



## *Adult Clothing Size Recommendation Using K-Nearest Neighbor and Support Vector Machine Algorithm*

### **Rekomendasi Ukuran Baju Dewasa Menggunakan Algoritma K-Nearest Neighbor dan Support Vector Machine**

Binta Bailina Mumtazah<sup>1\*</sup>, Sulisty Dwi Sancoko<sup>2</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Sains dan Teknologi  
Universitas Teknologi Yogyakarta, Indonesia

<sup>2</sup>Program Studi Sains Data, Fakultas Sains dan Teknologi  
Universitas Teknologi Yogyakarta, Indonesia

E-Mail: <sup>1</sup>bintamumtazah04@gmail.com, <sup>2</sup>sulisty.dwisancoko@staff.uty.ac.id

Received Sep 20th 2024; Revised Oct 19th 2024; Accepted Oct 25th 2024  
Corresponding Author: Binta Bailina Mumtazah

#### **Abstract**

*The emergence of online shopping platforms has transformed the way people shop, including for clothing. While online shopping offers convenience and accessibility, it often leads to issues with size mismatches, which can result in consumer dissatisfaction. To address this problem, this study developed a clothing size recommendation system using attributes such as gender, body shape, height, weight, and chest circumference. The system employs two algorithms, namely K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM), which were evaluated to determine the most effective model for providing accurate size recommendations. The experimental results indicate that the SVM method with an RBF kernel and a One-vs-Rest approach, using an 80:20 data split, achieved the highest accuracy of 76%, with a precision of 78%, recall of 76%, and F1 score of 76%. Consequently, this model was selected for implementation in the recommendation website. Functional testing of the website, utilizing black box testing methods, demonstrated that the system fulfilled all required functionalities, including user registration, login, data input, size recommendation, and logout. The overall findings of this study reveal that the SVM algorithm outperformed K-NN in classifying clothing sizes, ensuring the system's reliability in providing accurate and user-appropriate size recommendations.*

*Keyword: Clothing Size Recommendation, K-Nearest Neighbor, Machine Learning, Support Vector Machine.*

#### **Abstrak**

Kemunculan platform toko online telah mengubah cara masyarakat berbelanja, termasuk dalam pembelian pakaian. Kendati menawarkan kemudahan dan aksesibilitas, berbelanja pakaian secara online seringkali menghadirkan masalah ketidakcocokan ukuran, yang menyebabkan ketidaknyamanan bagi konsumen. Untuk mengatasi masalah tersebut, penelitian ini mengembangkan sistem rekomendasi ukuran baju menggunakan atribut meliputi jenis kelamin, bentuk badan, tinggi badan, berat badan, dan lingkar dada. Sistem ini menggunakan dua algoritma, yaitu K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), yang dievaluasi untuk menentukan performa terbaik dalam memberikan rekomendasi ukuran. Hasil eksperimen menunjukkan bahwa algoritma SVM dengan kernel RBF dan pendekatan One-vs-Rest pada proporsi data 80:20 memberikan akurasi tertinggi sebesar 76%, presisi 78%, recall 76%, dan F1 score 76%, sehingga terpilih sebagai model yang diimplementasikan pada website rekomendasi. Pengujian fungsional pada website dengan algoritma black box testing menunjukkan bahwa sistem telah memenuhi seluruh persyaratan, termasuk registrasi, login, input data, rekomendasi ukuran, dan logout. Keseluruhan hasil penelitian ini menunjukkan bahwa algoritma SVM memiliki performa lebih unggul dibandingkan K-NN dalam klasifikasi ukuran baju, serta memastikan keandalan fungsional sistem dalam memberikan rekomendasi yang sesuai bagi pengguna.

Kata Kunci: K-Nearest Neighbor, Machine Learning, Rekomendasi Ukuran Baju, Support Vector Machine

#### **1. PENDAHULUAN**

Kemunculan *platform* toko online memudahkan kegiatan jual beli [1], sehingga masyarakat kini lebih cenderung memilih berbelanja secara online karena aksesnya yang praktis. Penggunaan *platform* toko online

memudahkan konsumen dalam berbelanja tanpa datang ke tempat langsung. Toko online menawarkan berbagai pilihan produk yang lebih luas daripada toko offline. Konsumen dapat dengan mudah membandingkan harga dan kualitas dari berbagai penjual dalam waktu singkat. Toko online juga sering kali menawarkan diskon dan promosi yang menarik, dari beberapa kelebihan tersebut tidak dipungkiri bahwa masyarakat kini beralih untuk membeli barang melalui toko online daripada melalui toko offline [2].

