metode ini disajikan dalam bentuk dendogram atau semacam pohon. Secara umum metode ini dibagi menjadi dua yaitu *Agglomerative* (Penggabungan) dan *Divisive* (Pembagian).

2. *Non-Hierarchical (Partitional) Clustering* adalah metode yang mempartisi data dan menentukan jumlah *cluster* terlebih dahulu. Metode ini lebih efisien untuk data yang lebih banyak. Metode yang banyak digunakan dalam *Non-Hierarchical Clustering* yaitu *K-Means* dan *K-Medoids*.

2.2.6 *K-Means Clustering*

K-Means Clustering merupakan metode pengelompokan data dengan mempartisi sekumpulan data (*dataset*) ke dalam sejumlah k *cluster*. Tujuannya untuk meminimalkan kemiripan karakteristik data antar *cluster* dan memaksimalkan kemiripan karakteristik data dalam satu *cluster* (Astuti *et al.*, 2018).



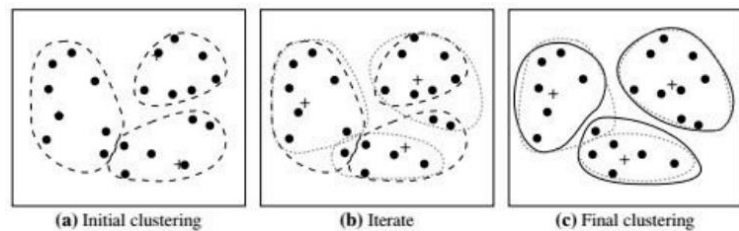
Gambar 4. *K-Means Clustering*

(Sumber: <https://www.trivusi.web.id/2022/06/algorithm-kmeans-clustering.html>)

Secara umum, berikut tahapan *K-Means Clustering*:

1. Tentukan nilai k sebagai jumlah *cluster*. Pilih k data sebagai pusat *cluster* (*centroid*) awal secara acak.

2. Setelah k *centroid* terbentuk, hitung jarak setiap data ke *centroid* terdekat menggunakan rumus *Euclidean Distance*.
3. Kemudian didapatkan anggota *cluster* berdasarkan jarak data ke *centroid* bernilai paling kecil (jarak pendek).
4. Hitung *centroid* baru.
5. Ulangi langkah 2 sampai 4 hingga tidak ada lagi data yang berpindah ke *cluster* lain. Dapat diilustrasikan seperti pada gambar 1 sebagai berikut:



Gambar 5. Ilustrasi proses *K-Means Clustering*

(Han *et al.*, 2012)

Rumus *Euclidean Distance* digunakan dalam *K-Means Clustering* untuk menghitung jarak setiap data ke *centroid* sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

x merupakan data dan y merupakan *centroid*.

Rumus untuk mencari *centroid* sebagai berikut:

$$centroid_i = \frac{\sum_{j=1}^n x_j}{n}$$

$centroid_i$ merupakan *centroid cluster* ke- i , x_j merupakan data *point* ke- j yang termasuk dalam *cluster* ke- i , dan n merupakan jumlah data *point* yang termasuk dalam *cluster* ke- i .

2.2.7 Metode *Elbow*

Metode *Elbow* merupakan metode untuk mendapatkan jumlah *cluster* yang optimal (Muningsih & Kiswati, 2018). Metode ini digunakan untuk evaluasi *cluster*. Dikatakan jumlah *cluster* terbaik apabila membentuk siku pada suatu titik. Untuk mendapatkan perbandingan dilakukan perhitungan *Sum of Square Error* (SSE):

$$SSE = \sum_{k=1}^k \sum_{x \in C_k} \|x - \mu_k\|_2^2$$

Dengan keterangan:

k = banyaknya *cluster*

x = data pada masing-masing *cluster*

c_k = *cluster* ke- k

μ_k = *centroid cluster* ke- k

2.2.8 *Unsupervised Learning*

Unsupervised learning merupakan teknik dari *machine learning* yang tidak perlu memberi label pada data dan digunakan untuk pengelompokan/*clustering* (Roihan *et al.*, 2020). Algoritma dalam *unsupervised learning* yaitu *K-Means*, *Fuzzy*, *Apriori*, dan *DBSCAN*. Dalam *unsupervised learning* model bekerja sendiri untuk menemukan struktur dan pola dari data.

