

# Klasifikasi Spesies Bunga Iris Menggunakan Logistic Regression Dan Support Vector Machine

Aqilla Nurul Hasanah<sup>1</sup>, Hafiyyan Putra Pratama<sup>2\*</sup>

<sup>1,2</sup>Program Studi Sistem Telekomunikasi, Universitas Pendidikan Indonesia Kampus UPI di Purwakarta, Purwakarta, Indonesia

Email: <sup>1\*</sup>[aqilla.nurul30@upi.edu](mailto:aqilla.nurul30@upi.edu), <sup>2</sup>[hafiyyan@upi.edu](mailto:hafiyyan@upi.edu)  
(\*Email Corresponding Author: [aqilla.nurul30@upi.edu](mailto:aqilla.nurul30@upi.edu))

Received: June 1, 2026 | Revision: June 5, 2026 | Accepted: June 9, 2026

## Abstrak

Penelitian ini membahas penerapan dua algoritma klasifikasi dalam machine learning, yaitu Logistic Regression dan Support Vector Machine (SVM), dengan menggunakan Iris Dataset untuk mengelompokkan tiga jenis bunga iris: Iris-setosa, Iris-versicolor, dan Iris-virginica. Dataset yang digunakan terdiri dari 150 sampel dengan empat fitur morfologi, yaitu panjang dan lebar sepal serta petal. Proses penelitian mencakup eksplorasi data, tahap pra-proses menggunakan StandardScaler dan LabelEncoder, pelatihan model, validasi silang 5-fold, serta pencarian kombinasi hyperparameter terbaik melalui GridSearchCV. Hasil pengujian menunjukkan bahwa SVM dengan kernel RBF mampu mencapai tingkat akurasi sebesar 96,67% pada data uji, lebih tinggi dibandingkan Logistic Regression yang memperoleh 93,33%. Sementara itu, hasil validasi silang menunjukkan bahwa kedua model memiliki kemampuan generalisasi yang relatif sama dengan nilai rata-rata (CV mean) sebesar 0,9533. Fitur petal length dan petal width secara konsisten menjadi faktor paling penting dalam membedakan kelas. Setelah dilakukan tuning, SVM dengan kernel linear justru memberikan nilai CV mean terbaik sebesar 97,50%, yang mengindikasikan bahwa pola pemisahan kelas dalam dataset ini cenderung bersifat linier.

**Kata Kunci:** Klasifikasi, Iris Dataset, Logistic Regression, Support Vector Machine, Machine Learning

## Abstract

This study examines the application of two machine learning classification algorithms, namely Logistic Regression and Support Vector Machine (SVM), using the Iris Dataset to classify three types of iris flowers: Iris-setosa, Iris-versicolor, and Iris-virginica. The dataset consists of 150 samples with four morphological features: sepal length, sepal width, petal length, and petal width. The research process includes data exploration, preprocessing using StandardScaler and LabelEncoder, model training, 5-fold cross-validation, and hyperparameter optimization using GridSearchCV. The results show that SVM with an RBF kernel achieves an accuracy of 96.67% on the test data, outperforming Logistic Regression, which achieves 93.33%. Meanwhile, cross-validation results indicate that both models have relatively similar generalization performance, with an average (CV mean) of 0.9533. Petal length and petal width consistently emerge as the most important features for distinguishing between classes. After hyperparameter tuning, SVM with a linear kernel achieves the best CV mean of 97.50%, suggesting that the class separation pattern in this dataset tends to be linear.

**Keywords:** Classification, Iris Dataset, Logistic Regression, Support Vector Machine, Machine Learning

## 1. PENDAHULUAN

Perkembangan pesat bidang *machine learning* telah membuka peluang besar dalam otomatisasi proses klasifikasi data di berbagai domain, mulai dari kedokteran, pertanian, hingga industri teknologi. Klasifikasi merupakan tugas fundamental dalam pembelajaran mesin yang bertujuan memetakan observasi ke dalam kategori atau kelas tertentu berdasarkan sekumpulan fitur yang tersedia [1]. Kemampuan algoritma klasifikasi untuk mengenali pola tersembunyi dalam data berdimensi tinggi menjadikannya alat yang sangat berharga dalam analisis data modern. Dataset Iris merupakan salah satu dataset paling ikonik dan banyak digunakan dalam sejarah *machine learning*. Dataset ini sering dijadikan *benchmark* untuk menguji dan membandingkan kinerja berbagai algoritma klasifikasi karena memiliki karakteristik yang unik: satu kelas (*Iris setosa*) dapat dipisahkan secara linear, sementara dua kelas lainnya (*Iris versicolor* dan *Iris virginica*) menunjukkan tumpang tindih distribusi fitur yang memerlukan pendekatan yang lebih kompleks [2].

