

**ANALISIS PERBANDINGAN KINERJA ALGORITMA *SUPPORT VECTOR MACHINE* , *DECISION TREE*, DAN *NAIVE BAYES* UNTUK
PENGELOMPOKAN SPESIES *IRIS* DALAM KONTEKS
KLASIFIKASI**

SKRIPSI



RISKI RONALDO REJEKI TUA

20201000027

TEKNIK INFORMATIKA

**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025

**ANALISIS PERBANDINGAN KINERJA ALGORITMA *SUPPORT VECTOR MACHINE* , *DECISION TREE*, DAN *NAIVE BAYES* UNTUK
PENGELOMPOKAN SPESIES *IRIS* DALAM KONTEKS
KLASIFIKASI**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh Gelar Sarjana pada
Program Studi Teknik Informatika**



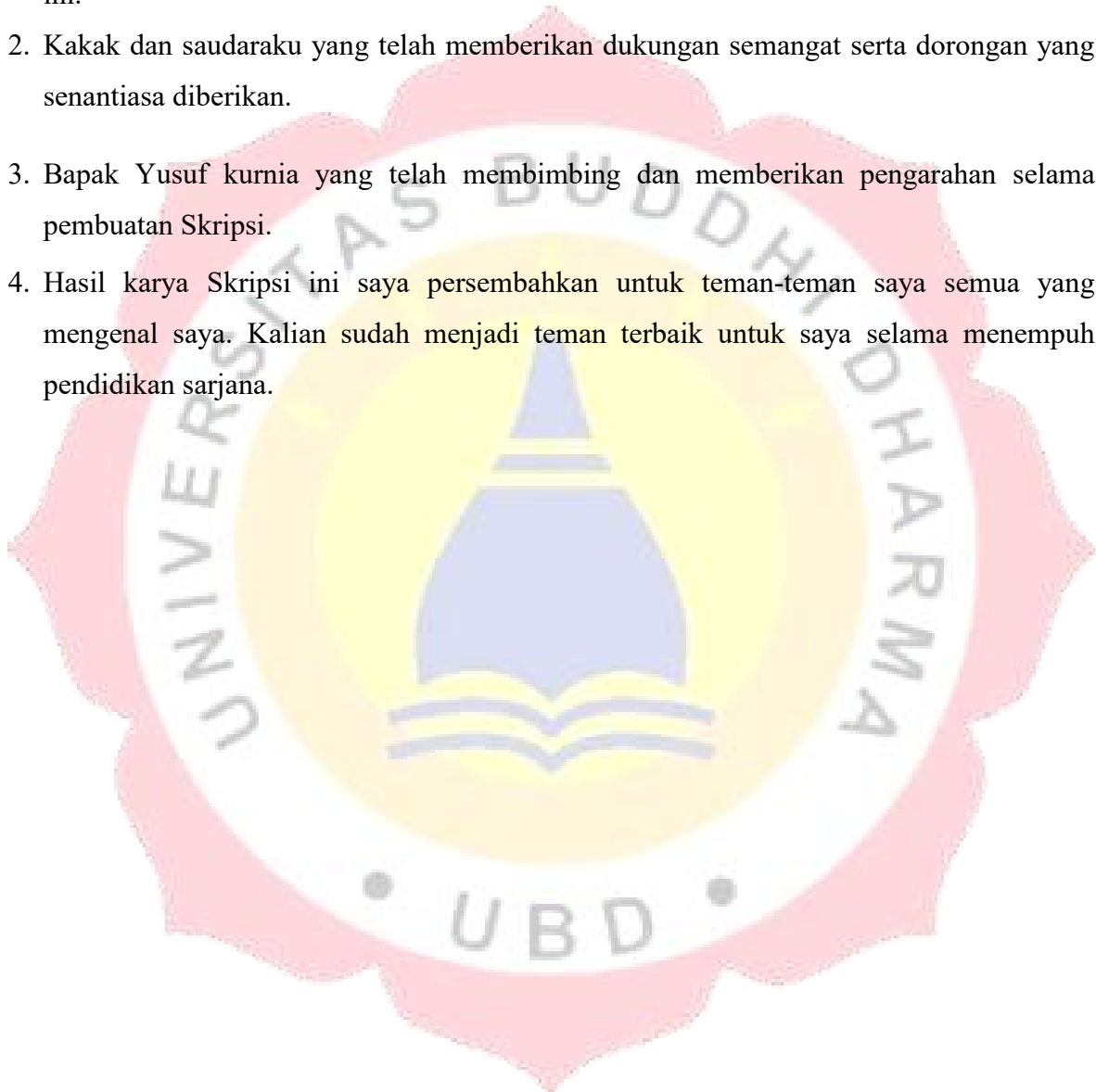
**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS BUDDHI DHARMA
TANGERANG**

2025

LEMBAR PERSEMBAHAN

Dengan mengucap puji syukur kepada Tuhan Yang Maha Esa, Skripsi ini kupersembahkan untuk:

1. Skripsi ini saya persembahkan untuk Ayah dan Ibu, terima kasih atas doa, semangat, motivasi, pengorbanan, nasihat serta kasih sayang yang tidak pernah henti sampai saat ini.
2. Kakak dan saudaraku yang telah memberikan dukungan semangat serta dorongan yang senantiasa diberikan.
3. Bapak Yusuf kurnia yang telah membimbing dan memberikan pengarahan selama pembuatan Skripsi.
4. Hasil karya Skripsi ini saya persembahkan untuk teman-teman saya semua yang mengenal saya. Kalian sudah menjadi teman terbaik untuk saya selama menempuh pendidikan sarjana.



UNIVERSITAS BUDDHI DHARMA
LEMBAR PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini.

NIM : 20201000027

Nama : Riski Ronaldo Rejeki Tua

Jenjang Studi : Strata 1

Program Studi : Teknik Informatika

Peminatan : *Database Development*

Dengan ini saya menyatakan bahwa:

1. Skripsi ini adalah asli dan belum pernah diajukan untuk mendapat gelar akademik Sarjana atau kelengkapan studi, baik di Universitas Buddhi Dharma maupun di Perguruan Tinggi lainnya.
2. Skripsi ini saya buat sendiri tanpa bantuan dari pihak lain, kecuali arahan dosen pembimbing.
3. Dalam Skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dan dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan daftar pustaka.
4. Dalam Skripsi ini tidak terdapat pemalsuan (kebohongan), seperti buku, artikel, jurnal, data sekunder, pengolahan data, dan pemalsuan tanda tangan dosen atau Ketua Program Studi Universitas Buddhi Dharma yang dibuktikan dengan keasliannya.
5. Lembar pernyataan ini saya buat dengan sesungguhnya, tanpa paksaan dan apabila dikemudian hari atau pada waktu lainnya terdapat penyimpangan dan ketidak benaran dalam pernyataan ini, saya bersedia menerima sanksi akademik berupa pencabutan gelar akademik yang telah saya peroleh karena Skripsi ini serta sanksi lainnya sesuai dengan peraturan dan norma yang berlaku.

Tangerang, 4 Februari 2025

Yang membuat pernyataan,



Riski Ronaldo Rejeki Tua
NIM: 20201000027

UNIVERSITAS BUDDHI DHARMA

LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini.

Nim : 20201000027
Nama : Riski Ronaldo Rejeki Tua
Jenjang Studi : Strata 1
Program Studi : Teknik Informatika
Konsentrasi : Database Development

Dengan ini menyetujui untuk memberikan ijin kepada pihak Universitas Buddhi Dharma Fakultas Sains & Teknologi, Hak bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right*) atas karya ilmiah saya yang berjudul "*Analisis Perbandingan Kinerja Algoritma Support Vector Machine , Decision Tree dan Naïve Bayes Untuk Pengelompokan Spesies iris dalam Konteks Klasifikasi*".

Dengan Hak Bebas Royalti Non-Eksklusif ini pihak, Universitas Buddhi Dharma berhak menyimpan, mengalih-media atau format-kan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis atau pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Universitas Buddhi Dharma, segala bentuk tuntutan hukum yang timbul atas Hak Cipta atas karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 4 Februari 2025

Yang membuat pernyataan,



Riski Ronaldo Rejeki Tua
NIM: 20201000027

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN PEMBIMBING

ANALISIS PERBANDINGAN KINERJA ALGORITMA *SUPPORT*
***VECTOR MACHINE, DECISION TREE, DAN NAÏVE BAYES* UNTUK**
PENGELOMPOKAN SPESIES *IRIS* DALAM KONTEKS
KLASIFIKASI

Dibuat Oleh:

NIM : 20201000027

Nama : Riski Ronaldo Rejeki Tua

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian

Komprehensif

Program Studi Teknik Informatika

Database Development

2025

Disahkan oleh,

Tangerang, 4 Februari 2025

Pembimbing,


Yusuf Kurnia, S.Kom., M.Kom.
NIDN: 0419128701

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN TIM PENGUJI

Nama : Riski Ronaldo Rejeki Tua

NIM : 20201000027

Fakultas : Sains & Teknologi

Judul Skripsi : Analisis Perbandingan Kinerja Algoritma *Support Vector Machine*, *Decision Tree*, dan *Naïve Bayes* Untuk Pengelompokan Spesies *iris* Dalam Konteks Klasifikasi.

Dinyatakan LULUS setelah mempertahankan di depan Tim Penguji pada hari Selasa, 04-02-2025.

Nama Penguji:

Tanda Tangan:

Ketua Sidang : **Dram Renaldi, S.kom., M.kom**
0411019001

Penguji 1 : **Muhammad Subhana,**
S.kom., M.kom
0401089401

Penguji 2 : **Yusuf Kurnia, S.kom., M.kom**
0419128701

Mengetahui,

Dekan Fakultas Sains dan Teknologi


Dr. Yakub, S.Kom., M.Kom., M.M.
NIDN: 0304056901

UNIVERSITAS BUDDHI DHARMA
LEMBAR PENGESAHAN SKRIPSI

**ANALISIS PERBANDINGAN KINERJA ALGORITMA *SUPPORT VECTOR MACHINE* , *DECISION TREE*, DAN *NAIVE BAYES* UNTUK
PENGELOMPOKAN SPESIES *IRIS* DALAM KONTEKS KLASIFIKASI**

Dibuat Oleh:

NIM : 20201000027

Nama : Riski Ronaldo Rejeki Tua

Telah disetujui untuk dipertahankan di hadapan Tim Penguji Ujian
Komprehensif

Program Studi Teknik Informatika

Database Development

2024/2025

Disahkan oleh,

Tangerang, 4 Februari 2025

Dekan,

Ketua Program Studi,


Dr. Yakub, S.Kom., M.Kom., M.M.
NIDN: 0304056901


Hartana Wijaya, S.Kom., M.Kom.
NIDN: 0412058102

KATA PENGANTAR

Puji dan Syukur kepada Tuhan Yang Maha Esa saya panjatkan, yang telah memberi hikmah dan kesehatan bagi saya sehingga saya dapat menyusun dan menyelesaikan karya ilmiah ini (Skripsi). Dengan judul *“Analisis Perbandingan Kinerja Algoritma Support Vector Machine, Decision Tree, dan Naïve Bayes Untuk Pengelompokan Spesies iris Dalam Konteks Klasifikasi”*.

Tujuan utama dari penyusunan karya ilmiah ini (Skripsi) adalah sebagai satu syarat untuk menyelesaikan program pendidikan Strata 1 (Sarjana) dengan Program Studi Teknik Informatika di Universitas Buddhi Dharma.

Dalam proses penyusunan karya ilmiah ini (Skripsi) saya banyak menerima masukan dan bantuan dari berbagai pihak, maka pada kesempatan kali ini saya akan menyampaikan banyak terimakasih yang sebesar – besarnya kepada:

1. Ibu. Dr. Limajatini, SE., MM., BKP, Selaku Rektor Universitas Buddhi Dharma.
2. Bapak. Dr. Yakub, S.Kom., M.Kom., M.M. Sebagai Dekan Fakultas Sains dan Teknologi Universitas Buddhi Dharma.
3. Bapak. Rudy Arijanto, S.Kom., M.Kom, Selaku Wakil Dekan Fakultas Sains & Teknologi Universitas Buddhi Dharma.
4. Bapak. Hartana Wijaya, S.Kom., M.Kom, Selaku Ketua Program Studi Teknik Informatika Universitas Buddhi Dharma.
5. Ibu. Desiyanna Lasut, S.Kom., M.Kom, Selaku Dosen Wali saya yang telah membantu dan mendukung saya.
6. Bapak. Yusuf Kurnia, S.Kom., M.Kom, Selaku Dosen Pembimbing saya yang telah membantu dan mendukung saya sehingga saya dapat menyelesaikan karya ilmiah ini.
7. Kedua Orang tua saya serta keluarga yang telah memberi dukungan kepada saya, sehingga saya berhasil menyelesaikan karya ilmiah ini.
8. Serta Teman – teman saya yang selalu membantu serta memberi semangat kepada saya, sehingga saya berhasil menyelesaikan karya ilmiah ini.

Saya sebagai penulis dan penyusun karya ilmiah ini (Skripsi) menyadari bahwa masih banyak kekurangan dalam penyusunan karya ilmiah ini, baik yang disengaja maupun tidak disengaja. Maka dari itulah saya sebagai penulis memohon maaf atas segala kekurangan tersebut. Dan saya sebagai penulis dan penyusun karya ilmiah ini (Skripsi) berharap semoga dengan selesainya penyusunan karya ilmiah ini dapat membantu serta bermanfaat bagi instansi, dan masyarakat luas. Tidak lupa saya ucapkan terimakasih banyak.

Tangerang, 4 Februari 2024

Penulis

ANALISIS PERBANDINGAN KINERJA ALGORITMA SUPPORT VECTOR
MACHINE, DECISION TREE, DAN NAIVE BAYES UNTUK PENGELOMPOKAN
SPESIES IRIS DALAM KONTEKS KLASIFIKASI

101 Halaman + xv / 13 Tabel / 31 Gambar

ABSTRAK

Kemampuan untuk mengelola dan menganalisis data menjadi aspek penting dalam era digital saat ini. Klasifikasi merupakan salah satu metode yang memungkinkan pengelompokan objek berdasarkan pola tertentu. Penelitian ini bertujuan membandingkan kinerja tiga algoritma klasifikasi, yaitu *Support Vector Machine (SVM)*, *Decision Tree*, dan *Naive Bayes*, dalam mengklasifikasikan spesies bunga *iris*. Dataset *iris* dipilih karena karakteristiknya yang sering digunakan dalam penelitian klasifikasi, mencakup tiga spesies utama: *iris setosa*, *iris versicolor*, dan *iris virginica*. Penelitian ini menggunakan metodologi berbasis pendekatan terstruktur yang mencakup pemahaman masalah, eksplorasi data, pemrosesan, pemodelan, evaluasi, dan penerapan. Implementasi algoritma dilakukan menggunakan *Python* dengan pustaka *scikit-learn*, sedangkan aplikasi berbasis web dikembangkan menggunakan framework *PHP* dan *Laravel* untuk mendukung pengujian hasil model. Hasil penelitian menunjukkan bahwa di antara ketiga algoritma yang dibandingkan, *SVM* memiliki kinerja paling unggul, terutama dalam hal akurasi, *presisi*, *recall*, dan *F1-score*. Keunggulan ini disebabkan oleh kemampuan *SVM* dalam menangani data yang tidak terpisah secara linear dengan menggunakan *kernel trick*, sehingga mampu membentuk *hyperplane* optimal yang memisahkan kelas dengan margin maksimal. Sebaliknya, *Decision Tree* cenderung mengalami *overfitting* pada data yang kompleks, sementara *Naive Bayes* mengasumsikan independensi fitur yang tidak selalu sesuai dengan karakteristik dataset *iris*. Aplikasi berbasis web yang dibangun mampu memproses gambar bunga yang diunggah untuk menghasilkan prediksi spesies dengan tingkat keakuratan yang tinggi. Temuan ini menunjukkan potensi integrasi algoritma pembelajaran mesin dengan aplikasi berbasis web sebagai solusi praktis dalam pengelompokan dan klasifikasi data secara efisien.

Kata kunci: Klasifikasi, *Machine Learning*, *Spesies iris*, *Support Vector Machine*.

COMPARATIVE ANALYSIS OF THE PERFORMANCE OF THE SUPPORT VECTOR ALGORITHM

MACHINE, DECISION TREE, AND NAIVE BAYES ALGORITHMS FOR THE CLASSIFICATION OF

IRIS SPECIES IN THE CONTEXT OF CLASSIFICATION

101 Pages + xv / 13 Tables / 31 Figures

ABSTRACT

The ability to manage and analyze data is an important aspect in today's digital era. Classification is one method that allows grouping of objects based on certain patterns. This study aims to compare the performance of three classification algorithms, namely Support Vector Machine (SVM), Decision Tree, and Naive Bayes, in classifying iris species. The iris dataset was chosen due to its characteristics that are often used in classification research, including three main species: iris setosa, iris versicolor, and iris virginica. This research uses a methodology based on a structured approach that includes problem understanding, data exploration, processing, modeling, evaluation, and deployment. Algorithm implementation was conducted using Python with the scikit-learn library, while the web-based application was developed using PHP and Laravel frameworks to support testing of model results. The results show that among the three algorithms compared, SVM has the most superior performance, especially in terms of accuracy, precision, recall, and F1-score. This superiority is due to SVM's ability to handle data that is not linearly separated by using kernel trick, so that it is able to form an optimal hyperplane that separates classes with maximum margin. In contrast, Decision Tree tends to experience overfitting on complex data, while Naive Bayes assumes feature independence which is not always appropriate.

Keywords: Classification, Machine Learning, iris Species, Support Vector Machine.