*Platform* toko online telah merubah cara masyarakat dalam berbelanja, termasuk dalam pembelian pakaian. Salah satu alasan utama mengapa masyarakat cenderung lebih suka membeli baju di toko online, karena toko *online* menyediakan pilihan model yang lebih banyak daripada yang tersedia di toko offline. Selain itu, toko online juga lebih cepat dalam memperbarui trend baju terkini yang memungkinkan konsumen untuk selalu mendapatkan akses ke koleksi terbaru tanpa harus menunggu perubahan stok di toko offline. Namun, pembelian pakaian melalui *platform* toko online seringkali menimbulkan masalah, terutama dalam hal ketidakcocokan ukuran. Kurangnya referensi atau ulasan dari orang yang telah memakai, membuat konsumen rentan mendapatkan pakaian dengan ukuran yang tidak sesuai [3]. Tentu saja hal ini dapat menimbulkan kekecewaan dan ketidaknyamanan saat barang diterima dan dipakai. Oleh karena itu, untuk mengatasi permasalahan tersebut maka dibangun sistem rekomendasi ukuran baju dengan memanfaatkan atribut jenis kelamin, bentuk badan, tinggi badan, berat badan, lingkar dada, dan ukuran baju meliputi ukuran S, M, L, XL, XXL, XXXL. Selain itu, algoritma yang digunakan adalah K-Nearest Neighbor (K-NN) dan Support Vector Machine (SVM), di mana penelitian ini juga bertujuan untuk mengevaluasi dan membandingkan performa kedua algoritma tersebut dalam mengklasifikasi ukuran baju secara akurat.

Algoritma K-NN merupakan algoritma yang mengklasifikasikan suatu objek berdasarkan data pelatihan yang paling mirip atau paling dekat dengan objek tersebut. Pendekatan ini sederhana dan mudah dipahami [4]. Sedangkan untuk algoritma Support Vector Machine (SVM) merupakan algoritma yang digunakan untuk mengklasifikasikan data. Algoritma ini efektif dalam mengatasi masalah overfitting dan cocok untuk dataset yang berukuran kecil [5]. Alasan pemilihan kedua algoritma ini terkait dengan karakteristik dataset penelitian yang memiliki multikelas. K-NN dipilih karena kesederhanaannya dan kemampuannya yang baik dalam mengklasifikasikan data multikelas berdasarkan kemiripan objek dengan data pelatihan terdekat, sehingga mudah dipahami serta cocok untuk konteks klasifikasi ukuran baju. Sementara itu, SVM dipilih karena efektivitasnya dalam mengatasi masalah overfitting dan performanya yang baik dalam klasifikasi multikelas pada dataset berukuran kecil, sebagaimana ditunjukkan oleh hasil penelitian sebelumnya [6] dan [7].

Pada penelitian [3] membangun sistem rekomendasi ukuran baju menggunakan algoritma Decision Tree. Dataset penelitian ini berisi 207 data pelatihan, dengan variabel jenis kelamin, berat badan, tinggi badan, dan bentuk tubuh. Pengujian model menggunakan *Confusion Matrix* pada 21 data uji acak menghasilkan akurasi sebesar 67%. Selanjutnya pada penelitian [8] membangun sistem rekomendasi ukuran baju dengan algoritma Naïve Bayes, menggunakan dataset yang mencakup jenis kelamin, tinggi badan, berat badan, lingkar dada, dan ukuran baju. Model ini menghitung probabilitas dari data pengguna dan merekomendasikan ukuran berdasarkan pelatihan sebelumnya. Pengujian blackbox menunjukkan aplikasi berjalan baik dan memberikan hasil yang sesuai. Pada penelitian [9], telah dibangun sistem rekomendasi ukuran baju pria yang menggantikan fungsi kamar pas di toko offline. Sistem menggunakan kamera webcam dan Raspberry Pi untuk mendeteksi tinggi dan lebar bahu, kemudian mengklasifikasikannya dengan algoritma K-Nearest Neighbor. Rekomendasi ukuran ditampilkan pada layar Liquid Crystal Display (LCD) 16x2. Sistem ini mencapai akurasi 92% dengan waktu komputasi 3,08 detik. Penelitian [10] membangun aplikasi Augmented Reality Virtual Dressing Room untuk platform Android. Pengguna dapat mencoba baju secara virtual melalui kamera handphone, dan aplikasi memberikan rekomendasi kecocokan ukuran. Pengujian menunjukkan jarak efektif untuk mendeteksi marker adalah 30-50 cm, dan jarak ideal bagi pengguna untuk menyesuaikan ukuran baju adalah 1 meter.