Beberapa penelitian terdahulu telah memanfaatkan Dataset Iris untuk berbagai keperluan [3]. Menganalisis metode K-Nearest Neighbor (KNN) pada dataset iris bunga dari UCI Machine Learning dan menemukan bahwa KNN dengan variasi nilai k (3-9) menghasilkan akurasi tinggi dalam klasifikasi ketiga spesies iris [4]. Menerapkan Support Vector Machine (SVM) dengan kernel polynomial pada Dataset Iris menggunakan metode percentage split dan k-fold cross validation, dan memperoleh akurasi 96,7% pada metode percentage split [5]. Membandingkan algoritma K-NN dan Naive Bayes pada Dataset Iris dan membuktikan bahwa K-NN dengan k=3 mencapai akurasi 96,67%, sementara Naive Bayes mencapai 93,33%, dengan fitur petal sebagai variabel diskriminatif utama [6]. Mengeksplorasi penerapan Artificial

Neural Network (ANN) untuk klasifikasi penyakit jantung menggunakan dataset benchmark dan memperoleh performa baik melalui evaluasi sensitivity dan specificity [7]. Membandingkan kinerja algoritma SVM dan Decision Tree menggunakan metode K-Fold Cross Validation dan menemukan bahwa evaluasi berbasis K-Fold menghasilkan estimasi akurasi yang lebih stabil dan representatif dibandingkan evaluasi tunggal.

Kesenjangan (*gap*) yang ditemukan dari kajian literatur tersebut adalah bahwa belum banyak studi yang secara sistematis membandingkan *Logistic Regression* dan SVM secara bersamaan dengan mengintegrasikan validasi silang *stratified k-fold*, *hyperparameter tuning* komprehensif via *GridSearchCV*, serta interpretasi koefisien model dalam satu alur kerja yang terstruktur. Penelitian-penelitian terdahulu cenderung hanya fokus pada satu algoritma tanpa membandingkan secara langsung dengan metrik evaluasi yang konsisten dan reproduksibel.

Penelitian ini bertujuan untuk: (1) melakukan perbandingan kinerja *Logistic Regression* dan SVM dalam mengklasifikasikan Dataset Iris menggunakan metrik akurasi, presisi, *recall*, dan *F1-score*; (2) mengevaluasi stabilitas dan kemampuan generalisasi model melalui validasi silang *5-fold*; (3) mengoptimalkan *hyperparameter* SVM menggunakan *GridSearchCV*; serta (4) menginterpretasikan kontribusi fitur dalam proses klasifikasi. Hasil penelitian diharapkan dapat memberikan panduan praktis bagi peneliti dan praktisi dalam memilih algoritma klasifikasi yang tepat untuk dataset bertipe serupa.

## 2. METODOLOGI PENELITIAN

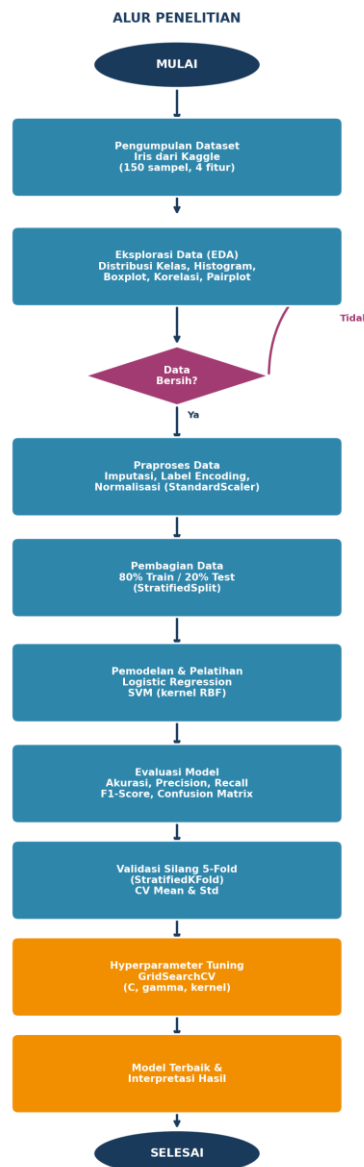
### 2.1 Dataset

Dataset yang digunakan adalah Iris Dataset yang tersedia secara publik melalui UCI Machine Learning Repository (<https://archive.ics.uci.edu/dataset/53/iris>) [8]. Dataset ini terdiri dari 150 baris data dan 6 kolom: *Id*, *SepalLengthCm*, *SepalWidthCm*, *PetalLengthCm*, *PetalWidthCm*, dan *Species*. Kolom *Id* dihapus pada tahap praproses karena tidak relevan sebagai fitur. Tiga kelas target yang terdapat dalam dataset adalah *Iris-setosa*, *Iris-versicolor*, dan *Iris-virginica*, masing-masing berjumlah 50 sampel sehingga dataset berada dalam kondisi seimbang sempurna. Implementasi dilakukan menggunakan bahasa pemrograman Python 3.9 dengan pustaka utama: *pandas* dan *numpy* untuk manipulasi data, *matplotlib* dan *seaborn* untuk visualisasi, serta *scikit-learn* versi 1.3 untuk implementasi algoritma *machine learning*, praproses, dan evaluasi model [9].