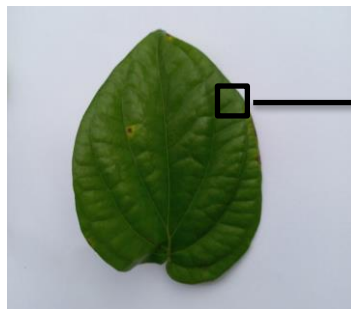
2.2.9 Pengolahan Citra Digital

Citra merupakan gambar dua dimensi yang disusun dengan sejumlah *pixel* yang memiliki koordinat. Citra digital merupakan representasi fungsi dua dimensi $f(x,y)$, dimana x dan y menunjukkan posisi piksel, dan fungsi f pada setiap

koordinat menunjukkan nilai intensitas warna citra (Kusuma & Ellyana, 2018). Citra dibagi menjadi 2 yaitu Citra RGB dan Citra *Grayscale*.

1. Citra RGB

Citra RGB merupakan citra berwarna yang terdiri dari 3 warna merah (*Red*), hijau (*Green*), dan biru (*Blue*). Setiap piksel pada citra RGB memiliki 3 nilai yang mewakili intensitas warna dengan skala 0 hingga 255.

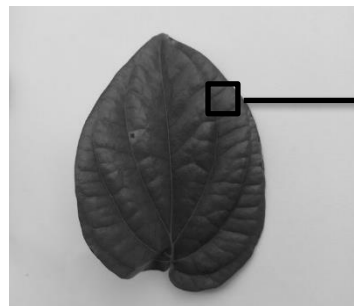


| | | |
|-------|-------|-------|
| R:169 | R:170 | R:171 |
| G:178 | G:179 | G:180 |
| B:185 | B:186 | B:187 |
| R:171 | R:170 | R:170 |
| G:180 | G:179 | G:179 |
| B:187 | B:186 | B:186 |
| R:171 | R:171 | R:171 |
| G:178 | G:178 | G:178 |
| B:186 | B:186 | B:186 |

Gambar 6. Citra RGB

2. Citra *Grayscale*

Citra *Grayscale* merupakan citra warna keabuan yang hanya memiliki satu nilai pada setiap pikselnya. Intensitas warna paling rendah dan paling tinggi pada citra *grayscale* diwakili dengan skala 0 (hitam) hingga 255 (putih).



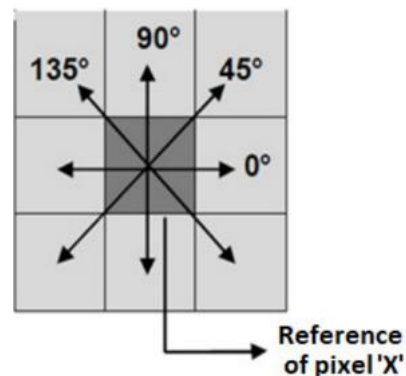
| | | |
|-----|-----|-----|
| 194 | 193 | 192 |
| 193 | 191 | 193 |
| 195 | 198 | 201 |

Gambar 7. Citra *Grayscale*

Pengolahan Citra Digital merupakan proses mengolah citra dan ditransformasikan menjadi citra lain secara digital menggunakan komputer. Pengolahan Citra Digital digunakan untuk menyamakan persepsi sistem visual manusia dengan komputer. Secara umum, teknik pengolahan citra digital yang banyak digunakan yaitu citra *grayscale*, citra biner, dan citra warna (RGB) (Jatmika & Purnamasari, 2014).

2.2.10 Gray Level Co-occurrence Matrix (GLCM)

Gray Level Co-occurrence Matrix merupakan metode ekstraksi fitur tekstur. GLCM merepresentasikan hubungan dua *pixel* bertetangga dengan intensitas warna keabuan dalam sudut dan jarak tertentu (Widyaningsih, 2017). Orientasi sudut berdasarkan arah 0° , 45° , 90° atau 135° .



Gambar 8. Arah sudut GLCM (Ben Atitallah *et al.*, 2018)

Untuk menentukan hasil ekstraksi fitur dilakukan perhitungan sebagai berikut:

- *Contrast*

$$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$$

- *Homogeneity*

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2}$$

- *Energy*

$$\sum_{i,j=0}^{N-1} P(i,j)^2$$

- *Correlation*

$$\sum_{i,j=0}^{N-1} P_{i,j} \left[\frac{(i - \mu_i)(i - \mu_j)}{\sigma_i \sigma_j} \right]$$

2.2.11 Python

Python merupakan bahasa pemrograman yang memiliki *source code* sederhana dan banyak digunakan dalam dunia pendidikan dan perusahaan besar terutama perusahaan yang membuat dan mengembangkan aplikasi (Retnoningsih & Pramudita, 2020). *Python* juga digunakan dalam bidang *data science* dan *machine learning*.