Penelitian [11] menghasilkan sistem rekomendasi ukuran baju tanpa harus mencoba langsung, relevan saat pandemi COVID-19 untuk mengurangi risiko penularan. Sistem menggunakan algoritma K-NN dan sensor ultrasonik untuk mendeteksi tinggi serta lebar tubuh konsumen. Hasil rekomendasi ditampilkan di layar LCD. Pengujian dengan 20 data menunjukkan akurasi 100% dan waktu rata-rata 2,1 detik. Pada penelitian [12] berhasil dibangun sistem pengukuran baju menggunakan human body estimation dengan algoritma Euclidean Distance dan dataset Mediapipe. Sistem ini relevan di masa pandemi COVID-19 karena memungkinkan pengukuran tanpa kontak langsung. Webcam mengambil gambar pengguna, lalu diproses untuk memberikan rekomendasi ukuran baju. Pengujian menunjukkan akurasi 95% dengan jarak optimal 120 cm dari kamera. Pada penelitian [13] membangun aplikasi pengukur pakaian berbasis mobile menggunakan image processing, yang relevan pada masa new normal untuk mengurangi kontak fisik di toko baju. Aplikasi ini menggunakan algoritma Canny Edge Detection. Pengguna memindai barcode pada tag baju, lalu aplikasi merekomendasikan ukuran pakaian.

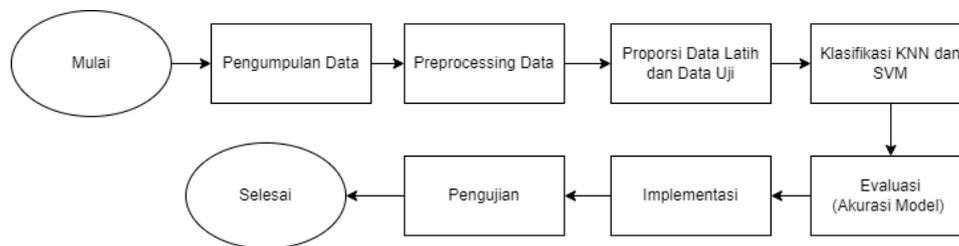
Penelitian ini memiliki perbedaan signifikan dengan penelitian sebelumnya karena memanfaatkan kombinasi atribut yakni, jenis kelamin, bentuk badan, tinggi badan, berat badan, lingkar dada, dan ukuran baju untuk membangun sistem rekomendasi ukuran baju dewasa. Atribut ini merupakan gabungan dari atribut-atribut yang telah digunakan pada penelitian-penelitian sebelumnya. Misalnya, penelitian [3] menggunakan algoritma Decision Tree dengan dataset yang terdiri dari 207 data pelatihan yang mencakup jenis kelamin,

berat badan, tinggi badan, dan bentuk tubuh, menghasilkan akurasi sebesar 67% pada 21 data uji acak menggunakan Confusion Matrix. Sementara itu, penelitian [4] menggunakan algoritma Naïve Bayes dengan atribut jenis kelamin, tinggi badan, berat badan, lingkar dada, dan ukuran baju, yang mana aplikasi hasil penelitian tersebut diuji dengan pengujian blackbox untuk menilai kelayakan sistem. Dengan adanya perbedaan dengan menggabungkan atribut dan juga perbedaan dalam penggunaan algoritma yakni K-NN dan SVM, penelitian ini diharapkan mampu memberikan rekomendasi ukuran baju yang lebih akurat.