DAFTAR ISI

LEMBAR JUDUL LUAR SKRIPSI	i
LEMBAR JUDUL DALAM SKRIPSI	ii
LEMBAR PERSEMBAHAN	iii
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	iv
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
LEMBAR PENGESAHAN PEMBIMBING	vi
LEMBAR PENGESAHAN SKRIPSI	vii
LEMBAR PENGESAHAN TIM PENGUJI	viii
KATA PENGANTAR	ix
ABSTRAK	x
<i>ABSTRACT</i>	xi
DAFTAR ISI	xii
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	2
1.3 Ruang Lingkup Penelitian	2
1.4 Tujuan dan Manfaat Penulisan	4
1.4.1 Tujuan	4

1.4.2	Manfaat	4
1.5	Sistematika Penulisan	5
BAB II	LANDASAN TEORI	6
2.1	Teori Khusus	6
2.1.1	Klasifikasi	6
2.1.2	<i>Support Vector Machine</i>	7
2.1.3	<i>Decision Tree</i>	9
2.1.4	<i>Naive Bayes</i>	10
2.1.5	<i>Flowchart</i>	13
2.2	Teori Perancangan	15
2.2.1	<i>Visual Studio Code</i>	15
2.2.2	<i>Python</i>	15
2.2.3	<i>Scikit-learn</i>	16
2.2.4	<i>Cross-industry Standard Process for Data Mining(CRISP-DM)</i>	17
2.2.5	<i>PHP (Hypertext Preprocessor)</i>	17
2.2.6	<i>Laravel</i>	18
2.2.7	<i>Javascript</i>	18
2.2.8	<i>CSS</i>	19
2.3	Tinjauan Studi	20
2.3.1	Penelitian Terdahulu	20
2.3.2	Rangkuman Model Penelitian	40
2.4	Kerangka Pemikiran	49
BAB III	METODOLOGI PENELITIAN	50
3.1	<i>CRISP-DM</i>	50
3.1.1	<i>Business Understanding</i>	50
3.1.2	<i>Data Understanding</i>	52
3.1.3	<i>Data Preparation</i>	54
3.1.4	<i>Modeling</i>	57
3.1.5	<i>Evaluation</i>	61
3.1.6	<i>Deployment</i>	62
3.2	Perancangan <i>Website</i>	64
3.3	<i>IRD (Initial Requirement Document)</i>	65
3.4	Diagram Alur Penelitian	67
BAB IV	ANALISIS & HASIL	69

4.1	Hitung Manual	69
4.2	Spesifikasi Hardware & Software	70
4.2.1	Perangkat Keras	70
4.2.2	Perangkat Lunak	70
4.3	Implementasi Algoritma	71
4.3.1	<i>Support Vector Machine (SVM)</i>	72
4.3.2	<i>Decision Tree</i>	72
4.3.3	<i>Naive Bayes</i>	73
4.4	Impelentasi Hasil dan Dokumentasi Aplikasi	74
4.4.1	Hasil Metrik Evaluasi Kinerja	74
4.4.2	Diskusi Hasil	75
4.4.3	Dokumentasi Aplikasi	76
4.4.4	Tampilan Registrasi	77
4.4.5	Tampilan <i>Login</i>	78
4.4.6	Tampilan Menu Utama	79
4.4.7	Tampilan pengaturan	80
4.4.8	Tampilan Hasil Pengujian	81
4.5	Pengujian <i>Blackbox Testing</i>	83
4.6	Analisis Data Kuesioner	83
4.6.1	Hasil Kuesioner dan Demografi Responden	83
BAB V	SIMPULAN & SARAN	88
5.1	Simpulan	88
5.2	Saran	89
DAFTAR PUSTAKA		90
LAMPIRAN		91

DAFTAR GAMBAR

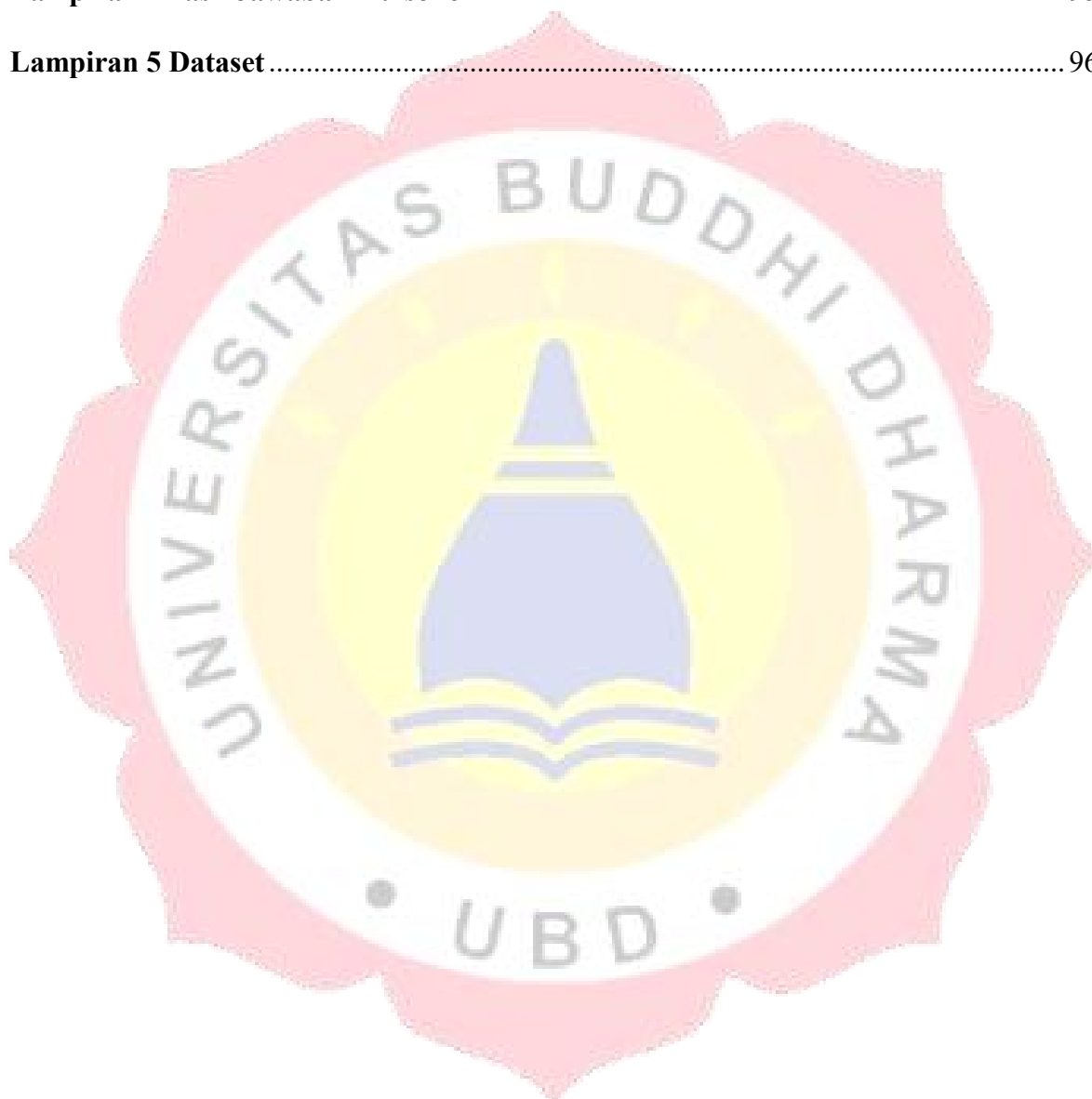
Gambar 2.1 <i>Ilustrasi Hyperplane</i>	8
Gambar 2.2 <i>Flowchart</i>	14
Gambar 2.3 Kerangka Pemikiran	49
Gambar 3. 1 Codingan Dataset	53
Gambar 3.2 Sample Dataset	53
Gambar 3.3 <i>import libray scikit-learn</i>	57
Gambar 3.4 <i>identifikasi dataset</i>	57
Gambar 3.5 hasil <i>identifikasi</i>	57
Gambar 3.6 <i>check dataset</i>	58
Gambar 3.7 hasil <i>check dataset</i>	58
Gambar 3.8 Penghilangan <i>header</i>	58
Gambar 3.9 Hasil Penghilangan <i>Header</i>	58
Gambar 3.10 <i>Konversi label ke numerik</i>	59
Gambar 3.11 Hasil <i>Konversi</i>	59
Gambar 3.12 Pembagian Data	59
Gambar 3.13 Hasil Pembagian Data	59
Gambar 3.14 Pelatihan Algoritma	60
Gambar 3.15 Hasil Pelatihan	60
Gambar 3.16 Metrik Evaluasi	61
Gambar 3.17 Hasil Evaluasi	62
Gambar 3.18 Hasil <i>Visualisasi</i>	62
Gambar 3.19 <i>Wireframe Website</i>	64
Gambar 3.20 Diagram Metodologi Penelitian	67
Gambar 4.1 <i>Registrasi</i>	78
Gambar 4.2 <i>Login</i>	78
Gambar 4.3 Menu Utama Aplikasi	79
Gambar 4.4 Pengaturan	80
Gambar 4.5 <i>Input Gambar</i>	81
Gambar 4.6 Hasil Klasifikasi	82
Gambar 4.7 Hasil Grafik Rata-rata Skor Kuisoner	84
Gambar 4.8 Hasil Grafik Skor Setiap Pertanyaan	85
Gambar 4.9 Hasil Grafik Keseluruhan	86

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	20
Tabel 2.2 Rangkuman Penelitian Terdahulu	40
Tabel 3.1 Struktur Dataset	55
Tabel 3.2 Dataset tambahan	56
Tabel 4.1 Sample Dataset Hitung Manual	69
Tabel 4.2 Spesifikasi Perangkat Keras	70
Tabel 4.3 Spesifikasi Perangkat Lunak	70
Tabel 4.4 Pembagian Dataset	71
Tabel 4.5 Implementasi <i>SVM</i>	72
Tabel 4.6 Implementasi <i>Decision Tree</i>	72
Tabel 4.7 Implementasi <i>Naive Bayes</i>	73
Tabel 4.8 Hasil Perbandingan Algoritma	74
Tabel 4.9 Hasil Pengujian <i>Blackbox</i>	83

DAFTAR LAMPIRAN

Lampiran 1 Daftar Riwayat Hidup	92
Lampiran 2 Kartu Bimbingan Skripsi	93
Lampiran 3 Link Uji Coba Kuisner	94
Lampiran 4 Hasil Jawaban Kuisner	95
Lampiran 5 Dataset	96



BAB I

PENDAHULUAN

1.1 Latar Belakang

Di zaman virtual saat ini, kemampuan untuk mengelola dan menganalisis data menjadi sangat krusial. Salah satu metode yang digunakan untuk mengategorikan data adalah klasifikasi. Pengelompokan data merupakan salah satu teknik yang paling sering diterapkan dalam analisis data untuk menemukan pola dan mengelompokkan objek ke dalam kategori yang tepat. Dalam beberapa tahun terakhir, pengelompokan data telah menjadi sangat signifikan di berbagai bidang, termasuk biologi, medis, dan teknologi. Klasifikasi adalah bagian dari *Machine Learning* yang bertujuan untuk memprediksi label atau kategori dari data baru berdasarkan model yang telah dilatih menggunakan data dengan label yang sudah diketahui (Asniar, Maulidevi, & Surendro, 2022). Salah satu tugas terpenting dalam ilmu data dan pembelajaran mesin adalah klasifikasi data. Contoh spesifik dari klasifikasi adalah klasifikasi spesies bunga *iris*, yang memiliki beberapa spesies yang dapat dibedakan berdasarkan karakteristik fisiknya, seperti lebar *sepal*, panjang *sepal*, panjang *kelopak*, dan lebar *kelopak* (Santoso et al., 2023). Salah satu dataset yang paling terkenal dan sering digunakan sebagai tolok ukur dalam penelitian klasifikasi adalah *iris*. Dataset ini mencakup tiga spesies *iris* (*iris setosa*, *iris versicolor*, dan *iris virginica*) dengan karakteristik sebagai berikut: lebar *sepal*, panjang *sepal*, panjang *kelopak*, dan lebar *kelopak* (Tuntun et al., 2022).

Dalam penelitian ini, algoritma klasifikasi seperti *Support Vector Machine* (*SVM*), *Decision Tree*, dan *Naive Bayes* dibandingkan. *SVM* bekerja dengan mencari *hyperplane* terbaik yang dapat merepresentasikan kumpulan data dengan margin yang besar. *Decision Tree* menggunakan struktur pohon untuk membuat kesimpulan berdasarkan fitur data. Sebaliknya, *Naive Bayes* adalah algoritma probabilistik

berdasarkan teori Bayes, yang mengasumsikan bahwa setiap kecocokan tidak bergantung pada yang lain (Tuntun et al., 2022). Tujuan dari penelitian ini adalah untuk mengevaluasi kinerja algoritma *SVM*, *Decision Tree*, dan *Naive Bayes* dalam klasifikasi spesies *iris* dan menemukan cara untuk meningkatkan kinerja algoritma yang disebutkan di atas. Untuk menentukan algoritma mana yang paling efektif, evaluasi akan dilakukan dengan menggunakan metrik seperti akurasi, *presisi*, *recall*, dan *F1-score* (Hoffman, n.d.).

1.2 Identifikasi Masalah

Identifikasi masalah yang dapat diidentifikasi dalam penelitian ini yaitu:

1. Cara *SVM*, *Decision Tree*, dan *Naive Bayes* dalam mengklasifikasikan *Spesies iris*?
2. Algoritma mana yang memiliki kinerja terbaik berdasarkan metrik *F1-skor*, *recall*, *akurasi*, dan *presisi*?
3. Bagaimana proses pengolahan data menggunakan metodologi *CRISP-DM*?
4. Apa kelebihan dan kekurangan masing-masing algoritma dalam konteks *Klasifikasi Speies iris*?
5. Bagaimana penggunaan *Python* dan *library scikit-learn* dalam implementasi algoritma *Klasifikasi*?
6. Bagaimana cara pengembangan aplikasi berbasis web ?

1.3 Ruang Lingkup Penelitian

Berdasarkan identifikasi masalah diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Cara (*SVM*, *DT*, dan *NB*) mengklasifikasikan spesies *iris* berdasarkan metrik *F1-skor*, *recall*, *akurasi*, dan *presisi*?

2. Algoritma mana yang menunjukkan kinerja terbaik dalam *Klasifikasi Spesies iris*?
3. Bagaimana penerapan metodologi *CRISP-DM* dapat membantu dalam proses pengolahan data dan *klasifikasi*?
4. Apa saja kelebihan dan kekurangan dari algoritma *SVM*, *Decision Tree*, dan *Naïve Bayes* dalam konteks *Klasifikasi Spesies iris*?
5. Bagaimana cara penggunaan *Python* dan *library scikit-learn* dalam mengimplementasikan algoritma-algoritma tersebut untuk *Klasifikasi*?
6. Bagaimana cara pengembangan aplikasi analisis pengelompokan bunga *iris* dalam konteks *klasifikasi* ?

Berikut ini adalah ruang lingkup yang lebih spesifik:

1. Pemilihan Dataset : Penelitian ini akan menggunakan dataset *iris* Yang terdiri dari tiga spesies bunga *iris* Memberikan informasi yang berguna dalam tugas serupa
2. Implementasi : Penelitian ini akan mengimplementasikan algoritma *SVM*, *Decision Tree*, dan *Naïve Bayes* untuk *Klasifikasi Spesies irish*.
3. Evaluasi : Penelitian ini akan Mengevaluasi kinerja ketiga algoritma berdasarkan hasil evaluasi.
4. Analisis : Penelitian ini akan menganalisis perbandingan kinerja algoritma berdasarkan hasil evaluasi.
5. Metodologi : Penelitian ini akan menggunakan Metodologi *CRISP-DM* sebagai kerangka kerja penelitian
6. Bahasa Pemrograman : Penelitian ini akan memakai *Python* dalam pengolahan dan implementasi algoritma mengembangkan algoritma terpilih dari hasil analisis perbandingan menjadi aplikasi analisa pengelompokan spesies bunga *iris* berbasis web menggunakan bahasa pemrograman *PHP*, *laravel*, *Javascript*, *CSS*.

1.4 Tujuan dan Manfaat Penulisan

1.4.1 Tujuan

Berfokus pada analisis perbandingan algoritma *SVM*, *Decision Tree*, & *Naive Bayes* Mengevaluasi Performa dan Akurasi Algoritma Menilai dan membandingkan tingkat akurasi dan performa dari *SVM*, *DT*, dan *NB* melakukan klasifikasi pada dataset bunga *iris*:

1. Membandingkan kinerja ketiga algoritma.
2. Menerapkan metodologi *CRISP-DM*.
3. Mengembangkan aplikasi berbasis web dari hasil perbandingan kinerja algoritma yang terpilih.

1.4.2 Manfaat

Manfaat penelitian membandingkan algoritma *klasifikasi SVM*, *DT*, & *NB* Dengan pemahaman yang lebih baik tentang perbedaan performa antara algoritma klasifikasi yang berbeda, penelitian ini berpotensi memberikan kontribusi yang signifikan bagi pengembangan teknologi informasi serta aplikasi dalam berbagai bidang industri dan ilmu pengetahuan:

1. Menambah wawasan mengenai *SVM*, *DT*, & *NB* untuk *Klasifikasi*.
2. Memberikan informasi yang berguna dalam tugas serupa.
3. Menyediakan dokumentasi proses pengolahan data menggunakan metodologi *CRISP-DM*.
4. Menjadi referensi bagi penelitian selanjutnya dalam pengembangan dan perbandingan algoritma *Klasifikasi*.
5. Memberikan panduan praktis dalam penggunaan *Python* dan *Library Scikit-learn* untuk implementasi algoritma *Machine Learning*.
6. Memberikan panduan pengembangan aplikasi berbasis web dalam konteks klasifikasi untuk pengelompokan spesies bunga *iris*.

1.5 Sistematika Penulisan

Sebagai acuan urutan rangkaian penulisan secara sistematis yang berguna untuk mempermudah dalam proses penulisan serta memudahkan pembaca dalam memahami konsep penelitian ini:

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai Latar Belakang, Identifikasi Masalah, Tujuan Penelitian, Manfaat Penelitian, Ruang Lingkup Penelitian, dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Bab ini menjelaskan tentang Teori khusus, Teori perancangan, Tinjauan studi, dan Kerangka pemikiran.

BAB III METODELOGI PENELITIAN

Bab ini menjelaskan mengenai, Metodologi penelitian, Perancangan aplikasi web, *IRD*, Diagram alur penelitian.

BAB IV ANALISIS DAN HASIL

Bab ini menjelaskan mengenai hasil dari eksperimen yang dilakukan serta pengembangan aplikasi berbasis web untuk pengelompokan spesies *iris* dalam konteks klasifikasi.

BAB V SIMPULAN DAN SARAN

Bab ini menguraikan dan menyajikan simpulan serta saran berdasarkan temuan dan analisis yang dihasilkan dari penelitian ini.

BAB II

LANDASAN TEORI

2.1 Teori Khusus

2.1.1 Klasifikasi

Klasifikasi adalah teknik yang digunakan untuk mengidentifikasi objek data dan mengelompokkannya sesuai dengan atribut atau kriteria tertentu ke dalam suatu model dengan menggunakan data jangka panjang yang telah diberi label atau nilai target. Untuk mempermudah proses klasifikasi data, maka diperlukan sebuah sistem yang menggunakan text mining sebagai alternatif. Salah satu teknik yang digunakan untuk klasifikasi adalah text mining, yang merupakan variasi dari data mining yang dilakukan menganalisis data tekstual dalam jumlah besar dan mengimplementasikan hyperparameter, atau variabel yang mempengaruhi model keluaran (Ramadhan, Andarsyah, & Awangga, 2022).

Klasifikasi adalah langkah proses untuk menemukan suatu model atau fungsi yang dapat menjelaskan atau memisahkan berbagai konsep atau kelas data. Hal ini bertujuan untuk memprediksi kelas dari objek yang tidak memiliki label atau identifikasi sebelumnya (Febriani & Sulistiani, 2021).

Dalam konteks klasifikasi, data dipilah-pilah berdasarkan pada kriteria atau atribut tertentu yang diberikan, dan setiap entitas data telah ditandai atau dikategorikan ke dalam kelas-kelas yang sudah ada. Proses ini memungkinkan sistem atau model yang digunakan untuk mengenali pola atau hubungan antara atribut-atribut yang ada dengan kelas-kelas atau label yang sudah terdefinisi (Djamaludin et al., 2022).

Dengan memanfaatkan data yang telah dilabeli, algoritma atau model klasifikasi dapat dilatih untuk memahami dan mengeksplorasi relasi antara fitur-fitur dalam data dengan kelas-kelas atau label yang telah ditetapkan sebelumnya. Setelah melalui tahap pelatihan ini, model tersebut dapat diterapkan pada data baru yang belum memiliki label, dan dengan menggunakan pengetahuan yang telah diperoleh, mampu memprediksi atau mengklasifikasikan entitas baru ke dalam kelas atau label yang sesuai berdasarkan karakteristik atau atribut yang dimiliki (Wahyono et al., 2020).