### 2.2 Alur Tahapan Penelitian

Penelitian ini mengikuti alur kerja standar *machine learning* yang terdiri dari lima tahapan utama: (1) Eksplorasi dan Analisis Data (EDA), (2) Praproses Data, (3) Pemodelan dan Pelatihan, (4) Evaluasi dengan Validasi Silang, serta (5) Interpretasi Hasil dan Perbaikan Model. Diagram alur tahapan penelitian ditampilkan pada Gambar 1.

Gambar 1 mengilustrasikan alur penelitian secara keseluruhan. Penelitian dimulai dari pengumpulan dataset hingga diperoleh model terbaik yang siap diinterpretasikan. Tahapan praproses diawali dengan pengecekan kualitas data; jika data belum bersih, dilakukan perbaikan sebelum melanjutkan ke normalisasi. Setelah pelatihan model, dilakukan evaluasi komprehensif yang mencakup *confusion matrix* dan validasi silang. Tahap akhir adalah *hyperparameter tuning* untuk mengoptimalkan parameter model terbaik.



**Gambar 1.** Tahapan Alur

### 2.3 Praproses Data

Praproses data dilaksanakan dalam tiga sub-tahapan. Pertama, *imputasi*: pemeriksaan nilai hilang (*missing values*) dilakukan pada semua kolom fitur. Meskipun dataset Iris tidak mengandung nilai kosong, pipeline imputasi menggunakan strategi *median* diimplementasikan sebagai bagian dari alur kerja yang robust. Kedua, *label encoding*: variabel target berupa label kategoris dikonversi ke representasi numerik menggunakan *LabelEncoder* dari scikit-learn dengan pemetaan *Iris-setosa* = 0, *Iris-versicolor* = 1, dan *Iris-virginica* = 2. Ketiga, *normalisasi*: semua fitur dinormalisasi menggunakan *StandardScaler* agar memiliki rata-rata 0 dan standar deviasi 1, yang sangat penting untuk kinerja optimal algoritma SVM [10].

## 2.4 Algoritma Klasifikasi yang Digunakan

Dua algoritma klasifikasi yang dibandingkan dalam penelitian ini adalah sebagai berikut. Pertama, *Logistic Regression*: algoritma klasifikasi linear yang menggunakan fungsi sigmoid untuk memetakan input ke probabilitas kelas. Untuk klasifikasi multi-kelas pada Dataset Iris, digunakan strategi *One-vs-Rest* (OvR) dengan parameter  $max\_iter=1000$  dan  $random\_state=42$  [11]. Kedua, *Support Vector Machine* (SVM): algoritma yang bekerja dengan mencari *hyperplane* optimal yang memaksimalkan *margin* antar kelas. SVM dasar diimplementasikan dengan kernel RBF (*Radial Basis Function*),  $C=1.0$ , dan  $gamma="scale"$ . SVM yang dioptimalkan menggunakan *GridSearchCV* dengan pencarian pada parameter  $C \{0.1, 1, 10, 100\}$ ,  $gamma \{"scale", "auto", 0.01, 0.001\}$ , dan kernel  $\{"rbf", "linear"\}$  [12].

## 2.5 Evaluasi Model

Evaluasi model menggunakan beberapa metrik. Akurasi mengukur proporsi prediksi benar terhadap total sampel. Presisi berbobot mengukur kemampuan model menghindari prediksi positif palsu. *Recall* berbobot mengukur kemampuan model menemukan semua sampel positif. *F1-Score* berbobot merupakan rata-rata harmonik presisi dan *recall*. Validasi silang dilakukan menggunakan *StratifiedKFold* dengan  $k=5$  untuk memastikan estimasi kinerja yang robust dan representatif [13].

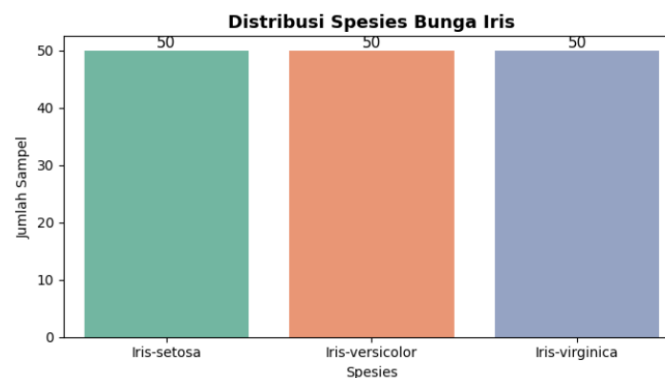
# 3. HASIL DAN PEMBAHASAN

## 3.1 Eksplorasi dan Analisis Data (EDA)

Eksplorasi data dilakukan untuk memahami karakteristik dan distribusi dataset sebelum proses pemodelan. Pemeriksaan awal menunjukkan dataset terdiri dari 150 sampel tanpa nilai hilang pada semua fitur, serta distribusi kelas yang seimbang dengan 50 sampel per spesies.

