

**IMPLEMENTASI *SUPPORT VECTOR MACHINE* (SVM) UNTUK
PENGELOMPOKAN DAUN TUMBUHAN OBAT BINAHONG
(*ANREDERA CORDIFOLIA*) DAN SIRIH HIJAU (*PIPER BETLE*)**

(SKRIPSI)

Oleh

**AKBAR SETIA GEMA MUSLIMIN
NPM 1617051073**



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

**IMPLEMENTASI *SUPPORT VECTOR MACHINE* (SVM) UNTUK
PENGELOMPOKAN DAUN TUMBUHAN OBAT BINAHONG
(*ANREDERA CORDIFOLIA*) DAN SIRIH HIJAU (*PIPER BETLE*)**

Oleh

AKBAR SETIA GEMA MUSLIMIN

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER

Pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam



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

ABSTRAK

IMPLEMENTASI *SUPPORT VECTOR MACHINE* (SVM) UNTUK PENGELOMPOKAN DAUN TUMBUHAN OBAT BINAHONG (*ANREDERA CORDIFOLIA*) DAN SIRIH HIJAU (*PIPER BETLE*)

Oleh

Akbar Setia Gema Muslimin

Tumbuhan obat merupakan tumbuhan yang memiliki khasiat untuk mengatasi berbagai penyakit. Klasifikasi citra digital tumbuhan obat dapat dilakukan dengan menggunakan beberapa algoritma, salah satu nya adalah algoritma *Support Vector Machine* (SVM). Klasifikasi menggunakan algoritma ini dapat dilakukan dengan mengekstrak beberapa fitur diantaranya, warna, tekstur dan bentuk. Hasil kinerja klasifikasi pada penelitian ini akan dibandingkan dengan algoritma *Convolutional Neural Network* (CNN). Data yang digunakan dalam penelitian ini merupakan data primer sebanyak 450 citra binahong dan 450 citra sirih. Data terbagi menjadi data latih dan data uji, perbandingan data latih dan data uji yang digunakan yaitu , 70:30, 75:25, dan 80:20 dengan kombinasi jarak piksel 1,2,3,4 dan 5. Hasil terbaik didapatkan pada perbandingan 75:25 dengan nilai akurasi 92.44%.

Kata Kunci : Tumbuhan obat, Klasifikasi, *Support Vector Machine* (SVM).

Abstract

THE IMPLEMENTATION OF SUPPORT VECTOR MACHINE (SVM) FOR CLASSIFICATION OF THE LEAVES OF MEDICINAL PLANTS BINAHONG (ANREDERA CORDIFOLIA) AND SIRIH HIJAU (PIPER BETLE).

by

Akbar Setia Gema Muslimin

Medicinal plants are plants that have the properties to cure various diseases. One of the easily found medicinal plants in Indonesia is the binahong and betel plants. Binahong and betel plants are known to have useful properties for society. The properties possessed by these two plants are different, so it is important to be able to differentiate between the two leaves. Computers can differentiate between the two leaves using machine learning, known as classification. Classification can be performed by taking digital images of the medicinal plants. Digital image classification can be done using several algorithms, one of which is the Support Vector Machine (SVM) algorithm. Classification using this algorithm can be done by extracting some features such as color, texture, and shape. The performance of the classification results in this study will be compared with the Convolutional Neural Network (CNN) algorithm. The data used in this study is primary data consisting of 450 binahong and 450 betel images. The features that will be extracted from the data will go through preprocessing stages, such as resizing, labeling, and segmentation. The data is divided into training data and testing data, the comparison of training data and testing data used is 70:30, 75:25, and 80:20 with a combination of pixel distances 1,2,3,4, and 5. The best result was obtained in the comparison of 75:25 with an accuracy of 92.44%..

Keywords: Medicinal plants; Classification; Support Vector Machine (SVM).

Judul Skripsi : **IMPLEMENTASI SUPPORT VECTOR MACHINE (SVM) UNTUK PENGELOMPOKAN DAUN TUMBUHAN OBAT BINAHONG (ANREDERA CORDIFOLIA) DAN SIRIH HIJAU (PIPER BETLE)**

Nama : Akbar Setia Gema Muslimin

Nomor Pokok Mahasiswa : 1617051073

Program Studi : Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam



1. **Komisi Pembimbing,**

Rizky Prabowo, M.Kom.
NIP. 19880807 201903 1 011

Mengetahui,

2. **Ketua Jurusan Ilmu Komputer,**

Didik Kurniawan, S.Si., MT.
NIP. 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : Rizky Prabowo, M.Kom.



Penguji Bukan Pembimbing : Prof. Admi Syarif, Ph.D.



Penguji Bukan Pembimbing : Bambang Hermanto, S.Kom., M.Cs.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Hori Satria, S.Si., M.Si.
NIP. 197110012005011002


Tanggal Lulus Ujian Skripsi : 9 Februari 2023

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul "**IMPLEMENTASI *SUPPORT VECTOR MACHINE (SVM)* UNTUK PENGELOMPOKAN DAUN TUMBUHAN OBAT BINAHONG (*ANREDERA CORDIFOLIA*) DAN SIRIH HIJAU (*PIPER BETLE*)**" ini merupakan hasil karya sendiri dan bukan hasil karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar akademik yang telah saya terima.

Bandar Lampung, 9 Februari 2023




Akbar Setia Gema Muslimin
1617051073

RIWAYAT HIDUP



Penulis dilahirkan pada 7 Agustus 1998 di Tanjung Enim Kabupaten Muara Enim Provinsi Sumatera Selatan, merupakan putra pertama dari pasangan Bapak Subarkah dan Ibu Wahyu Kurniati. Penulis menyelesaikan pendidikan formal pertamanya di Taman Kanak-Kanak (TK) Darul Akhyar pada tahun 2004, melanjutkan sekolah dasar di SDIT Syifa Fikriya Kabupaten Serang dan selesai pada tahun 2010. Kemudian, penulis melanjutkan pendidikan sekolah menengah pertama di SMPIT Al-Izzah Kota Serang yang diselesaikan pada tahun 2013, serta menyelesaikan pendidikan sekolah menengah kejuruan pada tahun 2016 di MAN 2 Kabupaten Tangerang dengan Jurusan IPA.

Pada tahun 2016, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung dengan jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN). Selama menjadi mahasiswa, penulis mengikuti beberapa kegiatan antara lain:

- 1 Anggota Abacus Himpunan Mahasiswa Jurusan Ilmu Komputer periode 2016/2017.
- 2 Anggota muda BEM Universitas Lampung periode 2016/2017 .
- 3 Bulan Januari 2017 penulis melaksanakan karya wisata ilmiah di Desa Margosari, Kecamatan Pagelaran Utara, Kabupaten Pringsewu.
- 4 Ketua Kaderisasi ROIS FMIPA Unila tahun 2018.
- 5 Ketua Komisi III DPM FMIPA Unila tahun 2019.
- 6 Bulan Januari 2019 penulis melaksanakan kerja praktik di PT. Indonesia Connets Plus (ICON+) di Palembang Sumatera Selatan.
- 7 Bulan Januari 2020 penulis melaksanakan KKN di Desa Tambah Subur, Kecamatan Way Bungur, Kabupaten Lampung Timur.

MOTTO

“Selesaikan satu urusan, lalu pindah ke suatu urusan lain”

SANWANCANA

Puji syukur kehadiran Allah Subhanahu wa ta'ala yang telah melimpahkan rahmat, hidayah serta inayah-Nya sehingga penulis dapat menyelesaikan skripsi ini berjudul “Implementasi *Support Vector Machine* (SVM) untuk Pengelompokan Daun Tumbuhan Obat Binahong (*anredera cordifolia*) dan Sirih Hijau (*piper betle*)”. Tidak lupa shalawat serta salam kepada Nabi Muhammad Shallahu ‘alaihi wasallam, yang kita nanti-nantikan syafaatnya di yaumul akhir kelak.

Penulis mengucapkan terimakasih kepada semua pihak yang telah membantu dan memiliki peran besar dalam penyusunan skripsi ini, yaitu:

- 1 Kedua orang tuaku tercinta, ayahku Subarkah dan ibuku Wahyu Kurniati yang selalu memberikan semua dukungan, kasih sayang dan segala doa yang tiada hentinya.
- 2 Kedua adikku yang selalu mendukung dan membantu kakakmu dalam menyelesaikan skripsi.
- 3 Istri tercinta yang selalu support dalam proses menyelesaikan skripsi.
- 4 Bapak Rizky Prabowo, M.Kom. selaku dosen pembimbing utama atas kesediaannya dan kesabarannya untuk memberikan dukungan, bimbingan, kritik, dan saran dalam proses penyelesaian skripsi.
- 5 Bapak Prof. Admi Syarif, Ph.D., selaku pembahas I skripsi atas bimbingan, arahan serta kritik yang membangun dalam proses skripsi.
- 6 Bapak Bambang Hermanto, S.Kom., M.Kom., selaku pembahas II skripsi atas bimbingan, arahan dan kritik yang membangun dalam proses skripsi.
- 7 Bapak Didik Kurniawan, S.Si., MT., selaku Ketua Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
- 8 Bapak Dekan Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan FMIPA Universitas Lampung.

- 9 Bapak Dr. rer. nat. Akmal Junaidi, M.Sc., selaku Sekertaris Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
- 10 Bapak dan Ibu Dosen Jurusan Ilmu Komputer Universitas Lampung yang telah memberikan ilmu dan pengetahuan hidup selama penulis menjadi mahasiswa.
- 11 Agosto dan Yudi, *you are the apple of my eyes*.
- 12 Teman-teman angkatan 2016 Ilmu Komputer.
- 13 Teman-teman kontrakan d3/14 GGMI terimakasih unutk waktu 4 tahun bersama dalam satu atap.
- 14 Teman-teman QH yang senantiasa support diakhir masa studi.
- 15 Semua pihak yang secara langsung maupun tidak langsung yang telah membantu menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, akan tetapi semoga skripsi ini dapat membawa manfaat dan keberkahan bagi perkembangan ilmu pengetahuan terutama bagi semua civitas Ilmu Komputer Universitas Lampung.