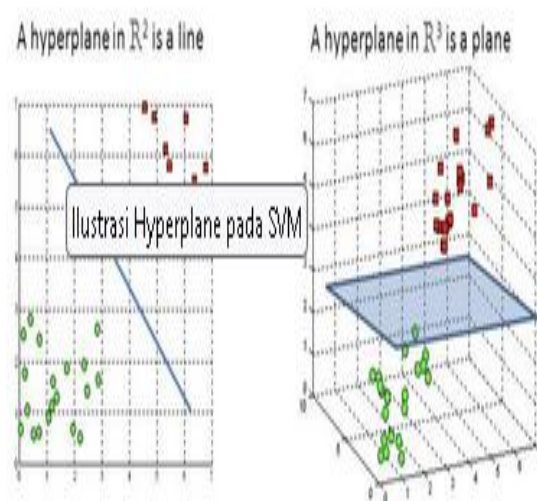
2.1.2 *Support Vector Machine*

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 sebagai teknik klasifikasi yang efisien untuk masalah nonlinier. SVM juga dikenal sebagai teknik pembelajaran mesin, yang paling populer setelah teknik pembelajaran mesin seperti jaringan saraf (NN) digunakan. Baik SVM dan NN telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pola data input dan output yang diinginkan. Konsep SVM secara sederhana dapat dijelaskan sebagai hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada ruang input. SVM bertujuan untuk memaksimalkan jarak antar kelas agar dapat memahami fungsi hyperplane.

Melalui metode ini, SVM dapat memberikan kemampuan generalisasi yang tinggi untuk data yang akan dikumpulkan. Metode yang paling efektif untuk klasifikasi adalah Kelebihan Support Vector Machine (Suyanto, 2017). Support Vector Machine (SVM) muncul sebagai algoritma machine learning yang populer dan banyak digunakan, yang dikenal karena keefektifannya dalam menangani masalah klasifikasi linier dan non-linier. Pada bagian ini,

Akan membahas pentingnya *SVM* dan latar belakangnya, dengan menyoroti penggunaannya dalam aplikasi berbasis data saat ini. *Support Vector Machine (SVM)* adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. *SVM* mencari *hyperplane* terbaik yang dapat menunjukkan dua set data dengan margin yang signifikan. *SVM* juga dapat digunakan untuk menganalisis data *non-linear* dengan menggunakan kernel (Iman & Ujianto, 2020).

Konsep Hyperplane: Tujuan *SVM* adalah untuk mengidentifikasi *hyperplane* terbaik yang menggabungkan dua kelas dalam area yang sesuai. *Hyperplane* adalah dimensi yang sedikit lebih besar dari ruang fitur yang memiliki dua kelas. Fungsi *hyperplane* ini dalam kasus klasifikasi *biner* adalah untuk meminimalkan margin antara dua kelas. *Hyperplane* adalah batas keputusan yang menyematkan dua kelas dalam *SVM*. Titik data yang jatuh di kedua sisi *hyperplane* dapat berhubungan dengan kelas yang berbeda. Dimensi *hyperplane* bergantung pada fitur yang ada di dalam dataset, sehingga jika terdapat dua fitur, maka *hyperplane* akan berbentuk garis lurus. Selain itu, jika ada tiga dimensi, *hyperplane* akan menjadi objek dua dimensi (www.trivusi.web.id, 2022).



Gambar 2. 1 Ilustrasi Hyperplane

2.1.3 *Decision Tree*

Algoritma *Decision Tree*, sebagaimana namanya, menggunakan pohon keputusan untuk melakukan klasifikasi. Titikteratas adalah *root*/akar, kemudian bercabang-cabang, dan ujungnya adalah *leaf*/daun. Setiap cabangmenggambarkan aturan keputusan sementara daunmenggambarkan keputusan akhir yang didapatkan, ketika suatu data baru masuk melalui akar, ada berbagai kemungkinan jalur yang bisa dilewati untuk mencapaisalah satu daun. Pemilihan jalurnya tentu berkaitandengan karakteristik data yang masuk tersebut, jugaberkaitan dengan aturan jika-maka yang sudah ada. Pada akhirnya, data akan mencapai suatu daun sehingga kitamengetahui suatu data masuk ke golongan yang mana (Panca Budiarto, 2018).

Decision Tree digunakan secara luas di banyak bidang ilmu komputer dan bidang-bidang terkait: sarana untuk representasi pengetahuan, sebagai pengklasifikasi, dan sebagai algoritma untuk menyelesaikan berbagai masalah optimasi kombinatorial, geometri komputasi, (Azad, Chikalov, Hussain, & Moshkov, 2022).

Konsep *Decision Tree*:

- *Node dan Leaf* (Daun): *Node* Mewakili keputusan atau pemisahan pada suatu atribut. Setiap *node* mengandung sebuah atribut dan kriteria pemisahan. *Leaf* Terminal *node* yang menghasilkan prediksi kelas. Setiap *leaf* berhubungan dengan satu kelas atau nilai prediksi (Sutoyo, 2018).
- *Root Node* (*Node Akar*): *Node* pertama dalam pohon keputusan yang memulai proses pemisahan. Ini mengandung atribut dan kriteria pemisahan pertama (Prasetia, Syahril, & Kustini, 2022).
- *Atribut dan Nilai Atribut*: Atribut Fitur atau variabel yang digunakan

untuk memisahkan data. Misalnya, pada dataset pelanggan, atribut dapat menjadi "Usia" atau "Pendapatan." , Nilai Atribut Nilai spesifik yang dapat diambil oleh suatu atribut. Misalnya, untuk atribut "Usia," nilai atribut mungkin berupa "Muda," "Sedang," atau "Tua." (Dewi, Parwita, & Setiawan, 2021).

- **Kriteria Pemisahan:** Pada setiap *node*, *Decision Tree* memutuskan cara terbaik untuk memisahkan data berdasarkan nilai atribut tertentu. Contoh kriteria pemisahan termasuk "nilai lebih kecil dari X" atau "sama dengan nilai tertentu." (Fhadli & Usman, 2023).
- **Impurity dan Ketidakhomogenan:** *Gini Impurity* Ukuran ketidakhomogenan di suatu *node*. Semakin rendah nilai Gini, semakin homogen data di dalam *node*. *Entropy* Mengukur tingkat ketidakpastian atau kejernihan informasi di suatu *node*. Semakin rendah entropy, semakin homogen data di dalam *node*. *Error Rate* Proporsi kesalahan klasifikasi di dalam suatu *node* (Fhadli & Usman, 2023).
- **Proses Pembentukan Pohon:** *Decision Tree* membangun pohon secara *rekursif* dengan memilih atribut dan kriteria pemisahan terbaik pada setiap *node* (Muhammad, 2023).
- **Prediksi:** Setelah pohon dibangun, prediksi dilakukan dengan melewati data melalui pohon, mengikuti aturan pemisahan di setiap *node*, hingga mencapai *leaf* yang memberikan prediksi kelas (Awomuti et al., 2023).

2.1.4 Naïve Bayes

Naïve Bayes telah terbukti sebagai metode yang mudah digunakan dan efisien untuk klasifikasi dalam analisis *multivariat*. Namun, fitur-fitur biasanya berkorelasi,

sebuah fakta yang melanggar asumsi independensi bersyarat *Naïve Bayes*, dan dapat memperburuk kinerja metode ini. Selain itu, dataset sering kali dicirikan oleh sejumlah besar fitur, yang dapat mempersulit interpretasi hasil serta memperlambat eksekusi metode (Blanquero, Carrizosa, Ramírez-Cobo, & Sillero-Denamiel, 2021).

Metode penelitian terbaik untuk klasifikasi data yang menerapkan algoritma pada *teorema bayes* adalah *Naive Bayes Classifier*. Metode klasifikasi *Naive Bayes* telah banyak digunakan dalam domain penelitian yang bertujuan untuk mengklasifikasi jumlah data yang besar. Suatu jurnal internasional menyebutkan bahwa *Teorema bayes* adalah suatu konsep aturan kemungkinan yang benar dan salah untuk dapat diolah menjadi informasi atau pengetahuan tambahan (Susanti, Walid, & Hoiriyah, 2022).

Berikut rumus teorema bayes: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$

Naive Bayes (NB) terus menjadi salah satu dari 10 algoritma data mining teratas karena kesederhanaan, efisiensi, dan keampuhannya. Dari sekian banyak proposal untuk meningkatkan akurasi *Naive Bayes* dengan melemahkan asumsi independensi fiturnya, pendekatan pembobotan fitur kurang mendapat perhatian dari para peneliti. Selain itu, semua pendekatan pembobotan fitur yang ada hanya memasukkan bobot fitur yang dipelajari ke dalam rumus klasifikasi *Naive Bayes* dan tidak memasukkan bobot fitur yang dipelajari ke dalam estimasi probabilitas bersyaratnya sama sekali (Jiang, Li, Wang, & Zhang, 2016).

Konsep *Naive Bayes* :

- *Teorema bayes: Naive Bayes* didasarkan pada *Teorema bayes*, yang menyatakan hubungan antara probabilitas kondisional dari dua peristiwa (Susanti et al., 2022).
- *Probabilitas Kondisional: Naive Bayes* memanfaatkan probabilitas

kondisional untuk memprediksi kelas suatu instans berdasarkan atribut-atributnya (Zakaria, Julianto, & Bernada, 2019).

- Asumsi "*Naive*": Asumsi dasar dalam *Naive Bayes* adalah independensi yang dikenal sebagai asumsi "*Naive*". Ini berarti setiap atribut dianggap independen terhadap yang lain, meskipun ini seringkali tidak sepenuhnya terjadi di dunia nyata (Informatika & Informasi, 2023).
- Model Klasifikasi: *Naive Bayes* digunakan untuk membangun model klasifikasi di mana setiap kelas diwakili oleh distribusi probabilitas atribut. Model ini memperkirakan probabilitas suatu instans masuk ke dalam kelas tertentu (Zakaria et al., 2019).
- Rumus *Teorema bayes*: Dalam konteks klasifikasi, *Teorema bayes* dapat ditulis sebagai berikut menurut (Susanti et al., 2022) $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.
 $P(A|B)$ = Probabilitas kelas A diberikan atribut B.
 $P(B|A)$ = Probabilitas atribut B diberikan kelas A.
 $P(A)$ = Probabilitas kelas A.
 $P(B)$ = Probabilitas atribut B.
- Estimasi Parameter: *Naive Bayes* memerlukan estimasi parameter seperti $P(B|A)$ dan $P(A)$ dari data latih untuk melakukan prediksi pada data uji (Susanti et al., 2022)
- Jenis *Naive Bayes*: Multinomial *Naive Bayes* Cocok untuk data yang dideskripsikan oleh distribusi multinomial (mis., data kategorikal). Gaussian *Naive Bayes* Cocok untuk data numerik yang terdistribusi normal. Bernoulli *Naive Bayes* Cocok untuk data biner (0/1) (Sabrani,

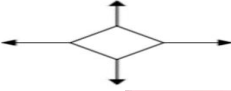
Gede Putu Wirarama Wedashwara, & Bimantoro, 2020).

- Prediksi Kelas: *Naive Bayes* sering digunakan dalam klasifikasi teks, khususnya dalam analisis sentimen dan kategorisasi dokumen (Susanti et al., 2022).
- Pengolahan Teks: *Naive Bayes* sering digunakan dalam klasifikasi teks, khususnya dalam analisis sentimen dan kategorisasi dokumen (Zakaria et al., 2019).

2.1.5 *Flowchart*

Flowchart adalah Teknik analitis bergambar yang di gunakan untuk menjelaskan beberapa aspek dari sistem informasi secara jelas, ringkas, dan logis. Flowchart mencatat cara proses bisnis dilakukan dan cara dokumen mengalir melalui organisasi (Tuasamu et al., 2023).

Dalam *flowchart*, setiap simbol atau bentuk geometris mewakili langkah-langkah atau aktivitas dalam suatu proses. Garis-garis penghubung antara simbol-simbol tersebut mengindikasikan arus atau aliran dari satu langkah ke langkah berikutnya, menjelaskan urutan dari suatu tindakan atau kejadian. *Flowchart* juga dapat mencatat bagaimana dokumen atau informasi mengalir melalui berbagai tahapan atau unit dalam organisasi. Berikut adalah simbol-simbol yang digunakan dalam *flowchart*:

SIMBOL	FUNGSI
 Simbol Terminal	Digunakan untuk menunjukkan awal dan akhir dari serangkaian proses yang berhubungan dengan komputer.
 Simbol <i>Input/output</i>	Digunakan untuk menunjukkan operasi <i>input</i> atau <i>output</i> apapun.
 Simbol pemrosesan komputer	Digunakan untuk menunjukkan proses apa pun yang dilakukan oleh sistem komputer.
 Simbol pemrosesan yang ditentukan	Digunakan untuk menunjukkan proses yang tidak secara khusus didefinisikan dalam diagram alur.
 Simbol komentar	Digunakan menuliskan pernyataan penjelasan yang diperlukan untuk memperjelas sesuatu.
 Simbol garis alur	Digunakan untuk menghubungkan simbol-simbol.
 Simbol <i>input/output</i> dokumen	Digunakan ketika <i>input</i> berasal dari dokumen dan <i>output</i> dikirim ke dokumen.
 Simbol keputusan	Digunakan untuk menunjukkan titik dalam proses dimana keputusan harus dibuat untuk menentukan tindakan lebih lanjut
 Simbol penghubung	Digunakan untuk menghubungkan bagian dari diagram alur yang berlanjut pada halaman yang sama .
 Simbol penghubung antar halaman	Digunakan untuk menghubungkan bagian dari diagram alur yang berlanjut ke halaman terpisah.

Gambar 2. 2 Flowchart

(Chaudhuri, 2020)

2.2 Teori Perancangan

2.2.1 *Visual Studio Code*

Dikembangkan oleh *Microsoft* untuk berbagai sistem operasi, adalah *editor* code yang dapat diperluas dan bersifat open source di bawah lisensi *MIT*.

Sebagai proyek *Microsoft* dengan jumlah kontributor tertinggi di *GitHub*, *VS Code* telah meningkatkan reputasi *Microsoft* dan memposisikannya sebagai pemain utama dalam komunitas perangkat lunak. Pengembang dapat menambahkan bahasa pemrograman baru ke lingkungan ini, seperti *Python*, dengan mengunduh dan menginstal *plugin* yang sesuai untuk menyesuaikan kebutuhan mereka. *VS Code* dilengkapi dengan fitur-fitur canggih seperti integrasi mesin penyelesaian kode otomatis, konsol debugging, dan terminal untuk menjalankan perintah server. Dirancang dengan sangat baik, keuntungan utama dari *VS Code* adalah arsitektur berbasis ekstensi yang memungkinkan penambahan komponen tambahan sesuai kebutuhan. Selain itu, sebagai *Integrated Development Environment (IDE)* dari *Microsoft* yang cukup ringan, *VS Code* dapat diperluas dengan mudah (Ismail Setiawan, 2022).

2.2.2 *Python*

Python, sebagai bahasa pemrograman tingkat tinggi, telah menjadi salah satu bahasa yang paling populer dan paling banyak digunakan di dunia. Diciptakan pertama kali oleh Guido van Rossum pada akhir 1980-an dan dirilis pertama kali pada tahun 1991, *Python* dirancang untuk menjadi bahasa yang mudah dibaca, sederhana, dan mudah digunakan. Nama *Python* sendiri bukanlah merujuk pada hewan ular, melainkan menghormati acara komedi televisi Inggris "*Monty Python's Flying Circus*" yang disukai oleh van Rossum (R, 2020). *Python* menonjol dengan sintaksis yang bersih dan ekspresif, yang memungkinkan pengembang untuk mengekspresikan konsep

pemrograman dalam beberapa baris kode. Karakteristik ini menjadikan Python sangat ramah bagi pemula dan memfasilitasi pengembangan solusi kompleks. Selain itu,

Python adalah bahasa yang bersifat multiparadigma, mendukung pemrograman berorientasi objek, pemrograman prosedural, dan pemrograman fungsional (Fávero et al., 2022).

2.2.3 *Scikit-learn*

Scikit-learn, atau disebut juga *sklearn*, adalah perpustakaan (*library*) sumber terbuka (*open-source*) untuk pemrosesan data dan pembelajaran mesin (*machine learning*) yang dikembangkan menggunakan bahasa pemrograman *Python*. *Scikit-learn* dirancang untuk memberikan alat-alat yang efisien dan mudah digunakan untuk analisis data dan pemodelan prediktif. Perpustakaan ini didukung oleh komunitas yang besar dan aktif, peneliti di bidang kecerdasan buatan dan analisis data (VanderPlas, 2022). Proyek ini dimulai pada tahun 2007 sebagai proyek *Google Summer of Code* oleh David Cournapeau. Kemudian pada tahun yang sama, Matthieu Brucher mulai mengerjakan proyek ini sebagai bagian dari tesisnya. Pada tahun 2010, Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, dan Vincent Michel dari *INRIA* mengambil alih kepemimpinan proyek ini dan melakukan rilis publik pertama pada tanggal 1 Februari 2010. Sejak saat itu, beberapa rilis telah muncul mengikuti siklus 3 bulan, dan komunitas internasional yang berkembang pesat telah memimpin pengembangannya (Hackeling, 2017). *Scikit-learn*, sebagai salah satu perpustakaan pemrosesan data dan pembelajaran mesin yang populer dalam ekosistem *Python*, telah mendapatkan pengakuan positif dari para ahli dan praktisi di bidang kecerdasan buatan dan analisis data.

2.2.4 Cross-industry Standard Process for Data Mining(CRISP-DM)

Crisp-DM (Cross-Industry Standard Process for Data Mining) adalah sebuah metodologi yang dirancang untuk mengelola proyek analisis data dan penemuan pengetahuan. *Crisp-DM* dibuat oleh sebuah tim yang terdiri dari para ahli di bidang analisis data dan penemuan pengetahuan. Tim ini dibentuk oleh sebuah organisasi bernama "*And The Data Mining Knowledge Discovery (KDD) Working Group*" pada tahun 1995. Tujuan awal dari tim ini adalah untuk menciptakan sebuah metodologi yang dapat digunakan secara luas oleh para ahli di bidang analisis data dan penemuan pengetahuan (Nick Hotz, 2024).

Tujuan utama dari *Crisp-DM* adalah untuk memberikan sebuah metodologi yang dapat digunakan secara luas untuk mengelola proyek analisis data dan penemuan pengetahuan. Metodologi ini dirancang untuk membantu para ahli dalam mengidentifikasi masalah, mengumpulkan data, menganalisis data, dan menemukan pengetahuan yang berguna dari data tersebut (Schröer, Kruse, & Gómez, 2021).

2.2.5 PHP (Hypertext Preprocessor)

Bahasa pemrograman yang berorientasi pada pengembangan aplikasi web dinamis. Diciptakan oleh Rasmus Lerdorf pada tahun 1994, *PHP* awalnya digunakan untuk melacak pengunjung ke situs pribadinya. Kini, *PHP* menjadi salah satu bahasa pemrograman *server-side* paling populer di dunia.

PHP adalah bahasa pemrograman yang sederhana, fleksibel, dan bertenaga untuk pengembangan aplikasi web. Dengan komunitas global yang besar, *PHP* terus berkembang untuk memenuhi kebutuhan teknologi modern.