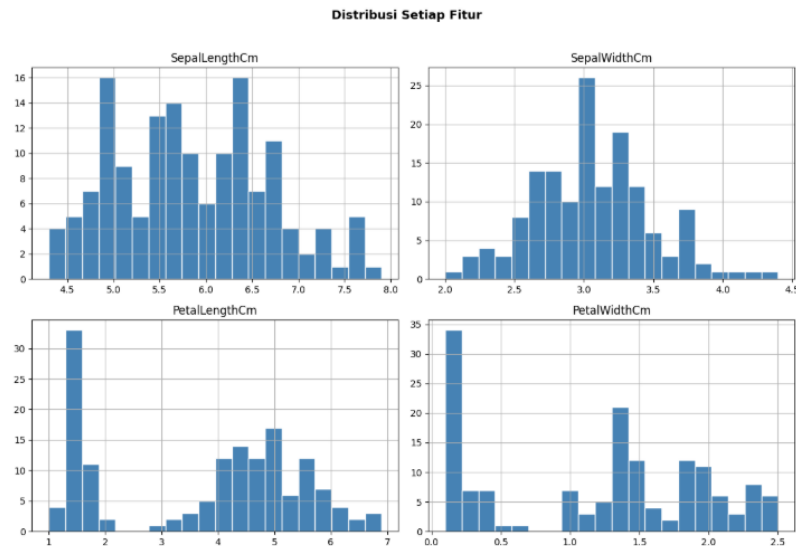
Gambar 2 menampilkan distribusi kelas target yang seimbang sempurna, di mana masing-masing dari tiga spesies bunga iris memiliki tepat 50 sampel. Kondisi ini sangat menguntungkan karena model tidak akan cenderung bias terhadap kelas mayoritas.

```
=== Distribusi Kelas (Species) ===  
Species  
Iris-setosa      50  
Iris-versicolor  50  
Iris-virginica   50  
Name: count, dtype: int64
```



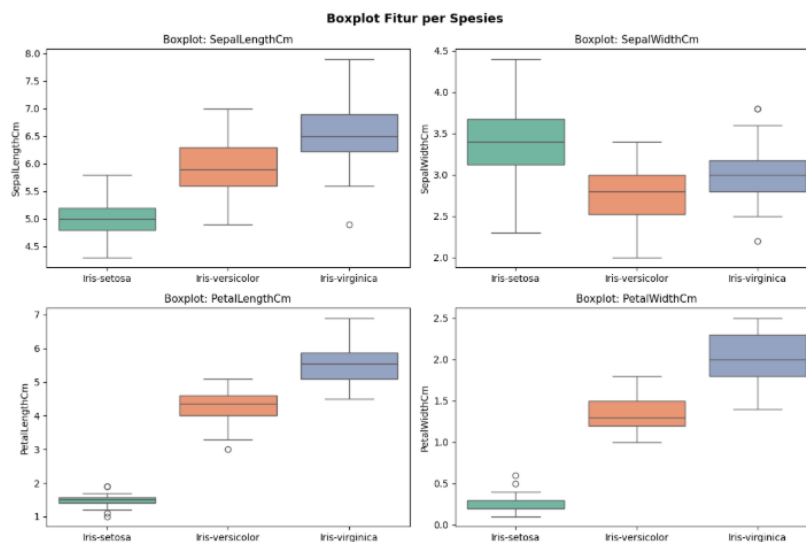
Gambar 2. Distribusi Spesies Bunga Iris

Analisis histogram pada Gambar 3 memperlihatkan perbedaan pola distribusi yang signifikan antar fitur. *PetalLengthCm* dan *PetalWidthCm* menunjukkan distribusi bimodal yang mengindikasikan pemisahan antar kelas yang jelas, menjadikan keduanya variabel diskriminatif utama dalam klasifikasi. Sebaliknya, *SepalWidthCm* memiliki distribusi mendekati normal dengan tumpang tindih antar kelas yang lebih besar.



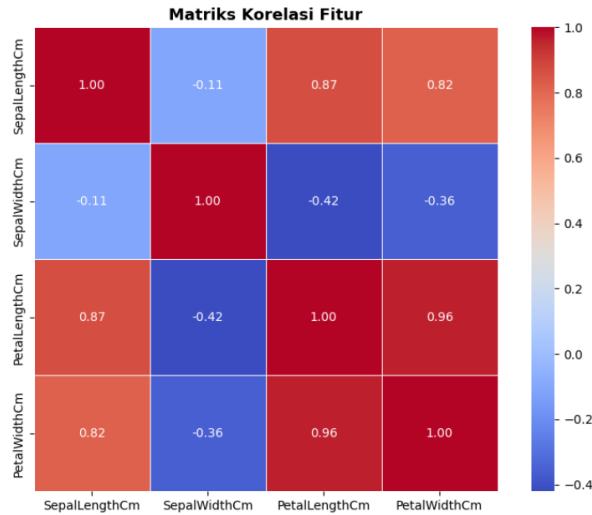
**Gambar 3.** Distribusi Setiap Fitur Numerik