Bandar Lampung, 9 Februari 2023

Akbar Setia Gema Muslimin

1617051073

DAFTAR ISI

	Halaman
DAFTAR ISI	ii
DAFTAR TABEL	v
DAFTAR GAMBAR	vi
DAFTAR <i>PSEUDOCODE</i>	viii
I. PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
II. TINJAUAN PUSTAKA	
2.1 Penelitian Terdahulu.....	4
2.2 Tumbuhan Obat.....	7
2.3 Binahong.....	8
2.4 Sirih.....	8
2.5 <i>Machine Learning</i>	9
2.6 <i>Support Vector Machine</i>	10
2.6.1 <i>Support Vector Machine Linear</i>	11
2.6.2 <i>Support Vector Machine Non-Linear</i>	14
2.7 <i>Sequential Training SVM</i>	16

2.8 <i>Cross Validation</i>	18
2.9 Nilai Evaluasi.....	19
III. METODOLOGI PENELITIAN	
3.1 Tempat dan Waktu Penelitian.....	22
3.2 Data dan Alat.....	23
3.2.1 Data.....	23
3.2.2 Alat.....	24
3.3 Metode.....	25
3.3.1 Studi Literatur.....	26
3.3.2 Pengumpulan Data.....	27
3.3.3 <i>Preprocessing</i>	27
3.3.4 Ekstraksi Fitur.....	28
3.3.5 Membagi Dataset.....	30
3.3.6 Model kernel klasifikasi SVM.....	30
3.3.7 <i>Cross Validation</i>	30
3.3.8 Hasil.....	31
3.3.9 Evaluasi.....	31
IV. HASIL DAN PEMBAHASAN	
4.1 <i>Preprocessing</i>	32
4.1.1 <i>Resize</i>	32
4.1.2 <i>Labeling</i>	33
4.1.3 Segmentasi.....	34
4.2 Ekstraksi Fitur.....	37
4.2.1 Warna.....	37
4.2.2 Bentuk.....	40

4.2.3 Tekstur.....	41
4.3 <i>Eksport Data ke Spreadsheet</i>	42
4.4 Pembagian Data.....	42
4.5 Panggil <i>Dataset</i>	43
4.6 Model Kernel SVM.....	43
4.7 <i>Cross Validation</i>	43
4.8 Hasil.....	44
4.9 Evaluasi.....	45
4.10 Pembahasan.....	45
4.10.1 Pemodelan 1.....	46
4.10.2 Pemodelan 2.....	49
4.10.3 Pemodelan 3.....	53
4.11 Perbandingan Model.....	56
4.12 Perbandingan dengan Penelitian Lain.....	57
4.13 <i>Deploymen of Machine Learning</i>	57
V. KESIMPULAN	
5.1 Kesimpulan.....	59
5.2 Saran.....	59
DAFTAR PUSTAKA	

DAFTAR TABEL

	Halaman
Tabel 1. Penelitian Terdahulu.....	4
Tabel 2. <i>confusion matrix</i>	20
Tabel 3. <i>timeline</i> penelitian.....	22
Tabel 4. Formula pada HSV.....	29
Tabel 5. Pengujian dengan <i>K-Fold Cross Validation</i> pada Pemodelan 1.....	46
Tabel 6. Pengujian dengan <i>Confusion Matrix</i> pada Pemodelan 1.....	47
Tabel 7. Hasil Evaluasi Kinerja Klasifikasi pada Pemodelan 1.....	48
Tabel 8. Pengujian dengan <i>K Fold Cross Validation</i> pada Pemodelan 2.....	49
Tabel 9. Pengujian dengan <i>Confusion Matrix</i> pada Pemodelan 2.....	50
Tabel 10. Hasil Evaluasi Kinerja Klasifikasi pada Pemodelan 2.....	52
Tabel 11. Pengujian dengan <i>K-Fold Cross Validation</i> pada Pemodelan 3.....	53
Tabel 12. Pengujian dengan <i>Confusion Matrix</i> pada Pemodelan 3.....	54
Tabel 13. Hasil Evaluasi Kinerja Klasifikasi pada Pemodelan 3.....	55
Tabel 14. Hasil Perbandingan Kedua Metode.....	57

DAFTAR GAMBAR

	Halaman
Gambar 1. Binahong.....	8
Gambar 2. Sirih.....	9
Gambar 3. Konsep <i>supervised learning</i>	10
Gambar 4. Konsep SVM.....	11
Gambar 5. SVM <i>Non Linear</i>	15
Gambar 6. <i>K-Fold Cross Validation</i>	19
Gambar 7. Data Daun Sirih.....	23
Gambar 8. Data Daun Binahong.....	23
Gambar 9. Alur Penelitian.....	26
Gambar 10. Proses <i>Resize</i> Citra.....	33
Gambar 11. Hasil <i>Resize</i> Citra.....	33
Gambar 12. Citra <i>Grayscale</i>	35
Gambar 13. Citra <i>Thresholding</i>	35
Gambar 14. Proses Dilasi.....	36
Gambar 15. Proses Erosi.....	36
Gambar 16. Citra <i>Cropping</i>	37
Gambar 17. Citra RGB	38
Gambar 18. Citra HSV.....	38
Gambar 19. Warna Dominan Citra RGB.....	39
Gambar 20. Warna Dominan Citra HSV.....	39

Gambar 21. Nilai Fitur HSV.....	40
Gambar 22. Nilai Fitur <i>Eccentricity</i> dan <i>Metric</i>	41
Gambar 23. Nilai Fitur GLCM.....	42
Gambar 24. <i>Plot</i> Nilai Hasil.....	45
Gambar 25. Tampilan <i>Deployment of Machine Learning</i>	58

DAFTAR *PSEUDOCODE*

	Halaman
<i>Pseudocode 1. Kode Program labeling.....</i>	34
<i>Pseudocode 2. Kode Program Preprocessing Segmentasi.....</i>	34
<i>Pseudocode 3. Kode Program Ekstraksi Fitur Warna.....</i>	38
<i>Pseudocode 4. Kode Program Ekstraksi Fitur Bentuk.....</i>	40
<i>Pseudocode 5. Kode Program Ekstraksi Fitur Tekstur.....</i>	42
<i>Pseudocode 6. Kode Program Ekspor Data Ke spreadsheet.....</i>	42
<i>Pseudocode 7. Kode Program Panggil Dataset ke Jupyter Notebook.....</i>	43
<i>Pseudocode 8. Kode Program Model Kernel SVM.....</i>	43
<i>Pseudocode 9. Kode Program Cross Validation.....</i>	44
<i>Pseudocode 10. Kode Program Hasil.....</i>	44
<i>Pseudocode 11. Kode Program Evaluasi Confusion Matrix.....</i>	45

I. PENDAHULUAN

1.1 Latar Belakang

Penggunaan obat tradisional semakin berkembang pesat bersamaan dengan maraknya semboyan *back to nature*, hal ini dibuktikan dengan industri farmasi yang memproduksi obat tradisional juga semakin meningkat. Tumbuhan obat di Indonesia berperan besar terhadap pembuatan obat-obatan di dunia. Amerika Serikat menghasilkan 45 obat penting dari tumbuhan tropis, 14 diantaranya berasal dari Indonesia. Melimpahnya tumbuhan di Indonesia telah lama dimanfaatkan oleh masyarakat untuk kebutuhan sehari-hari dan meningkatkan kualitas hidup. Sekitar 9.606 jenis tumbuhan memiliki kandungan yang dapat dimanfaatkan sebagai kebutuhan obat (Rizki & Leilani, 2017).

Obat tradisional terbukti memiliki beberapa kelebihan diantaranya, dalam satu racikan menghasilkan efek yang saling mendukung, memiliki sedikit efek samping bahkan sampai tidak menimbulkan efek samping sama sekali, dan sebagian besar bahan kimia yang terdapat pada tanaman obat mampu dimetabolisme dengan baik (Cahyaningrum & Putri, 2017). Penggunaan tanaman obat dapat dikonsumsi dengan beberapa cara yaitu diminum, ditempel, dan dihirup sehingga memenuhi konsep bagaimana reseptor sel bekerja saat menerima bahan kimia. Indonesia merupakan kawasan tropis dimana tanaman obat bisa tumbuh dengan mudah, diantaranya tanaman yang mudah dijumpai adalah binahong dan sirih. Binahong (*anredera cordifolia*) sering digunakan masyarakat untuk mengobati luka dengan cara menempelkan daun binahong yang telah ditumbuk pada luka (Siregar dkk., 2020).

Menurut Ariani dkk., (2013), sebagai obat luka daun binahong memiliki kandungan bahan kimia seperti asam linoleat, asam askorbat, saponin, flavonoid, dan protein. Tumbuhan obat sirih (*piper betle* linn) dikenal memiliki kemampuan

anti bakteri yang kuat, beberapa kandungan kimia ada pada tumbuhan ini yaitu minyak atsiri, terpenoid, polifenol, steroid, dan tanin. Sirih biasa digunakan pada bagian daunnya untuk merawat gigi, keputihan wanita, mimisan, sariawan, gatal-gatal, bau badan dan bau mulut (Sundari & Almasyhuri, 2019).

Daun binahong dan daun sirih memiliki kemiripan dari segi bentuk seperti hati berujung runcing, dan berwarna hijau. Binahong dan sirih memiliki khasiat yang berbeda walaupun memiliki bentuk daun yang sekilas mirip, maka penting untuk mengelompokkan tanaman binahong dan sirih. Pengelompokan daun menggunakan bantuan teknologi komputer dikenal dengan metode pengolahan citra digital, hal ini bisa dilakukan berdasarkan struktur daun. Bidang pengelompokan citra digital beberapa algoritma dapat digunakan, diantaranya *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Convolutional Neural Network* (CNN) (Maulana & Rochmawati, 2019).