Meskipun banyak bahasa lain bermunculan, *PHP* tetap relevan karena kemampuannya yang luas dan dukungan ekosistem yang kuat ('Lab 5 : Basic PHP Tutorial', n.d.).

2.2.6 *Laravel*

Salah satu *framework PHP open-source* yang dirancang untuk mempermudah proses pengembangan aplikasi web modern. *Framework* ini pertama kali diperkenalkan oleh Taylor Otwell pada tahun 2011 dengan tujuan menyediakan alat yang elegan dan efisien untuk membangun aplikasi berbasis web.

Laravel adalah *framework PHP* modern yang menawarkan berbagai fitur untuk mempercepat dan mempermudah pengembangan aplikasi web. Dengan dukungan komunitas yang kuat, dokumentasi yang lengkap, dan fitur bawaan yang canggih, *Laravel* menjadi salah satu pilihan terbaik bagi pengembang untuk membangun aplikasi web yang elegan dan tangguh. (Malinowski et al., 2019).

2.2.7 *Javascript*

Bahasa pemrograman yang digunakan secara luas untuk mengembangkan aplikasi web interaktif. Diperkenalkan pada tahun 1995 oleh Brendan Eich, *Javascript* awalnya bertujuan untuk menambahkan elemen dinamis ke situs web statis. Kini, *Javascript* menjadi fondasi utama teknologi web modern bersama dengan *HTML* dan *CSS*.

Javascript adalah bahasa pemrograman serbaguna yang menjadi inti dari pengembangan web modern. Kemampuannya untuk berjalan di browser maupun server menjadikannya pilihan utama bagi pengembang dalam membangun aplikasi yang dinamis dan interaktif. Dengan komunitas global yang besar dan ekosistem yang luas,

Javascript terus berkembang dan tetap relevan di dunia teknologi yang selalu berubah(R, 2020).

2.2.8 CSS

Cascading Style Sheet (CSS) merupakan sebuah kode program yang memiliki tujuan untuk memperindah serta mengelola tata letak atau desain visual dari halaman web agar terlihat lebih estetik dan menarik. *CSS* merupakan dokumen yang berdiri sendiri dan bisa diintegrasikan ke dalam kode *HTML* atau digunakan sebagai acuan dalam mendefinisikan gaya atau style pada *HTML*. *CSS* memiliki kemampuan yang sangat beragam dibandingkan dengan bahasa pemrograman inti seperti *HTML* dan *PHP*.

Saat menggunakan *CSS*, kita bisa mengontrol berbagai aspek seperti warna teks, jenis font, jarak antar baris, lebar kolom, dan jenis latar belakang yang digunakan. *CSS* juga memungkinkan untuk mendesain tata letak, menyesuaikan tampilan pada berbagai perangkat yang berbeda, serta menerapkan berbagai efek di dalam website. Mudah dipelajari, namun kuat dalam mengatur tampilan dari dokumen *HTML*, mulai dari yang sederhana hingga yang kompleks. Tidak mengherankan jika saat ini *CSS* banyak digunakan di berbagai website, digabungkan dengan *HTML* dan *PHP* untuk menghasilkan tampilan yang menarik dan bervariasi (Noviana, 2022).

2.3 Tinjauan Studi

2.3.1 Penelitian Terdahulu

1. *Comparison of SVM, Naïve Bayes, and Perceptron Algorithms for Sentiment Analysis on Product Reviews on Tokopedia.*

Tabel 2. 1 Penelitian Terdahulu

No	Journal/Paper Data	Keterangan
1	Judul	Perbandingan Algoritma <i>SVM</i> , Naïve Bayes, dan Perceptron untuk Analisis Sentimen pada Ulasan Produk di Tokopedia.
2	Journal	Budidarma Informatics Media
3	Volume dan halaman	7, Nomor 4, dan Page 1850-1859
4	Tanggal & Tahun	01/10/2023
5	Penulis	Muhammad Aulia, Arief Hermawan
6	Publisher	STMIK Budi Darma
7	Research Objectives	membandingkan kinerja tiga algoritma analisis sentimen yang berbeda (<i>Support Vector Machine</i> , Multinomial Naïve Bayes, dan Perceptron) dalam konteks analisis sentimen terkait harga produk di platform <i>e-commerce</i> Tokopedia.
8	Lokasi dan Subjek penelitian	platform <i>E-commerce</i> Tokopedia dan Sentimen Ulasan Produk
9	Perancangan Sistem	Web Scraping, Labeling, Preprocessing Text, Case Folding, Normalize, Remove Stopwords, Tokenizing, Stemming, Model Algoritma Klasifikasi.
10	Hasil Penelitian	The system successfully detects positive and negative emotions in Multinomial Naïve Bayes with 84% accuracy, <i>Support Vector Machine</i> with 80% accuracy, and Perceptron with 81% accuracy.
11	Kekuatan Penelitian	Multinomial Naïve Bayes adalah algoritme yang efektif untuk mengklasifikasikan atribut produk dengan sensitivitas tinggi.
12	Kelemahan Penelitian	berkaitan dengan penemuan informasi kata-kata dalam kategori sentimen.

13	Kesimpulan	Secara keseluruhan, analisis sentimen produk <i>e-commerce</i> menggunakan teknik Multinomial Naïve Bayes telah menunjukkan hasil yang sangat membantu dalam mengidentifikasi sentimen ulasan. Dengan peningkatan akurasi dan konsistensi, sistem ini dapat menjadi alat yang efektif untuk memahami persepsi dan sikap konsumen terhadap produk <i>e-commerce</i> dan membantu pengembangan rencana bisnis yang lebih sukses.
----	------------	--

2. Perbandingan *NB* & *SVM* untuk Klasifikasi Pelanggan Indihome

No	Journal/Paper Data	Keterangan
1	Judul	Perbandingan <i>Naive Bayes</i> dan <i>Support Vector Machine</i> untuk Klasifikasi Pelanggan Indihome
2	Journal	STRING
3	Volume dan halaman	6, Nomor 1
4	Tanggal & Tahun	01/08/2021
5	Penulis	Aan Rohanah, Dwi Latifah Rianti, Betha Nurina Sari
6	Penerbit	Singaperbangsa Karawang University
7	Tujuan Penelitian	Comparing two algorithms (<i>Support Vector Machine</i> , Naïve Bayes,) in the context of indihome customer review classification.
8	Lokasi dan Subjek penelitian	PT.TELKOM dan Klasifikasi Ulasan Pelanggan Indihome
9	Perancangan Sistem	Metode Pengumpulan Data, Pendekatan Analisis dan Pengolahan Data, Pemahaman Bisnis, Pemahaman Data, Persiapan Data, Pemodelan, Penilaian, dan Implementasi.
10	Research Results	Menurut temuan penelitian, IndiHome menerima 45,56% opini positif dan 54,44 persen opini negatif. Kata yang sering muncul dalam ulasan positif adalah “nonton”, sedangkan kata yang sering muncul dalam ulasan negatif adalah “admin”.
11	Kekuatan Penelitian	<i>Support Vector Machine</i> lebih baik dengan nilai accuracy 82,11%, Precision 76,44%, recall 88,01%, dan nilai AUC 0,909.
12	Kelemahan Penelitian	terkait dengan temuan informasi kata-

		kata pada ulasan
13	Kesimpulan	Naïve Bayes and <i>Support Vector Machine</i> algorithms can be used to classify Indihome customer reviews from Twitter. Vector Machine can be used for the classification of Indihome customer reviews from Twitter social media.

3. *Comparison of Decision Tree and Support Vector Machine Methods for Sentiment Analysis on Instagram Regarding PSSI Performance*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Comparison of <i>Decision Tree</i> and <i>Support Vector Machine</i> Methods for Sentiment Analysis on Instagram Regarding PSSI Performance
2	Jurnal	e-Proceeding of Engineering
3	Volume dan halaman	7, No.3 Page 9936
4	Tanggal & Tahun	01/12/2020
5	Penulis	Muhammad Fadli Asshiddiqi , Kemas Muslim Lhaksmana
6	Penerbit	Singaperbangsa Karawang University
7	Tujuan Penelitian	membandingkan dua algoritma (<i>Support Vector Machine</i> , Naïve Bayes,)dalam konteks klasifikasi ulasan pelanggan indihome
8	Lokasi dan Subjek penelitian	Telkom University
9	Perancangan Sistem	membandingkan dua algoritma <i>DCS</i> & <i>NB</i>
10	Hasil Penelitian	Pengumpulan Data, Pelabelan Data, Preprocessing, Term weighting, Klasifikasi, Confusion Matrix.
11	Kekuatan Penelitian	Based on the test results, data sharing with a composition of 80% for training and 20% for testing provides the best results. In the <i>Decision Tree</i> algorithm using TF-IDF, the accuracy obtained reached 87.45%, with a precision value of 87.72% and F-Measure of 89.69%. Meanwhile, in the <i>SVM</i> algorithm with TF-IDF, the same data composition resulted in an accuracy of 94.36%, precision of 96.78%, <i>recall</i> of 94.30%, and <i>F1-score</i> of 95.53%.
12	Kelemahan Penelitian	terkait dengan penggunaan data

13	Kesimpulan	The <i>SVM</i> algorithm with TF-IDF shows superior performance compared to DT using TF-IDF. However, both algorithms have provided excellent results as the accuracy exceeds 80%. Therefore, for the case of sentiment analysis on Instagram comments, the use of <i>SVM</i> with TF-IDF is more recommended because it produces the best performance.
----	------------	---

4. *Email Sentiment Analysis Using Labeling with K-Means & Classification by Support Vector Machine*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Email Sentiment Analysis Using Labeling with K-Means & Classification by <i>Support Vector Machine</i>
2	Jurnal	An International Journal
3	Volume and page	VOL. 49, NO. 3, 181–199
4	Date & Year	23 May 2018
5	Author	Sisi Liu dan Ickjai Lee
6	Penerbit	Cybernetics and Systems
7	Tujuan Penelitian	Berfokus pada bagian klasifikasi sentimen dari kerangka kerja, sebuah metode identifikasi fitur dasar menggunakan daftar kata opini yang telah ditentukan
8	Lokasi dan Subjek penelitian	Email dan analisis sentiment.
9	Perancangan Sistem	pemilihan fitur, en, pengelompokan sentimen dan pendekatan klasifikasi sentimen dibandingkan, dengan model TF-IDF, pengelompokan k-means, dan pengklasifikasian <i>SVM</i> yang diidentifikasi sebagai teknik yang paling sesuai untuk analisis sentimen Email

10	Hasil Penelitian	pendekatan klasifikasi sentimen, yang menunjukkan bahwa TF-IDF mengungguli metode ekstraksi fitur lainnya, pengelompokan k-means berkinerja baik untuk pengelompokan sentimen Email, dan SVM secara konsisten menghasilkan hasil yang baik untuk klasifikasi sentimen Email.
11	Kekuatan Penelitian	Secara keseluruhan, penelitian ini memberikan kontribusi yang signifikan pada bidang analisis sentimen dengan mengatasi kurangnya studi menyeluruh tentang analisis sentimen Email dan menyediakan kerangka kerja yang kuat untuk menganalisis sentimen dalam data Email.
12	Kelemahan Penelitian	Kelemahan dari penelitian ini terletak pada terbatasnya pembahasan mengenai tantangan dan kesulitan spesifik yang dihadapi selama implementasi kerangka kerja analisis sentimen hibrida yang diusulkan untuk data Email.
13	Kesimpulan	Analisis Sentimen Email dan mengusulkan kerangka kerja hibrida untuk menganalisis sentimen dalam email melalui pelabelan k-means dan klasifikasi SVM. Studi ini menunjukkan efektivitas penggunaan model pembobotan TF-IDF untuk ekstraksi fitur, pengelompokan k-means untuk pengelompokan sentimen, dan SVM untuk klasifikasi sentimen

5. *Sentiment Analysis on Short Texts Using SVM and VSM-based Multiclass Semantic Classification*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Sentiment Analysis on Short Texts Using SVM and VSM-based Multiclass Semantic Classification
2	Jurnal	Kecerdasan Buatan Terapan
3	Volume dan halaman	Volume 38, 2024 - Issue 1
4	Tanggal & Tahun	01/03/2024

5	Penulis	Ata Jahangir Moshayedi, K. Suresh Kumar, A.S. Radha Mani, T. Ananth Kumar, Ahmad Jalili, Mehdi Gheisari, Yasir Malik, and Hsing-Chung Chen
6	Penerbit	Cybernetics and Systems
7	Tujuan Penelitian	Berfokus pada bagian klasifikasi sentimen dari kerangka kerja, sebuah metode identifikasi fitur dasar menggunakan daftar kata opini yang telah ditentukan
8	Lokasi dan Subjek penelitian	Email dan analisis sentiment.
9	Perancangan Sistem	Pemilihan dan Persiapan Dataset Ekstraksi Fitur Menggunakan EVSM Pemilihan Fitur Implementasi Model Ruang Vektor yang Disempurnakan Metode Peningkatan Bobot Penyisipan Kata Mesin Vektor Pendukung Hibrida (Hybrid Support Vector Machine (HSVM))
10	Hasil Penelitian	Akurasi yang diperoleh adalah 92,78% dengan tingkat sentimen positif 91,33% dan tingkat sentimen negatif 97,32%.
11	Kekuatan Penelitian	Sentimen diklasifikasikan ke dalam kategori positif, negatif, dan netral.
12	Kelemahan Penelitian	Kelemahan dari penelitian ini terletak pada terbatasnya pembahasan mengenai tantangan dan kesulitan spesifik yang dihadapi selama implementasi kerangka kerja analisis sentimen hibrida yang diusulkan untuk data Email.
13	Kesimpulan	Mempertimbangkan metode-metode mutakhir, peningkatan yang luar biasa dalam akurasi dicapai dan ditingkatkan menjadi 92,78%. Dan dengan mempertimbangkan polaritas sentimen, hasil maksimum dicapai untuk sentimen positif dan negatif. Keakuratan tingkat sentimen positif dan tingkat sentimen negatif dimaksimalkan menjadi 91,33% dan 97,32.

6. *Comprehensive Classification of iris Flower Species: A Machine learning Approach*

No	Data Jurnal / Makalah	Keterangan
----	-----------------------	------------

1	Judul	Comprehensive Classification of <i>iris</i> Flower Species: A <i>Machine learning</i> Approach
2	Jurnal	Indonesian Journal of Computer Science
3	Volume dan halaman	Vol. 13 No. 3 ISSN 2549-7286
4	Tanggal & Tahun	01/02/2024
5	Penulis	Renas Rajab Asaad, Adnan Mohsin Abdulazeez
6	Penerbit	Indonesian Journal of Computer Science (IJCS)
7	Tujuan Penelitian	Berfokus pada bagian klasifikasi sentimen dari kerangka kerja, sebuah metode identifikasi fitur dasar menggunakan daftar kata opini yang telah ditentukan
8	Lokasi dan Subjek penelitian	Klasifikasi bunga <i>iris</i>
9	Perancangan Sistem	Preprocessing: Normalization, Feature Selection, Classifiers, Dataset, Evaluation Metrics, Results and discussion
10	Hasil Penelitian	Hasil ini menggarisbawahi potensi transformatif dari metode analitik yang canggih di bidang analisis spesies bunga <i>iris</i> , membuka jalan baru untuk perawatan tanaman yang lebih baik dan hasil pemantauan yang lebih baik.
11	Kekuatan Penelitian	kemampuan luar biasa dari algoritma pembelajaran mesin, terutama XGBoost, dalam klasifikasi akurat dari berbagai kondisi bunga <i>iris</i> dengan menggunakan data spesies <i>iris</i> .
12	Kelemahan Penelitian	Kelemahan dari penelitian ini terletak pada terbatasnya pembahasan mengenai tantangan dan kesulitan spesifik.
13	Kesimpulan	Studi ini secara efektif menunjukkan bagaimana model pembelajaran mesin dapat dengan mahir menavigasi aspek-aspek diagnosis tanaman yang kompleks dan terperinci, khususnya dalam studi spesies bunga <i>iris</i> .

7. Klasifikasi Bunga *iris* dengan Algoritma *Random Forest*

No	Data Jurnal / Makalah	Keterangan
----	-----------------------	------------

1	Judul	Klasifikasi Bunga <i>iris</i> dengan Algoritma Random Forest
2	Jurnal	Indonesian Journal of Computer Science
3	Volume dan halaman	ISSN 2757-7422 Vol. 2 (No. 1), pp. 7-14
4	Tanggal & Tahun	feb 16 2022
5	Penulis	Hilmi Cenk Bayrakçı , Abdullah Burak Keşkekçi, Recep Arslan.
6	Penerbit	Advances in Artificial Intelligence Research (AAIR)
7	Tujuan Penelitian	pengklasifikasi pembelajaran mesin jaringan saraf tiruan dengan tingkat akurasi yang tinggi. Sebagai hasilnya klasifikasi, kinerja model yang dilatih dievaluasi menurut kebingungan matriks, sensitivitas, spesifisitas, akurasi, skor <i>FI</i> , kurva ROC, dan kriteria evaluasi AUC.
8	Lokasi dan Subjek penelitian	Klasifikasi bunga <i>iris</i>
9	Perancangan Sistem	<i>iris</i> Dataset, RF Algorithm, <i>Support Vector Machine</i> Algorithm, Artificial Neural Networks (ANN), Performance Evaluation Criteria.
10	Hasil Penelitian	Hasil ini menggarisbawahi potensi transformatif dari metode analitik yang canggih di bidang analisis spesies bunga <i>iris</i> , membuka jalan baru untuk perawatan tanaman yang lebih baik dan hasil pemantauan yang lebih baik.
11	Kekuatan Penelitian	Dengan hasil yang diperoleh dalam penelitian ini, model berbasis kecerdasan buatan telah diusulkan untuk klasifikasi bunga <i>iris</i> . Dengan model yang diusulkan ini, diharapkan dapat memberikan kontribusi pada literatur akademis untuk aplikasi kecerdasan buatan di bidang pertanian.
12	Kelemahan Penelitian	Kelemahan dari penelitian ini terletak pada terbatasnya pembahasan mengenai tantangan dan kesulitan spesifik.