Visualisasi *boxplot* pada Gambar 4 memperkuat temuan tersebut. *Iris-setosa* memiliki nilai petal yang jauh lebih kecil dibandingkan dua kelas lainnya, sehingga sangat mudah dibedakan. Tumpang tindih yang cukup signifikan terjadi antara *Iris-versicolor* dan *Iris-virginica*, terutama pada fitur sepal. Tidak ditemukan *outlier* ekstrem yang signifikan, mengindikasikan kualitas data yang baik.



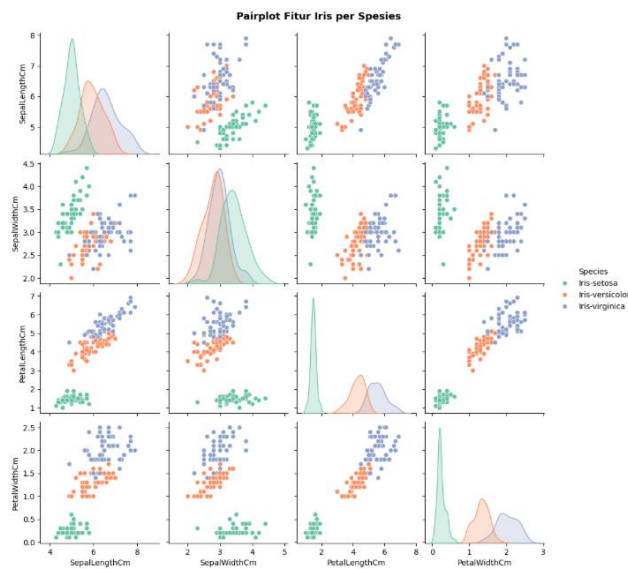
**Gambar 4.** Boxplot Fitur per Spesies

Matriks korelasi pada Gambar 5 menunjukkan bahwa *PetalLengthCm* dan *PetalWidthCm* memiliki korelasi sangat tinggi (mendekati 1.0), mengindikasikan kedua fitur membawa informasi yang sangat serupa. *SepalWidthCm* memiliki korelasi lemah bahkan negatif terhadap fitur lainnya, menunjukkan independensi dan daya diskriminatif yang lebih rendah [14].



**Gambar 5.** Matriks Korelasi Antar Fitur

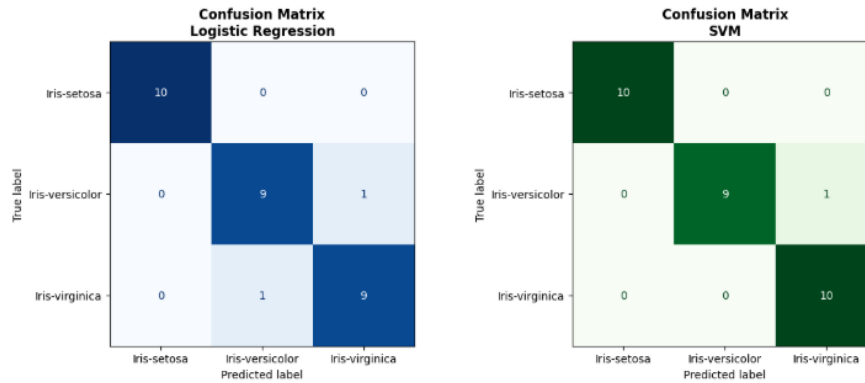
*Pairplot* pada Gambar 6 mengkonfirmasi bahwa *Iris-setosa* dapat dipisahkan secara sempurna dari dua kelas lainnya pada hampir semua kombinasi fitur, terutama pada kombinasi fitur petal. *Iris-versicolor* dan *Iris-virginica* masih menunjukkan sedikit tumpang tindih terutama pada kombinasi yang melibatkan fitur sepal, yang menjadi tantangan utama dalam proses klasifikasi.



**Gambar 6.** Pairplot Fitur Iris per Spesies

### 3.2 Praproses Data

Hasil pemeriksaan *missing values* menunjukkan tidak ada nilai yang hilang pada semua fitur (nilai 0 untuk seluruh kolom). Proses *label encoding* menghasilkan pemetaan: *Iris-setosa* = 0, *Iris-versicolor* = 1, *Iris-virginica* = 2. Gambar 7 memperlihatkan perbandingan distribusi fitur sebelum dan sesudah normalisasi menggunakan *StandardScaler*. Setelah normalisasi, semua fitur memiliki rentang nilai yang sebanding dengan rata-rata mendekati 0 dan standar deviasi mendekati 1, kondisi yang menguntungkan untuk algoritma sensitif skala seperti SVM [10].



