

## Penerapan Support Vector Machine dalam Klasifikasi Ukuran Udang Putih (*Litopenaeus vannamei*) Berdasarkan Parameter Panjang

Ervina Kartika Sari<sup>1</sup>

<sup>1</sup>Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

Email: [ervinakartikasari75@gmail.com](mailto:ervinakartikasari75@gmail.com)<sup>1</sup>

### ABSTRAK

Proses klasifikasi ukuran udang yang dilakukan secara konvensional dapat memiliki keterbatasan dalam konsistensi hasil dan efisiensi proses pengelompokan. Penelitian ini bertujuan untuk menerapkan metode Support Vector Machine (SVM) dalam mengklasifikasikan ukuran udang putih (*Litopenaeus vannamei*) ke dalam kategori Kecil, Sedang, dan Besar berdasarkan parameter panjang tubuh. Dataset penelitian terdiri dari 1.000 data panjang tubuh udang putih yang dikumpulkan secara langsung. Setelah dilakukan proses pembersihan dan pengolahan data, diperoleh 996 data yang dapat digunakan dalam proses klasifikasi. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data, penskalaan menggunakan StandardScaler, validasi model menggunakan Stratified K-Fold Cross Validation, klasifikasi menggunakan metode SVM, dan evaluasi model. Evaluasi dilakukan menggunakan Confusion Matrix, Accuracy, Precision, Recall, dan F1-Score. Hasil pengujian menggunakan Stratified K-Fold Cross Validation dengan k=5 memperoleh rata-rata akurasi sebesar 98,39% dengan standar deviasi sebesar 0,58%. Sementara itu, pengujian menggunakan Percentage Split dengan rasio 80:20 menghasilkan akurasi sebesar 98,00% pada 200 data testing, dengan 196 data berhasil diklasifikasikan dengan benar dan 4 data mengalami kesalahan klasifikasi. Berdasarkan hasil evaluasi, nilai Precision sebesar 99,07%, Recall sebesar 96,67%, dan F1-Score sebesar 97,78%. Hasil tersebut menunjukkan bahwa metode SVM dapat digunakan untuk melakukan klasifikasi ukuran udang putih berdasarkan parameter panjang dengan kinerja yang baik.

**Kata Kunci:** Support Vector Machine, klasifikasi ukuran udang putih, *Litopenaeus vannamei*, machine learning, panjang udang.

### ABSTRACT

Conventional shrimp size classification processes can suffer from limitations regarding result consistency and grouping efficiency. This study aims to apply the Support Vector Machine (SVM) method to classify the size of whiteleg shrimp (*Litopenaeus vannamei*) into Small, Medium, and Large categories based on body length parameters. The research dataset consists of 1,000 body length measurements of whiteleg shrimp collected directly. Following data cleaning and processing, 996 data points remained suitable for the classification process. The research stages included data collection, data preprocessing, scaling using StandardScaler, model validation using Stratified K-Fold Cross-Validation, classification using the SVM method, and model evaluation. Evaluation was conducted using a confusion matrix, accuracy, precision, recall, and F1-score. Testing results using Stratified K-Fold Cross-Validation (with k=5) yielded an average accuracy of 98.39% with a standard deviation of 0.58%. Meanwhile, testing using a Percentage Split with an 80:20 ratio resulted in an accuracy of 98.00% on 200 test samples, with 196 samples correctly classified and 4 misclassified. Based on the evaluation results, the precision was 99.07%, recall was 96.67%, and the F1-score was 97.78%. These results demonstrate that the SVM method can effectively classify whiteleg shrimp sizes based on length parameters with high performance.

**Keywords:** Support Vector Machine, white shrimp size classification, *Litopenaeus vannamei*, machine learning, shrimp length.

## 1 PENDAHULUAN

Komoditas perikanan merupakan salah satu sektor strategis yang berkontribusi terhadap penyediaan pangan, peningkatan pendapatan masyarakat, serta pertumbuhan ekonomi nasional.

Salah satu komoditas unggulan pada sektor budidaya perikanan adalah udang putih (*Litopenaeus vannamei*), yang memiliki produktivitas tinggi dan permintaan pasar yang terus meningkat, baik untuk kebutuhan dalam negeri maupun ekspor [1], [2]. Tingginya nilai ekonomi udang putih menjadikan proses pengendalian mutu sebagai tahapan yang sangat penting sebelum produk dipasarkan. Salah satu parameter utama dalam penentuan mutu tersebut adalah ukuran tubuh udang, karena ukuran berpengaruh langsung terhadap harga jual, standar kualitas, serta klasifikasi produk yang ditetapkan oleh industri pengolahan dan pasar ekspor [3]. Oleh sebab itu, proses pengelompokan ukuran (grading) menjadi bagian yang tidak dapat dipisahkan dalam rantai produksi udang putih.

Pada praktiknya, proses grading udang di berbagai unit budidaya maupun industri pengolahan masih banyak dilakukan secara manual dengan mengandalkan pengamatan tenaga kerja. Cara tersebut memiliki beberapa keterbatasan, antara lain hasil klasifikasi yang kurang konsisten, membutuhkan waktu yang relatif lama, serta rentan terhadap human error akibat perbedaan tingkat ketelitian dan subjektivitas masing-masing pekerja [4], [5]. Seiring meningkatnya volume produksi udang putih setiap tahun, proses penyortiran secara manual menjadi semakin kurang efisien dan berpotensi menurunkan produktivitas. Kondisi tersebut menunjukkan perlunya suatu sistem klasifikasi yang mampu bekerja secara otomatis, objektif, dan konsisten sehingga proses grading dapat dilakukan dengan lebih cepat dan akurat [6].

Perkembangan teknologi machine learning memberikan peluang besar dalam menyelesaikan berbagai permasalahan klasifikasi berdasarkan pola yang terdapat pada data [7]. Berbagai algoritma klasifikasi telah banyak diterapkan, di antaranya K-Nearest Neighbor (KNN), Random Forest (RF), dan Support Vector Machine (SVM) [8], [9]. Penelitian sebelumnya menunjukkan bahwa algoritma tersebut mampu memberikan performa yang baik pada berbagai kasus klasifikasi di bidang pertanian, peternakan, maupun perikanan [10]. Namun, sebagian besar penelitian di bidang perikanan masih berfokus pada identifikasi spesies, deteksi penyakit, estimasi berat, maupun klasifikasi berbasis citra digital. Penelitian yang memanfaatkan parameter morfometrik berupa panjang tubuh sebagai fitur utama untuk mengklasifikasikan ukuran udang putih menggunakan metode Support Vector Machine (SVM) masih relatif terbatas. Kondisi tersebut membuka peluang untuk mengembangkan model klasifikasi berbasis data numerik yang lebih sederhana, mudah diterapkan, serta memiliki potensi implementasi yang tinggi pada proses grading di lapangan.

Dalam kegiatan perdagangan udang putih, panjang tubuh merupakan salah satu parameter morfometrik yang paling umum digunakan untuk menentukan kelas ukuran. Parameter ini dipilih karena mudah diperoleh melalui proses pengukuran, memiliki hubungan yang erat dengan ukuran komersial udang, serta mampu merepresentasikan karakteristik ukuran secara objektif [11]. Pemanfaatan data panjang sebagai fitur utama diharapkan dapat mengurangi subjektivitas dalam proses penilaian manual sekaligus meningkatkan konsistensi pengelompokan ukuran udang ke dalam kategori kecil, sedang, dan besar [12].