Flask adalah *web framework* dalam bahasa python yang berjenis *micro-framework*. *Flask* termasuk *micro-framework* dikarenakan tidak memerlukan tools atau library tertentu dalam penggunaannya. *Flask* menyediakan library untuk membangun website secara terstruktur (Irsyad, 2018). Untuk menginstall *flask* masukan perintah “\$ pip install *Flask*” pada text editor. *Flask* memiliki fitur Jinja Template Engine dan WSGI toolkit.

Penelitian ini menggunakan bahasa pemrograman *python* dikarenakan *python* memiliki algoritma dan *library* untuk

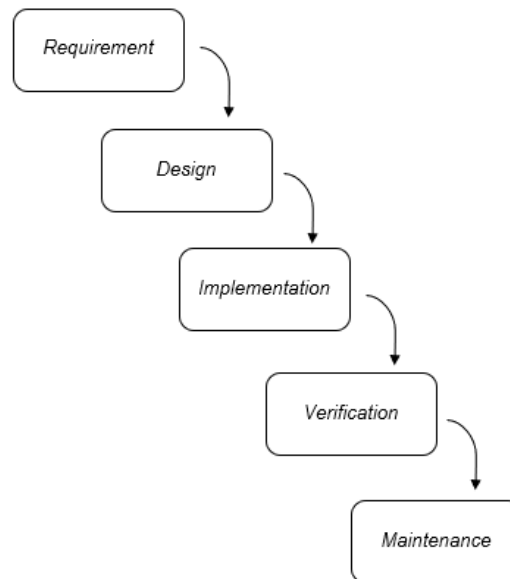
machine learning dan bersifat *opensource*. *Library* yang akan digunakan adalah sebagai berikut:

- *NumPy*
NumPy merupakan *library* yang digunakan untuk perhitungan *scientific* seperti aljabar, statistik, matriks, dan sebagainya.
- *Matplotlib*
Matplotlib merupakan *library* yang digunakan untuk memvisualisasikan data seperti membuat grafik, plot, histogram, dan sebagainya.
- *Scikit Learn*
Scikit learn merupakan *library* yang digunakan untuk *data mining* dan analisis data yang memanfaatkan algoritma *clustering*, *naïve bayes*, *regresi*, *classification*, dan sebagainya.
- *OpenCV*
OpenCV merupakan *library* yang digunakan untuk mengolah citra yang diambil oleh kamera lalu dikonversi ke digital melalui komputer. *OpenCV* digunakan untuk pengolahan citra dan *computer vision*.

2.2.12 Metode Waterfall

Metode *Waterfall* merupakan metode pengembangan sistem yang dilakukan secara berurutan. Proses metode ini mengalir ke bawah seperti air terjun. Pada metode *waterfall* proses yang dilakukan, yaitu menganalisis kebutuhan sistem (*requirement*) berdasarkan kebutuhan fungsional dan non fungsional. Perancangan sistem (*design*) untuk membentuk arsitektur sistem sesuai dengan kebutuhan sistem. *Implementation* merupakan proses realisasi *design* ke dalam

bentuk kode program. *Verification* merupakan proses pengujian sistem apakah sistem sudah sesuai dengan arsitektur pada tahap perancangan. *Maintenance* merupakan tahap perbaikan kesalahan pada sistem (Putri *et al.*, 2017). Tahapan metode *waterfall* ditunjukkan pada Gambar 9.



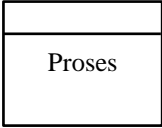
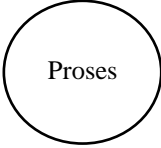


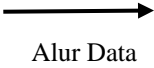
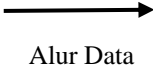
Gambar 9. Tahapan *waterfall* (Tristiyanto *et al.*, 2021)

2.2.13 Data Flow Diagram (DFD)

DFD merupakan diagram yang menggambarkan aliran data dari sebuah sistem. DFD dibuat untuk mengetahui *input* – proses – *output* pada data dalam suatu sistem (Soufitri, 2019). DFD memiliki empat simbol, yaitu ditunjukkan pada Tabel 2.

Tabel 2. Simbol DFD

| No. | Simbol | | Keterangan |
|-----|---------------|-------------------|--|
| | Gane & Sarson | Yourdon & DeMarco | |
| 1. | Terminator | Terminator | Terminator merupakan entitas eksternal dapat |

| | | | |
|----|---|--|---|
| | | | berupa orang atau organisasi, yang berinteraksi dengan sistem. |
| 2. |  |  | Proses menunjukkan kegiatan yang dilakukan oleh orang atau komputer terhadap aliran data. Proses harus menerima <i>input</i> dan menghasilkan <i>output</i> . |
| 3. |  |  | <i>Data Store</i> merupakan penyimpanan data secara komputerisasi maupun manual. <i>Data store</i> hanya berhubungan dengan proses. |
| 4. |  |  | Alur data menunjukkan arah data berupa <i>input</i> maupun <i>output</i> dari sebuah proses. Alur data diberi nama sesuai dengan data/informasi yang jelas. |

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian dilaksanakan di Lab. Komputasi Dasar Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung yang beralamat di Jalan Prof. Dr. Ir. Sumantri Brojonegoro No.1, Gedong Meneng, Kota Bandar Lampung. Penelitian dilakukan pada Januari-Juli 2023.