Salah satu metode klasifikasi pengolahan citra digital yaitu *Support Vector Machine* (SVM), metode ini meningkatkan akurasi pada pengolahan pola dan keunggulan lain dari metode ini adalah proses pembelajaran yang cepat (Budianto dkk., 2018). Penelitian yang sudah pernah dilakukan oleh Noprianto dkk., (2017), terhadap 21 kendaraan dengan jarak 1, 3, dan 5 meter menggunakan metode *Support Vector Machine* (SVM). Nilai akurasi yang didapat pada jarak 1 meter akurasi 89.77%, pada jarak 3 meter akurasi 82.86% ,dan pada jarak 5 meter akurasi 65.22%. Penelitian lain juga dilakukan oleh (Sholehudin, 2016), dalam mendeteksi penyakit pada tumbuhan apel berdasarkan citra daun dengan menggunakan metode *Support Vector Machine* (SVM) hasil akurasi sebesar 90%.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat diambil perumusan masalah pada penelitian ini sebagai berikut:

1. Bagaimana mengklasifikasikan citra digital daun binahong dan sirih serta melihat hasil akurasi nya?

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah:

1. Data yang digunakan merupakan citra digital (*image*).
2. Ukuran pada data citra digital 600 x 800 piksel.
3. Kernel yang digunakan yaitu, *linear*, *polynomial*, *rbf*, dan *sigmoid*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan metode *Support Vector Machine* (SVM) untuk mengelompokkan tumbuhan obat binahong dan sirih.
2. Melihat kernel dan fitur terbaik pada metode *Support Vector Machine* (SVM) untuk mengelompokkan tumbuhan obat binahong dan sirih.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dapat membantu peneliti lainnya yang menggunakan metode klasifikasi yang serupa.
2. Model klasifikasi pada penelitian ini dapat digunakan untuk membuat sistem prediksi (*deployment of machine learning*).

II. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini membutuhkan beberapa penelitian sebelumnya yang berkaitan dengan penelitian ini, guna membandingkan hasil akurasi pada penelitian tersebut. Berikut adalah ringkasan dari penelitian terdahulu yang pernah dilakukan.

Tabel 1. Penelitian Terdahulu

Peneliti	Judul	Data	Metode	Hasil
(Budianto dkk., 2018)	Perbandingan <i>K-Nearest Neighbor</i> (Knn) dan <i>Support Vector Machine</i> (SVM) dalam Pengenalan Karakter Plat Kendaraan Bermotor.	Terdapat dua data yaitu data <i>training</i> dan <i>testing</i> , data latih terdiri dari 540 citra, dan data uji sebanyak 20 citra.	<i>K-Nearest Neighbor</i> (KNN) dan <i>Support Vector Machine</i> (SVM).	Hasil akurasi sebesar 95%.
(Wiradinata dkk., 2019)	Klasifikasi 12 Motif Batik Banten Menggunakan <i>Support Vector Machine</i> (SVM).	Data latih sebanyak 120 citra dan data uji 60 citra dari 12 motif batik Banten.	<i>Support Vector Machine</i> (SVM).	Akurasi tertinggi pada ordo 2 RBF 0,2 sebesar 96,7% .

(Safitri & Wulanningrum, 2020)	Aplikasi Pengenalan Pola Tulisan Tangan Menggunakan Metode <i>Support Vector Machine</i> (SVM).	Data sebanyak 50 sampel tulisan dari 10 orang yang berbeda.	<i>Support Vector Machine</i> (SVM).	Hasil akurasi terbaik diperoleh pada skenario uji 2 dengan tingkat akurasi sebesar 85%.
(Dino & Abdulrazzaq, 2019)	<i>Facial Expression Classification Based on SVM, KNN and MLP Classifiers.</i>	Menggunakan dataset CK+ pada 210 wajah orang dewasa, total 634 citra dengan 8 label ekspresi.	SVM, K-NN dan MLP.	Hasil akurasi terbaik diperoleh SVM sebesar 93.53%.
(Noprianto dkk., 2017)	<i>Long Distance Automatic Number Plate Recogniting under Perspective Distortion using Zonal Density and Support Vector Machine.</i>	21 kendaraan yang diambil citranya masing-masing pada jarak 1,3, dan 5 meter.	<i>Support Vector Machine</i> (SVM).	Hasil akurasi terbaik didapatkan pada jarak 1 meter sebesar 89.77%.

(Vijayarajeswari dkk., 2019)	<i>Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform</i>	Terdapat 95 gambar <i>mammogram</i> yang digunakan dalam penelitian ini.	<i>Support Vector Machine</i> (SVM).	Hasil akurasi didapatkan sebesar 94%.
------------------------------	--	--	--------------------------------------	---------------------------------------

Penelitian yang terdapat pada Tabel 1 dilakukan oleh (Budianto dkk., 2018), membahas tentang perbandingan metode *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM) pada pengenalan pola plat kendaraan bermotor. Terdapat dua data yang digunakan dalam penelitian yaitu data *training* dan *testing*. Data *training* terdiri dari citra A-Z dan 0-9, masing-masing memiliki 15 citra. Total data *training* sebanyak 540 citra, sedangkan data *testing* sebanyak 20 citra. Penelitian ini menghasilkan akurasi sebesar 95%.

Penelitian kedua dilakukan oleh (Wiryadinata dkk., 2019), penelitian ini bertujuan untuk mengklasifikasikan 12 batik motif Banten menggunakan *Support Vector Machine* (SVM) pada citra digital. Terdapat dua data digunakan pada penelitian ini yaitu data *training* dan *testing*. Data *training* didapat sebanyak 120 citra dari 12 motif batik Banten, sedangkan data *testing* 60 citra batik Banten. Hasil akurasi terbaik didapat orde 1 pada nilai RBF 0,2 sebesar 95%.

Penelitian selanjutnya dilakukan oleh (Safitri & Wulanningrum, 2020) dengan menggunakan metode *Support Vector Machine* (SVM) dalam implementasi pengenalan pola pada tulisan tangan. Pengujian aplikasi pengenalan pola yang dilakukan terhadap 50 sampel tulisan yang berbeda dengan bentuk tulisan yang sama. Pengujian ini dilakukan dengan tiga skenario, pada skenario 1 menghasilkan akurasi 70%, skenario 2 menghasilkan akurasi 85%, dan pada skenario 3 menghasilkan akurasi sebesar 80%. Akurasi terbaik didapatkan pada skenario 2 dengan akurasi sebesar 85%.

Penelitian selanjutnya mengklasifikasi ekspresi wajah dengan menggunakan metode *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), dan *Multi Layer Perceptron* (MLP). Penelitian yang dilakukan oleh (Dino & Abdulrazzaq, 2019) menggunakan dataset CK+ pada 210 wajah orang dewasa dengan total 634 citra dengan 8 label ekspresi. Hasil akurasi terbaik pada penelitian ini didapatkan dengan metode SVM dengan akurasi sebesar 93.53%, sedangkan metode KNN sebesar 79.97% dan metode MLP sebesar 82.57%.

Penelitian selanjutnya dilakukan oleh (Noprianto dkk., 2017) dengan menggunakan *Support Vector Machine* (SVM) dalam aplikasi pengenalan gambar pada plat kendaraan. Penelitian ini dilakukan terhadap 21 citra plat kendaraan dengan jarak 1, 3, dan 5 meter. Pada penelitian ini hasil akurasi terbaik didapatkan pada jarak 1 meter dengan akurasi sebesar 91.98%, pada jarak 3 meter akurasi sebesar 83.33%, dan pada jarak 5 meter akurasi yang didapatkan sebesar 51.85%.

Penelitian selanjutnya dilakukan oleh (Vijayarajeswari dkk., 2019) dengan menggunakan metode *Support Vector Machine* (SVM) yang dilakukan untuk mengklasifikasikan *mammogram* pada kanker payudara. Penelitian ini menggunakan 95 citra *mammogram* dengan ukuran 1024 x 1024 piksel. Hasil akurasi terbaik yang dilakukan menggunakan metode *Support Vector Machine* (SVM) didapatkan sebesar 94%.

2.2 Tumbuhan Obat

Tumbuhan obat merupakan jenis tumbuhan yang menghasilkan satu atau lebih zat aktif yang digunakan dalam perawatan kesehatan serta pengobatan atau tumbuhan yang diketahui memiliki khasiat untuk pengobatan (Tudjuka dkk., 2014). Tumbuhan obat merupakan jenis tumbuhan yang telah digunakan dalam pengobatan penyakit berdasarkan pengalaman dalam jangka waktu lama. Penggunaan tumbuhan obat biasa dilakukan dengan cara direbus untuk diminum, atau ditumbuk lalu ditempelkan pada bagian luka. Indonesia dikenal memiliki banyak tumbuhan obat, terdapat sekitar 20.000 jenis tumbuhan obat dimana 1.000

diantaranya telah didokumentasikan dan 300 telah digunakan sebagai obat (Ningsih, 2016).

2.3 Binahong

Binahong (*anredera cordifolia*) merupakan tanaman obat potensial yang dapat mengatasi berbagai jenis penyakit. Binahong termasuk dalam keluarga *Basellaceae* yang berasal dari negara Amerika Selatan dan telah dikenal sebagai tanaman obat sejak ratusan tahun lalu. Tanaman binahong tersebar di wilayah Indonesia dan wilayah lain yang memiliki iklim tropis, binahong tumbuh menjalar dengan panjangnya sekitar 5 meter. Di Indonesia binahong memiliki sebutan yang beragam diantaranya genjorat, gendolak, garang-garang, kandula, gendol, uci-uci, duyumu dan gendolak (Sundari & Almasyhuri, 2019). Binahong memiliki daun dengan tekstur tipis, yang berwarna hijau, serta bentuknya seperti jantung, dan berujung runcing. Daun binahong memiliki kandungan flavonoid, senyawa yang berperan menghilangkan, membersihkan serta menahan efek radikal bebas. Secara empiris tumbuhan binahong telah terbukti khasiatnya, untuk mengatasi berbagai macam penyakit seperti luka bakar, jantung dan kanker (Selawa dkk., 2013).