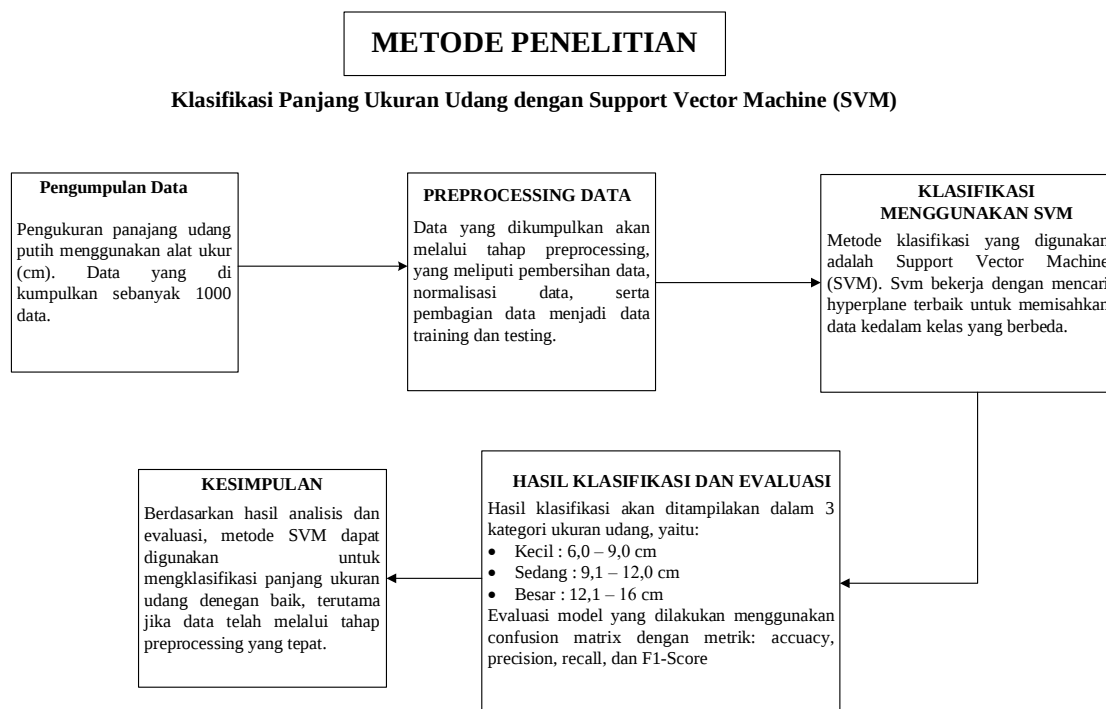
Salah satu algoritma klasifikasi yang banyak digunakan dalam penelitian machine learning adalah Support Vector Machine (SVM). Algoritma ini bekerja dengan membentuk hyperplane optimal yang mampu memisahkan data ke dalam beberapa kelas berdasarkan margin maksimum [13]. Selain memiliki kemampuan generalisasi yang baik, SVM juga dikenal efektif dalam menangani data numerik dengan jumlah fitur yang relatif sedikit serta mampu menghasilkan performa klasifikasi yang tinggi pada berbagai penelitian sebelumnya [14]. Oleh karena itu, metode SVM dinilai sesuai untuk diterapkan pada penelitian ini yang menggunakan panjang tubuh sebagai parameter utama dalam proses klasifikasi ukuran udang putih. Untuk mengetahui tingkat kinerja model yang dihasilkan, proses evaluasi dilakukan menggunakan Confusion Matrix dengan parameter accuracy, precision, dan recall sehingga kemampuan model dapat diukur secara objektif [15].

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan model klasifikasi ukuran udang putih (*Litopenaeus vannamei*) berdasarkan data panjang tubuh menggunakan

metode Support Vector Machine (SVM). Tahapan penelitian meliputi pengumpulan dataset, preprocessing data, normalisasi menggunakan metode Min-Max Scaling, proses pelatihan dan pengujian model SVM, serta evaluasi menggunakan Confusion Matrix. Hasil penelitian diharapkan dapat memberikan alternatif solusi dalam proses klasifikasi ukuran udang secara otomatis sehingga mampu meningkatkan efisiensi, konsistensi, dan objektivitas proses grading, sekaligus menjadi referensi bagi pengembangan penerapan machine learning pada bidang perikanan.

## 2 METODE PENELITIAN

Penelitian ini menerapkan metode Support Vector Machine (SVM) untuk mengklasifikasikan ukuran udang putih (*Litopenaeus vannamei*) berdasarkan data panjang tubuhnya. Proses penelitian dilakukan melalui beberapa tahapan yang dikerjakan secara berurutan, meliputi pengumpulan data, pra-pemrosesan data, klasifikasi menggunakan metode SVM, dan evaluasi hasil. Alur keseluruhan tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1 Alur Penelitian Klasifikasi Ukuran Udang Putih Menggunakan SVM**

### 2.1 Pengumpulan Dataset

Pada tahap ini dilakukan pengumpulan dataset yang berisi data ukuran panjang udang sebanyak 1000 sampel data numerik. Dataset ini digunakan sebagai sumber data utama dalam eksperimen komputasi. Data panjang udang digunakan sebagai parameter fitur tunggal untuk menentukan kategori ukuran udang yang dibagi ke dalam tiga kelas, yaitu kecil, sedang, dan besar. Batasan rentang nilai untuk penentuan kategori ukuran udang berbasis data kuantitatif ini ditetapkan sebagai berikut:

1. Kecil: Udang dengan ukuran panjang 6,0 cm – 9,0 cm.
2. Sedang: Udang dengan ukuran panjang 9,1 cm – 12,0 cm.
3. Besar: Udang dengan ukuran panjang 12,1 cm – 16 cm.

Contoh representasi dari beberapa sampel data panjang udang beserta pelabelan kategorinya yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1. Tabel 1. Ukuran Udang .

**Tabel 1. Contoh Sampel Dataset Panjang Udang Putih**

| No | Panjang Udang (cm) | Kategori |
|----|--------------------|----------|
|----|--------------------|----------|

|   |                  |        |
|---|------------------|--------|
| 1 | 6,0 cm – 9,0 cm  | Kecil  |
| 2 | 9,1 cm – 12,0 cm | Sedang |
| 3 | 12,1 cm – 16 cm  | Besar  |

Penentuan batas rentang panjang untuk masing-masing kategori dalam penelitian ini didasarkan pada data morfometrik panjang tubuh udang putih yang diperoleh dari hasil pengukuran langsung di lapangan. Penggunaan panjang tubuh sebagai parameter ukuran didukung oleh penelitian Setiawan dkk. (2022) yang melaporkan bahwa panjang rata-rata udang vaname pada masa panen adalah 12,3 cm [5]. Selain itu, González dkk. (2024) menjelaskan bahwa pengukuran panjang total tubuh merupakan salah satu variabel morfologi yang digunakan dalam pengukuran karakteristik udang pada bidang akuakultur [1].

Berdasarkan hasil pengukuran pada dataset penelitian, panjang udang yang digunakan berada pada rentang 6,0–15,7 cm. Selanjutnya, data dikelompokkan menjadi tiga kategori berdasarkan rentang panjang yang ditetapkan dalam penelitian, yaitu kategori Kecil sebesar 6,0–9,0 cm, kategori Sedang sebesar 9,1–12,0 cm, dan kategori Besar sebesar 12,1–16,0 cm. Pembagian tersebut digunakan sebagai dasar pelabelan data dalam proses klasifikasi menggunakan metode Support Vector Machine (SVM).