13	Kesimpulan	Studi ini secara efektif menunjukkan bagaimana model pembelajaran mesin dapat dengan mahir menavigasi aspek-aspek diagnosis tanaman yang kompleks dan terperinci, khususnya dalam studi spesies bunga <i>iris</i> .
----	------------	---

8. Klasifikasi Dataset Bunga *iris* menggunakan Algoritma yang Berbeda

No	Data Jurnal / Makalah	Keterangan
1	Judul	Klasifikasi Dataset Bunga <i>iris</i> menggunakan Algoritma yang Berbeda
2	Jurnal	International Journal of Scientific Research in Mathematical and Statistical Sciences
3	Volume dan halaman	Volume-9,E-ISSN: 2348-4519
4	Tanggal & Tahun	01/12/2020
5	Penulis	U. Honey, S.B. Sogir, S. Hossain, S. Akter, dan S.A. Mithy
6	Penerbit	Jahangirnagar University
7	Tujuan Penelitian	Membandingkan Kinerja Algoritma Klasifikasi: Menganalisis dan mengevaluasi keefektifan berbagai algoritma klasifikasi (<i>machine learning</i>) seperti <i>Support Vector Machine (SVM)</i> , <i>Decision Tree</i> , <i>Naive Bayes</i> , atau lainnya dalam mengklasifikasikan spesies bunga <i>iris</i> berdasarkan fitur dataset.
8	Lokasi dan Subjek penelitian	lokasi: Penelitian dilakukan menggunakan dataset <i>iris</i> yang merupakan dataset standar, tersedia di platform terbuka seperti <i>UCI Machine learning Repository</i> . Subjek: Dataset <i>iris</i> yang terdiri dari 150 data sampel spesies bunga (<i>iris-setosa</i> , <i>iris-versicolor</i> , <i>iris-virginica</i>), masing-masing dengan 4 atribut fitur: panjang dan lebar <i>sepal</i> serta <i>petal</i> .
9	Perancangan Sistem	Data diolah menggunakan algoritma klasifikasi seperti <i>Support Vector Machine (SVM)</i> , <i>Decision Tree</i> , dan <i>Naive Bayes</i> . Tahapan meliputi: Preprocessing dataset untuk memastikan data bersih. Penerapan algoritma dengan parameter tertentu. Evaluasi menggunakan metrik seperti akurasi, <i>presisi</i> , dan <i>recall</i> .
10	Hasil Penelitian	Akurasi Algoritma: <i>SVM</i> sering kali

		memberikan akurasi tertinggi dibandingkan algoritma lain. <i>Decision Tree</i> menunjukkan kinerja yang baik pada data kecil namun rentan terhadap overfitting. <i>Naive Bayes</i> memiliki akurasi moderat tetapi unggul pada kecepatan komputasi.
11	Kekuatan Penelitian	Menggunakan dataset standar yang teruji secara akademis. Membandingkan beberapa algoritma secara komprehensif dengan metrik evaluasi yang relevan. Memberikan wawasan tentang efektivitas algoritma untuk klasifikasi multi-kelas.
12	Kelemahan Penelitian	Dataset yang digunakan relatif kecil (hanya 150 sampel), sehingga hasil mungkin kurang representatif untuk data skala besar. Tidak semua algoritma pembelajaran mesin yang lebih kompleks (seperti Random Forest atau Neural Networks) dianalisis.
13	Kesimpulan	Penelitian menunjukkan bahwa algoritma <i>SVM</i> memiliki kinerja terbaik dalam klasifikasi dataset <i>iris</i> berdasarkan akurasi. Dataset <i>iris</i> tetap menjadi benchmark sederhana namun efektif untuk mengevaluasi algoritma pembelajaran mesin. Hasil dapat digunakan sebagai referensi untuk memilih algoritma yang sesuai dengan kebutuhan klasifikasi multi-kelas.

9. Pengenalan Spesies Bunga *iris* Berbasis Pendekatan Menggunakan Pengklasifikasi Pembelajaran Mesin

No	Data Jurnal / Makalah	Keterangan
1	Judul	Pengenalan Spesies Bunga <i>iris</i> Berbasis Pendekatan Menggunakan Pengklasifikasi Pembelajaran Mesin
2	Jurnal	Jurnal Akademik Universitas Nawroz (AJNU)
3	Volume dan halaman	Vol.11 e-ISSN: 2520-789X
4	Tanggal & Tahun	01/06/2023

5	Penulis	Amar Yehya Hussien, Universitas Van Yuzuncu Yil, Departemen Statistik
6	Penerbit	Jurnal Internasional Ilmu Sosial Humaniora dan Manajemen (IJHSSM)
7	Tujuan Penelitian	Berfokus pada bagian klasifikasi sentimen dari kerangka kerja, sebuah metode identifikasi fitur dasar menggunakan daftar kata opini yang telah ditentukan
8	Lokasi dan Subjek penelitian	menggunakan dataset Fisher's <i>iris</i> Dataset, yang berisi data tentang tiga spesies bunga <i>iris</i> (<i>Setosa</i> , <i>Versicolor</i> , dan <i>Virginica</i>), dengan fokus pada panjang dan lebar kelopak (<i>petal</i>) serta <i>sepal</i> sebagai parameter klasifikasi
9	Perancangan Sistem	Model dikembangkan dengan algoritma pembelajaran mesin seperti RF, <i>SVM</i> , dan ANN. Dataset dibagi menjadi data pelatihan dan pengujian. Metode preprocessing seperti normalisasi data juga dilakukan. Evaluasi kinerja model dilakukan menggunakan metrik seperti ROC-AUC, akurasi, <i>presisi</i> , sensitivitas, dan <i>F1-score</i>
10	Hasil Penelitian	Algoritma RF menghasilkan akurasi tertinggi (97%) dibandingkan dengan <i>SVM</i> dan ANN. Model yang dikembangkan berhasil menunjukkan efisiensi dalam klasifikasi dan potensi untuk diterapkan dalam pengembangan aplikasi web untuk identifikasi spesies bunga
11	Kekuatan Penelitian	Penelitian ini menggunakan metrik evaluasi yang komprehensif untuk membandingkan algoritma yang berbeda. Selain itu, penggunaannya dalam aplikasi berbasis web memungkinkan integrasi yang lebih praktis dan relevan
12	Kelemahan Penelitian	Penelitian ini terbatas pada dataset <i>iris</i> yang bersifat kecil dan sederhana, sehingga mungkin kurang relevan untuk masalah klasifikasi yang lebih kompleks. Selain itu, studi tidak membahas potensi bias atau implikasi etis dari penerapan model

13	Kesimpulan	Penelitian ini menunjukkan bagaimana pembelajaran mesin dapat digunakan untuk klasifikasi bunga <i>iris</i> secara efektif. Algoritma seperti RF memiliki potensi besar untuk diintegrasikan dalam aplikasi otomatis untuk mendukung bidang botani dan ilmu lingkungan. Studi ini juga memberikan dasar yang kuat untuk penelitian lebih lanjut dalam klasifikasi berbasis data
----	------------	---

10. Perbandingan Empiris Algoritma Klasifikasi untuk Dataset Bunga *iris*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Perbandingan Empiris Algoritma Klasifikasi untuk Dataset Bunga <i>iris</i>
2	Jurnal	jurnal International Journal of Scientific Research in Mathematical and Statistical Sciences
3	Volume dan halaman	Volume 9, Edisi 6, 1-10
4	Tanggal & Tahun	01/12/2022
5	Penulis	S.A. Mithy, S. Hossain, S. Akter, U. Honey, dan S.B. Sogir
6	Penerbit	IJCSE, Indore, Indiax`
7	Tujuan Penelitian	Penelitian bertujuan untuk membandingkan kinerja berbagai algoritma klasifikasi, seperti <i>Support Vector Machine (SVM)</i> , Logistic Regression, k-Nearest Neighbors (KNN), Random Forest, K-Means Clustering, dan K-Medoids dalam mengklasifikasikan dataset <i>iris</i>
8	Lokasi dan Subjek penelitian	Penelitian ini dilakukan pada dataset <i>iris</i> yang berasal dari repositori UCI, dengan data yang mencakup tiga spesies bunga (<i>Setosa</i> , <i>Versicolor</i> , <i>Virginica</i>), masing-masing dengan 50 sampel
9	Perancangan Sistem	Penelitian menggunakan <i>Python</i> dan pustaka <i>Scikit-learn</i> untuk mengimplementasikan berbagai algoritma. Sistem dirancang untuk mengevaluasi performa klasifikasi berdasarkan metrik seperti akurasi, <i>presisi</i> , dan <i>recall</i> , serta menggunakan analisis visualisasi untuk mendeteksi pola data

10	Hasil Penelitian	Logistic Regression dan <i>Support Vector Machine</i> menunjukkan hasil terbaik dalam akurasi. Algoritma K-Means dan K-Medoids kurang efektif dibandingkan model supervised seperti Random Forest dan KNN
11	Kekuatan Penelitian	Penelitian menggunakan pendekatan empiris dengan membandingkan beberapa algoritma secara langsung pada dataset yang terkenal. Analisis metrik evaluasi yang menyeluruh memberikan wawasan akurat tentang performa setiap model
12	Kelemahan Penelitian	Studi hanya berfokus pada dataset kecil (<i>iris</i>), sehingga hasilnya mungkin tidak dapat digeneralisasi untuk dataset besar atau kompleks. Penggunaan K-Means dan K-Medoids tidak relevan untuk tugas klasifikasi karena mereka adalah algoritma unsupervised
13	Kesimpulan	Penelitian menyimpulkan bahwa model supervised seperti Logistic Regression dan <i>SVM</i> adalah pilihan terbaik untuk tugas klasifikasi pada dataset <i>iris</i> . Meskipun demikian, algoritma seperti Random Forest dan KNN juga menunjukkan performa yang kompetitif. Penelitian ini memberikan panduan bagi peneliti lain untuk memilih algoritma yang tepat berdasarkan kebutuhan klasifikasi mereka

11. *Iris Flower Species Identification Using Machine learning Approach*

No	Data Jurnal / Makalah	Keterangan
1	Judul	<i>iris Flower Species Identification Using Machine learning Approach</i>
2	Jurnal	Jurnal Internasional Penelitian Ilmiah dalam Ilmu Matematika dan Ilmu Statistik
3	Volume and page	Volume 9, Edisi 6, 1-10
4	Tanggal & Tahun	01/01/2022
5	Penulis	Shilpi Jain
6	Penerbit	IJCSE, Indore, India

7	Tujuan Penelitian	Penelitian ini bertujuan untuk mengidentifikasi spesies bunga <i>iris</i> menggunakan pendekatan <i>machine learning</i> , dengan fokus pada pengklasifikasian berdasarkan pengukuran atribut bunga seperti ukuran <i>petal</i> dan <i>sepal</i> .
8	Lokasi dan Subjek penelitian	Penelitian ini menggunakan dataset bunga <i>iris</i> yang umum tersedia, yang terdiri dari tiga kelas spesies: <i>Setosa</i> , <i>Versicolor</i> , dan <i>Virginica</i> . Dataset ini berisi 150 instance dengan 50 instance untuk masing-masing kelas.
9	Perancangan Sistem	segmentasi: Memisahkan objek bunga dari latar belakang. Ekstraksi Fitur: Mengambil ukuran <i>petal</i> dan <i>sepal</i> sebagai fitur untuk klasifikasi. Klasifikasi: Menggunakan algoritma <i>machine learning</i> seperti Neural Network, Logistic Regression, <i>Support Vector Machine (SVM)</i> , dan k-Nearest Neighbors (KNN) untuk mengklasifikasikan spesies.
10	Hasil Penelitian	model <i>machine learning</i> yang dilatih dapat memprediksi spesies bunga <i>iris</i> dengan akurasi tinggi. Algoritma <i>SVM</i> dan Neural Network menunjukkan performa terbaik dalam klasifikasi.
11	Kekuatan Penelitian	Akurasi Tinggi: Penggunaan berbagai algoritma <i>machine learning</i> memberikan hasil yang akurat dalam identifikasi spesies. Metodologi Jelas: Pendekatan sistematis dalam segmentasi, ekstraksi fitur, dan klasifikasi.
12	Kelemahan Penelitian	Ketergantungan pada Dataset: Hasil sangat bergantung pada kualitas dataset yang digunakan. Keterbatasan Algoritma: Beberapa algoritma mungkin tidak optimal untuk semua jenis data.
13	Kesimpulan	Penelitian ini menunjukkan potensi <i>machine learning</i> dalam mengidentifikasi spesies bunga <i>iris</i> dengan akurasi yang signifikan. Dengan menggunakan dataset yang tepat dan algoritma yang sesuai, identifikasi spesies dapat dilakukan secara otomatis, memberikan kontribusi penting bagi taksonomi botani dan aplikasi dalam bidang kecerdasan buatan.

12. Studi Komparatif tentang Algoritma Klasifikasi untuk Dataset *iris*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Studi Komparatif tentang Algoritma Klasifikasi untuk Dataset <i>iris</i>
2	Jurnal	Jurnal Internasional Teknik Elektro dan Komputer (IJECE)
3	Volume and page	Volume 14, 1959-1968
4	Tanggal & Tahun	01/04/2022
5	Penulis	Institute of Advanced Engineering and Science (IAES)
6	Penerbit	Institute of Advanced Engineering and Science (IAES)
7	Tujuan Penelitian	Penelitian ini bertujuan untuk membandingkan efektivitas beberapa algoritma klasifikasi dalam mengidentifikasi spesies bunga <i>iris</i> menggunakan dataset <i>iris</i> , yang terdiri dari tiga spesies: <i>Setosa</i> , <i>Versicolor</i> , dan <i>Virginica</i> .
8	Lokasi dan Subjek penelitian	Subjek penelitian adalah dataset <i>iris</i> yang terkenal, yang berisi 150 data dengan empat fitur (panjang dan lebar <i>sepal</i> serta panjang dan lebar <i>petal</i>).
9	Perancangan Sistem	Sistem dirancang dengan membagi dataset menjadi dua bagian: data pelatihan dan data pengujian. Dua algoritma klasifikasi utama yang digunakan adalah Gaussian Naïve Bayes (GNB) dan <i>Decision Tree</i> (DT). Kinerja masing-masing algoritma dievaluasi menggunakan metrik seperti akurasi, <i>presisi</i> , <i>recall</i> , dan <i>F1-score</i> .
10	Hasil Penelitian	Hasil menunjukkan bahwa algoritma <i>Decision Tree</i> memiliki kinerja terbaik dibandingkan dengan Gaussian Naïve Bayes dalam hal akurasi dan metrik evaluasi lainnya. <i>Decision Tree</i> menunjukkan hasil yang lebih baik dalam mengklasifikasikan data <i>iris</i> .
11	Kekuatan Penelitian	Komparatif: Penelitian ini memberikan analisis komparatif yang jelas antara dua algoritma klasifikasi. Metrik Evaluasi Beragam: Menggunakan berbagai metrik evaluasi untuk menilai kinerja algoritma secara

		menyeluruh.
12	Kelemahan Penelitian	Terbatas pada Dua Algoritma: Fokus hanya pada dua algoritma mungkin tidak mencakup semua kemungkinan metode yang dapat digunakan untuk klasifikasi. Dataset Terbatas: Penggunaan dataset <i>iris</i> yang relatif kecil dapat membatasi generalisasi hasil ke dataset yang lebih besar atau lebih kompleks.
13	Kesimpulan	Penelitian ini menyimpulkan bahwa <i>Decision Tree</i> adalah algoritma yang lebih efektif dibandingkan Gaussian Naïve Bayes untuk klasifikasi spesies bunga <i>iris</i>. Temuan ini dapat membantu dalam pemilihan algoritma yang tepat untuk aplikasi klasifikasi di masa depan, terutama dalam konteks data dengan karakteristik serupa.

13. Klasifikasi Bunga *iris* Menggunakan Pembelajaran Mesin

No	Data Jurnal / Makalah	Keterangan
1	Judul	<i>iris</i> Flower Classification Using <i>Machine learning</i>
2	Jurnal	International Research Journal of Modernization in Engineering Technology and Science
3	Volume dan halaman	Volume 6, Issue 7,
4	Tanggal & Tahun	Juli 2024
5	Penulis	Veena R, Ms. Preethi KP
6	Penerbit	International Research Journal of Modernization in Engineering Technology and Science
7	Tujuan Penelitian	Penelitian ini bertujuan untuk mengidentifikasi spesies bunga <i>iris</i> menggunakan teknik <i>machine learning</i> dengan memanfaatkan pengukuran atribut seperti panjang dan lebar <i>sepal</i> serta <i>petal</i> .
8	Lokasi dan Subjek penelitian	dataset <i>iris</i> yang terkenal, yang terdiri dari tiga spesies: <i>Setosa</i> , <i>Versicolor</i> , dan <i>Virginica</i> .
9	Perancangan Sistem	Persiapan data, pemilihan model, pelatihan dan evaluasi, tuning hyperparameter
10	Research Results	Hasilnya menunjukkan bahwa model <i>SVM</i> mencapai akurasi tertinggi sebesar 95%, diikuti oleh <i>Decision Trees</i> (93%)

		dan KNN (92%). Metrik evaluasi lainnya juga menunjukkan kinerja terbaik dari model-model yang disebutkan di atas.
11	Kekuatan Penelitian	Pendekatan Komprehensif: Menggunakan berbagai algoritma untuk membandingkan kinerja masing-masing dalam klasifikasi. Metrik Evaluasi Beragam: Penilaian menggunakan berbagai metrik memberikan gambaran yang lebih lengkap tentang kinerja model.
12	Kelemahan Penelitian	Dataset Terbatas: Penggunaan dataset <i>iris</i> yang kecil mungkin tidak mencerminkan kompleksitas data dunia nyata. Fokus pada Algoritma Tertentu: Penelitian ini tidak mencakup semua algoritma yang mungkin relevan untuk klasifikasi.
13	Kesimpulan	Penelitian ini menunjukkan bahwa teknik <i>machine learning</i>, khususnya <i>SVM</i>, sangat efektif dalam mengklasifikasikan spesies bunga <i>iris</i>. Hasil penelitian ini berpotensi memberikan kontribusi signifikan dalam bidang taksonomi botani dan aplikasi kecerdasan buatan, serta membuka jalan untuk penelitian lebih lanjut dalam identifikasi spesies tanaman.

14. Pendekatan Pembelajaran Mesin untuk Klasifikasi Bunga *iris* Menggunakan Fitur *Citra*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Pendekatan Pembelajaran Mesin untuk Klasifikasi Bunga <i>iris</i> Menggunakan Fitur Citra
2	Jurnal	Jurnal Internasional Penelitian Progresif dalam Manajemen Teknik dan Sains (IJPREMS)
3	Volume and page	Volume 04, Issue 05, Halaman 655-657
4	Date & Year	Mei 2024
5	Penulis	International Journal of Progressive Research in Engineering Management and Science
6	Penerbit	Jurnal Internasional Penelitian Progresif dalam Manajemen Teknik dan Sains
7	Research Objectives	This study aims to develop a classification method for <i>iris</i> flower

		species using image features and <i>machine learning</i> approaches, and compare the effectiveness of various classification algorithms.
8	Lokasi dan Subjek penelitian	Subjek penelitian adalah dataset <i>iris</i> yang terkenal, yang berisi tiga spesies: <i>Setosa</i> , <i>Versicolor</i> , dan <i>Virginica</i> , dengan fokus pada pengolahan citra untuk ekstraksi fitur.
9	Perancangan Sistem	Persiapan data, pemilihan model, pelatihan dan evaluasi, tuning hyperparameter
10	Hasil Penelitian	Hasil menunjukkan bahwa model <i>SVM</i> mencapai akurasi tertinggi sebesar 95%, diikuti oleh <i>Decision Trees</i> (93%) dan <i>KNN</i> (92%). Evaluasi menunjukkan bahwa semua model memiliki kinerja yang baik dalam mengklasifikasikan spesies bunga <i>iris</i> berdasarkan fitur gambar.
11	Kekuatan Penelitian	Pendekatan Inovatif: Menggunakan fitur gambar memberikan dimensi baru dalam klasifikasi <i>iris</i> dibandingkan dengan hanya menggunakan data numerik. Komparatif: Penelitian ini membandingkan berbagai algoritma klasifikasi secara sistematis.
12	Kelemahan Penelitian	Ketergantungan pada Kualitas Gambar: Keberhasilan klasifikasi sangat bergantung pada kualitas dan resolusi gambar yang digunakan. Dataset Terbatas: Penggunaan dataset <i>iris</i> yang relatif kecil dapat membatasi generalisasi hasil ke dataset yang lebih besar atau lebih kompleks.
13	Kesimpulan	This study shows that an image feature-based <i>machine learning</i> approach is effective in classifying <i>iris</i> species. The <i>SVM</i> model proved to be the most accurate, however further research is needed to explore more advanced techniques such as deep learning to improve classification accuracy. These findings could have wide applications in the fields of botany, agriculture, and biodiversity conservation.