3.2 Sumber Data

Data dalam penelitian ini merupakan data sekunder yang diperoleh dari penelitian sebelumnya. Peneliti tersebut mengambil citra secara langsung menggunakan kamera *handphone*. Data yang digunakan berjumlah 900 data citra tanaman obat dari 3 jenis tanaman yaitu daun sirih hijau, daun binahong, dan daun keji beling.

3.3 Alat Pendukung

Alat pendukung yang digunakan dalam penelitian ini yaitu:

3.3.1 Perangkat Keras (*Hardware*)

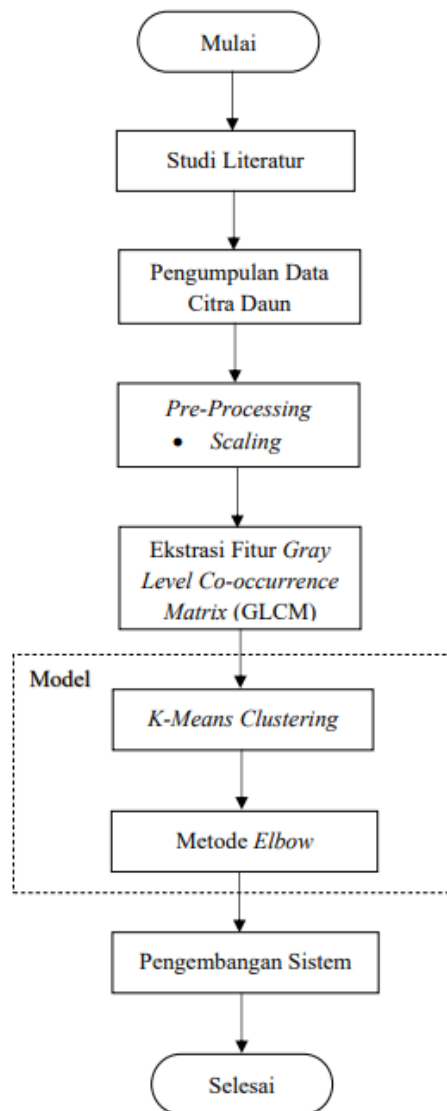
Perangkat keras yang digunakan dalam penelitian ini sebagai berikut:

- a. *Processor* : Intel(R) Core(TM) i5-8265U CPU @ 1.60GHz 1.80 GHz
- b. RAM : 4GB
- c. VGA : NVIDIA GeForce MX230

3.3.2 Perangkat Lunak (*Software*)

Perangkat Lunak yang digunakan dalam penelitian ini yaitu Sistem Operasi : Windows 11 Home Single Language 64-bit, Python 3, Jupyter Notebook, Flask, VSCode, Library (*Numpy, Matplotlib, Scikit Learn, OpenCV*).

3.4 Tahapan Penelitian



Gambar 10. Tahapan Penelitian

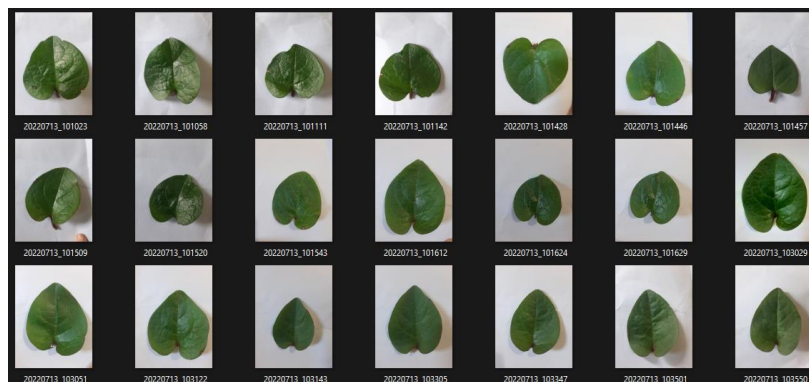
Berdasarkan Gambar 10. berikut penjelasan setiap tahapannya:

3.4.1 Studi Literatur

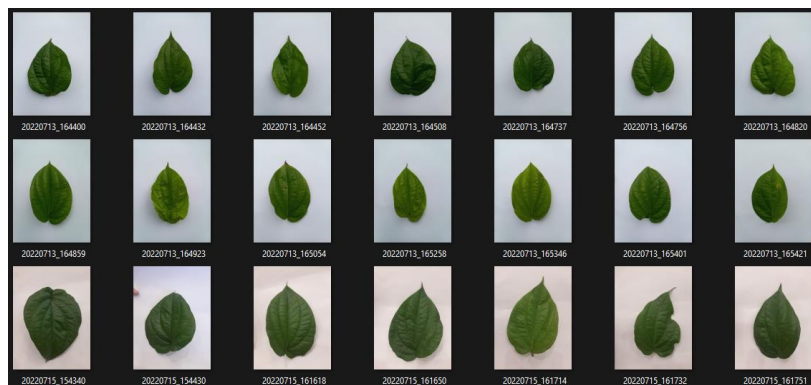
Tahap pertama yang dilakukan dalam penelitian ini adalah mencari dan mengumpulkan jurnal penelitian terdahulu, artikel ilmiah, dan buku yang berkaitan dengan penelitian ini sebagai referensi, pendukung penelitian dan informasi yang dibutuhkan.

3.4.2 Pengumpulan Data Citra Daun

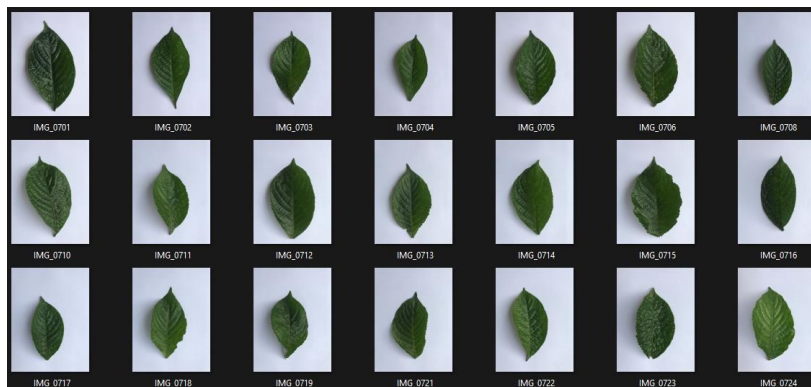
Pada penelitian ini, pengumpulan data diperoleh dari penelitian sebelumnya. Peneliti tersebut mengambil citra secara langsung menggunakan kamera *handphone*. Data yang akan digunakan berjumlah 900 data dari 3 jenis tanaman obat.



Gambar 11. Data Daun Binahong



Gambar 12. Data Daun Sirih Hijau



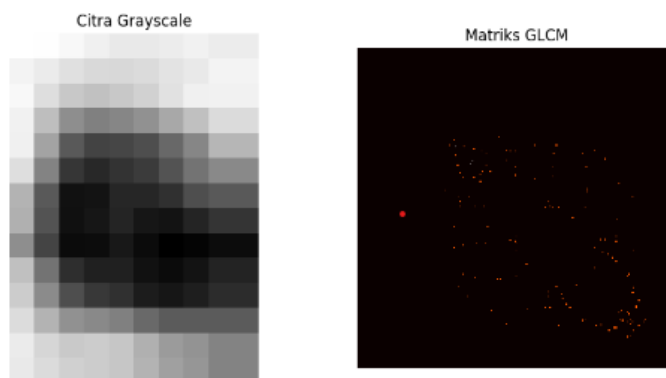
Gambar 13. Data Daun Keji Beling

3.4.3 *Pre-Processing*

Tahap *pre-processing* dilakukan *scaling* pada citra atau mengubah ukuran piksel menjadi 600 x 800 agar menghasilkan ukuran yang sama setiap data. *Pre-processing* berguna untuk memastikan kualitas data dan menghasilkan data yang lebih siap untuk diolah ke proses selanjutnya.

3.4.4 Ekstrasi Fitur

Dalam penelitian ini menggunakan ekstrasi fitur *Gray Level Co-occurrence Matrix (GLCM)* sebagai matrix pengambilan ciri tekstur dari *citra* dengan parameter *correlation*, *homogeneity*, *contrast*, dan *energy*. Arah sudut yang digunakan yaitu 0° , 45° , 90° atau 135° . Ekstrasi fitur GLCM ditunjukkan pada Gambar 14.