## 2.2 Preprocessing Data Pelatihan dan Pengujian

Sebelum data dimasukkan ke dalam proses pelatihan model, dilakukan tahapan pra-pemrosesan (preprocessing) terlebih dahulu. Tujuan dari tahap ini adalah untuk memastikan kualitas dan keseragaman data sehingga proses pembentukan model klasifikasi SVM dapat berjalan secara optimal. Pra-pemrosesan data dalam penelitian ini mencakup dua langkah utama, yaitu pembersihan data dan normalisasi data.

Langkah pertama adalah pembersihan data (data cleaning). Pada tahap ini, dilakukan pemeriksaan terhadap seluruh sampel data untuk memastikan tidak terdapat nilai yang kosong (missing value) pada kolom panjang udang. Data yang tidak lengkap akan dihapus agar tidak mengganggu proses pelatihan model.

Langkah kedua adalah normalisasi data menggunakan metode Min-Max Scaling. Normalisasi dilakukan karena algoritma SVM sangat sensitif terhadap perbedaan skala nilai fitur masukan. Seluruh nilai panjang udang ditransformasikan ke dalam rentang [0, 1] menggunakan persamaan berikut.

$$x^1 = \frac{x - x_{\min}}{x_{\max}} \quad (1)$$

Keterangan:  $x$  adalah nilai panjang udang asli,  $x_{\min}$  adalah nilai minimum dari seluruh data,  $x_{\max}$  adalah nilai maksimum dari seluruh data, dan  $x^1$  adalah nilai panjang hasil normalisasi. Normalisasi ini bertujuan untuk membantu algoritma SVM dalam menentukan batas pemisah antar kelas secara lebih optimal. Setelah pra-pemrosesan selesai, data dibagi menggunakan rasio 80:20, yaitu 80% sebagai data latih (training) dan 20% sebagai data uji (testing).

## 2.3 Metode Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan metode machine learning terarah (supervised learning) yang bekerja dengan cara mencari hyperplane optimal, yaitu bidang pemisah yang memiliki jarak margin terluas antara kelas-kelas data pada ruang fitur. Kelebihan dari metode SVM adalah kemampuannya dalam menghasilkan generalisasi yang baik, efektif pada data dengan jumlah fitur yang terbatas, serta mampu memberikan performa klasifikasi yang tinggi.

Pada penelitian ini, SVM diterapkan untuk mengklasifikasikan data panjang udang ke dalam tiga kategori ukuran, yaitu Kecil, Sedang, dan Besar. Karena melibatkan lebih dari dua kelas, pendekatan yang digunakan adalah multi-class classification dengan strategi One-vs-One (OvO), di mana model membangun satu classifier untuk setiap pasangan kelas secara berpasangan.

Prinsip kerja SVM adalah membangun batas keputusan (hyperplane) yang memisahkan data ke

dalam kelas yang berbeda dengan margin terlebar. SVM membagi dataset menjadi dua kelas yang dipisahkan oleh hyperplane dengan nilai sebagai berikut.

$$X_i \cdot W + b \geq 1 \text{ untuk } Y_i = 1 \quad (2)$$

$$X_i \cdot W + b \leq -1 \text{ untuk } Y_i = -1 \quad (3)$$

Rumus yang digunakan untuk menghitung hasil prediksi dengan hyperplane adalah sebagai berikut.

$$f(x) = \text{sign}(w \cdot \varphi(x) + b) \quad (4)$$

Keterangan:  $X_i$  adalah data ke- $i$ ,  $W$  adalah nilai bobot support vector yang tegak lurus dengan hyperplane,  $b$  adalah nilai bias, dan  $Y_i$  adalah kelas data ke- $i$ . Untuk menentukan nilai optimal dari hyperplane, digunakan persamaan berikut.

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

Model SVM dalam penelitian ini diimplementasikan menggunakan library Scikit-Learn pada bahasa pemrograman Python melalui Google Colab. Data latih sebanyak 800 sampel digunakan untuk melatih model agar dapat mengenali pola panjang udang dari setiap kategori, sedangkan data uji sebanyak 200 sampel digunakan untuk menguji kemampuan model dalam mengklasifikasikan data baru.

## 2.4 Evaluasi Hasil

Proses evaluasi hasil dilakukan menggunakan instrumen Confusion Matrix, yaitu tabel yang merangkum hasil prediksi model dibandingkan dengan data aktual [9]. Dari Confusion Matrix tersebut, kemampuan model diukur berdasarkan tiga parameter evaluasi utama yaitu akurasi, presisi, dan recall.

Akurasi merupakan angka prediksi yang benar dibuat oleh model dari keseluruhan data uji. Akurasi dihitung menggunakan persamaan berikut.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN)} \quad (6)$$

Presisi merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Presisi dihitung menggunakan persamaan berikut.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (7)$$

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar-benar positif. Recall dihitung menggunakan persamaan berikut.

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (8)$$

Keterangan: TP (True Positive) adalah jumlah data yang diprediksi benar pada kelas positif, TN (True Negative) adalah jumlah data yang diprediksi benar pada kelas negatif, FP (False Positive) adalah jumlah data yang diprediksi positif namun sebenarnya negatif, dan FN (False Negative) adalah jumlah data yang diprediksi negatif namun sebenarnya positif [9]. Ketiga parameter evaluasi tersebut digunakan bersama-sama untuk memberikan gambaran yang menyeluruh terhadap kinerja model klasifikasi SVM dalam penelitian ini.

### 3 HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini terdiri dari 1.000 data hasil pengukuran panjang tubuh udang putih (*Litopenaeus vannamei*). Data tersebut kemudian melalui tahap pembersihan dan penyesuaian kategori ukuran berdasarkan rentang panjang yang telah ditetapkan. Kategori ukuran terdiri dari Kecil, Sedang, dan Besar. Metode klasifikasi yang digunakan adalah Support Vector Machine (SVM). Sebelum dilakukan pengujian menggunakan Percentage Split, model terlebih dahulu divalidasi menggunakan metode Stratified K-Fold Cross Validation sebanyak 5 fold.

#### 3.1 Implementasi Sistem Klasifikasi

Pada subbab ini dijelaskan tahapan pengolahan data yang digunakan dalam penelitian, mulai dari deskripsi dataset, pra-pemrosesan data, validasi model menggunakan K-Fold Cross Validation, hingga pembagian data training dan testing menggunakan Percentage Split. Tahapan tersebut dilakukan untuk memastikan data yang digunakan sesuai dengan kategori yang telah ditentukan sebelum masuk ke proses klasifikasi menggunakan SVM.

##### 3.1.1 Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini merupakan data primer hasil pengukuran langsung terhadap panjang fisik tubuh udang putih (*Litopenaeus vannamei*). Total data yang dikumpulkan berjumlah 1.000 sampel. Setiap sampel berisi nilai panjang tubuh udang dalam satuan sentimeter (cm) yang kemudian dikelompokkan ke dalam tiga kategori ukuran, yaitu Kecil, Sedang, dan Besar.

Pada penelitian ini, kategori ukuran udang ditentukan berdasarkan rentang panjang yang telah ditetapkan, yaitu kategori Kecil dengan panjang 6,0–9,0 cm, kategori Sedang dengan panjang 9,1–12,0 cm, dan kategori Besar dengan panjang 12,1–16,0 cm.

Setelah dilakukan proses pembersihan data dan penentuan kategori berdasarkan rentang tersebut, diperoleh 996 data yang dapat digunakan dalam proses klasifikasi. Dari 1.000 data awal, terdapat 4 data yang tidak dapat diproses sebagai nilai panjang numerik sehingga tidak digunakan dalam proses klasifikasi.