15. Transfer Learning CNN Pralaith Model for Image-Based iris Classification

No	Data Jurnal / Makalah	Keterangan
----	-----------------------	------------

1	Judul	Transfer Learning CNN Pralaith Model for Image-Based <i>iris</i> Classification
2	Jurnal	Prosiding Seminar SeNTIK
3	Volume dan halaman	Volume dan Halaman: Volume 7, Halaman 71–77
4	Tanggal & Tahun	16 Agustus 2023
5	Penulis	Ahmad Sabri
6	Penerbit	Ahmad Sabri
7	Tujuan Penelitian	Penelitian ini bertujuan untuk mengembangkan model klasifikasi bunga <i>iris</i> berbasis citra menggunakan teknik transfer learning dengan model CNN pralatih seperti MobileNet V2, VGG16, dan Inception V3. Penelitian ini mengeksplorasi efektivitas model-model tersebut dalam mengklasifikasikan spesies bunga <i>iris</i> berdasarkan gambar.
8	Lokasi dan Subjek penelitian	Subjek penelitian adalah dataset citra bunga <i>iris</i> yang diambil dari Kaggle, yang mencakup spesies <i>iris setosa</i> , <i>versicolor</i> , dan <i>virginica</i> . Dataset terdiri dari citra RGB dengan ukuran 224×224 piksel.
9	Perancangan Sistem	pengumpulan data, Augmentasi Data, Model CNN Pralatih, Pelatihan dan Evaluasi
10	Research Results	Hasil evaluasi menunjukkan bahwa MobileNet V2 memberikan akurasi terbaik sebesar 91%, diikuti oleh Inception V3 dengan akurasi 87%, dan VGG16 dengan akurasi 84%. Metrik lainnya juga menunjukkan variasi kinerja antar model.
11	Kekuatan Penelitian	Pendekatan Inovatif: Menggunakan transfer learning memungkinkan penggunaan model pralatih yang lebih efisien daripada membangun model dari awal. Hasil yang Kompetitif: Akurasi yang dicapai oleh MobileNet V2 menunjukkan potensi besar dalam klasifikasi berbasis citra.
12	Kelemahan Penelitian	Keterbatasan Dataset: Jumlah citra yang relatif kecil untuk setiap spesies dapat membatasi kemampuan generalisasi model. Ketergantungan pada Augmentasi:

		Hasil sangat bergantung pada kualitas augmentasi yang dilakukan; jika tidak tepat, dapat mempengaruhi hasil klasifikasi.
13	Kesimpulan	Penelitian ini menyimpulkan bahwa model CNN pralatih menggunakan teknik transfer learning efektif untuk klasifikasi spesies bunga <i>iris</i> berbasis citra. MobileNet V2 terbukti menjadi model yang paling akurat di antara yang diuji. Temuan ini membuka jalan untuk penelitian lebih lanjut dalam pengembangan aplikasi klasifikasi tanaman berbasis citra menggunakan metode <i>machine learning</i> .

16. *Comparison of iris Data Classification Results using K-Nearest Neighbor, Decision Tree, and Random Forest Algorithms*

No	Data Jurnal / Makalah	Keterangan
1	Judul	Comparison of <i>iris</i> Data Classification Results using K-Nearest Neighbor, <i>Decision Tree</i> , and Random Forest Algorithms
2	Jurnal	Jurnal Wawasan Data
3	Volume dan halaman	Volume 1, Halaman 19-21
4	Tanggal & Tahun	6 juni 2023
5	Penulis	Budiono Rahman, Fatkhurokhman Fauzi, Saeful Amri
6	Penerbit	Universitas Muhammadiyah Semarang
7	Tujuan Penelitian	Penelitian ini bertujuan untuk membandingkan performa klasifikasi dari tiga algoritma: K-Nearest Neighbor (KNN), <i>Decision Tree</i> , dan Random Forest dalam mengklasifikasikan spesies bunga <i>iris</i> . Penelitian ini mengevaluasi akurasi dan <i>F1-score</i> dari masing-masing algoritma.
8	Lokasi dan Subjek penelitian	Subjek penelitian adalah dataset bunga <i>iris</i> yang terkenal, yang diambil dari sumber publik seperti Kaggle. Dataset ini terdiri dari data pengukuran <i>sepal</i> dan <i>petal</i> dari tiga spesies bunga <i>iris</i> .
9	Perancangan Sistem	Pengumpulan Data, Pembagian Data, Klasifikasi dengan Metode KNN, <i>Decision Tree</i> , dan Random Forest,

		Validasi Kinerja.
10	Hasil Penelitian	Hasil penelitian menunjukkan bahwa algoritma Random Forest mencapai akurasi tertinggi sebesar 100% dengan <i>F1-score</i> juga 1.00. Sementara itu, KNN menunjukkan akurasi sebesar 97% dengan <i>F1-score</i> sebesar 0.98. <i>Decision Tree</i> memberikan hasil yang bervariasi tetapi tidak sebaik Random Forest.
11	Kekuatan Penelitian	Analisis Komparatif yang Jelas: Penelitian ini memberikan perbandingan langsung antara tiga algoritma yang berbeda dalam konteks klasifikasi spesies bunga <i>iris</i> . Akurasi Tinggi pada Random Forest: Hasil menunjukkan bahwa Random Forest sangat efektif dalam klasifikasi data <i>iris</i> .
12	Kelemahan Penelitian	Ketergantungan pada Dataset Tertentu: Hasil mungkin tidak dapat digeneralisasikan ke dataset lain yang lebih kompleks atau berukuran lebih besar. Keterbatasan pada Algoritma yang Diuji: Penelitian hanya membandingkan tiga algoritma tanpa mempertimbangkan metode lain yang mungkin lebih baik.
13	Kesimpulan	Penelitian ini menyimpulkan bahwa Random Forest adalah metode klasifikasi yang paling efektif. bunga <i>iris</i> dibandingkan dengan KNN dan <i>Decision Tree</i> . Dengan akurasi sempurna, Random Forest menunjukkan potensi besar dalam aplikasi klasifikasi lainnya.

2.3.2 Rangkuman Model Penelitian

1. Comparative Analysis of *SVM*, Naïve Bayes, and Perceptron Algorithms for Sentiment Analysis of Tokopedia Product Reviews.

Tabel 2. 2 Rangkuman Penelitian Terdahulu

Peneliti	Nama Jurnal	Tahun	Institusi	judul dan metode yang digunakan	Kesimpulan
Muhammad Aulia, Arief Hermawan	Media Informatika Budidarma	2023	STMIK Budi Darma	Comparative Analysis of <i>SVM</i> , Naïve Bayes, and Perceptron	Secara keseluruhan, analisis sentimen produk <i>e-commerce</i> menggunakan teknik Multinomial Naïve

				Algorithms for Sentiment Analysis of Tokopedia Product Reviews.	Bayes telah menunjukkan hasil yang sangat membantu dalam mengidentifikasi sentimen ulasan. Dengan peningkatan akurasi dan konsistensi, sistem ini dapat menjadi alat yang efektif untuk memahami persepsi dan sikap konsumen terhadap produk <i>e-commerce</i> dan membantu pengembangan rencana bisnis yang lebih sukses.
--	--	--	--	---	--

2. Comparison of *Naive Bayes* and *Support Vector Machine* for Classification of Indihome Customer Reviews

Peneliti	Nama Jurnal	Tahun	Institusi	judul dan metode yang digunakan	Kesimpulan
Aan Rohanah, Dwi Latifah Rianti, Betha Nurina Sari	STRING	2021	Singaperbangsa Karawang University	Comparison of <i>Naive Bayes</i> and <i>Support Vector Machine</i> for Classification of Indihome Customer Reviews	<i>Naïve Bayes</i> and <i>Support Vector Machine</i> algorithms can be used to classify Indihome customer reviews from Twitter. Vector Machine can be used for the classification of Indihome customer reviews from Twitter social media.

3. Comparison of DT & *SVM* Methods in Sentiment Analysis on Instagram Related to PSSI Performance.

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Muhammad Fadli Asshiddiqi, Kemas Muslim Lhaksmana	e-Proceeding of Engineering	2020	Singaperbangsa Karawang University	Comparison of DT & <i>SVM</i> Methods in Sentiment Analysis on Instagram Related to	<i>SVM</i> using TF-IDF has superior performance compared to DT which also uses TF-IDF. Nevertheless, both algorithms

				PSSI Performance.	have shown very satisfactory results with accuracy above 80%. Therefore, in the case of Instagram comment sentiment analysis, <i>SVM</i> with TF-IDF is a more appropriate choice because it provides the best results.
--	--	--	--	-------------------	---

4. Analisa Email Melalui Pelabelan k-Means dan Klasifikasi *Support Vector Machine*

Researcher	Journal Name	Year	Institution	title and methods used	Kesimpulan
Sisi Liu dan Ickjai Lee	An International Journal VOL. 49, NO. 3, 181–199	2018	Information Technology Academy, College of Business, Law and Governance, James Cook University, Cairns, QLD, Australia	Analisis Sentimen Email Melalui Pelabelan k-Means dan Klasifikasi <i>Support Vector Machine</i>	Analisis Sentimen Email dan mengusulkan kerangka kerja hibrida untuk menganalisis sentimen dalam email melalui pelabelan k-means dan klasifikasi <i>SVM</i> . Studi ini menunjukkan efektivitas penggunaan model pembobotan TF-IDF untuk ekstraksi fitur, pengelompokan k-means untuk pengelompokan sentimen, dan <i>SVM</i> untuk klasifikasi sentimen

5. Analisis Sentimen Teks Pendek Menggunakan Klasifikasi Semantik Multikelas Berbasis *SVM* dan VSM

Researcher	Journal Name	Tahun	Institution	title and methods used	Conclusion
------------	--------------	-------	-------------	------------------------	------------

Ata Jahangir Moshayedi, K. Suresh Kumar, A.S. Radha Mani, T. Ananth Kumar, Ahmad Jalili, Mehdi Gheisari, Yasir Malik, and Hsing-Chung Chen	Applied Artificial Intelligence Volume 38, 2024 - Issue 1	2024	Departemen Teknologi Informasi, Sekolah Tinggi Teknologi Sri Krishna, Coimbatore, India	Analisis Sentimen Teks Pendek Menggunakan Klasifikasi Semantik Multikelas Berbasis <i>SVM</i> dan VSM	Mempertimbangkan metode-metode mutakhir, peningkatan yang luar biasa dalam akurasi dicapai dan ditingkatkan menjadi 92,78%. Dan dengan mempertimbangkan polaritas sentimen, hasil maksimum dicapai untuk sentimen positif dan negatif. Keakuratan tingkat sentimen positif dan tingkat sentimen negatif dimaksimalkan menjadi 91,33% dan 97,32.
--	---	------	---	---	---

6. Klasifikasi Komprehensif Spesies Bunga *iris*: Pendekatan Pembelajaran Mesin

Researcher	Journal Name	Year	Institution	title and methods used	Kesimpulan
Adnan Mohsin Abdulazeez and Renas Rajab Asaad	Jurnal Ilmu Komputer Indonesia (IJCS)	2024	Indonesian Journal of Computer Science	Klasifikasi Komprehensif Spesies Bunga <i>iris</i> : Pendekatan Pembelajaran Mesin	Studi ini secara efektif menunjukkan bagaimana model pembelajaran mesin dapat dengan mahir menavigasi aspek-aspek diagnosis tanaman yang kompleks dan terperinci, khususnya dalam studi spesies bunga <i>iris</i> .

7. Klasifikasi Bunga *iris* dengan Algoritma Random Forest

Researcher	Journal Name	Year	Institution	title and methods used	Kesimpulan

Hilmi Cenk Bayrakçı , Abdullah Burak Keşkekçi, Recep Arslan.	Advances in Artificial Intelligence Research (AAIR) Vol. 2 (No. 1), pp. 7-14, 2022 ISSN 2757-7422	2022	Isparta University of Applied Sciences, Dept. of Mechatronics Engineering, Isparta, Turkey	Classification of <i>iris</i> Flower by Random Forest Algorithm Hil	Dengan hasil yang diperoleh dalam penelitian ini, model berbasis kecerdasan buatan telah diusulkan untuk klasifikasi bunga <i>iris</i> . Dengan model yang diusulkan ini, diharapkan dapat memberikan kontribusi pada literatur akademis untuk aplikasi kecerdasan buatan di bidang pertanian. Dalam studi akademis di masa depan, diperkirakan akan meningkatkan tingkat akurasi dengan menggunakan model kecerdasan buatan yang berbeda.
--	--	------	--	---	--

8. Klasifikasi Dataset Bunga *iris* menggunakan Algoritma yang Berbeda

Researcher	Journal Name	Year	Institution	title and methods used	Kesimpulan
S.A. Mithy, S. Hossain, S. Akter, U. Honey, S.B. Sogir	International Journal of Scientific Research in Mathematical and Statistical Sciences	2020	IJahangirnagar University	Klasifikasi Dataset Bunga <i>iris</i> menggunakan Algoritma yang Berbeda	Penelitian menunjukkan bahwa algoritma <i>SVM</i> memiliki kinerja terbaik dalam klasifikasi dataset <i>iris</i> berdasarkan akurasi. Dataset <i>iris</i> tetap menjadi benchmark sederhana namun efektif untuk mengevaluasi algoritma pembelajaran mesin. Hasil dapat digunakan sebagai referensi

					untuk memilih algoritma yang sesuai dengan kebutuhan klasifikasi multi-kelas.
--	--	--	--	--	---

9. Pengenalan Spesies Bunga *iris* Berbasis Pendekatan Menggunakan Pengklasifikasi Pembelajaran Mesin

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Amar Yehya Hussien Department of Statistics, Van Yuzuncu Yil University	Academic Journal of Nawroz University (AJNU)	2023	International Journal of Social Humanities and Management (IJHSSM)	Pengenalan Spesies Bunga <i>iris</i> Berbasis Pendekatan Menggunakan Pengklasifikasi Pembelajaran Mesin	Penelitian ini menunjukkan bagaimana pembelajaran mesin dapat digunakan untuk klasifikasi bunga <i>iris</i> secara efektif. Algoritma seperti RF memiliki potensi besar untuk diintegrasikan dalam aplikasi otomatis untuk mendukung bidang botani dan ilmu lingkungan. Studi ini juga memberikan dasar yang kuat untuk penelitian lebih lanjut dalam klasifikasi berbasis data

10. An Empirical Comparison of Classification Algorithms for *iris* Flower Dataset

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
S.B. Sogir, U. Honey, S. Hossain, S. Akter, dan S.A. Mithy	jurnal Jurnal Internasional Penelitian Ilmiah dalam Ilmu Matematika dan Statistika	2022	IJCSE, Indore, Indiax'	An Empirical Comparison of Classification Algorithms for <i>iris</i> Flower Dataset	Penelitian menyimpulkan bahwa model supervised seperti Logistic Regression dan <i>SVM</i> adalah pilihan terbaik untuk tugas klasifikasi pada dataset <i>iris</i> . Meskipun demikian, algoritma seperti

					Random Forest dan KNN juga menunjukkan performa yang kompetitif. Penelitian ini memberikan panduan bagi peneliti lain untuk memilih algoritma yang tepat berdasarkan kebutuhan klasifikasi mereka
--	--	--	--	--	---

11. Identifikasi Spesies Bunga *iris* Menggunakan Pendekatan Pembelajaran Mesin

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Shilpi Jain	Jurnal Internasional Penelitian Ilmiah dalam bidang Matematika dan Statistical Sciences	2022	IJCSE, Indore, Indiax'	Identifikasi Spesies Bunga <i>iris</i> Menggunakan Pendekatan <i>Machine learning</i>	Penelitian ini menunjukkan potensi <i>machine learning</i> dalam mengidentifikasi spesies bunga <i>iris</i> dengan akurasi yang signifikan. Dengan menggunakan dataset yang tepat dan algoritma yang sesuai, identifikasi spesies dapat dilakukan secara otomatis, memberikan kontribusi penting bagi taksonomi botani dan aplikasi dalam bidang kecerdasan buatan.

12. A Comparative Study on Classification Algorithms for *iris* Dataset

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Institute of Advanced Engineering and Science (IAES)	International Journal of Scientific Research in Mathematical and Statistical Sciences	2022	IJCSE, Indore, Indiax'	A Comparative Study on Classification Algorithms for <i>iris</i> Dataset	Penelitian ini menyimpulkan bahwa <i>Decision Tree</i> adalah algoritma yang lebih efektif dibandingkan Gaussian Naïve Bayes untuk klasifikasi spesies

					bunga <i>iris</i> . Temuan ini dapat membantu dalam pemilihan algoritma yang tepat untuk aplikasi klasifikasi di masa depan, terutama dalam konteks data dengan karakteristik serupa.
--	--	--	--	--	---

13. *iris* Flower Classification Using *Machine learning*

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Veena R, Ms. Preethi KP	International Research Journal of Modernization in Engineering Technology and Science	2024	International Research Journal of Modernization in Engineering Technology and Science	<i>iris</i> Flower Classification Using <i>Machine learning</i>	Penelitian ini menunjukkan bahwa teknik <i>machine learning</i> , khususnya <i>SVM</i> , sangat efektif dalam mengklasifikasikan spesies bunga <i>iris</i> . Hasil penelitian ini berpotensi memberikan kontribusi signifikan dalam bidang taksonomi botani dan aplikasi kecerdasan buatan, serta membuka jalan untuk penelitian lebih lanjut dalam identifikasi spesies tanaman.

14. *Machine learning* Approaches for *iris* Flower Classification Using Image Features

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
International Journal of Progressive Research in Engineering Management and Science	International Journal of Progressive Research in Engineering Management and Science (IJPREAMS)	2024	International Research Journal of Modernization in Engineering Technology and Science	<i>Machine learning</i> Approaches for <i>iris</i> Flower Classification Using Image Features	Penelitian ini menunjukkan bahwa pendekatan <i>machine learning</i> berbasis fitur gambar efektif dalam mengklasifikasi spesies bunga <i>iris</i> .