Gambar 1. Binahong

2.4 Sirih

Sirih (*pipper batle l*) merupakan salah satu jenis tumbuhan yang biasa digunakan untuk pengobatan. Tumbuhan ini termasuk dalam keluarga *piperaceae*. Tumbuh seperti tanaman binahong merambat dan menjalar dengan panjang mencapai sekitar 10 meter, tergantung pada letak dan pertumbuhannya. Tumbuhan ini sering dimanfaatkan untuk pengobatan, bagian dari tumbuhan sirih yang digunakan untuk pengobatan yaitu daun, akar, dan biji. Memiliki daun dengan bentuk berujung runcing seperti jantung, batang kasar ketika disentuh, bertangkai, batang berselang seling, dan mengeluarkan bau yang sedap. Sirih (*pipper betle l*) tumbuh subur di wilayah tropis seperti Asia hingga Afrika Timur dan hampir menyebar di seluruh Afrika (Carolia & Noventi, 2016).



Gambar 2. Sirih

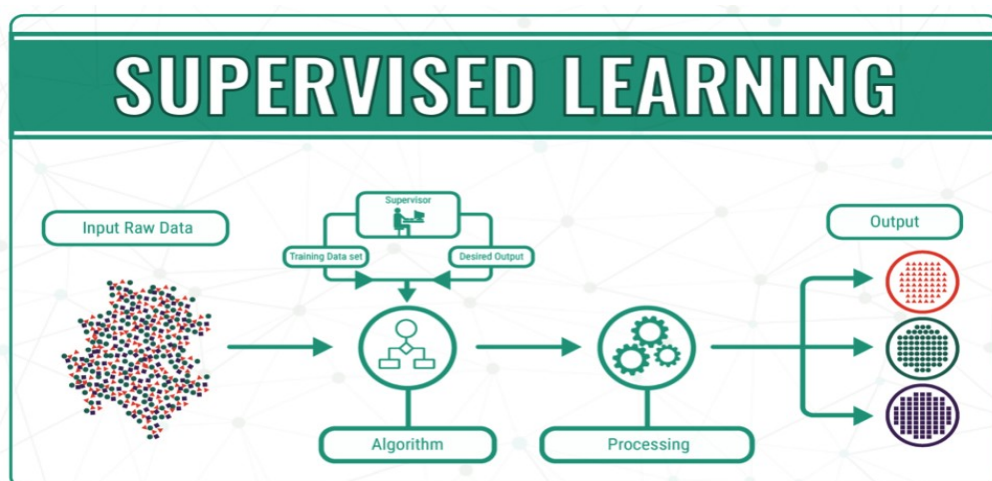
2.5 *Machine Learning*

Machine learning merupakan salah satu bagian cabang ilmu kecerdasan buatan (*artificial intelligence*) yang menjadikan komputer berperilaku cerdas seperti manusia melalui data yang sebelumnya dipelajari oleh komputer. *Machine Learning* biasa diterapkan pada beberapa bidang seperti medis, keuangan, keamanan dan lain-lain. *Machine learning* memiliki fokus terhadap

pengembangan sistem yang mampu belajar sendiri dengan data *training* dan *testing* untuk mengambil keputusan tanpa harus berulang kali diprogram oleh manusia (Retnoningsih & Pramudita, 2020).

Algoritma pada *machine learning* pada umumnya dirancang untuk menyelesaikan tiga kebutuhan, yaitu prediksi, klasifikasi, dan *clustering*. Prediksi merupakan kegiatan menebak nilai *output* berdasarkan data yang masuk. Klasifikasi adalah kegiatan untuk mengelompokkan data berdasarkan kategori tertentu serta mendeteksi kelompok dari suatu data. *Clustering* merupakan kegiatan pengelompokan data berdasarkan tingkat kemiripan. Menurut Dike dkk., (2018) Secara umum ada tiga tipe teknik dalam *machine learning* yaitu *supervised*, *unsupervised* dan *reinforcement learning*.

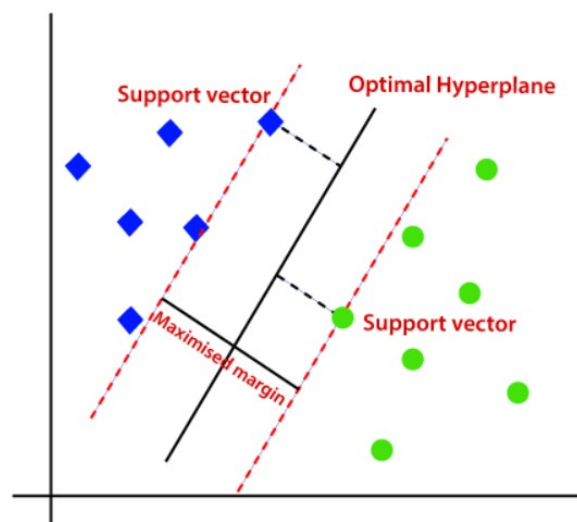
Supervised Learning merupakan tipe algoritma dalam *machine learning* yang menggunakan *dataset* untuk membuat klasifikasi atau prediksi. Dalam *supervised* data diberikan label untuk digunakan dalam pembelajaran koefisien vektor, pada gambar 2.3 digambarkan *output* yang dihasilkan sudah diarahkan pada label. Beberapa algoritma pembelajaran *supervised* yaitu *decision tree*, *k-nearest neighbors*, *logistic regression*, *SVM*, dan *naive bayes*.



Gambar 3. Konsep *supervised learning*
Sumber (Santoso dkk., 2021)

2.6 Support Vector Machine

Support Vector Machine (SVM) adalah sistem pembelajaran dengan fungsi *linear* yang dilatih oleh algoritma pembelajaran berdasarkan teori optimasi. Prinsip kerja SVM pada dasarnya adalah pengklasifikasi *linear*, kemudian dikembangkan agar mampu menyelesaikan masalah pada klasifikasi *non-linear*. Klasifikasi dalam svm dapat ditransformasikan dengan mencari garis (*hyperplane*) yang memisahkan kelompok. *Hyperplane* terbaik untuk pemisah antar dua kelompok diperoleh dengan menghitung *margin hyperlane* dan menemukan titik maksimumnya. *Margin hyperplane* adalah jarak antara *hyperplane* dengan pola terdekat dari masing-masing kelompok.



Gambar 4. Konsep SVM
Sumber (Medium.com)

Konsep SVM dapat digambarkan sebagai upaya untuk menemukan pemisah terbaik (*hyperplane optimal*) dari dua buah kelompok. Pada gambar 2.3 pola yang paling dekat *hyperplane* disebut *support vector*. Pencarian pemisah terbaik inilah yang menjadi inti dari SVM. Akan tetapi, dalam menemukan *hyperplane* akan

ada permasalahan baru berupa formula yang sulit dipecahkan, permasalahan ini disebut *quadratic programming* (Munawarah dkk., 2016).

2.6.1 *Support Vector Machine Linear*

SVM *linear* biasanya berlaku untuk klasifikasi data yang dapat dipisahkan secara linier (*linearly separable data*). Misalnya $x_i = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^n$ dan $y_i \in \{-1, +1\}$ dimana x_i adalah sebuah pola dan y_i adalah kelas dari pola x_i . Hal pertama yang harus dilakukan dalam SVM adalah menemukan *hyperplane* menggunakan fungsi linier yang dapat didefinisikan:

$$f(x) = w \cdot x_i + b \quad (1)$$

W adalah *support vector* atau vector tegak lurus terhadap *hyperplane* yang dapat didefinisikan:

$$w = \sum_{i=1}^n \alpha_i y_i x_i \quad (2)$$

Keterangan:

x_i = pola ke-i

y_i = kelas pola ke-i

α_i = nilai α dari pola ke-i

B merupakan nilai bias (*threshold*)

$$b = -\frac{1}{2} (w \cdot x^+ + w \cdot x^-) \quad (3)$$

x^+ adalah kelas positif dari pola *support vector* dan x^- adalah data *support vector* negatif terbesar dari nilai alpha. $\text{Sign}(f(x))$ yang merupakan fungsi keputusan klasifikasi digunakan dalam klasifikasi data pada kelas positif atau negatif.

$$f(x) = \sum_{i=1}^m \alpha_i y_i K(x, x_i) + b \quad (4)$$

Keterangan:

m = jumlah *support vector*

α_i = nilai α dari data ke- i

$K(x, x_i)$ = fungsi kernel

Dalam menentukan *hyperplane* dari dua kelas, persamaan berikut perlu digunakan untuk memaksimalkan margin:

$$\text{Minimize } J_i[w] = \frac{1}{2} \|w\|^2 \quad (5)$$

Permasalahan optimasi *constrain* pada pencarian *hyperplane optimal* diubah kedalam permasalahan *langrangian* dengan menggunakan persamaan *langrangian multiplier*. Persamaan *langrangian multiplier* didefinisikan sebagai berikut:

$$\min_{w,b} L_p(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^n \alpha_i \quad (6)$$

Nilai *vector* w seringkali besar dan α_i bernilai tak terhingga maka permasalahan *langrangian* atau pada bentuk persamaan 6 diubah menjadi bentuk *dual problem* (LD) dengan persamaan sebagai berikut:

$$L_d = \sum_{i=1}^N \alpha_i - \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (7)$$

Syarat

$$0 \leq \alpha_i \leq C \text{ dan } \sum_{i=1}^N \alpha_i y_i = 0$$

Keterangan:

L_d = Dualitas *lagrange multiplier*

x = Titik data

y = Kelas data

c = Konstanta

N = Banyak data

Formula dalam pencarian pemisah terbaik (*hyperplane*) adalah permasalahan *quadratic programming*, nilai maksimum global α_i dapat selalu ditemukan dengan *quadratic programming*, setelah nilainya (α_i) ditemukan dengan permasalahan *quadratic programming* maka kelas data pengujian x dapat ditentukan dengan nilai dari fungsi keputusan:

$$f(x_d) = \sum_{i=1}^{ns} \alpha_i y_i x_i \cdot x_d + b \quad (8)$$

2.6.2 *Support Vector Machine Non-Linear*

Beberapa permasalahan data tidak bisa dipisahkan menggunakan *Support Vector Machine (SVM) linear*, sehingga perlu menambahkan *variable slack* untuk membantu mengatasi kondisi data non-linier.

$$\min \frac{1}{2} \|w\|^2 + C (\sum_{i=1}^n \xi_i)$$

$$\begin{aligned} \text{Dengan } y_i(x_i \cdot w + b) - 1 + \xi_i &\geq 0 \\ \xi_i &\geq 0, i = 1, \dots, n \end{aligned} \quad (9)$$

Keterangan:

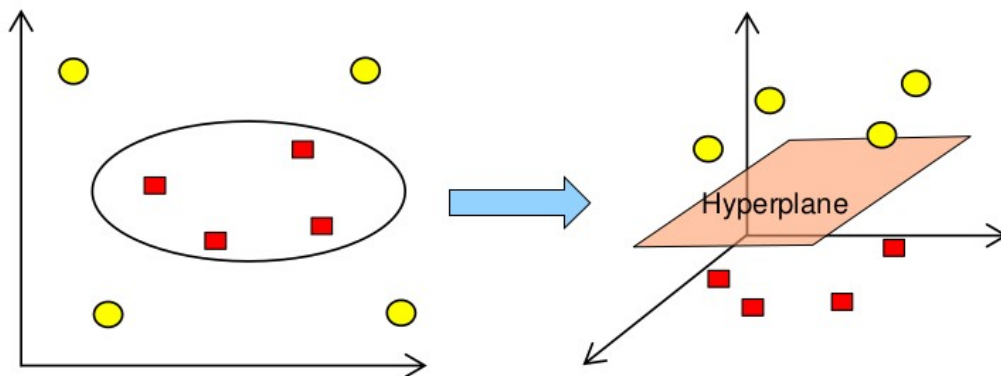
- x_i = data ke-i
- y_i = kelas data ke-i
- b = nilai bias
- w = bobot *support vector*
- ξ_i = *slack variable* (mengukur error dari data)
- C = parameter *user* bernilai positif

Sebelum mengoptimalkan *quadratic programming* nilai C atau disebut *complexity* dipilih terlebih dahulu. Nilai C berkisar dari 0 hingga jumlah tak

terbatas dari bilangan positif ($0 \leq C \leq \infty$). Fungsi dari nilai C adalah untuk meminimalkan kesalahan dan mengurangi nilai *slack variable*. Lebar margin terhadap bidang pembatas akan menjadi maksimal ketika nilai C mendekati nol, sementara pada waktu yang bersamaan banyak data latih yang berada dalam margin tidak akan diperhitungkan. Data latih yang tidak diperhitungkan akan mengurangi tingkat akurasi pada proses *training*, sehingga data uji tidak dapat diklasifikasikan dengan benar.

$$\min_{w,b} L_p(w,b,a) = \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n \xi_i \right) - \sum_{i=1}^n \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} - \sum_{i=1}^n \mu_i \xi_i \quad (10)$$

Persamaan 9 akan meminimalkan nilai $\|w\|^2$ serta memaksimalkan margin antar kelas. Permasalahan non-linier juga dapat dirubah kedalam persamaan *langrange*, sehingga menjadi bentuk *primal problem* pada persamaan 10.



Gambar 5. SVM Non Linear
Sumber (Cholissodin & Soebroto, 2019)

Data (kiri) pada gambar 5 merupakan klasifikasi yang tidak dapat di pisah kan secara *linear*. Pemetaan perlu dilakukan dimensi input (*feature space*) kedalam ruang dimensi yang lebih tinggi, sehingga membuat data mampu dipisahkan secara linear kedalam ruang dimensi baru. Metode ini disebut *kernel tricks* yang

menjadikan fungsi pemetaan tidak dapat diketahui secara pasti. Beberapa fungsi *kernel* yang dapat digunakan, sebagai berikut:

1. *kernel linear* didefinisikan pada persamaan 11

$$K(x, y) = x^T \cdot y + c \quad (11)$$

2. *kernel polynomial* didefinisikan pada persamaan 12

$$K(x, y) = (\alpha x^T y + c)^d \quad (12)$$

3. *kernel gaussian RBF* didefinisikan pada persamaan 13

$$K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (13)$$

4. *kernel hyperbolic tangent (sigmoid)* didefinisikan pada persamaan 14

$$K(x, y) = \tanh(\alpha x^T \cdot y + c) \quad (14)$$

kernel linear digunakan ketika mengklasifikasikan data yang dapat dipisahkan dengan garis atau *hyperplane*, sedangkan *kernel non-linear* digunakan ketika mengklasifikasikan data yang dapat dipisahkan dengan garis lengkung atau pada ruang yang berdimensi tinggi.

2.7 Sequential Training SVM

Beberapa proses dilakukan untuk melatih data pada SVM yaitu *quadratic programming (QP)*, *Sequential Minimal Optimization (SMO)* dan *Sequential Training*. Mencari garis pemisah terbaik *hyperplane optimal* pada SVM dapat dihitung menggunakan QP, akan tetapi QP memakan waktu lama, rentan terhadap numerik yang tidak stabil, dan algoritma yang rumit. Algoritma SMO yang merupakan pengembangan dari metode QP lebih umum dengan masalah optimasi kecil pada setiap tahap, akan tetapi sama seperti QP algoritma SMO cukup rumit. Metode *Sequential Training* memiliki algoritma yang lebih sederhana

dibandingkan dengan metode sebelumnya, dan waktu yang diperlukan lebih cepat (Cholissodin & Soebroto, 2019). Berikut adalah tahapan pada metode *Sequential Training*.

1. Proses pertama yang dilakukan dalam metode *Sequential Training* setelah melakukan perhitungan kernel adalah inisialisasi untuk parameter SVM, α $i = 0$ dan parameter lainnya (misal $\lambda = -0,0298$, $\gamma = 0,239$, $C = 24,8$ dan $\epsilon = 0,1$).

Keterangan:

α_i = Alfa / *Lagrange Multiplier*.

λ = Koefisien *Lagrange Multiplier*.

γ = Konstanta *Gamma*.

C = Konstanta C.

ϵ = Epsilon.

Fungsi *Lagrange Multiplier* (α_i) digunakan untuk mencari *support vector* dan λ adalah koefisien *Lagrange Multiplier*. Proses pembelajaran yang digunakan untuk mengontrol kecepatan adalah *Gamma* (γ). Konstanta *Gamma* (γ) digunakan untuk mengontrol kecepatan pada proses pembelajaran, sedangkan konstanta C (Nilai C), digunakan untuk membatasi nilai *Lagrange Multiplier* (α_i), dan nilai epsilon (ϵ) digunakan untuk ukuran kesalahan pada proses klasifikasi.

2. Menampung matriks kernel dengan menggunakan matriks Hessian (D_{ij}).
Persamaan matriks Hessian didefinisikan sebagai berikut:

$$D_{ij} = y_i y_j (\lambda^2 + K(x_i, x_j)) \quad (15)$$

Keterangan:

$K(x_i, x_j)$ = Fungsi kernel.

x_j = Data ke-j.

x_i = Data ke-i.

y_j = Kelas data ke-j.

y_i = Kelas data ke-i.

n = Jumlah data.

3. Menggunakan persamaan berikut untuk menghitung setiap nilai $i, j = 1, 2, \dots, n$:

$$a) \quad E_i = \sum_{j=1}^n \alpha_j D_{ij} \quad (16)$$

Keterangan:

α_j = Alpha ke- j .

E_i = *Error Rate*.

D_{ij} = Matriks Hessian.

$$b) \quad \delta\alpha_i = \min \{ \max [\gamma (1 - E_i), -\alpha_i], C - \alpha_i \} \quad (17)$$

Keterangan

$\delta\alpha_i$ = Variabel tunggal (bukan bentuk perkalian δ dan α_i)

γ = *Learning Rate* (menjaga kecepatan pembelajaran).

E_i = *Error rate*.

Inisialisasi parameter SVM untuk melihat nilai konstanta, berikut bentuk persamaannya:

$$\gamma = \frac{\text{konstanta}}{\text{Max } \{i\} D_{ij}} \quad (18)$$

$\text{Max } \{i\} D_{ij}$ merupakan nilai maksimal yang didapat dari nilai diagonal matriks.

$$c) \quad \alpha_i = \alpha_i + \delta\alpha_i \quad (19)$$

Perubahan pada fungsi *Lagrange Multiplier* dapat dilihat dengan fungsi konvergensi ($\delta\alpha_i$). Iterasi akan berhenti ketika data latih telah mencapai nilai konvergen ($\max(|\delta\alpha_i|) < \epsilon$), dan nilai maximum iterasi mencapai nilai yang ditentukan.

- d) Nilai *Lagrange Multiplier* (α) akan berhenti ketika mencapai nilai konvergen, dan nilai konvergen dapat diuraikan pada tahap perubahan nilai α .
- e) Persamaan $SV = (\alpha_i > \text{Threshold SV})$ dapat menghitung nilai *Support Vector* (SV). Nilai *Threshold SV* merupakan hasil dari percobaan, biasanya menggunakan $\text{Threshold SV} \geq 0$.