Gambar 14. GLCM

3.4.5 *K-Means Clustering*

Tahap ini, citra yang sudah diolah dan ekstraksi dilakukan pengelompokkan data menggunakan metode *K-Means*. Pada metode *K-Means* akan ditentukan terlebih dahulu jumlah *cluster* dan data sebagai pusat *cluster (centroid)* awal. Lalu, menghitung jarak tiap data dengan *centroid* menggunakan perhitungan *Euclidean Distance*. Data yang memiliki jarak minimum terhadap *centroid* akan dikelompokkan ke dalam sebuah *cluster*. Hitung *centroid* baru dan cek data hingga tidak ada lagi yang berpindah *cluster*. Untuk menentukan jumlah *cluster* optimal akan dilakukan perhitungan menggunakan metode *elbow*.

3.4.6 *Metode Elbow*

Metode *elbow* digunakan untuk menentukan jumlah *cluster* yang optimal. Jumlah *cluster* yang optimal ditentukan berdasarkan perbandingan nilai *SSE*. Semakin besar jumlah *cluster* maka semakin kecil nilai *SSE*. Hasil perbandingan nilai *SSE* dapat ditunjukkan menggunakan *plot* dan nilai optimal akan membentuk siku.

3.4.7 *Pengembangan Sistem*

Pengembangan sistem ini menggunakan metode *waterfall* dengan tahapan sebagai berikut:

3.4.7.1 *Requirement*

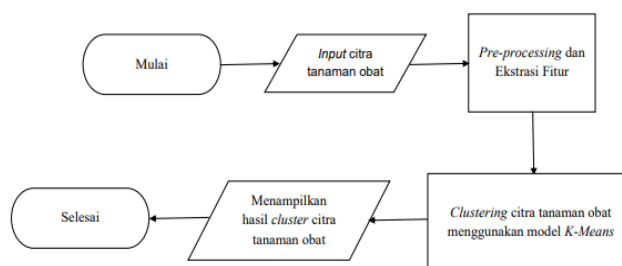
Tahap *requirement* dilakukan analisis kebutuhan fungsional dan non fungsional untuk pengembangan sistem. Kebutuhan fungsional pada sistem ini yaitu

pengguna dapat menginput data citra tanaman obat pada system. Sistem akan menentukan *cluster* citra yang diinputkan lalu memunculkan hasil *clustering* citra sesuai dengan *centroid* terdekat. Kebutuhan non fungsional pada sistem ini yaitu terdapat pada bagian alat pendukung.

3.4.7.2 Design

Design merupakan tahap perancangan sistem sesuai dengan kebutuhan yang telah ditentukan. Pada tahap ini dibagi menjadi 3 tahapan yaitu:

- *Flowchart Diagram*



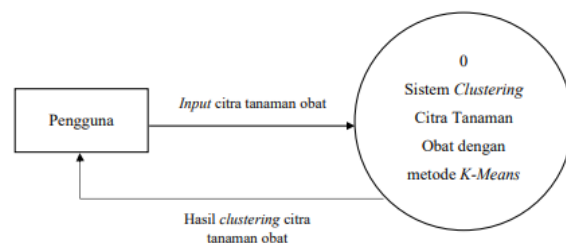
Gambar 15. *Flowchart* Sistem

Flowchart pada Gambar 15. menggambarkan alur kerja sistem. Berikut penjelasan dari *flowchart* di atas:

- 1) Pengguna menyiapkan citra tanaman obat yang akan digunakan dalam penelitian.
- 2) Pengguna menginput citra tanaman obat pada sistem.
- 3) Sistem melakukan *pre-processing*. Warna citra RGB ke *grayscale*. Melakukan ekstrasi fitur menggunakan GLCM.
- 4) Sistem melakukan *clustering* citra tanaman obat menggunakan model *K-Means*.

- 5) Sistem menampilkan hasil *cluster* citra tanaman obat.
- *Data Flow Diagram* (DFD)
 1. *Context Diagram*

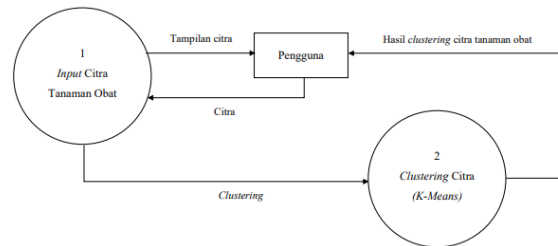
Context diagram menggambarkan sistem secara umum. Tentukan entitas luar yang terlibat yaitu Pengguna. Selanjutnya, tentukan *input* dan *output* untuk entitas luar. Berikut gambaran *context diagram* ditunjukkan pada Gambar 16.