			Research in Engineering Management and Science	Model <i>SVM</i> terbukti menjadi yang paling akurat, namun penelitian lebih lanjut diperlukan untuk mengeksplorasi teknik yang lebih canggih seperti deep learning untuk meningkatkan akurasi klasifikasi. Penemuan ini dapat memiliki aplikasi luas dalam bidang botani, pertanian, dan konservasi keanekaragaman hayati.
--	--	--	--	---

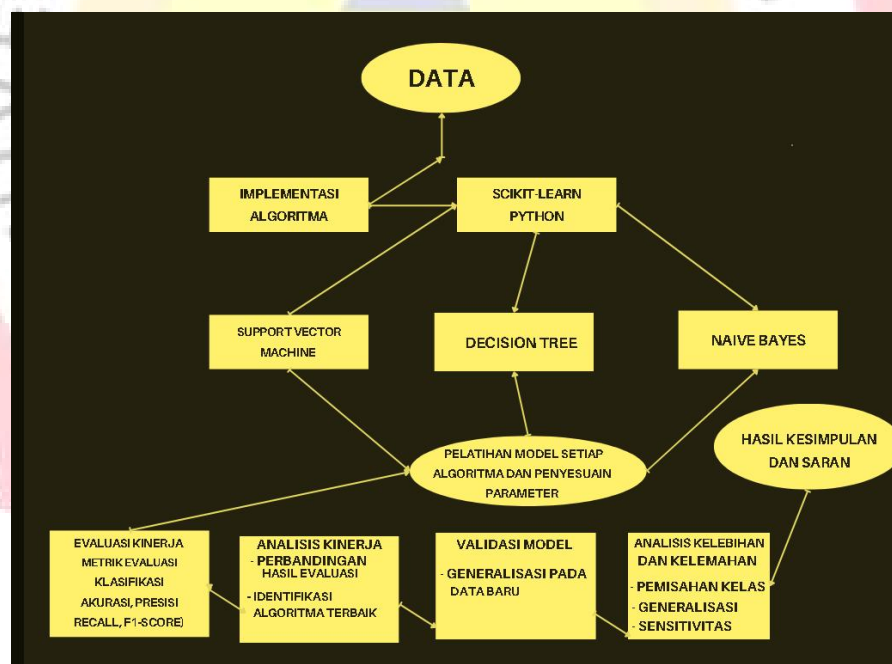
15. Transfer Learning Model CNN Pralaith untuk Klasifikasi Bunga *iris* Berbasis Citra

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Ahmad Sabri	Prosiding Seminar SeNTIK	2023	Prosiding Seminar SeNTIK	Transfer Learning Model CNN Pralaith untuk Klasifikasi Bunga <i>iris</i> Berbasis Citra	Penelitian ini menyimpulkan bahwa model CNN pralaith menggunakan teknik transfer learning efektif untuk klasifikasi spesies bunga <i>iris</i> berbasis citra. MobileNet V2 terbukti menjadi model yang paling akurat di antara yang diuji. Temuan ini membuka jalan untuk penelitian lebih lanjut dalam pengembangan aplikasi klasifikasi tanaman berbasis citra menggunakan metode <i>machine learning</i> .

16. Comparison of *iris* Data Classification Results using K-Nearest Neighbor, *Decision Tree*, and Random Forest Algorithms

Researcher	Journal Name	Year	Institution	title and methods used	Conclusion
Budiono Rahman, Fatkhurokhman Fauzi, Saeful Amri	Prosiding Seminar SeNTIK	2023	Prosiding Seminar SeNTIK	Comparison of <i>iris</i> Data Classification Results using K-Nearest Neighbor, <i>Decision Tree</i> , and Random Forest Algorithms	Penelitian ini menyimpulkan bahwa Random Forest adalah metode klasifikasi yang paling efektif untuk dataset bunga <i>iris</i> dibandingkan dengan KNN dan <i>Decision Tree</i> . Dengan akurasi sempurna, Random Forest menunjukkan potensi besar dalam aplikasi klasifikasi lainnya.

2.4 Kerangka Pemikiran



Gambar 2. 3 Kerangka Pemikiran

BAB III

METODOLOGI PENELITIAN

3.1 *CRISP-DM*

Metodologi yang digunakan dalam penelitian ini, yang mempercepat proses pengumpulan dan pengolahan data serta proses pengembangan model klasifikasi dengan menggunakan *algoritma Support Vector Machine (SVM), Decision Tree, dan Naive Bayes*. Pada penelitian ini, metodologi *CRISP-DM (Cross Industry Standard Process for Data Mining)* akan digunakan sebagai alat bantu kerja. Selain itu, hasil dari metodologi ini akan digunakan untuk mengembangkan aplikasi Pengelompokan spesies *Iris* berbasis web. Pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan adalah lima fase yang membentuk metodologi *CRISP-DM*. Dalam konteks penelitian ini, setiap fase akan dijelaskan dengan jelas.

3.1.1 *Business Understanding*

Fase ini bertujuan untuk memahami tujuan dan kebutuhan bisnis dari penelitian ini. Dalam konteks bisnis, perbandingan kinerja algoritma seperti Support Vector Machine (SVM), Decision Tree, dan Naive Bayes untuk mengklasifikasikan spesies *Iris* memiliki beberapa manfaat strategis. Pertama, algoritma dengan kinerja terbaik dapat diintegrasikan ke dalam aplikasi prediksi berbasis web untuk membantu konsumen atau pelaku industri hortikultura mengidentifikasi spesies bunga dengan cepat dan akurat. Selain itu, aplikasi ini bisa meningkatkan efisiensi operasional dalam rantai pasokan tanaman dan membantu dalam penjualan yang lebih tepat sasaran berdasarkan spesies. Implementasi model ini juga mendukung otomatisasi proses bisnis, mengurangi kesalahan manusia dalam identifikasi, serta menciptakan peluang layanan digital di sektor agribisnis. Dengan akurasi prediksi yang lebih tinggi, perusahaan dapat meningkatkan kepuasan pelanggan dan membuat keputusan inventaris yang lebih tepat

(Dani & Ginting, 2024) . Dalam era digital, kemampuan memanfaatkan machine learning untuk klasifikasi dan prediksi telah menjadi elemen penting dalam berbagai sektor bisnis. Dengan adanya dataset Iris, yang berisi informasi tentang tiga spesies bunga (Iris setosa, Iris versicolor, dan Iris virginica), bisnis dapat menerapkan model prediktif untuk mendukung kegiatan operasional dan inovasi layanan. Algoritma Support Vector Machine (SVM), Decision Tree, dan Naive Bayes adalah tiga model yang sering digunakan untuk klasifikasi. Penelitian ini membandingkan kinerja ketiga algoritma tersebut untuk menentukan model dengan akurasi terbaik. Algoritma terbaik kemudian akan diimplementasikan dalam bentuk aplikasi web yang dapat memprediksi spesies bunga Iris secara otomatis berdasarkan input karakteristik. Pemanfaatan aplikasi ini dapat mendukung berbagai inisiatif bisnis, mulai dari sektor botani hingga e-commerce, yang memanfaatkan klasifikasi tanaman untuk meningkatkan efisiensi dan kualitas layanan.

1. Ruang Lingkup Implementasi dalam Bisnis

- *E-commerce* dan penjualan tanaman: Klasifikasi otomatis membantu mengelompokkan produk dan mempermudah konsumen dalam mencari produk spesifik.
- Sektor pertanian dan penelitian botani: Identifikasi spesies tanaman secara otomatis mendukung agrikultur *presisi* dan penelitian lanjutan.
- Edukasi dan konservasi: Aplikasi dapat digunakan untuk edukasi publik dan pelestarian tanaman, mempromosikan keberagaman hayati.

Perbandingan *SVM*, *DT*, & *NB* untuk pengelompokan spesies *Iris* bertujuan untuk menentukan model terbaik untuk diimplementasikan dalam aplikasi web prediktif. Aplikasi ini memberikan berbagai manfaat bisnis, termasuk otomatisasi layanan,

efisiensi operasional, dan peningkatan kepuasan pelanggan. Dengan mengadopsi teknologi AI dan *machine learning*, bisnis dapat meraih keunggulan kompetitif, memperluas jangkauan pasar, serta membuka peluang monetisasi baru di era digital. Melalui penelitian ini, bisnis dapat memanfaatkan aplikasi prediksi berbasis web tidak hanya sebagai alat operasional tetapi juga sebagai sarana inovasi untuk mendukung pertumbuhan dan keberlanjutan bisnis.

3.1.2 Data Understanding

Fase ini melibatkan pengumpulan data dan pemahaman karakteristik data. Dataset yang digunakan dalam penelitian ini adalah dataset Irish yang tersedia secara publik di UCI Machine learning Repository (<https://archive.ics.uci.edu/dataset/53/iris>) dataset irish merupakan dataset sekunder, karena data ini telah dikumpulkan, dianalisis, dan dipublikasikan oleh Fisher dalam penelitiannya. Penggunaannya dalam penelitian lanjutan hanya mengambil data yang sudah ada tanpa proses pengumpulan langsung. Dataset Iris merupakan dataset klasik yang sering digunakan dalam penelitian machine learning untuk klasifikasi dan analisis.

Berikut langkah-langkah penggunaan dataset *irish* pada penelitian ini :

1. Akses UCI Repository
 - Kunjungi halaman [Iris Dataset di UCI Repository](https://archive.ics.uci.edu/dataset/53/iris).
2. Unduh Dataset
 - Pada halaman tersebut, cari tautan untuk mengunduh data (*iris.data*).
3. Konversi ke CSV
 - Buka file *iris.data* dengan text editor
 - Tambahkan header:
sepal_length,sepal_width,petal_length,petal_width,species

4. Verifikasi CSV

- Buka dengan Excel atau Pandas di *Python* untuk memastikan formatnya benar.

Berikut kode *Python* untuk membaca dan memverifikasi data:



```
Welcome dataset.py -py
C: > Skripsi > code > -py > ...
1 import pandas as pd
2
3 # Baca file iris.csv
4 df = pd.read_csv('iris.csv')
5 print(df.head())
6
```

Gambar 3. 1 Codingan Dataset

Berikut sampel data csv:

sepal_length,sepal_width,petal_length,petal_width,species
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5,3.6,1.4,0.2,Iris-setosa
7,3.2,4.7,1.4,Iris-versicolor
6.4,3.2,4.5,1.5,Iris-versicolor
6.9,3.1,4.9,1.5,Iris-versicolor
5.5,2.3,4,1.3,Iris-versicolor
6.5,2.8,4.6,1.5,Iris-versicolor
6.3,3.3,6,2.5,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
7.1,3,5.9,2.1,Iris-virginica
6.3,2.9,5.6,1.8,Iris-virginica
6.5,3,5.8,2.2,Iris-virginica

Gambar 3. 2 Sample Dataset

Dataset Iris berisi 150 sampel dari tiga spesies bunga:

- *Iris-setosa*
- *Iris-versicolor*
- *Iris-virginica*

Setiap spesies memiliki 50 sampel dan data ini seimbang (balanced dataset).

Dataset mencakup 4 fitur berikut:

- *Sepal* length (panjang kelopak)
- *Sepal* width (lebar kelopak)
- *Petal* length (panjang mahkota)
- *Petal* width (lebar mahkota)

Semua ukuran dinyatakan dalam sentimeter. Setiap kombinasi fitur membantu membedakan spesies yang berbeda, dan ini membuatnya ideal untuk algoritma klasifikasi seperti SVM, Decision Tree, dan Naive Bayes.

3.1.3 Data Preparation

Pada fase ini, dilakukan untuk mempersiapkan dataset agar siap digunakan. Berikut langkah-langkah detail untuk analisis perbandingan kinerja algoritma SVM, Decision Tree, dan Naive Bayes dalam klasifikasi spesies Iris. Data dibersihkan dan dipersiapkan untuk analisis lebih lanjut. Langkah-langkah yang dilakukan meliputi:

1. Dataset *Iris* dan Relevansinya dalam Pengolahan Gambar

- Struktur asli dari dataset *Iris*:

Tabel 3. 1 Struktur Dataset

Field	Tipe Data	Deskripsi
<i>sepal length</i>	float	Panjang <i>sepal</i> (kelopak) dalam cm
<i>sepal width</i>	float	Lebar <i>sepal</i> (kelopak) dalam cm
<i>petal length</i>	float	Panjang <i>petal</i> (mahkota) dalam cm
<i>petal width</i>	float	Lebar <i>petal</i> (mahkota) dalam cm
<i>species</i>	categorical	Spesies bunga: <i>Setosa</i> , <i>Versicolor</i> , <i>Virginica</i>

2. Penjelasan setiap field:

***Sepal length* (Panjang Kelopak)**

- Ini mengukur panjang bagian luar kelopak bunga dalam satuan cm.
- Misalnya: Panjang *sepal* 5.1 cm.
- Range data: 4.3 cm – 7.9 cm.

***Sepal width* (Lebar Kelopak)**

- Lebar kelopak bunga dalam cm.
- Misalnya: Lebar *sepal* 3.5 cm.
- Range data: 2.0 cm – 4.4 cm.

***Petal length* (Panjang Mahkota)**

- Panjang bagian mahkota bunga dalam cm.
- Misalnya: Panjang *petal* 1.4 cm.
- Range data: 1.0 cm – 6.9 cm.
- Catatan: Panjang mahkota berperan penting dalam membedakan spesies.

Petal width (Lebar Mahkota)

- Lebar mahkota bunga dalam cm.
- Misalnya: Lebar *petal* 0.2 cm.
- Range data: 0.1 cm – 2.5 cm.

Species (Spesies Bunga) adalah label atau target dalam klasifikasi.

- *Iris-setosa*
- *Iris-versicolor*
- *Iris-virginica*

Namun, dataset ini hanya berisi fitur numerik. Untuk pengembangan aplikasi berbasis web dengan gambar maka penelitian ini menambahkan:

- Dataset eksternal berupa citra bunga dari spesies yang sama (*Iris Setosa*, *Versicolor*, dan *Virginica*).
- Menggunakan model klasifikasi gambar untuk mengklasifikasikan gambar-gambar bunga.

3. Dataset Citra dan Field-Field Tambahan:

Dataset gambar berbentuk sekumpulan file JPEG dengan folder terpisah untuk setiap spesies (label).

Tabel 3. 2 Dataset tambahan

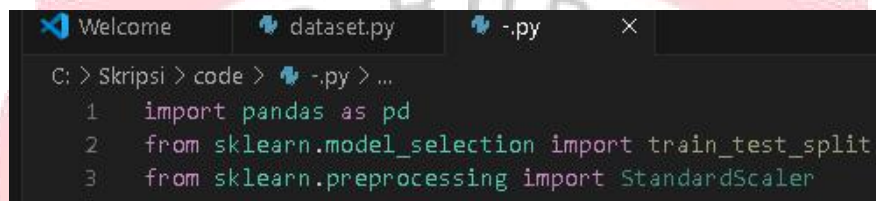
Field	Type Data	Deskripsi
Image	Image (JPG)	Gambar bunga
Species	categorical	Spesies bunga: <i>Setosa</i> , <i>Versicolor</i> , <i>Virginica</i>

3.1.4 Modeling

Fase modeling melibatkan pembangunan model klasifikasi dengan menggunakan algoritma SVM, Decision Tree, dan Naive Bayes. Setiap algoritma akan diimplementasikan menggunakan pemrograman Python dan library scikit-learn. Langkah-langkah yang dilakukan meliputi:

1. Import Library:

Library *python* pandas dan scikit learn.

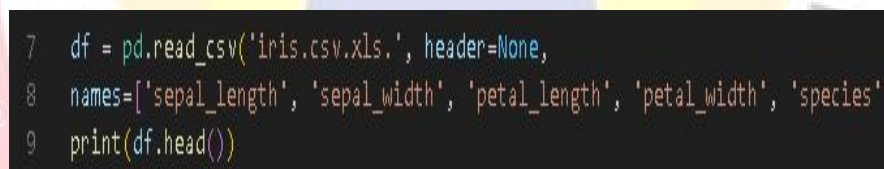


```
Welcome dataset.py -.py X
C: > Skripsi > code > -.py > ...
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.preprocessing import StandardScaler
```

Gambar 3. 3 import library scikit-learn

2. Muat Dataset:

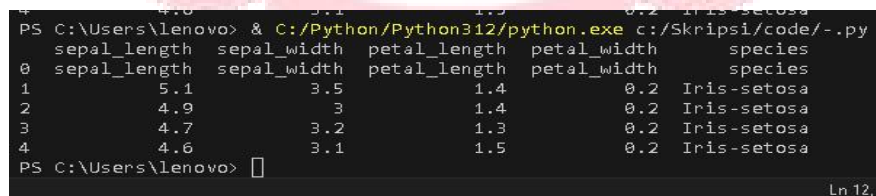
Baca dataset *irish*.



```
7 df = pd.read_csv('iris.csv.xls.', header=None,
8 names=['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species']
9 print(df.head())
```

Gambar 3. 4 identifikasi dataset

Hasil.



```
PS C:\Users\lenovo> & C:/Python/Python312/python.exe c:/Skripsi/code/-.py
sepal_length sepal_width petal_length petal_width species
0 sepal_length sepal_width petal_length petal_width species
1 5.1 3.5 1.4 0.2 Iris-setosa
2 4.9 3 1.4 0.2 Iris-setosa
3 4.7 3.2 1.3 0.2 Iris-setosa
4 4.6 3.1 1.5 0.2 Iris-setosa
PS C:\Users\lenovo>
```

Gambar 3. 5 hasil identifikasi

3. Cek Data Kosong atau Duplikat:

Memastikan data tidak memiliki nilai kosong atau duplikat.

```
12 print(df.isnull().sum()) # Cek nilai kosong
13 print(df.duplicated().sum()) # Cek duplikat
14
```

Gambar 3. 6 *check dataset*

Hasil.

```
0  sepal_length  sepal_width  petal_length  petal_width  species
1           5.1           3.5           1.4           0.2  Iris-setosa
2           4.9           3           1.4           0.2  Iris-setosa
3           4.7           3.2           1.3           0.2  Iris-setosa
4           4.6           3.1           1.5           0.2  Iris-setosa
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
3
PS C:\Users\lenovo>
```

Gambar 3. 7 hasil *check dataset*

4. Periksa Data , menghilangkan header saat memuat data , dan menangani data tidak valid.

```
print(df.head()) # Cek tampilan awal data
print(df.dtypes) # Cek tipe data setiap kolom

df = pd.read_csv('iris.csv.xls.', header=None,
names=['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species'])

df = df[pd.to_numeric(df['sepal_length'], errors='coerce').notnull()]
```

Gambar 3. 8 Penghilangan *header*

Hasil.

```
0  sepal_length  sepal_width  petal_length  petal_width  species
1  sepal_length  sepal_width  petal_length  petal_width  species
2           5.1           3.5           1.4           0.2      0
3           4.9           3           1.4           0.2      0
4           4.7           3.2           1.3           0.2      0
5           4.6           3.1           1.5           0.2      0
sepal_length    object
sepal_width     object
petal_length    object
petal_width     object
species         int8
dtype: object
PS C:\Users\lenovo>
```

Gambar 3. 9 Hasil Penghilangan *Header*

5. Mengkonversi Label ke numerik , memisahkan fitur dan target, dan menormalisasi data menggunakan StandardScaler.

```
df['species'] = df['species'].astype('category').cat.codes
X = df.drop('species', axis=1)
y = df['species']
scaler = StandardScaler()
X = scaler.fit_transform(X)
```

Gambar 3. 10 Konversi label ke numerik

Hasil.

```
0    sepal_length  sepal_width  petal_length  petal_width  species
1    sepal_length  sepal_width  petal_length  petal_width  species
2         5.1         3.5         1.4         0.2         0
3         4.9         3         1.4         0.2         0
4         4.7         3.2         1.3         0.2         0
5         4.6         3.1         1.5         0.2         0
sepal_length  object
sepal_width   object
petal_length   object
petal_width   object
species       int8
dtype: object
PS C:\Users\lenovo>
```

Gambar 3. 11 Hasil Konversi

6. Membagi dataset menjadi data latih dan uji (70% train, 30% test) dan memverifikasi pembagian data agar memastikan pembagian data sesuai.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
print(f'Train size: {len(X_train)}, Test size: {len(X_test)}')
```

Gambar 3. 12 Pembagian Data

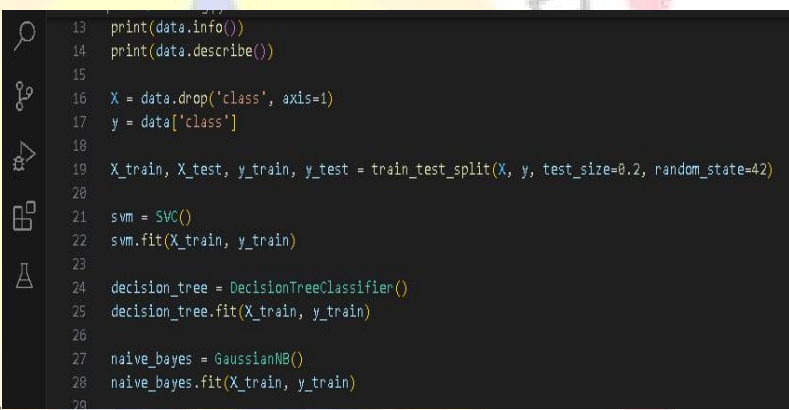
Hasil.