Berdasarkan 996 data yang digunakan, kategori Kecil berjumlah 100 data (10,04%), kategori Sedang berjumlah 699 data (70,18%), dan kategori Besar berjumlah 197 data (19,78%). Distribusi data berdasarkan kategori ukuran udang putih dapat dilihat pada Tabel 2.

**Tabel 2. Distribusi Data per Kategori Ukuran Udang Putih**

| Kategori | Jumlah Data | Persentase | Panjang Min | Panjang Max | Rata-rata |
|----------|-------------|------------|-------------|-------------|-----------|
| Kecil    | 100 data    | 10,04%     | 6,0 cm      | 9,0 cm      | 8,11      |
| Sedang   | 699 data    | 70,18%     | 9,1 cm      | 12,0 cm     | 10,94     |
| Besar    | 197 data    | 19,78%     | 12,1 cm     | 16 cm       | 13,25     |
| Total    | 996 data    | 100%       | 6,0 cm      | 15,7 cm     |           |

Berdasarkan Tabel 2, dari 996 data yang digunakan dalam penelitian, kategori Kecil berjumlah 100 data (10,04%) dengan panjang 6,0–9,0 cm. Kategori Sedang memiliki jumlah data terbanyak, yaitu 699 data (70,18%) dengan panjang 9,1–12,0 cm. Sementara itu, kategori Besar berjumlah 197 data (19,78%) dengan panjang 12,1–15,7 cm. Secara keseluruhan, data yang digunakan memiliki panjang udang antara 6,0–15,7 cm.

##### 3.1.2 Hasil Pra-Pemrosesan Data

Pra-pemrosesan data dilakukan sebelum data digunakan dalam proses klasifikasi menggunakan metode Support Vector Machine (SVM). Tahap ini dilakukan untuk memastikan data panjang udang dapat diproses dengan baik pada tahap pemodelan.

Dari 1.000 data awal, diperoleh 996 data yang dapat digunakan setelah dilakukan pemeriksaan dan pengolahan pada data panjang udang. Sebanyak 4 data tidak dapat diproses sebagai nilai panjang numerik sehingga tidak digunakan dalam proses klasifikasi. Dengan demikian, jumlah data yang digunakan pada tahap selanjutnya adalah 996 data.

Selanjutnya, nilai panjang udang yang masih menggunakan satuan cm dan tanda koma sebagai pemisah desimal dikonversi menjadi nilai numerik. Setelah itu, data dikelompokkan berdasarkan rentang ukuran yang telah ditentukan, yaitu kategori Kecil dengan panjang 6,0–9,0 cm, kategori Sedang dengan panjang 9,1–12,0 cm, dan kategori Besar dengan panjang 12,1–16,0 cm.

Berdasarkan hasil pengelompokan, terdapat 100 data pada kategori Kecil, 699 data pada kategori Sedang, dan 197 data pada kategori Besar. Data yang telah melalui tahap pembersihan dan pengelompokan tersebut kemudian digunakan sebagai data masukan dalam proses klasifikasi.

Sebelum dilakukan proses klasifikasi menggunakan SVM, data panjang udang dilakukan penskalaan menggunakan StandardScaler. Penskalaan ini bertujuan untuk menyamakan skala nilai fitur sehingga proses pemodelan SVM dapat dilakukan dengan lebih baik. StandardScaler melakukan standardisasi berdasarkan nilai rata-rata dan standar deviasi dari data.

Dengan demikian, hasil pra-pemrosesan menghasilkan 996 data yang siap digunakan untuk tahap pengujian dan klasifikasi menggunakan metode SVM.

### 3.1.3 Percentage Split

Metode Percentage Split digunakan untuk membagi dataset menjadi data latih (training) dan data uji (testing). Pada penelitian ini digunakan rasio 80:20, yaitu 80% data digunakan sebagai data latih dan 20% data digunakan sebagai data uji. Pembagian dilakukan menggunakan stratified split agar setiap kategori tetap terwakili pada data training dan testing.

Sebelum dilakukan pembagian data menggunakan Percentage Split, model SVM terlebih dahulu diuji menggunakan metode Stratified K-Fold Cross Validation dengan nilai  $k = 5$ . Pada proses ini, dataset dibagi menjadi lima bagian (fold), kemudian setiap fold digunakan secara bergantian sebagai data uji, sedangkan empat fold lainnya digunakan sebagai data latih. Penggunaan Stratified K-Fold bertujuan untuk melihat konsistensi kinerja model pada beberapa pembagian data.

**Tabel 3. Hasil K-Fold Cross Validation (5-Fold) Model SVM**

| keterangan       | Fold 1                 | Fold 2 | Fold 3 | Fold 4 | Fold 5 |
|------------------|------------------------|--------|--------|--------|--------|
| Akurasi          | 97,50%                 | 97,99% | 98,99% | 98,99% | 98,49% |
| <b>Rata-rata</b> | <b>Standar Deviasi</b> |        |        |        |        |
| 98,39%           | ±0,58%                 |        |        |        |        |

Berdasarkan Tabel 3, hasil pengujian pada lima fold menunjukkan nilai akurasi antara 97,50% sampai 98,99%. Akurasi tertinggi diperoleh pada Fold 3 dan Fold 4, yaitu sebesar 98,99%, sedangkan akurasi terendah terdapat pada Fold 1 sebesar 97,50%. Rata-rata akurasi yang diperoleh adalah 98,39% dengan standar deviasi sebesar 0,58%. Hasil tersebut menunjukkan bahwa nilai akurasi antar fold memiliki perbedaan yang relatif kecil.

### Pembagian Data Training dan Testing

Setelah proses pembersihan data, terdapat 996 sampel data yang valid dari total 1.000 data yang dikumpulkan. Sebanyak 4 data terhapus karena terdeteksi sebagai nilai yang tidak dapat dibaca (missing value) pada saat proses cleaning. Dataset yang valid kemudian dibagi menggunakan rasio 80:20, yaitu 80% sebagai data latih (training) dan 20% sebagai data uji (testing). Pembagian dilakukan menggunakan metode stratified split agar proporsi setiap kategori terwakili secara proporsional pada kedua bagian data.

**Tabel 4. Pembagian Data Training dan Testing (Percentage Split 80:20)**

| Kategori | Total Data | Training (80%) | Testing (20%) |
|----------|------------|----------------|---------------|
| Kecil    | 100 data   | 80 data        | 20 data       |
| Sedang   | 699 data   | 559 data       | 140 data      |
| Besar    | 197 data   | 157 data       | 40 data       |
| Total    | 996 data   | 796 data       | 200 data      |