Gambar 16. *Context Diagram*

2. DFD Level 0

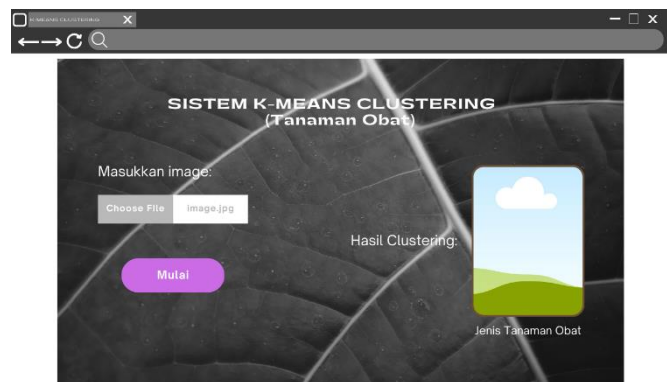
DFD level 0 menggambarkan sistem secara terperinci. Prosesnya yaitu pada proses 1 untuk *input* citra tanaman obat terdapat alur data citra dari pengguna ke sistem kemudian sistem menampilkan citra ke pengguna. Pada proses 2 untuk *clustering* citra menggunakan metode *K-Means* terdapat alur data dari proses 1 untuk *diclustering* ke sistem kemudian sistem mengeluarkan hasil *clustering* citra tanaman obat ke pengguna. DFD level 0 ditunjukkan pada Gambar 17.



Gambar 17. DFD Level 0

- *Design Interface*

Design interface ini berbentuk sederhana, hanya terdapat kolom *input* citra, *button* mulai, tampilan citra yang diinput, dan hasil *clustering*. *Design interface* ditunjukkan pada Gambar 18.

Gambar 18. *Design Interface*

3.4.7.3 *Implementation*

Implementation merupakan tahap realisasi *design* ke dalam bentuk kode program untuk membuat sistem *K-Means Clustering* Tanaman Obat berbasis *web* menggunakan *framework flask*.

3.4.7.4 *Verification*

Verification merupakan tahap pemeriksaan sistem apakah fungsi-fungsi sistem telah sesuai dengan kebutuhan yang ditetapkan.

3.4.7.5 *Maintenance*

Maintenance merupakan tahap pemeliharaan dan perbaikan sistem jika terdapat kesalahan dalam jangka waktu tertentu. *Maintenance* dilakukan setelah sistem di*hosting*.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian *Clustering* Citra Tanaman Obat menggunakan Metode *K-Means* ini dapat ditarik kesimpulan sebagai berikut:

1. Metode *K-Means Clustering* untuk mengcluster tanaman obat berhasil dilakukan. Namun, belum cukup baik dikarenakan penentuan *centroid* di awal berpengaruh pada hasil *cluster*.
2. Sistem *K-Means Clustering* Tanaman Obat berhasil diimplementasikan dengan tampilan hanya 1 halaman *web*. Batasan sistem hanya berlaku untuk citra daun sirih hijau, daun binahong, dan daun keji beling.

5.2 Saran

Saran yang diberikan pada penelitian ini yaitu:

1. Menambahkan data citra daun agar mendapatkan hasil *clustering* yang lebih baik.
2. Menambahkan ekstraksi fitur lainnya seperti HSV, RGB, dan Deteksi Tepi agar hasil *clustering* lebih baik.
3. Mengembangkan sistem dengan menambahkan berbagai fitur agar lebih interaktif dan menarik.

DAFTAR PUSTAKA

- Apriyanti, nur ridha, Nugroho, radityo adi, & Soesanto, O. (2016). Algoritma K-Means Clustering Dalam Pengolahan Citra Digital Landsat. *KLIK Kumpulan Jurnal Ilmu Komputer*, 2(2), 110–122.
- Astuti, S. I., Arso, S. P., & Wigati, P. A. (2018). Perbandingan Algoritma K-Means Dan Algoritma K-Medoids Dalam Pengelompokan Komoditas Tanaman Biofarmaka Di Provinsi Jawa Tengah Tahun 2018. In *Analisis Standar Pelayanan Minimal Pada Instalasi Rawat Jalan di RSUD Kota Semarang* (Vol. 3).
- Atina, A. (2017). Segmentasi Citra Paru Menggunakan Metode k-Means Clustering. *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)*, 3(2), 57. <https://doi.org/10.25273/jpfk.v3i2.1475>
- Ben Atitallah, M. A., Kachouri, R., Kammoun, M., & Mnif, H. (2018). An efficient implementation of GLCM algorithm in FPGA. *2018 International Conference on Internet of Things, Embedded Systems and Communications, IINTEC 2018 - Proceedings*, 147–152. <https://doi.org/10.1109/IINTEC.2018.8695275>
- Darnita, Y., Toyib, R., & Kurniawan, Y. (2020). Penerapan Metode K-Means Clustering Pada Aplikasi Android Pada Tanaman Obat Herbal. *Pseudocode*, 7(2), 105–114. <https://doi.org/10.33369/pseudocode.7.2.18-27>
- Desmanto, S., & Angreni, R. (2014). Penerapan Algoritma K-Means Clustering Untuk Pengelompokkan Citra Digital Dengan Ekstraksi Fitur Warna RGB. *Julyxxxx, x, No.x(x)*, 1–5.
- Han, J., Kamber, M., & Pei, J. (2012). Third Edition : Data Mining Concepts and Techniques. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). <http://library.books24x7.com/toc.aspx?bkid=44712>
- Jatmika, S., & Purnamasari, D. (2014). Rancang Bangun Alat Pendeteksi Kematangan Buah Apel Dengan Menggunakan Metode Image Processing Berdasarkan Komposisi Warna. *Jurnal Ilmiah Teknologi Informasi Asia*, 8(1),