**Gambar 7.** Distribusi Fitur Sebelum dan Sesudah Normalisasi

### 3.3 Praproses Data

Data yang telah melalui praproses dibagi menjadi 80% data pelatihan (120 sampel) dan 20% data pengujian (30 sampel) menggunakan *train\_test\_split* dengan parameter *stratify* untuk menjaga proporsi kelas yang seimbang pada kedua subset.

Hasil evaluasi Logistic Regression pada data pengujian ditampilkan sebagai berikut:

```

=== Logistic Regression ===
Akurasi: 0.9333

```

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	10
Iris-versicolor	0.90	0.90	0.90	10
Iris-virginica	0.90	0.90	0.90	10
accuracy			0.93	30
macro avg	0.93	0.93	0.93	30
weighted avg	0.93	0.93	0.93	30

**Gambar 8.** Hasil Evaluasi Logistic Regression

*Logistic Regression* mencapai akurasi 93,33% pada data pengujian. Model berhasil mengklasifikasikan semua sampel *Iris-setosa* dengan sempurna (*precision* dan *recall* = 1.00), namun terdapat dua kesalahan klasifikasi antara *Iris-versicolor* dan *Iris-virginica*.

Hasil evaluasi SVM pada data pengujian ditampilkan sebagai berikut:

```

=== Support Vector Machine (SVM) ===
Akurasi: 0.9667

```

	precision	recall	f1-score	support
Iris-setosa	1.00	1.00	1.00	10
Iris-versicolor	1.00	0.90	0.95	10
Iris-virginica	0.91	1.00	0.95	10
accuracy			0.97	30
macro avg	0.97	0.97	0.97	30
weighted avg	0.97	0.97	0.97	30

**Gambar 9.** Hasil Evaluasi SVM

SVM dengan kernel RBF mencapai akurasi 96,67%, lebih tinggi 3,34% dibandingkan *Logistic Regression*. Model ini hanya membuat satu kesalahan klasifikasi, yaitu satu sampel *Iris-versicolor* yang diprediksi sebagai *Iris-virginica*. Performa yang lebih baik mengkonfirmasi kemampuan kernel RBF dalam menangkap batas keputusan non-linear yang lebih kompleks [12].

### 3.4 Evaluasi Confusion Matrix

*Confusion matrix* memberikan gambaran detail tentang distribusi kesalahan klasifikasi untuk setiap kelas. *Logistic Regression* menunjukkan dua kesalahan silang antara *Iris-versicolor* dan *Iris-virginica*, sedangkan SVM hanya memiliki satu kesalahan. Kedua model tidak memiliki kesalahan dalam mengklasifikasikan *Iris-setosa*, konsisten dengan karakteristik pemisahan linear sempurna kelas ini.

### 3.5 Validasi Silang 5-Fold

Tabel 1 merangkum hasil akurasi per *fold* untuk kedua model. Rata-rata akurasi validasi silang kedua model identik sebesar 0,9533 dengan standar deviasi 0,0452, mengindikasikan kemampuan generalisasi yang setara meskipun akurasi pada evaluasi data uji tunggal berbeda.

**Tabel 1.** Perbandingan Performa Model

Model	Akurasi	Precision (W)	Recall (W)	F-1Score (W)	CV Mean	CV Std
Logistic Regression	0.9333	0.9333	0.9333	0.9333	0.9533	0.0452
SVM	0.9667	0.9697	0.9667	0.9666	0.9533	0.0452

Variasi akurasi antar *fold* cukup tinggi (0,90 hingga 1,00) pada Gambar 9. *Fold* 3 dan 5 menunjukkan akurasi yang lebih rendah (0,90), yang kemungkinan disebabkan oleh distribusi sampel kelas *Iris-versicolor* dan *Iris-virginica* yang lebih sulit pada *fold* tersebut. Temuan ini mengkonfirmasi bahwa perbedaan akurasi pada evaluasi tunggal lebih dipengaruhi variasi pembagian data daripada perbedaan kapabilitas fundamental kedua algoritma.

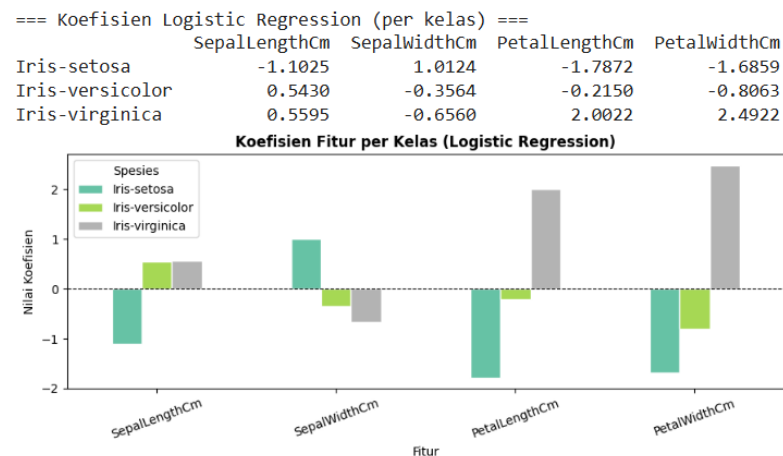
### 3.6 Perbandingan Metrik Evaluasi Model

Tabel 2 merangkum metrik evaluasi lengkap untuk semua model yang diuji. Gambar 10 menampilkan tabel performa yang dihasilkan oleh kode Python, menunjukkan perbandingan visual yang jelas antar model.