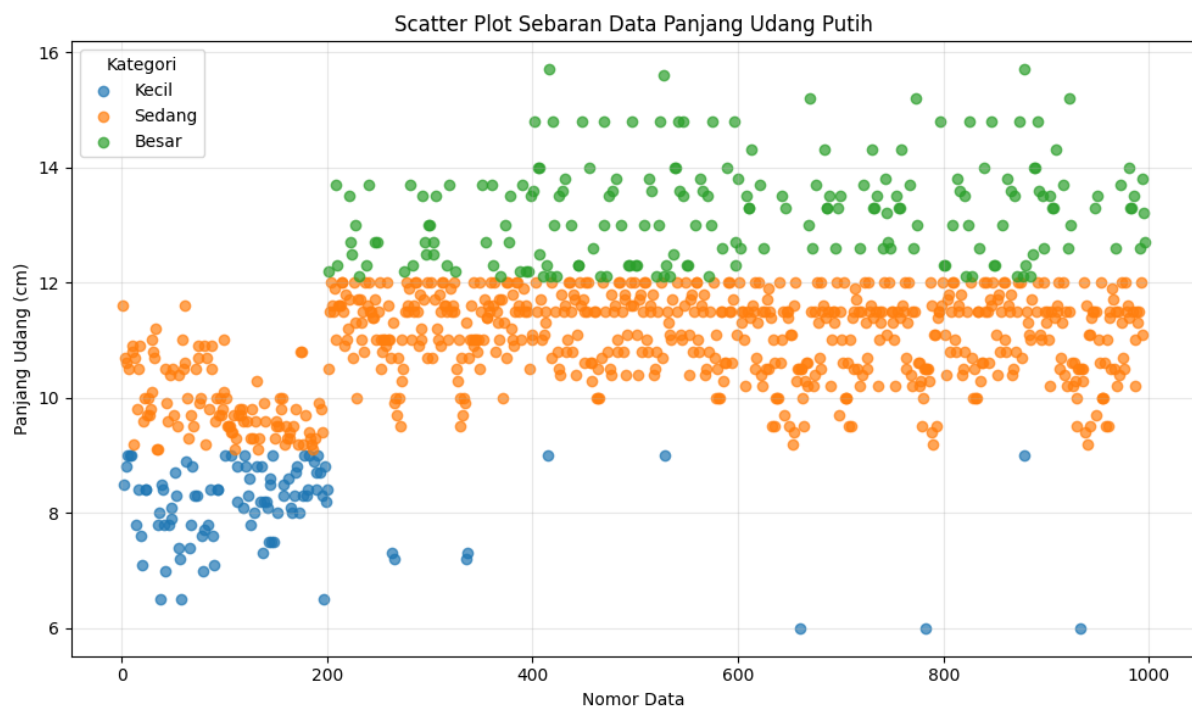
Berdasarkan Tabel 4, sebanyak 796 data digunakan sebagai data training dan 200 data digunakan sebagai data testing. Data training digunakan untuk membentuk model SVM berdasarkan data panjang udang yang telah dikategorikan, sedangkan data testing digunakan untuk melihat kemampuan model dalam mengklasifikasikan data yang tidak digunakan pada proses pelatihan.

### 3.2 Visualisasi Sebaran Data

Sebelum masuk ke proses klasifikasi, dilakukan visualisasi data untuk memberikan gambaran yang lebih jelas mengenai pola dan sebaran nilai panjang udang pada setiap kategori. Visualisasi ini membantu memahami karakteristik distribusi data dan mengevaluasi apakah batas antar kategori sudah cukup terpisah sehingga model SVM dapat bekerja dengan optimal.

#### 3.2.1 Scatter Plot

Scatter plot digunakan untuk menampilkan sebaran seluruh titik data panjang udang berdasarkan kategorinya. Setiap titik merepresentasikan satu sampel data udang, dengan warna berbeda untuk setiap kategori (Kecil, Sedang, Besar). Scatter plot memperlihatkan apakah kelompok data antar kategori sudah terpisah dengan jelas atau masih terdapat tumpang tindih di area batas.

**Gambar 2. Scatter Plot Sebaran Data Panjang Udang Putih per Kategori**

#### 3.2.2 Analisis Sebaran Data

Berdasarkan Gambar 2, scatter plot menunjukkan bahwa data panjang udang putih membentuk tiga kelompok berdasarkan kategori ukuran yang telah ditentukan. Kategori Kecil berada pada rentang panjang 6,0–9,0 cm, kategori Sedang berada pada rentang 9,1–12,0 cm, sedangkan kategori Besar berada pada rentang 12,1–16,0 cm. Dari keseluruhan data yang digunakan, terlihat bahwa kategori Sedang memiliki jumlah titik data yang lebih banyak dibandingkan kategori Kecil dan Besar.

Sebaran titik pada grafik menunjukkan adanya perbedaan nilai panjang antara ketiga kategori. Data pada kategori Kecil berada pada nilai panjang yang lebih rendah, sedangkan kategori Besar berada pada nilai panjang yang lebih tinggi. Kategori Sedang berada di antara kedua kategori tersebut. Karena pembagian kategori didasarkan pada nilai panjang, pola sebaran data terlihat mengikuti batas ukuran yang telah ditentukan.

Perbedaan jumlah data pada setiap kategori juga terlihat pada scatter plot. Kategori Sedang memiliki sebaran titik yang lebih banyak, sedangkan kategori Kecil memiliki jumlah titik yang lebih sedikit. Kondisi ini menunjukkan bahwa distribusi data yang digunakan dalam penelitian belum merata pada setiap kategori. Meskipun demikian, seluruh kategori tetap digunakan dalam proses pelatihan dan pengujian model SVM.

Berdasarkan hasil visualisasi tersebut, scatter plot memberikan gambaran awal mengenai pola sebaran data panjang udang putih sebelum dilakukan proses klasifikasi. Perbedaan nilai panjang pada setiap kategori dapat menjadi dasar bagi model Support Vector Machine (SVM) dalam membentuk batas keputusan (*decision boundary*) untuk membedakan kategori Kecil, Sedang, dan Besar. Selanjutnya, data yang telah melalui tahap pra-pemrosesan dan penskalaan menggunakan StandardScaler digunakan dalam proses klasifikasi SVM.

### 3.3 Hasil Klasifikasi Menggunakan SVM

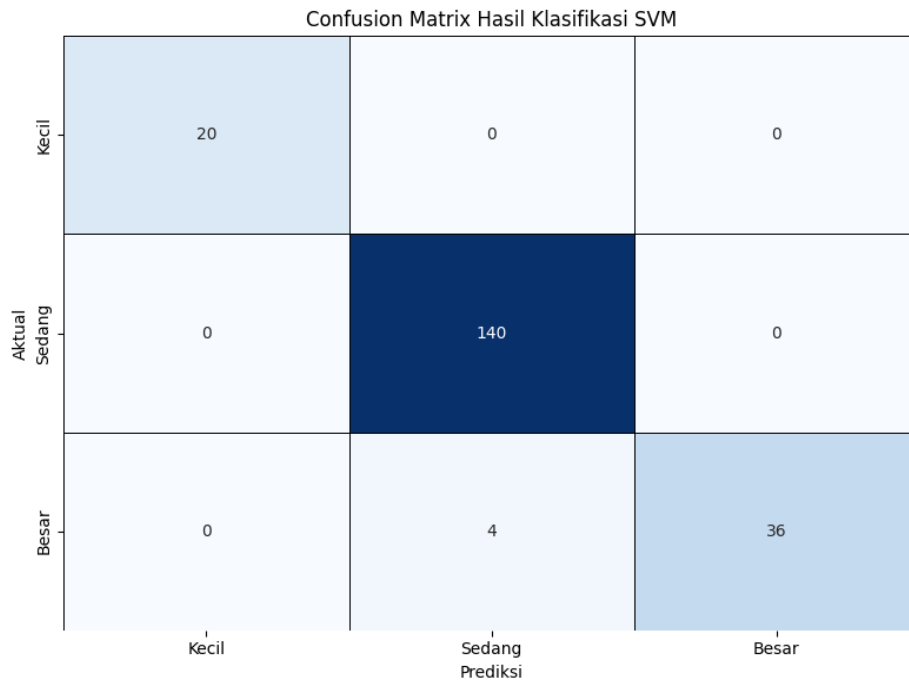
Model SVM diimplementasikan menggunakan library Scikit-Learn pada bahasa pemrograman Python melalui Google Colab. Model dilatih menggunakan 800 data training kemudian diuji menggunakan 200 data testing. Hasil klasifikasi ditampilkan dalam bentuk Confusion Matrix dan Classification Report.