51–58.

- Kusuma, A. W., & Ellyana, R. L. (2018). Penerapan Citra Terkompresi Pada Segmentasi Citra Menggunakan Algoritma K-Means. *Jurnal Terapan Teknologi Informasi*, 2(1), 65–74. <https://doi.org/10.21460/jutei.2018.21.65>
- Lestari, A. D. (2018). *Analisis Multivariat Clustering K-Means Pada Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indeks Pembangunan Manusia Tahun 2017 Dengan Bantuan Software SPSS*. <http://lib.unnes.ac.id/36913/>
- Marhaenanto, B., Soediby, D. W., & Farid, M. (2015). Penentuan lama Sangrai Kopi Berdasarkan Variasi Derajat Sangrai Menggunakan Model Warna Rgb Pada Pengolahan Citra Digital (Digital Image Processing). *Jurnal Agroteknologi*, 09(02), 102–111. <https://jurnal.unej.ac.id/index.php/JAGT/article/view/3536>
- Muningsih, E., & Kiswati, S. (2018). Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>
- Noventi, W. R.-4272-2-P. pdfa., & Carolia, N. (2016). Potensi Ekstrak Daun Sirih Hijau (Piper betle L .) sebagai Alternatif Terapi Acne vulgaris The Potential of Green Sirih Leaf (Piper betle L .) for Alternative Therapy Acne vulgaris. *Studi Pendidikan Dokter Fakultas Kedokteran Universitas Lampung, Vol. 5(1)*, Hal. 140.
- Nurhidayah, K., Fadraersada, J., & Rijai, L. (2015). *POTENSI EKSTRAK DAUN KEJI BELING (Strobilanthes crispus) SEBAGAI PENURUN KADAR GLUKOSA DARAH: UJI IN VIVO PADA TIKUS PUTIH (Rattus norvegicus)*. 43–49. <https://doi.org/10.25026/mpc.v2i1.38>
- Putri, C. N., Sadariawati, R., Si, M., Ganiardi, M. A., & Si, S. (2017). *Implementasi Algoritma K-Means Untuk Clustering Data Penjualan Pada Cv . Alfa Fresh*.
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biict.v7i2.1422>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning

- dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Sarno, S. (2019). Pemanfaatan Tanaman Obat (Biofarmaka) Sebagai Produk Unggulan Masyarakat Desa Depok Banjarnegara. *Abdimas Unwahas*, 4(2), 73–78. <https://doi.org/10.31942/abd.v4i2.3007>
- Soufitri, F. (2019). Perancangan Data Flow Diagram Untuk Sistem Informasi Sekolah (Studi Kasus Pada Smp Plus Terpadu). *Ready Star*, 2(1), 240–246.
- Tristiyanto, T., Hermanto, B., & Rachmadi Kartawiria, A. (2021). Aplikasi Pencarian Letak Gudang Koperasi Dengan Metode K-Means Cluster Menggunakan Framework Laravel. *Jurnal Pepadun*, 2(3), 319–326. <https://doi.org/10.23960/pepadun.v2i3.89>
- Widodo, R., Widodo, A. W., & Supriyanto, A. (2018). Pemanfaatan Ciri Gray Level Co-Occurrence Matrix (GLCM) Citra Buah Jeruk Keprok (*Citrus reticulata* Blanco) untuk Klasifikasi Mutu. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(11), 5769–5776. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3420>
- Widyaningsih, M. (2017). Identifikasi Kematangan Buah Apel Dengan Gray Level Co-Occurrence Matrix (GLCM). *Jurnal SAINTEKOM*, 6(1), 71. <https://doi.org/10.33020/saintekom.v6i1.7>