2.8 Cross Validation

Cross validation merupakan teknik validasi model yang bertujuan untuk memberikan nilai pada hasil analisis statistik. Hasil analisis statistik akan digeneralisasi kumpulan data independen. Teknik *Cross validation* banyak digunakan untuk membuat prediksi model dan menilai tingkat akurasi sebuah model prediksi. *K-Fold Cross Validation* merupakan salah satu teknik dalam validasi silang (*Cross Validation*), cara kerja teknik ini memecah data menjadi k set data dengan bagian yang sama. *K-Fold Cross Validation* berfungsi untuk memaksimalkan peluang semua data bisa menjadi data *training* dan data *testing* (Tempola dkk., 2018). Berikut gambaran validasi silang *K-Fold Cross Validation* dengan K=10 pada gambar 6.

Split / Fold	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10
Split 1	Test	Train	Train	Train	Train	Train	Train	Train	Train	Train
Split 2	Train	Test	Train	Train	Train	Train	Train	Train	Train	Train
Split 3	Train	Train	Test	Train	Train	Train	Train	Train	Train	Train
Split 4	Train	Train	Train	Test	Train	Train	Train	Train	Train	Train
Split 5	Train	Train	Train	Train	Test	Train	Train	Train	Train	Train
Split 6	Train	Train	Train	Train	Train	Test	Train	Train	Train	Train
Split 7	Train	Train	Train	Train	Train	Train	Test	Train	Train	Train
Split 8	Train	Train	Train	Train	Train	Train	Train	Test	Train	Train
Split 9	Train	Train	Train	Train	Train	Train	Train	Train	Test	Train
Split 10	Train	Train	Train	Train	Train	Train	Train	Train	Train	Test

Gambar 6. *K-Fold Cross Validation*

2.9 Nilai Evaluasi

Perhitungan nilai evaluasi pada pembelajaran SVM dapat dilakukan dengan *confusion matrix*. *Confusion matrix* merupakan sebuah *matrix* dari hasil prediksi yang akan dibandingkan dengan nilai asli dari data *testing*. *Confusion matrix* digunakan untuk menentukan nilai akurasi, presisi, *recall*, dan *F1 score* dalam suatu algoritma (Pratiwi dkk., 2021).

Tabel 2. *confusion matrix*

KELAS PREDIKSI	KELAS AKTUAL	
	TRUE	FALSE
TRUE	TRUE POSITIVE (TP)	FALSE POSITIVE (FP)
FALSE	FALSE NEGATIVE (FN)	TRUE NEGATIVE (TN)

Keterangan

TP = *True Positive*, jumlah data positif yang diklasifikasi dengan nilai benar.

TN = *True Negative*, jumlah data negatif yang diklasifikasi dengan nilai benar.

FN = *False Negative*, jumlah data negatif yang diklasifikasi dengan nilai salah.

FP = *False Positive*, jumlah data positif yang diklasifikasi dengan nilai salah.

$$\text{Akurasi} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \quad (20)$$

Menurut (Pratiwi dkk., 2021) akurasi pada klasifikasi merupakan presentase hasil pengujian klasifikasi berdasarkan ketepatan *record* data yang diklasifikasikan dengan benar. Nilai akurasi pada persamaan 20 merupakan perbandingan data yang bernilai benar dengan seluruh data.

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (21)$$

Presisi merupakan proporsi yang diprediksi positif dan positif benar pada data sebenarnya. Nilai presisi pada persamaan 21 menggambarkan perbandingan antara nilai positif benar dengan jumlah data yang diklasifikasikan positif.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (22)$$

Nilai *recall* adalah data kategori positif yang diklasifikasi dengan data sebenarnya.

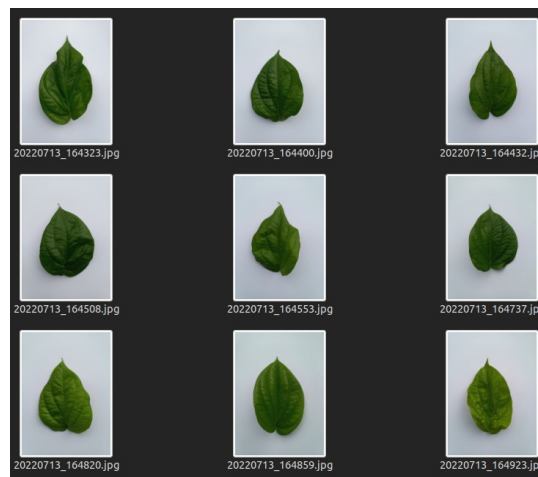
$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Presisi}}{\text{Recall} + \text{Presisi}} \quad (23)$$

Nilai *F1 Score* merupakan nilai rata-rata dari perbandingan presisi yang dijumlah dan dikalikan dengan recall.

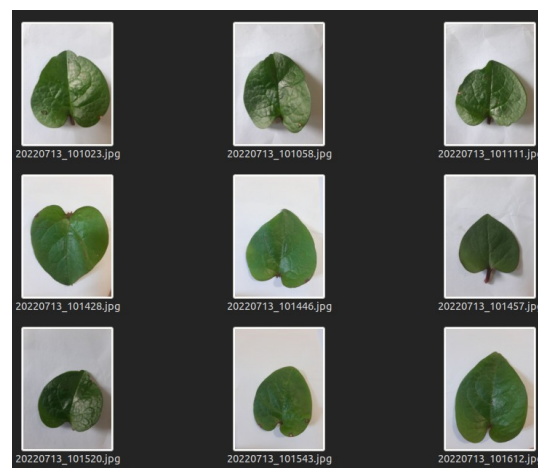
3.2 Data dan Alat

3.2.1 Data

Penelitian ini menggunakan data primer. Data diperoleh dari pengambilan gambar menggunakan kamera *handphone*. Data yang digunakan berjumlah 900 data citra daun tumbuhan obat, dimana 450 citra daun tumbuhan sirih dan 450 citra daun tumbuhan binahong. Berikut sebagian data citrasirih dan binahong pada gambar 7 dan 8.



Gambar 7. Data Daun Sirih



Gambar 8. Data Daun Binahong

3.2.2 Alat

Penelitian ini menggunakan perangkat keras (*Hardware*) dan perangkat lunak (*Software*).

a) Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam penelitian ini merupakan sebuah komputer dengan spesifikasi berikut.

- *Processor* : Intel^(R) Core^(TM) i3-4030U CPU 1.90 Ghz
- *RAM* : 6.0 GB

b) Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam penelitian ini adalah sebagai berikut.

- Sistem Operasi Ubuntu 20.04.3 LTS
- *Anaconda*

Anaconda merupakan aplikasi distribusi yang bersifat gratis (*open source*) khusus diciptakan untuk bahasa pemrograman *python* dan R.

Anaconda digunakan pada disiplin ilmu pembelajaran mesin (*machine learning*) dan ilmu tentang data (*data science*). Aplikasi distribusi ini juga bersifat lintas platform yang dapat diinstal pada beberapa sistem operasi seperti windows, linux dan MacOS (Bhavsar & Richa, 2019).

- *Jupyter Notebook*

Jupyter Notebook merupakan sebuah perangkat lunak yang populer untuk memproses dan berinteraksi dengan data pada *python*. *Jupyter Notebook* memungkinkan *user* untuk mengintegrasikan kode dan *output* secara interaktif kedalam suatu dokumen. Jupyter adalah nama sebuah organisasi yang bertujuan komersial (*non-profit*), adapun *Notebook* adalah salah satu software yang dibuat oleh *Jupyter* (Wang dkk., 2020).

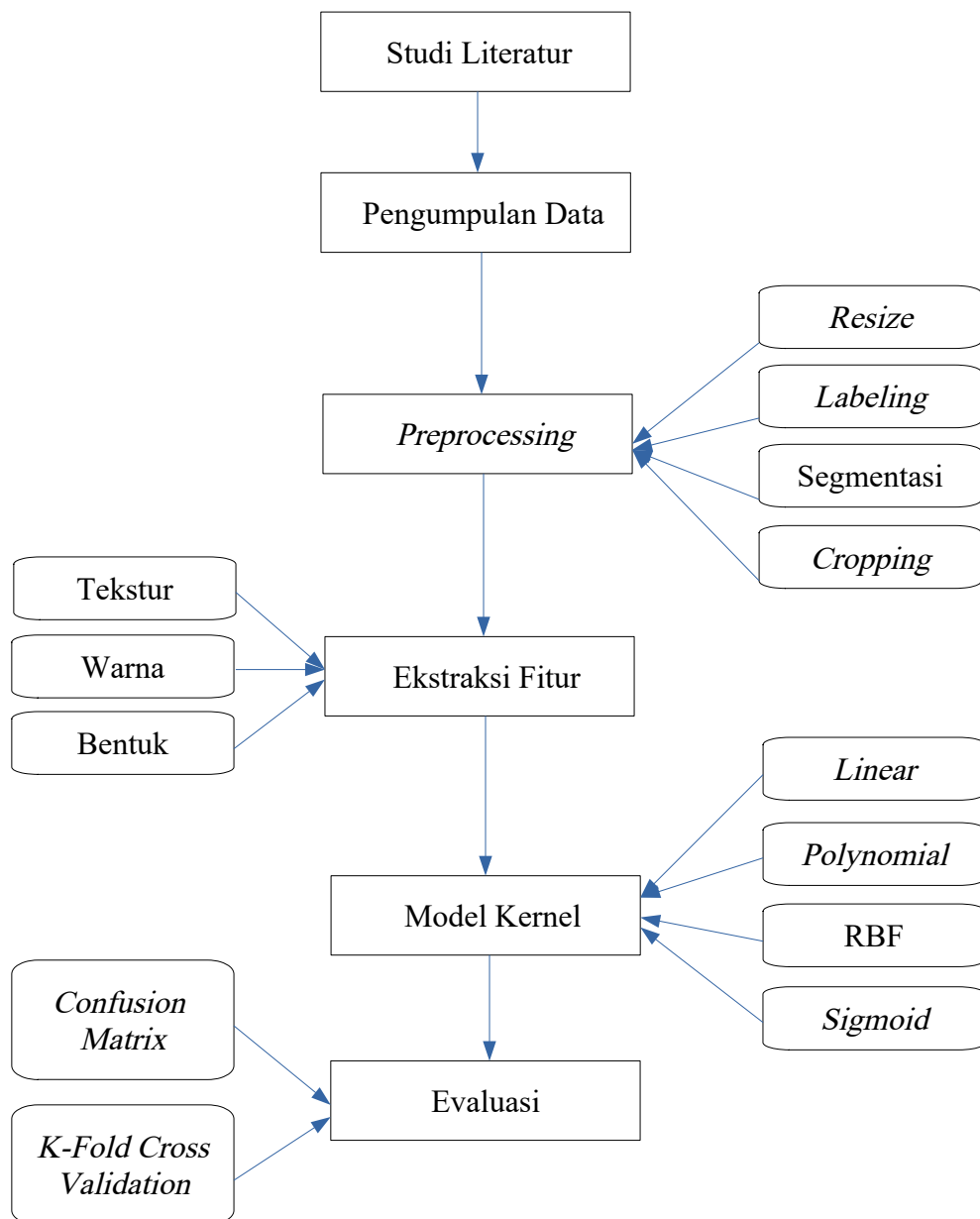
- *Python*

Penelitian ini menggunakan bahasa pemrograman *python* 3.8.3 untuk mengolah dan menganalisis, dan menampilkan data pada penelitian. Adapun *library* pada *python* yang dipakai dalam penelitian ini sebagai berikut.