#### 3.3.1 Confusion Matrik

Model SVM diimplementasikan menggunakan library Scikit-Learn pada bahasa pemrograman Python melalui Google Colab. Model dilatih menggunakan 796 data training kemudian diuji menggunakan 200 data testing. Hasil pengujian model ditampilkan dalam bentuk Confusion Matrix pada Tabel 5 berikut.

**Tabel 5. Confusion Matrix Hasil Klasifikasi SVM**

| Aktual/prediksi | Pred. Kecil | Pred. Sedang | Pred. Besar |
|-----------------|-------------|--------------|-------------|
| Aktual Kecil    | 20          | 0            | 0           |
| Aktual Sedang   | 0           | 140          | 0           |
| Aktual Besar    | 0           | 4            | 36          |



**Gambar 3. Visualisasi Confusion Matrix Hasil Klasifikasi SVM**

Berdasarkan Tabel 6 dan Gambar 3, dari 200 data testing sebanyak 196 data berhasil diklasifikasikan dengan benar dan 4 data diklasifikasikan secara salah. Seluruh 20 data Kecil berhasil diprediksi benar sebagai Kecil. Seluruh 140 data Sedang berhasil diprediksi benar sebagai Sedang. Dari 40 data Besar, sebanyak 36 diprediksi benar dan 4 diprediksi keliru sebagai Sedang. Kesalahan klasifikasi terjadi pada 4 data kategori Besar yang memiliki nilai panjang mendekati batas antar kategori Besar dan Sedang, yaitu mendekati nilai 12,1 cm, sehingga model mengalami sedikit kesulitan dalam memisahkan keduanya pada area perbatasan tersebut.

### 3.3.2 Classification Report

Classification Report menampilkan nilai presisi, recall, dan F1-Score secara terperinci untuk setiap kategori ukuran udang putih berdasarkan hasil pengujian model SVM.

**Tabel 6. Classification Report Hasil Klasifikasi SVM**

| Kategori                      | Presisi       | Recall  | F1-Score | Support    |
|-------------------------------|---------------|---------|----------|------------|
| Kecil                         | 100,00%       | 100,00% | 100,00%  | 20         |
| Sedang                        | 97,22%        | 100,00% | 98,59%   | 140        |
| Besar                         | 100,00%       | 90,00%  | 94,74%   | 40         |
| Rata-rata<br>(Macro Avg)      | 99.07%        | 96.67%  | 97,78%   | 200        |
| Rata – rata<br>(Weighted Avg) | 98.06%        | 98.00%  | 97,96%   | 200        |
| <b>Akurasi</b>                | <b>98,00%</b> | -       | -        | <b>200</b> |
| <b>Keseluruhan</b>            |               |         |          |            |

Berdasarkan Tabel 6, kategori Kecil memperoleh nilai presisi, recall, dan F1-Score masing-masing sebesar 100,00%. Hal ini menunjukkan bahwa seluruh data kategori Kecil pada data testing berhasil diklasifikasikan dengan benar oleh model SVM. Hasil tersebut juga sesuai dengan Confusion Matrix pada Tabel 5, di mana seluruh 20 data aktual kategori Kecil diprediksi sebagai Kecil.

Pada kategori Sedang, nilai presisi yang diperoleh sebesar 97,22%, recall sebesar 100,00%, dan F1-Score sebesar 98,59%. Nilai recall 100,00% menunjukkan bahwa seluruh 140 data aktual kategori

Sedang berhasil dikenali oleh model. Namun, terdapat 4 data dari kategori Besar yang diprediksi sebagai Sedang. Kesalahan tersebut menyebabkan nilai presisi kategori Sedang menjadi 97,22%.

Sementara itu, kategori Besar memperoleh nilai presisi sebesar 100,00%, recall sebesar 90,00%, dan F1-Score sebesar 94,74%. Nilai presisi 100,00% menunjukkan bahwa seluruh data yang diprediksi sebagai Besar memang termasuk dalam kategori Besar. Namun, nilai recall sebesar 90,00% menunjukkan bahwa masih terdapat 4 data aktual kategori Besar yang diprediksi sebagai Sedang. Kondisi ini menyebabkan nilai F1-Score kategori Besar lebih rendah dibandingkan kategori Kecil dan Sedang.

Secara keseluruhan, model SVM memperoleh akurasi sebesar 98,00% pada 200 data testing. Nilai macro average yang diperoleh adalah 99,07% untuk presisi, 96,67% untuk recall, dan 97,78% untuk F1-Score. Sementara itu, weighted average menghasilkan presisi sebesar 98,06%, recall sebesar 98,00%, dan F1-Score sebesar 97,96%. Hasil tersebut menunjukkan bahwa model SVM mampu melakukan klasifikasi ukuran udang putih dengan tingkat ketepatan yang tinggi berdasarkan parameter panjang.

Hasil pengujian menggunakan Stratified K-Fold Cross Validation dengan 5 fold juga menunjukkan rata-rata akurasi sebesar 98,39% dengan standar deviasi sebesar 0,58%. Nilai akurasi yang relatif konsisten pada setiap fold menunjukkan bahwa kinerja model tidak banyak berubah pada pembagian data yang berbeda. Dengan demikian, hasil pengujian Percentage Split sebesar 98,00% dan K-Fold Cross Validation sebesar 98,39% menunjukkan bahwa model SVM memberikan hasil klasifikasi yang konsisten pada data yang digunakan dalam penelitian.

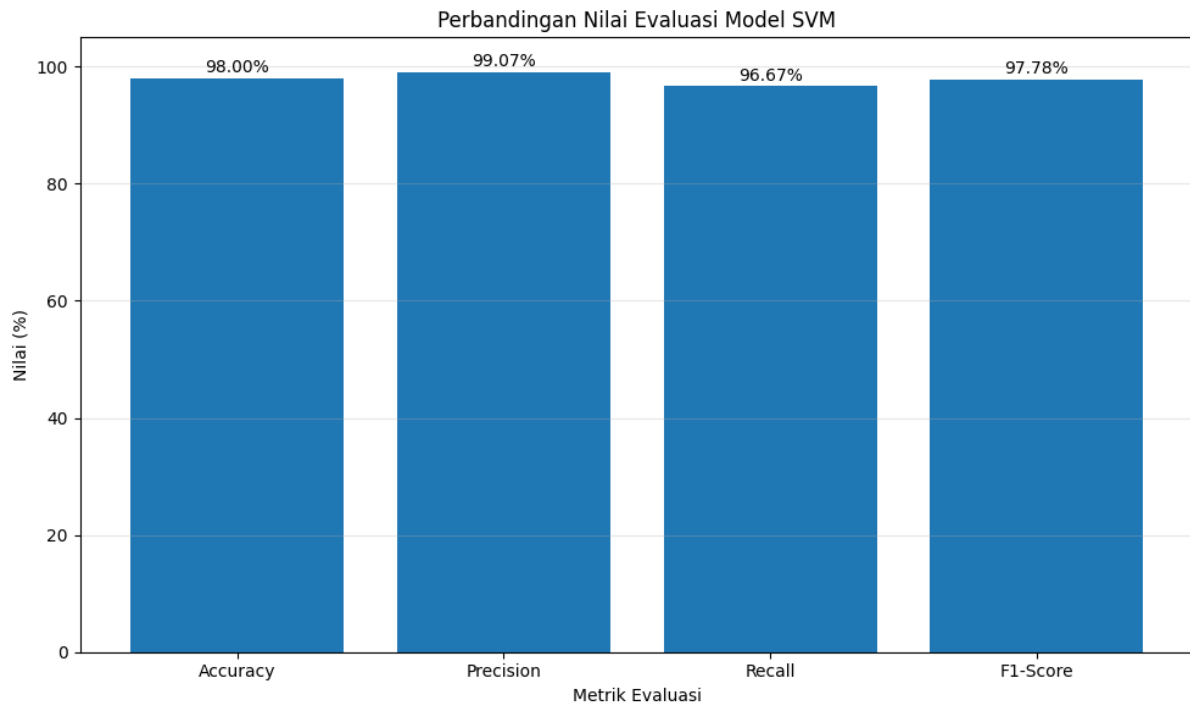
### 3.4 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan secara menyeluruh untuk mengukur performa model SVM dari berbagai sudut pandang, mencakup akurasi, presisi, recall, dan F1-Score yang ditampilkan melalui tabel dan grafik perbandingan.