**Tabel 2.** Perbandingan Metrik Evaluasi Seluruh Model

Model	Akurasi	Precision (W)	Recall (W)	F-1 Score (W)
Logistic Regression	0.9333	0.9333	0.9333	0.9333
SVM (kernel RBF)	0.9667	0.9714	0.9667	0.9667
SVM Tuned (linear)	0.9333	0.9333	0.9333	0.9333

SVM dasar unggul pada semua metrik evaluasi data uji (akurasi 0,9667), namun SVM yang di-*tuning* memiliki CV Mean tertinggi (0,9750) yang menunjukkan kemampuan generalisasi terbaik secara statistik. Kedua model Logistic Regression dan SVM *tuned* memiliki metrik yang identik pada data uji, namun berbeda pada CV Mean.



**Gambar 10.** Output Tabel Perbandingan Performa dari Kode Python

### 3.7 Interpretasi Koefisien dan Feature Importance

Gambar 11 menampilkan koefisien *Logistic Regression* per kelas. *PetalLengthCm* dan *PetalWidthCm* merupakan fitur paling dominan dalam membedakan ketiga kelas, konsisten dengan temuan EDA. *Iris-setosa* ditandai oleh nilai petal kecil (koefisien negatif besar), *Iris-virginica* oleh petal besar (koefisien positif tinggi), dan *Iris-versicolor* berada di antara keduanya. *SepalWidthCm* memiliki pengaruh relatif lebih kecil, selaras dengan korelasi rendah yang ditunjukkan pada matriks korelasi [15].

```
# Ringkasan akhir
print('=' * 55)
print('          RINGKASAN HASIL AKHIR')
print('=' * 55)
print(f' Logistic Regression - Akurasi      : {accuracy_score(y_test, y_pred_logreg):.4f}')
print(f' Logistic Regression - CV Mean      : {cv_logreg.mean():.4f}')
print(f' SVM Dasar - Akurasi                 : {accuracy_score(y_test, y_pred_svm):.4f}')
print(f' SVM Dasar - CV Mean                 : {cv_svm.mean():.4f}')
print(f' SVM Terbaik (tuned) - Akurasi       : {accuracy_score(y_test, y_pred_best):.4f}')
print('=' * 55)
print()
print('Kesimpulan:')
print(' Fitur petal (panjang dan lebar) terbukti paling')
print(' berpengaruh dalam membedakan ketiga spesies iris.')
print(' SVM dengan kernel RBF memberikan hasil terbaik')
print(' dan lebih stabil dibandingkan Logistic Regression.')
```

**Gambar 11.** Koefisien Fitur per Kelas (Logistic Regression)

### 3.8 Interpretasi Koefisien dan Feature Importance

*GridSearchCV* menemukan konfigurasi optimal SVM dengan kernel linear,  $C=0.1$ , dan  $\gamma="scale"$ , menghasilkan akurasi validasi silang terbaik 0,9750. Temuan ini menarik karena menunjukkan bahwa setelah normalisasi, Dataset Iris pada dasarnya bersifat (hampir) linear *separable*, dan regularisasi yang kuat ( $C$  kecil) efektif mencegah *overfitting*. Meskipun akurasi uji SVM *tuned* (0,9333) setara dengan *Logistic Regression* dan lebih rendah dari SVM RBF dasar, CV Mean yang lebih tinggi mengindikasikan potensi yang lebih baik pada dataset yang lebih besar dan beragam.