- *Open CV* (Andrekha & Huda, 2021), sebuah *library* yang dimanfaatkan untuk mengekstrak informasi dari sebuah gambar atau video.
- *Numpy* (Retnoningsih & Pramudita, 2020), sebuah *library* dimanfaatkan dalam kebutuhan matematis.
- *Matplotlib* (Nofiyanti dkk., 2021), sebuah *library* dalam membuat plot grafik serta menampilkan hasil dengan visualisasi.
- *Sckit learn* (Fahlepi & Widjaja, 2019), sebuah *library* yang dimanfaatkan dalam membuat metode dan algoritma pada *machine learning*.
- *Sckit Image* (Ravikumar & Arulmozhi, 2019), sebuah *library* pada *python* untuk ekstrak citra digital.

3.3 Metode

Metode yang digunakan dalam penelitian ini menggunakan *Support Vector Machine* (SVM), tahapan dalam penelitian meliputi studi literatur, pengumpulan data, *preprocessing* pada data asli, ekstraksi ciri pada citra, membuat model kernel *Support Vector Machine* (SVM), *cross validation* pada data *training*, melihat hasil dengan *confusion matrix* dan evaluasi kinerja klasifikasi. Berikut flowchart tahapan penelitian ini pada gambar 8.



Gambar 9. Alur Penelitian

3.3.1 Studi Literatur

Studi literatur merupakan tahapan mencari serta mengumpulkan informasi yang dibutuhkan untuk mendukung penelitian ini. Informasi yang dibutuhkan berupa artikel ilmiah, buku, dan jurnal penelitian mirip dengan penelitian ini yang pernah dilakukan. Studi literatur dilakukan dengan tujuan untuk mendukung serta menguatkan penelitian ini.

3.3.2 Pengumpulan Data

Tahap pengambilan data pada penelitian ini dilakukan dengan mengambil data berupa citra (*image*) secara langsung menggunakan kamera *handphone* dengan jarak 5-10 cm. Penelitian ini menggunakan 2 objek data yaitu daun sirih dan binahong. Data yang dikumpulkan total sebanyak 900 citra, masing-masing 450 daun sirih dan 450 daun binahong. Berikut dokumentasi salah satu pengambilan data.

3.3.3 *Preprocessing*

Tahapan *preprocessing* merupakan tahapan dimana data yang diambil berupa citra diolah sehingga menghasilkan data yang siap digunakan untuk proses klasifikasi. Tahapan ini akan mengolah data dengan melakukan *resize*, labeling, segmentasi dan *cropping* terhadap data.

Resize merupakan teknik untuk mengubah piksel dan dimensi pada citra sehingga menghasilkan piksel dan dimensi yang sama pada semua data. Tujuan dari tahapan ini adalah agar data *training* dan data *testing* memiliki ukuran yang sama. Tahapan ini menggunakan aplikasi pada website onlineimagetool.com.

Support Vector Machine (SVM) merupakan algoritma *supervised learning* sehingga memerlukan proses labeling pada tahap *preprocessing*. Labeling merupakan tahapan dimana data diberi label sesuai dengan kelas yang telah ditentukan, tujuannya agar komputer mampu mengenali kelas pada masing-masing data.

Segmentasi adalah tahapan memisahkan *background* dan *foreground* agar tahapan selanjutnya tepat dalam mengekstraksi objek. Tahapan segmentasi pada penelitian ini menggunakan *thresholding* dimana segmentasi citra dilakukan berdasarkan perbedaan nilai abu-abu citra. Nilai keabuan dari citra bisa didapatkan dari proses *grayscale*, rumus yang digunakan untuk mengubah nilai r, g, dan b adalah sebagai persamaan berikut.

$$S = \frac{R+G+B}{3} \quad (24)$$

Nilai *grayscale* (s) didapat dari rata-rata nilai r, g, dan b. Tahapan selanjutnya adalah memisahkan *background* dan objek dengan menggunakan *thresholding*, setelah dipisahkan gunakan teknik dilasi dan erosi untuk memperjelas batas objek dan *background*.

Tahapan terakhir pada proses *preprocessing* adalah menggunakan teknik *cropping*, teknik ini dibutuhkan untuk proses ekstraksi fitur warna agar mendapatkan nilai dominan dari warna objek, untuk dapat menghitungnya maka gambar akan di potong sesuai dengan batasan objek.

3.3.4 Ekstraksi Fitur

Tahapan ini melakukan ekstraksi fitur pada citra dengan tujuan mengambil informasi yang dibutuhkan untuk proses klasifikasi. Tahapan ini merujuk pada penelitian yang dilakukan oleh (Yana & Nafi'iyah, 2021) dengan judul Klasifikasi Jenis Pisang Berdasarkan Fitur Warna, Tekstur, Bentuk Citra Menggunakan SVM dan KNN. Penelitian ini menggunakan 3 fitur yaitu, tekstur, warna, dan bentuk.

Fitur tekstur pada penelitian ini menggunakan GLCM (*Gray-Level Co-occurrence Matrix*) dengan properti *dissimilarity*, *correlation*, *homogeneity*, *contrast*, *ASM* dan *energy* serta kombinasi *angles* 0°, 45°, 90°, dan 135° juga jarak piksel (d) = 1,2,3,4, dan 5. Properti pada penelitian ini merujuk pada penelitian yang dilakukan oleh (A'yun & Utaminingrum, 2022).

Ekstraksi fitur warna pada penelitian ini menggunakan HSV, properti pada HSV meliputi *Hue*, *Saturation*, dan *Value*. *Hue* digunakan untuk mencari nilai warna merah, hijau dan yang lainnya. *Saturation* digunakan untuk mencari kemurnian atau kekuatan pada warna citra dan *Value* digunakan dalam mencari nilai kecerahan dengan rentang 0-100 semakin mendekati 0, maka kecerahan semakin redup. Berikut formula HSV pada tabel 4.

Tabel 4. Formula pada HSV

HSV	Formula
<i>Hue</i>	$\tan \frac{3(G-B)}{(G-B)+(R+B)}$
<i>Saturation</i>	$1 - \frac{\min(R,G,B)}{V}$
<i>Value</i>	$\frac{R+G+B}{3}$

Ekstraksi bentuk pada penelitian ini menggunakan parameter *eccentricity* dan *metric* untuk membedakan bentuk objek satu citra dengan yang lainnya. Parameter *eccentricity* menghitung nilai perbandingan *foci ellips* minor dan mayor. *Eccentricity* memiliki rentang nilai 0 sampai 1, semakin mendekati 1 maka objeknya berbentuk garis lurus sedangkan nilai yang mendekati 0 maka objek semakin berbentuk lingkaran. Parameter *metric* menghitung perbandingan luas dan keliling objek, parameter ini memiliki rentang nilai 0 sampai 1. Objek yang berbentuk memanjang seperti garis lurus maka nilai nya mendekati 0, sebaliknya bentuk yang berbentuk lingkaran maka nilai nya mendekati 1. Berikut formula parameter *eccentricity* dan *metric* dapat dilihat pada persamaan 25 dan 26.

$$e = \sqrt{1 - \frac{b^2}{a^2}} \quad (25)$$

e = *eccentricity*

a = mayor

b = minor

$$M = \frac{4\pi \times A}{C^2} \quad (26)$$

M = *Metric*

A = *Area*

C = *Circumference*

3.3.5 Membagi Dataset

Penelitian ini membagi data merujuk pada penelitian (Maimunah & Primadewi, 2021), dengan judul Model *Machine Learning* untuk Klasifikasi Mutu Telur Ayam Ras Berdasarkan Kebersihan Kerabang. Penelitian ini membagi data *training* dan *testing* menjadi 3 bagian yaitu, 70:30, 75:25, dan 80:20.

3.3.6 Model kernel klasifikasi SVM

Model kernel yang digunakan dalam penelitian ini merujuk pada penelitian (Maulina & Sagara, 2018) yang berjudul Klasifikasi Artikel Hoax Menggunakan *Support Vector Machine Linear* dengan Pembobotan *Term Frequency – Inverse Document Frequency*. Penelitian ini menggunakan 4 kernel yaitu, *Linear*, *Sigmoid*, *Polynomial*, dan *RBF*.