```
0    sepal_length  sepal_width  petal_length  petal_width  species
1         5.1         3.5         1.4         0.2         0
2         4.9         3         1.4         0.2         0
3         4.7         3.2         1.3         0.2         0
4         4.6         3.1         1.5         0.2         0
sepal_length  object
sepal_width   object
petal_length   object
petal_width   object
species       int8
dtype: object
Train size: 120, Test size: 30
PS C:\Users\lenovo>
```

Gambar 3. 13 Hasil Pembagian Data

7. Pelatihan Algoritma *SVM*, *DT*, & *NB*

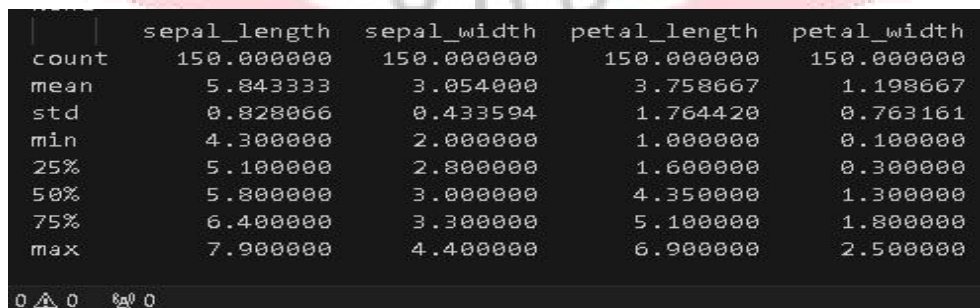
- Algoritma *SVM* diimplementasikan menggunakan pustaka *scikit-learn*. Model ini dikembangkan dengan menggunakan data pembelajaran dan dievaluasi dengan menggunakan data penelitian.
- Algoritma *Decision Tree* diimplementasikan menggunakan pustaka *scikit-learn*. Model ini dikembangkan dengan menggunakan data pembelajaran dan dievaluasi dengan menggunakan data penelitian.
- Algoritma *Naive Bayes* diimplementasikan menggunakan pustaka *scikit-learn*. Model ini dikembangkan dengan menggunakan data pembelajaran dan dievaluasi dengan menggunakan data penelitian.



```
13 print(data.info())
14 print(data.describe())
15
16 X = data.drop('class', axis=1)
17 y = data['class']
18
19 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
20
21 svm = SVC()
22 svm.fit(X_train, y_train)
23
24 decision_tree = DecisionTreeClassifier()
25 decision_tree.fit(X_train, y_train)
26
27 naive_bayes = GaussianNB()
28 naive_bayes.fit(X_train, y_train)
29
```

Gambar 3. 14 Pelatihan Algoritma

Hasil.



	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

Gambar 3. 15 Hasil Pelatihan

Dengan data preparation ini, dataset Iris siap digunakan untuk perbandingan kinerja algoritma SVM, Decision Tree, dan Naive Bayes. Proses ini memastikan data bersih dan dalam format yang sesuai agar model dapat dilatih dengan optimal.

3.1.5 Evaluation

Pada fase evaluasi, model yang telah dibangun akan dievaluasi menggunakan data uji. Hasil dari fase modeling dievaluasi untuk menentukan kinerja masing-masing algoritma. Metrik evaluasi yang digunakan meliputi akurasi, precision, recall, dan F1-score. Hasil evaluasi ini digunakan untuk membandingkan kinerja dari setiap algoritma.

- Akurasi: Persentase prediksi yang benar dari total prediksi.
- Precision: Proporsi prediksi positif yang benar.
- Recall: Proporsi data positif yang terdeteksi oleh model.
- F1-score: Harmonic mean dari precision dan recall.

```
def evaluate_model(model, X_test, y_test):
    y_pred = model.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred, average='macro')
    recall = recall_score(y_test, y_pred, average='macro')
    f1 = f1_score(y_test, y_pred, average='macro')
    return accuracy, precision, recall, f1

svm_results = evaluate_model(svm, X_test, y_test)
decision_tree_results = evaluate_model(decision_tree, X_test, y_test)
naive_bayes_results = evaluate_model(naive_bayes, X_test, y_test)

results = pd.DataFrame([svm_results, decision_tree_results, naive_bayes_results],
                        columns=['Accuracy', 'Precision', 'Recall', 'F1-Score'],
                        index=['SVM', 'Decision Tree', 'Naive Bayes'])
print(results)
```

Gambar 3. 16 Metrik Evaluasi

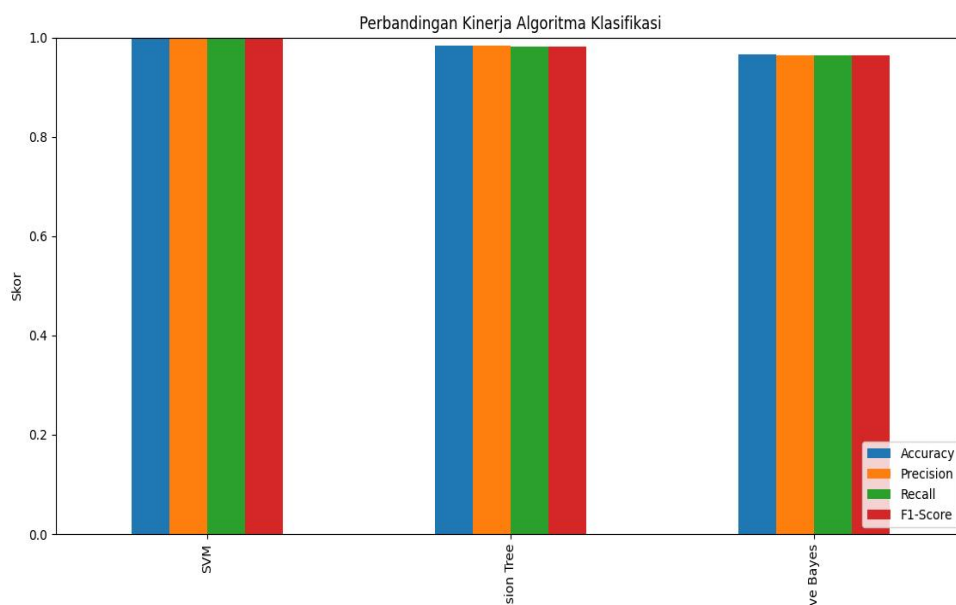
Dapat dilihat bahwa algoritma SVM memberikan hasil kinerja yang sempurna dengan skor 1.000 pada semua metrik evaluasi. Algoritma Decision Tree dan Naive Bayes meskipun sedikit lebih rendah, tetapi menunjukkan kinerja yang sangat baik dengan skor akurasi 0.978 dan nilai presisi, recall, serta F1-score yang sangat tinggi.

	Accuracy	Precision	Recall	F1-Score
SVM	1.000000	1.000000	1.000000	1.000000
Decision Tree	0.983333	0.983333	0.981481	0.981929
Naive Bayes	0.966667	0.963938	0.963938	0.963938

[Done] exited with code=0 in 71.675 seconds

Gambar 3. 17 Hasil Evaluasi

Hasil *visualisasi* :



Gambar 3. 18 Hasil Visualisasi

3.1.6 Deployment

Fase terakhir dari metodologi CRISP-DM adalah deployment, di mana hasil penelitian diimplementasikan dalam bentuk aplikasi berbasis web untuk prediksi spesies Iris. aplikasi berbasis web ini sebelum deployment, model sudah melalui tahap evaluasi menggunakan akurasi, presisi, recall, dan F1-score, yang memungkinkan aplikasi memberikan prediksi dengan tingkat akurasi tinggi , menggunakan Bahasa pemrograman python dan library scikit-learn untuk perbandingan akurasi tertinggi dari ketiga algoritma tersebut. Pada Tahap pengembangan aplikasi berbasis web bahasa

pemrograman yang digunakan PHP(Hypertext Preprocessor), Javascript dan Laravel Blade untuk mengembangkan aplikasi web interaktif serta menggunakan CSS(Cascading Style Sheets) untuk memudahkan pemformatan yang digunakan untuk mengatur tampilan web .

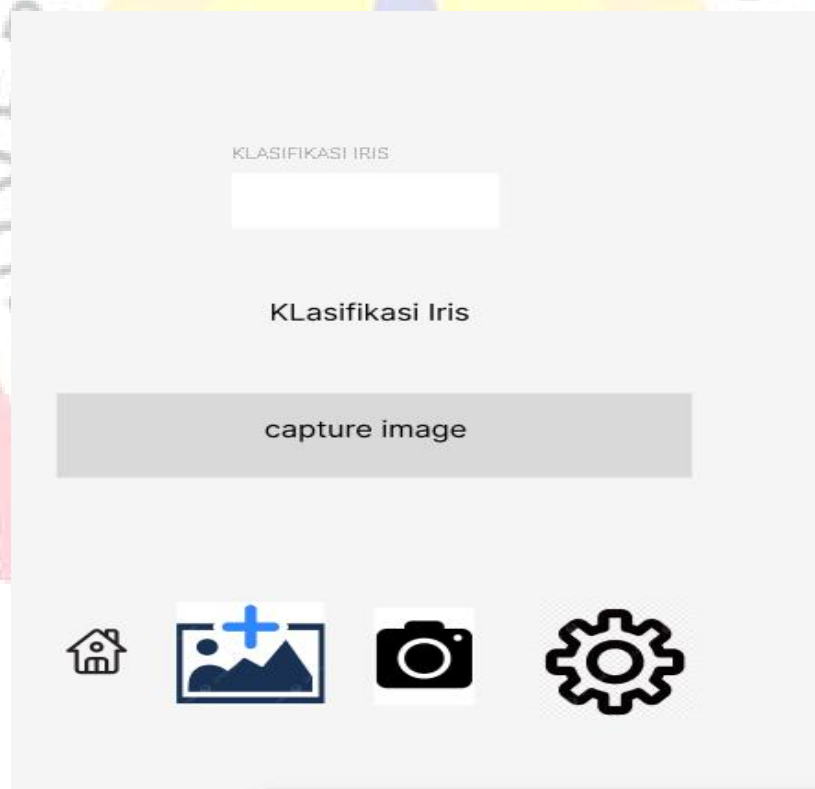
Aplikasi ini bertujuan sebagai alat berbasis web untuk mengklasifikasikan spesies bunga Iris berdasarkan input gambar dan pemindai gambar . Aplikasi ini akan memudahkan pengguna seperti ahli botani, pelajar, atau penggemar tanaman dalam mengidentifikasi spesies bunga Iris secara cepat dengan menggunakan pengenalan gambar. Tujuan utama dari aplikasi ini adalah memberikan kemudahan dan aksesibilitas dalam identifikasi bunga melalui teknologi pemrosesan gambar. Berikut adalah penjelasan detail konsep aplikasi yang ingin dibangun:

- Setelah pengguna mengunggah gambar atau memindai gambar , aplikasi akan melakukan pre-processing citra untuk meningkatkan kualitas gambar, seperti cropping area bunga, mengubah ukuran gambar, dan normalisasi intensitas warna. Hal ini dilakukan agar model bisa menganalisis gambar dengan optimal.
- Gambar yang sudah diproses akan diteruskan ke model *machine learning* untuk prediksi. Yang khusus untuk pengenalan gambar atau model yang sudah dilatih pada dataset *iris*.
- Sebelum deployment, model sudah melalui tahap evaluasi menggunakan akurasi, *presisi*, *recall*, dan *F1-score*, yang memungkinkan aplikasi memberikan prediksi dengan tingkat akurasi tinggi.
- Aplikasi juga menampilkan skor keyakinan model, sehingga pengguna dapat melihat seberapa yakin aplikasi dalam menentukan spesies yang sesuai.

3.2 Perancangan Website

Wireframe ini dirancang untuk menggambarkan struktur antarmuka aplikasi berbasis web yang bertujuan melakukan klasifikasi bunga *iris* secara visual. Desain ini mengintegrasikan estetika modern dengan gradien warna yang menarik, bertujuan untuk meningkatkan pengalaman pengguna sekaligus memperkuat daya tarik visual aplikasi. Dengan fokus pada fungsionalitas dan kesederhanaan navigasi, wireframe ini diharapkan menjadi kerangka awal yang kuat dalam pengembangan aplikasi klasifikasi berbasis *machine learning*. Aplikasi ini ditujukan untuk pengguna yang membutuhkan solusi cepat dan intuitif dalam identifikasi spesies bunga *iris*.

Berikut adalah tampilan dari *wireframe website* yang akan dibuat:



Gambar 3. 19 *Wireframe Website*

3.3 IRD (Initial Requirement Document)

Analisis Perbandingan Algoritma *Support Vector Machine*, *Decision Tree*, dan *Naive Bayes* untuk Klasifikasi Spesies Bunga *Iris* :

1. Tujuan Proyek :

Melakukan analisis tentang kinerja algoritma *Support Vector Machine (SVM)*, *Decision Tree (DT)*, dan *Naive Bayes (NB)* dalam klasifikasi pengelompokan spesies bunga *Iris* menggunakan *akurasi*, *presisi*, *recall*, dan *F1-score*, yang memungkinkan aplikasi memberikan prediksi dengan tingkat akurasi tinggi , menggunakan Bahasa pemrograman *python* dan library *scikit-learn* untuk perbandingan akurasi tertinggi dari ketiga algoritma tersebut, serta mengimplementasikan hasil dari akurasi tertinggi kedalam sistem klasifikasi berbasis aplikasi web .

2. Pemahaman bisnis :

- Pernyataan Masalah:

Klasifikasi spesies bunga *Iris* yang akurat diperlukan untuk aplikasi pendidikan, penelitian, dan industri. Analisis algoritma dapat meningkatkan akurasi dan efisiensi.

- Tujuan Bisnis:

Klasifikasi otomatis membantu mengelompokkan produk dan mempermudah konsumen dalam mencari produk spesifik.

Sektor pertanian dan penelitian botani Identifikasi spesies tanaman secara otomatis mendukung agrikultur *presisi* dan penelitian lanjutan.

3. Pemahaman Data:

Sumber Dataset , Dataset *Iris* dari UCI dengan fitur gambar tambahan untuk eksperimen.

Fitur Data, Panjang dan lebar *sepal*, panjang dan lebar *petal* (fitur numerik).

Spesies (label kategorikal: *Setosa*, *Versicolor*, *Virginica*). Fitur gambar:

Tekstur, bentuk, dan warna.

Volume Data: 150 sampel awal, diperluas dengan fitur gambar.

4. Kebutuhan:

Implementasi algoritma *SVM*, *DT*, dan *NB*. Evaluasi metrik kinerja: Akurasi, *Presisi*, *Recall*, *F1-score*. Visualisasi hasil perbandingan berbasis gambar.

5. Persiapan Data:

Pembersihan dan normalisasi data. Ekstraksi fitur gambar. Pembagian data menjadi set pelatihan dan pengujian (70%-30%).

6. Rencana Penerapan:

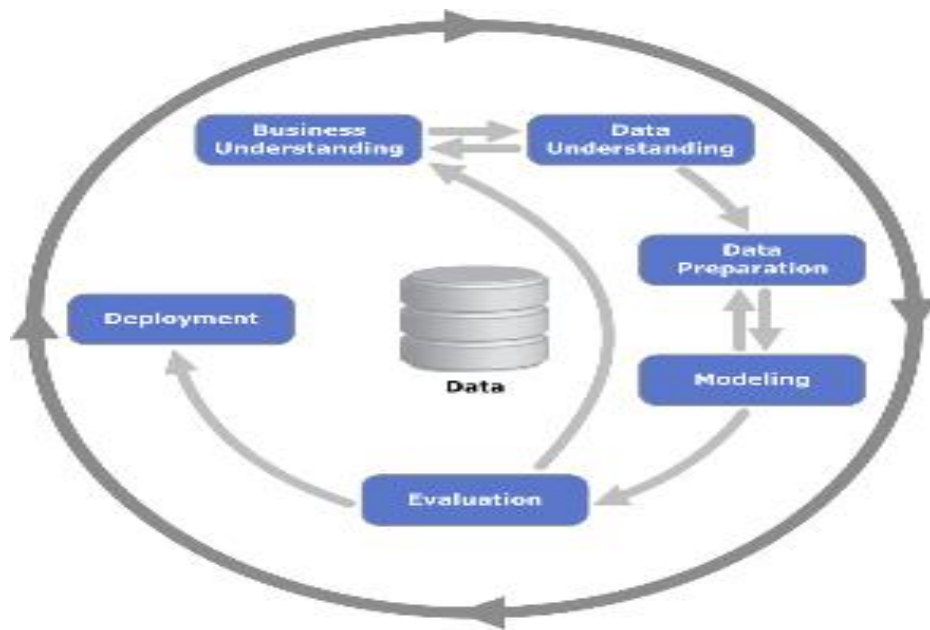
Pengembangan aplikasi berbasis web yang menyediakan *inputan* berupa gambar bunga *irish* dan pemindai , *output* hasil berupa gambar spesies *irish* yang diprediksi dengan akurasi.

Backend: *PHP/Laravel*

Modeling: *Scikit-learn/tensorflow*

Frontend: *Javascript,CSS*.

3.4 Diagram Alur Penelitian



Gambar 3. 20 Diagram Metodologi Penelitian

Berikut adalah diagram alur yang menggambarkan tahapan-tahapan dalam penelitian ini:

1. Business Understanding:

- Membandingkan performa *SVM*, *DT*, & *NB* pengelompokkan spesies *Iris*.
- Menentukan algoritma mana yang memiliki kinerja terbaik berdasarkan metrik akurasi, precision, *recall*, dan *F1-score*.
- Mengimplementasikan metodologi *CRISP-DM* untuk pengolahan data menggunakan *Python* dan *scikit-learn*.

2. Data Understanding

- Mengumpulkan data.
- Eksplorasi data awal.

3. Data Preparation:

- Pembersihan data.
- Transformasi data.
- Pembagian data menjadi data latih dan data uji.

4. Modeling:

- Implementasi algoritma *SVM*, *Decision Tree*, dan *Naive Bayes*.
- Tuning hyperparameter.

5. Evaluation:

- Evaluasi kinerja model menggunakan data uji.
- Analisis hasil evaluasi.

6. Deployment:

- Aplikasi Pengelompokan Spesies *Iris* Berbasis Web.