**Tabel 7. Rata-Rata Nilai Evaluasi Model SVM**

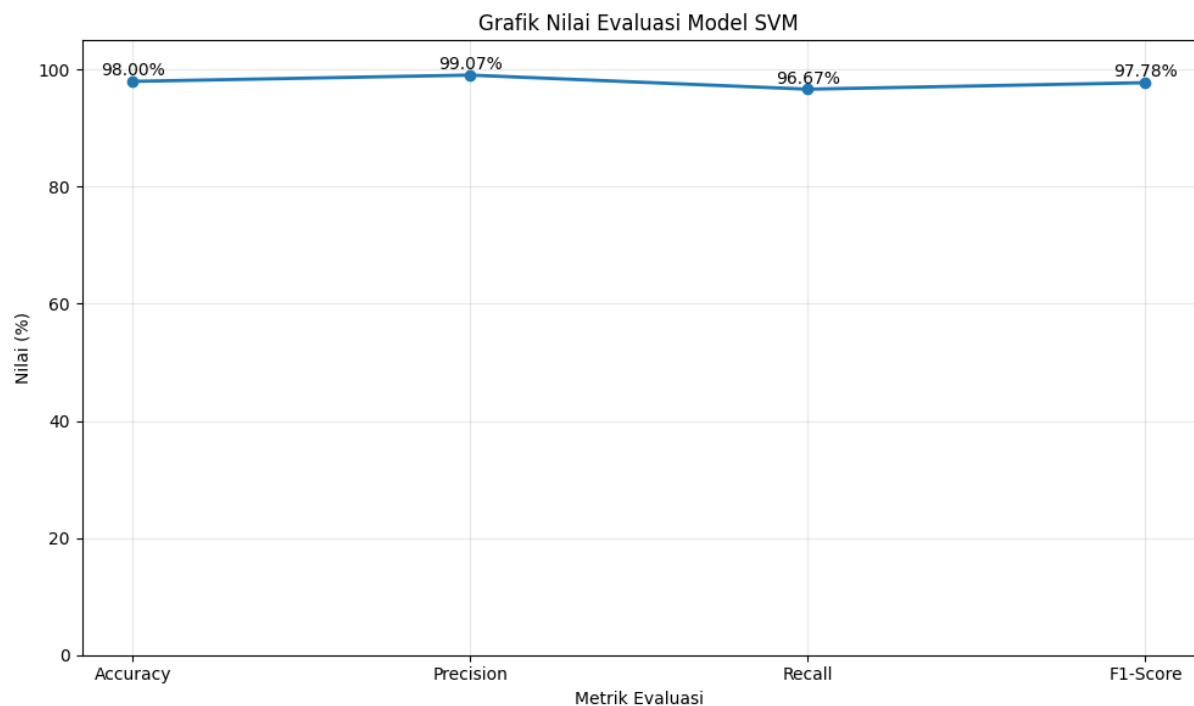
| <b>Metrik</b> | <b>Nilai Rata-rata</b> | <b>Kategori Penilaian</b> |
|---------------|------------------------|---------------------------|
| Accuracy      | 98,00%                 | Sangat Baik               |
| Precision     | 99,07%                 | Sangat Baik               |
| Recall        | 96,67%                 | Sangat Baik               |
| F1-Score      | 97,78%                 | Sangat Baik               |

Berdasarkan Tabel 7, nilai Accuracy yang diperoleh model SVM sebesar 98,00%, Precision sebesar 99,07%, Recall sebesar 96,67%, dan F1-Score sebesar 97,78%. Nilai tersebut kemudian divisualisasikan dalam bentuk grafik untuk melihat perbandingan kinerja model pada setiap metrik evaluasi.



**Gambar 4. Grafik Perbandingan Nilai Precision, Recall, dan F1-Score per Kategori**

Berdasarkan Gambar 4, terlihat bahwa nilai Precision memiliki nilai paling tinggi, yaitu sebesar 99,07%, sedangkan Recall memiliki nilai paling rendah, yaitu sebesar 96,67%. Accuracy memperoleh nilai sebesar 98,00% dan F1-Score sebesar 97,78%. Perbedaan nilai antar metrik tidak terlalu besar sehingga menunjukkan bahwa kinerja model SVM relatif konsisten dalam melakukan klasifikasi pada data testing.



**Gambar 5. Grafik Nilai Accuracy, Precision, Recall, F1-Score**

Pada Gambar 5, nilai setiap metrik evaluasi menunjukkan hasil yang berada pada tingkat tinggi. Precision sebesar 99,07% menunjukkan bahwa prediksi kategori yang dihasilkan model memiliki tingkat ketepatan yang tinggi. Accuracy sebesar 98,00% menunjukkan bahwa sebagian besar data testing berhasil diklasifikasikan dengan benar. Recall sebesar 96,67% menunjukkan kemampuan

model dalam mengenali data dari setiap kategori, sedangkan F1-Score sebesar 97,78% menunjukkan keseimbangan antara Precision dan Recall.

### 3.5 Pembahasan

Berdasarkan hasil pengujian yang telah dilakukan, model Support Vector Machine (SVM) mampu mengklasifikasikan ukuran udang putih berdasarkan parameter panjang dengan hasil yang cukup tinggi. Pada data testing sebanyak 200 data, diperoleh 196 data yang berhasil diklasifikasikan dengan benar dan 4 data mengalami kesalahan klasifikasi. Dengan demikian, nilai akurasi yang diperoleh adalah sebesar 98,00%.

Pada kategori Kecil, model memperoleh nilai presisi, recall, dan F1-Score masing-masing sebesar 100,00%. Hal ini menunjukkan bahwa seluruh data kategori Kecil pada data testing berhasil diklasifikasikan dengan benar. Sementara itu, kategori Sedang memperoleh presisi sebesar 97,22%, recall sebesar 100,00%, dan F1-Score sebesar 98,59%. Nilai recall yang mencapai 100,00% menunjukkan bahwa seluruh data aktual kategori Sedang berhasil dikenali oleh model.