3.3.7 Cross Validation

Tahapan ini melakukan pengujian pada data *training* untuk menilai performa pada data *training*. *Cross Validation* memberi kesempatan semua data *training* menjadi data *testing*, *Cross Validation* pada penelitian ini menggunakan *K-Fold Cross Validation* dengan nilai k=10. Nilai k=10 artinya membuat 10 kemungkinan yang akan terjadi pada data. Nilai k sama dengan banyaknya akurasi yang dihasilkan maka, nilai akhir akurasi akan dilihat dari nilai rata-rata akurasi dan standar deviasinya. Berikut formula standar deviasi pada persamaan

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (27)$$

S = Standar deviasi

x_i = Nilai x ke-i

\bar{x} = Nilai rata-rata data

n = Jumlah data

3.3.8 Hasil

Tahap hasil akurasi merupakan tahapan konfirmasi hasil uji pada penelitian ini, pada tahapan ini model SVM akan dikonfirmasi dengan data test untuk melihat performa model kernel SVM pada klasifikasi yang dilakukan sebelumnya.

Tahapan ini menggunakan *Confusion Matrix* dan menghasilkan nilai TP (*True Positive*), TN (*True Negative*), FP (*False Positive*), dan FN (*False Negative*).

TP adalah hasil prediksi benar pada nilai *positive*, TN hasil prediksi benar pada nilai *negative*, FP hasil prediksi salah pada nilai *positive*, dan FN hasil prediksi salah pada nilai *negative*. Secara umum hasil prediksi benar ada pada nilai TP dan TN sedangkan prediksi salah ada pada nilai FP dan FN.

3.3.9 Evaluasi

Tahap evaluasi pada penelitian ini dilakukan setelah mendapatkan nilai TP, TN, FP, dan FN. Tahapan ini menggunakan *Confusion Matrix*

V. KESIMPULAN

5.1 Kesimpulan

Penelitian implementasi *support vector machine* (SVM) untuk pengelompokan daun tumbuhan obat binahong (*anredera cordifolia*) dan sirih hijau (*piper betle*) dapat disimpulkan sebagai berikut.

1. Metode *support vector machine* (SVM) telah berhasil mengklasifikasikan daun tumbuhan obat binahong (*anredera cordifolia*) dan sirih hijau (*piper betle*) secara baik dengan nilai *Accuracy* sebesar 92.44%, *Precision* sebesar 92.61%, *Recall* sebesar 92.44%, dan *F1 Score* sebesar 92.43%.
2. Hasil terbaik diantara 3 pemodelan didapat pada pemodelan 2 dengan model kernel *linear*, menggunakan data *training* dan data *test* 75:25, dan jarak piksel (d) 5.

5.2 Saran

Saran yang diberikan dalam penelitian ini sebagai berikut.

1. Menambahkan jumlah data citra dan fitur pada penelitian selanjutnya agar algoritma mampu belajar lebih baik.
2. Penelitian ini bisa dikembangkan dengan menambahkan kelas atau menggunakan algoritma *clustering*, agar pada tampilan *deployment* gambar selain binahong dan sirih yang diinput dapat dikontrol.

DAFTAR PUSTAKA

- A'yun, Q., & Utamingrum, F. (2022). Rancang Bangun Deteksi Kemanisan Buah Semangka menggunakan Metode Gray Level Co-Occurrence Matrix dan Backpropagation Neural Network berbasis *Teknologi Informasi Dan Ilmu Komputer*, 6(2), 707–712.
- Andrekha, M. Z., & Huda, Y. (2021). Deteksi Warna Manggis Menggunakan Pengolahan Citra dengan Opencv Python. *Voteteknika (Vocational Teknik Elektronika Dan Informatika)*, 9(4), 27–33.
- Ariani, S., Loho, L., & Durry, M. F. (2013). Khasiat daun binahong (anredera cordifolia (ten.) steenis) terhadap pembedakan jaringan granulasi dan reepitalisasi penyembuhan luka terbuka kulit kelinci. *Jurnal E-Biomedik*, 1(2), 914–919.
- Bhavsar, H., & Richa, M. (2019). Sentiment analysis of twitter data using python. *International Research Journal of Engineering and Technology (IRJET)*, 06(03), 510–511.
- Budianto, A., Ariyuana, R., & Maryono, D. (2018). Perbandingan K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam pengenalan karakter plat kendaraan bermotor. *Jurnal Ilmiah Pendidikan Teknik Dan Kejuruan*, 11(1), 1–9.
- Cahyaningrum, E. D., & Putri, D. (2017). Perbedaan suhu tubuh anak demam sebelum dan setelah kompres bawang merah. *MEDISAINS Jurnal Ilmiah Ilmu-Ilmu Kesehatan*, 15(2), 66–74.
- Carolia, N., & Noventi, W. (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), 140–145.

- Cholissodin, I., & Soebroto, A. (2019). *Buku Ajar AI, Machine Learning & Deep Learning*.
- Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018). Unsupervised Learning Based On Artificial Neural Network: A Review. *2018 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, 322–327.
- Dino, H. I., & Abdulrazzaq, M. B. (2019). Facial Expression Classification Based on SVM, KNN and MLP Classifiers. *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 70–75.
- Fahlepi, M. reza, & Widjaja, A. (2019). Penerapan Metode Multiple Linear Regression Untuk Prediksi Harga Sewa Kamar Kost. *Jurnal Strategi*, 1(2), 615–629.
- Maimunah, & Primadewi, A. (2021). Model Machine Learning untuk Klasifikasi Mutu Telur Ayam Ras Berdasarkan Kebersihan Kerabang. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 386–391.
- Maulana, F. F., & Rochmawati, N. (2019). Klasifikasi citra buah menggunakan convolutional neural network. *Journal of Informatics and Computer Science (JINACS)*, 1(02), 104–108.
- Maulina, D., & Sagara, R. (2018). Klasifikasi Artikel Hoax Menggunakan Support Vector Machine Linear Dengan Pembobotan Term Frequency-Inverse Document Frequency. *Jurnal Mantik Penusa*, 2(1), 35–40.
- Munawarah, R., Soesanto, O., & Faisal, M. R. (2016). Penerapan metode support vector machine pada diagnosa hepatitis. *Kumpulan Jurnal Ilmu Komputer (KLIK)*, 04(01), 103–113.
- Ningsih, I. Y. (2016). Studi etnofarmasi penggunaan tumbuhan obat oleh suku Tengger di Kabupaten Lumajang dan Malang, Jawa Timur. *Pharmacy*, 13(1), 10–20.
- Nofiyanti, E., Maricha, E., & Nur Haryanto, O. (2021). Analisis Sentimen terhadap Penanggulangan Bencana di Indonesia. *Jurnal Ilmiah SINUS*, 19(2), 17–26.

- Noprianto, Wibirama, S., & Nugroho, H. A. (2017). Long distance automatic number Plate Recognition under perspective distortion using zonal density and support vector machine. *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, 159–164.
- Pratiwi, B. P., Handayani, A. S., & Sarjana, S. (2021). Pengukuran kinerja sistem kualitas udara dengan teknologi wsn menggunakan confusion matrix. *Jurnal Informatika Upgris*, 6(2), 66–75.
- Ravikumar, R., & Arulmozhi, V. (2019). Digital Image Processing-A Quick Review. *Digital Image Processing-A Quick Review*, 2(2), 16–24.
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal machine learning dengan teknik supervised dan unsupervised learning menggunakan python. *Bina Insani Ict Journal*, 7(2), 156–165.
- Rizki, & Leilani, I. (2017). Etnofarmakologi tumbuhan familia rhizophoraceae oleh masyarakat di Indonesia. *BioCONCETTA*, III(1), 51–60.
- Safitri, K. A., & Wulanningrum, R. (2020). Aplikasi pengenalan pola tulisan tangan menggunakan metode support vector machine. *Seminar Nasional Inovasi Teknologi*, 4(1), 201–206.
- Santoso, P., Abijono, H., & Anggreini, N. L. (2021). Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, 4(2), 315–318.
- Selawa, W., Runtuwene, M. R. J., & Citraningtyas, G. (2013). Kandungan flavonoid dan kapasitas antioksidan total ekstrak etanol daun binahong [anredera cordifolia(ten.)steenis]. *Jurnal Ilmiah Farmasi*, 2(1), 18–23.
- Sholehudin, B. (2016). Implementasi metode Support Vector Machine (SVM) untuk deteksi penyakit pada daun apel menggunakan citra digital [Universitas Brawijaya]. In *Jurnal Mahasiswa PTIIK UB* (Vol. 7, Issue 25).
- Siregar, R. S., Hadiguna, R. A., Kamil, I., Nazir, N., & Nofaldi, N. (2020). Permintaan dan penawaran tanaman obat tradisional Di Provinsi Sumatera Utara. *Jurnal Tumbuhan Obat Indonesia*, 13(1), 50–59.

- Sundari, D., & Almasyhuri, A. (2019). Uji aktivitas antiseptik ekstrak etanol daun sirih (*piper betle* linn.) dalam obat kumur terhadap *staphylococcus aureus* secara in vitro. *Jurnal Kefarmasian Indonesia*, 9(1), 10–18.
- Tempola, F., Muhammad, M., & Khairan, A. (2018). Perbandingan Klasifikasi Antara KNN dan Naive Bayes pada Penentuan Status Gunung Berapi dengan K-Fold Cross Validation. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(5), 577–584.
- Tudjuka, K., Ningsih, S., & Toknok, B. (2014). Keanekaragaman jenis tumbuhan obat pada kawasan hutan lindung di Desa Tindoli Kecamatan Pamona Tenggara Kabupaten Poso. *Warta Rimba*, 2(1), 120–128.
- Vijayarajeswari, R., Parthasarathy, P., Vivekanandan, S., & Basha, A. A. (2019). Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. *Measurement: Journal of the International Measurement Confederation*, 146, 800–805.
- Wang, J., Kuo, T. Y., Li, L., & Zeller, A. (2020). Assessing and Restoring Reproducibility of Jupyter Notebooks. *IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 138–149.
- Wiryadinata, R., Adli, M. R., Fahrizal, R., & Alfanz, R. (2019). Klasifikasi 12 motif batik Banten menggunakan support vector machine. *Jurnal EECCIS*, 13(1), 60–64.