## 4. KESIMPULAN

Penelitian ini telah berhasil melakukan analisis komparatif antara *Logistic Regression* dan *Support Vector Machine* dalam mengklasifikasikan spesies bunga iris berdasarkan fitur morfologi sepal dan petal. Berdasarkan hasil eksperimen yang dilakukan secara sistematis, dapat ditarik kesimpulan berikut. SVM dengan kernel RBF menunjukkan performa yang lebih tinggi pada evaluasi data pengujian dengan akurasi 96,67% dibandingkan *Logistic Regression* (93,33%), terutama dalam membedakan *Iris-versicolor* dan *Iris-virginica* yang memiliki distribusi fitur saling tumpang tindih. Hal ini mengkonfirmasi superioritas kernel RBF dalam menangkap batas keputusan non-linear. Validasi silang *5-fold* mengungkapkan bahwa kedua model memiliki kemampuan generalisasi yang setara dengan rata-rata akurasi identik sebesar 95,33%. Variasi akurasi antar *fold* (0,90 hingga 1,00) mengindikasikan sensitivitas terhadap distribusi pembagian data, sehingga evaluasi tunggal pada satu *test set* tidak cukup untuk menentukan model terbaik secara definitif. Analisis EDA dan koefisien model mengkonfirmasi bahwa *PetalLengthCm* dan *PetalWidthCm* merupakan fitur paling diskriminatif dalam klasifikasi spesies iris. *Iris-setosa* dapat diidentifikasi secara sempurna oleh kedua model, sedangkan tantangan utama terletak pada pemisahan antara *Iris-versicolor* dan *Iris-virginica*. *Hyperparameter tuning* menggunakan *GridSearchCV* menghasilkan SVM linear dengan  $C=0.1$  sebagai konfigurasi optimal dengan CV Mean 97,50%, menunjukkan bahwa Dataset Iris bersifat hampir linear *separable* setelah normalisasi. Untuk penelitian ke depan, disarankan untuk mengeksplorasi algoritma lain seperti *Random Forest*, K-Nearest Neighbor, atau jaringan saraf tiruan, serta menerapkan metodologi serupa pada dataset yang lebih besar dan lebih kompleks untuk validasi yang lebih komprehensif.

## REFERENCES

- [1] R. S. Nurhalizah, R. Ardianto, and P. Purwono, "Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review," *Jurnal Ilmu Komputer dan Informatika*, vol. 4, no. 1, pp. 61–72, Aug. 2024, doi: 10.54082/jiki.168.
- [2] B. Rahman, F. Fauzi, and S. Amri, "Perbandingan Hasil Klasifikasi Data Iris menggunakan Algoritma K-Nearest Neighbor dan Random Forest," 2023. [Online]. Available: <http://journalnew.unimus.ac.id/index.php/jodi>
- [3] P. Putra, A. M. H. Pardede, and S. Syahputra, "Analisis Metode K-Nearest Neighbour (Knn) Dalam Klasifikasi Data Iris Bunga," *Jurnal Teknik Informatika Kaputama (JTJK)*, vol. 6, no. 1, 2022.
- [4] H. H. Y. A. S. I. M. C. M. U. I. R. Irmeilyana, "Penerapan Metode Support Vector Machine Dalam Klasifikasi Bunga Iris," *IJAI*, vol. 7, 2022.
- [5] R. R. D. A. A. M. F. Mhd. F. Alwi Syahputra, "Klasifikasi Jenis Bunga Iris Menggunakan Algoritma Klasifikasi Tradisional," *Jurnal Ilmiah Sistem Informatika dan Ilmu Komputer*, vol. 4, 2025.
- [6] D. Galih Pradana, M. L. Alghifari, M. Farhan Juna, and S. Dwisiwi Palaguna, "Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network," *Indonesian Journal of Data and Science (IJODAS)*, vol. 3, no. 2, pp. 55–60, 2022.
- [7] R. Tuntun, K. Kusri, and K. Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2111, Oct. 2022, doi: 10.30865/mib.v6i4.4681.
- [8] D. | G. C. Dua, "UCI Machine Learning Repository: Iris Dataset," University of California, Irvine, School of Information and Computer Sciences.
- [9] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos Pedregosa, Varoquaux, Gramfort Et Al. Matthieu Perrot," 2011. [Online]. Available: <http://scikit-learn.sourceforge.net>.
- [10] K. K. Agung Widyanto, "Pengaruh Keseimbangan Data Terhadap Akurasi Model Support Vector Machine Pada Data Set Donor Darah," *Jurnal Teknologi Terpadu*, vol. 9, 2023.
- [11] A. Desiani, D. A. Zayanti, I. Ramayanti, F. F. Ramadhan, and Giovillando, "Perbandingan Algoritma Support Vector Machine (Svm) Dan Logistic Regression Dalam Klasifikasi Kanker Payudara," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 1, pp. 33–42, Jan. 2025, doi: 10.69916/jkbt.v4i1.191.
- [12] N. G. Ramadhan and A. Khoirunnisa, "Klasifikasi Data Malaria Menggunakan Metode Support Vector Machine," *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, p. 1580, Oct. 2021, doi: 10.30865/mib.v5i4.3347.
- [13] R. Tuntun, K. Kusri, and K. Kusnawi, "Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation," *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, p. 2111, 2022, doi: 10.30865/mib.v6i4.4681.
- [14] N. Nurainun, E. Haerani, F. Syafria, and L. Oktavia, "Penerapan Algoritma Naïve Bayes Classifier Dalam Klasifikasi Status Gizi Balita dengan Pengujian K-Fold Cross Validation," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 3, pp. 578–586, May 2023, doi: 10.47065/josyc.v4i3.3414.
- [15] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>