Pada kategori Besar, nilai presisi yang diperoleh sebesar 100,00%, recall sebesar 90,00%, dan F1-Score sebesar 94,74%. Nilai recall yang lebih rendah menunjukkan adanya 4 data aktual kategori Besar yang diprediksi sebagai kategori Sedang. Kesalahan klasifikasi tersebut terlihat pada Confusion Matrix, sedangkan seluruh data yang diprediksi sebagai kategori Besar berhasil diklasifikasikan dengan benar.

Hasil tersebut menunjukkan bahwa model SVM dapat membedakan kategori ukuran berdasarkan nilai panjang udang dengan cukup baik. Perbedaan nilai panjang antara kategori Kecil, Sedang, dan Besar menjadi dasar dalam proses pembentukan batas keputusan oleh model SVM. Meskipun jumlah data pada setiap kategori tidak sama, seluruh kategori tetap dapat digunakan dalam proses pelatihan dan pengujian model.

Selain pengujian menggunakan Percentage Split, model juga diuji menggunakan Stratified K-Fold Cross Validation dengan 5 fold. Hasil pengujian memperoleh rata-rata akurasi sebesar 98,39% dengan standar deviasi sebesar 0,58%. Nilai akurasi pada setiap fold berada pada rentang 97,50% hingga 98,99%. Hasil tersebut menunjukkan bahwa kinerja model relatif konsisten pada beberapa pembagian data yang berbeda.

Secara keseluruhan, hasil pengujian menunjukkan bahwa SVM dapat digunakan untuk melakukan klasifikasi ukuran udang putih berdasarkan parameter panjang. Nilai akurasi sebesar 98,00% pada data testing serta rata-rata akurasi sebesar 98,39% pada 5-Fold Cross Validation menunjukkan bahwa model memiliki kinerja yang baik pada data yang digunakan. Kesalahan klasifikasi yang masih terjadi terutama terdapat pada kategori Besar yang diprediksi sebagai Sedang, sehingga bagian tersebut dapat menjadi perhatian dalam pengembangan penelitian selanjutnya.

## 4 KESIMPULAN

Berdasarkan hasil eksperimen, analisis data, dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Machine (SVM) dapat digunakan untuk mengklasifikasikan ukuran udang putih (*Litopenaeus vannamei*) berdasarkan parameter panjang tubuh. Dari 1.000 data yang dikumpulkan, sebanyak 996 data dapat digunakan setelah melalui proses pembersihan dan pengolahan data. Dataset tersebut kemudian dibagi menggunakan metode Percentage Split dengan rasio 80:20, sehingga diperoleh 796 data training dan 200 data testing.

Hasil pengujian menggunakan Stratified K-Fold Cross Validation dengan  $k=5$  menghasilkan rata-rata akurasi sebesar 98,39% dengan standar deviasi sebesar 0,58%. Nilai tersebut menunjukkan bahwa performa model relatif konsisten pada setiap pembagian data. Selanjutnya, hasil pengujian menggunakan Percentage Split 80:20 menghasilkan akurasi sebesar 98,00%, dengan 196 dari 200 data testing berhasil diklasifikasikan dengan benar dan 4 data mengalami kesalahan klasifikasi.

Berdasarkan hasil evaluasi, kategori Kecil memperoleh presisi, recall, dan F1-Score masing-masing sebesar 100,00%. Kategori Sedang memperoleh presisi sebesar 97,22%, recall sebesar

100,00%, dan F1-Score sebesar 98,59%. Sementara itu, kategori Besar memperoleh presisi sebesar 100,00%, recall sebesar 90,00%, dan F1-Score sebesar 94,74%. Kesalahan klasifikasi terjadi pada 4 data kategori Besar yang diprediksi sebagai kategori Sedang.

Secara keseluruhan, hasil pengujian menunjukkan bahwa metode SVM mampu melakukan klasifikasi ukuran udang putih berdasarkan parameter panjang dengan hasil yang baik. Penggunaan StandardScaler dan pembagian data secara stratified membantu proses pemodelan dan pengujian dilakukan pada data dari setiap kategori. Untuk penelitian selanjutnya, penelitian dapat dikembangkan dengan menambahkan data atau fitur lain yang berkaitan dengan ukuran udang, sehingga informasi yang digunakan dalam proses klasifikasi menjadi lebih beragam.

## REFERENSI

- [1] R. González *et al.*, “IMASHRIMP : Automatic White Shrimp ( *Penaeus vannamei* ) Biometrical Analysis from Laboratory Images Using Computer Vision and Deep Learning,” 2024.
- [2] K. G. Liakos, V. Athanasiadis, E. Bozinou, and S. I. Lalas, “Machine Learning for Quality Control in the Food Industry : A Review,” pp. 1–35, 2025.
- [3] S. Bhusan and D. Debashish, *Leveraging artificial intelligence and advanced food processing techniques for enhanced food safety , quality , and security : a comprehensive review*. Springer International Publishing, 2025. doi: 10.1007/s42452-025-06472-w.
- [4] Y. Gültepe, “Modeling and predicting meat yield and growth performance using morphological features of narrow - clawed crayfish with machine learning techniques,” *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-69539-5.
- [5] A. Setiawan, H. Hadiyanto, and C. E. Widodo, “Revue d ’ Intelligence Artificielle Shrimp Body Weight Estimation in Aquaculture Ponds Using Morphometric Features Based on Underwater Image Analysis and Machine Learning Approach,” vol. 36, no. 6, pp. 905–912, 2022.
- [6] M. Çakir, M. Yilmaz, M. Atalay, H. Özgür, and O. Oral, “Journal of King Saud University – Science Accuracy assessment of RFerns , NB , SVM , and kNN machine learning classifiers in aquaculture,” vol. 35, 2023, doi: 10.1016/j.jksus.2023.102754.
- [7] L. Q. Hoang, N. Duong-trung, and C. Nguyen, “TOWARDS MACHINE LEARNING APPROACHES TO IDENTIFY SHRIMP,” pp. 1–8, 2019.
- [8] N. Amaya-tejera, M. Gamarra, I. V Jorge, and E. Zurek, “A distance-based kernel for classification via Support Vector Machines,” 2015.
- [9] I. Markoulidakis, “Probabilistic Confusion Matrix : A Novel Method for Machine Learning Algorithm Generalized Performance Analysis,” 2024.
- [10] V. Machine, R. Forest, L. Regression, L. Labb, P. Souˆ, and P. Pelissier, “Mohammadmehdi Saberioon 1, \* and Thierry Kerneis 2 1,” pp. 1–15, doi: 10.3390/s18041027.
- [11] A. Saleh, D. Jones, and D. Jerry, “MFLD - net : a lightweight deep learning network for fish morphometry using landmark detection,” *Aquat. Ecol.*, vol. 57, no. 4, pp. 913–931, 2023, doi: 10.1007/s10452-023-10044-8.
- [12] F. Monteiro *et al.*, “Classification of Fish Species Using Multispectral Data from a Low-Cost Camera and Machine Learning,” 2023.
- [13] M. O. Edeh, S. Dalal, I. C. Obagbuwa, B. V. V. S. Prasad, S. Z. Ninoria, and M. A. Wajid, “Bootstrapping random forest and CHAID for prediction of white spot disease among shrimp farmers,” *Sci. Rep.*, pp. 1–12, 2022, doi: 10.1038/s41598-022-25109-1.
- [14] M. Córdova, S. Torres, and A. Van Helmond, “Novelty Recognition : Fish Species Classification via Open-Set Recognition,” no. 1380, pp. 1–18, 2025.
- [15] B. Correia, O. Pacheco, and R. J. M. Rocha, “Image-Based Shrimp Aquaculture Monitoring,” pp. 1–18, 2025.