Ada beberapa penelitian sebelumnya yang telah melakukan klasifikasi menggunakan algoritma K-NN dan SVM. Seperti pada penelitian terkait Klasifikasi Pasien Penyakit Jantung [14] didapatkan akurasi tertinggi dari algoritma SVM. Kemudian pada penelitian mengenai klasifikasi kualitas udara di Jakarta [7] didapatkan akurasi algoritma SVM lebih unggul dari algoritma K-NN dengan selisih yang tidak terlalu banyak yakni 97% untuk SVM dan 96% untuk K-NN. Selanjutnya pada penelitian [15] yang mengidentifikasi kinerja pegawai, dalam penelitian ini algoritma K-NN lebih unggul dari algoritma SVM. Pada penelitian [6] yang membahas mengenai klasifikasi obesitas multi kelas didapatkan algoritma K-NN lebih unggul dari algoritma SVM dengan akurasi K-NN 99% dan algoritma SVM 96%.

## 2. METODOLOGI PENELITIAN

Alur tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian

Gambar 1. menunjukkan alur tahapan dalam Pembangunan model Machine Learning menggunakan algoritma K-Nearest Neighbor dan Support Vector Machine, berikut penjelasan untuk tiap alur dalam pembangunan model.

### 2.1. Pengumpulan Data

Data dalam penelitian ini dikumpulkan melalui kuesioner *Google Form* dan wawancara langsung. Kuesioner disebarakan melalui *WhatsApp* kepada responden berusia 18 tahun ke atas, dengan data meliputi jenis kelamin, bentuk tubuh, tinggi badan, berat badan, lingkar dada, dan ukuran baju. Data tersebut berjumlah 129 baris dengan 6 kelas meliputi ukuran S, M, L, XL, XXL, XXXL.

### 2.2. Preprocessing Data

Preprocessing adalah langkah penting dalam data mining untuk memastikan data siap diproses, terutama ketika data mengandung masalah seperti nilai hilang, outlier, atau format yang tidak sesuai. Tahap ini membantu mengatasi masalah yang dapat memengaruhi akurasi hasil klasifikasi data [16]. Pada tahap ini dilakukan proses sebagai berikut.

#### 2.2.1 Labeling

Pada tahap ini akan dilakukan perubahan data yang berbentuk string kedalam bentuk numerik, hal ini dilakukan untuk mempermudah dalam proses pengklasifikasian.

#### 2.2.2 One Hot Encoder

*One Hot Encoder* merupakan teknik yang akan mengubah dimana semua kolom akan bernilai 0, kecuali angka 1 yang berarti mewakili kata tersebut [17]. *One Hot Encoder* merupakan teknik dalam pengolahan data yang digunakan untuk mengubah data kategorikal atau string menjadi bentuk numerik yang dapat digunakan dalam algoritma Machine Learning salah satunya yakni K-Nearest Neighbor. Proses perubahan data dari kategorikal kedalam numerik dengan cara membuat kolom biner baru untuk setiap kategori unik dalam kolom aslinya. Setiap baris dalam data asli akan memiliki nilai 1 pada kolom yang sesuai dengan kategorinya dan 0 pada kolom lainnya.

#### 2.2.3 Normalisasi

Normalisasi data merupakan proses yang digunakan untuk mengubah rentang nilai pada variabel tertentu agar tidak ada data yang terlalu besar maupun terlalu kecil [18]. Normalisasi sendiri bertujuan untuk meningkatkan akurasi dari model yang dibangun. Berikut merupakan persamaan dari Standar Scaler [19]:

$$\text{Standar Scaler} = \frac{x - \bar{\mu}}{\sigma} \quad (1)$$

Pada persamaan (1)  $x$  merupakan nilai data tersebut,  $\bar{\mu}$  merupakan rata-rata dari fitur data, serta  $\sigma$  merupakan nilai standar deviasi dari fitur data.

### 2.3. Proporsi Data Latih dan Data Uji

Pembagian data latih dan data uji penting karena data latih digunakan untuk membangun model klasifikasi dengan mempelajari pola dalam dataset, sedangkan data uji digunakan untuk mengukur keakuratan dan validitas model tersebut pada data yang belum pernah dilihat. Hal ini memastikan bahwa model dapat menggeneralisasi dengan baik pada data baru dan tidak hanya menghafal data latih. Berdasarkan [20] akan digunakan proporsi dataset 70:30, 80:20, dan 60:40 untuk melihat akurasi yang terbaik.

### 2.4. Klasifikasi K-NN dan SVM

#### 2.4.1 K-Nearest Neighbor

K-Nearest Neighbor adalah algoritma *Supervised Learning* dimana hasil dari data yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Algoritma ini digunakan untuk mengklasifikasikan obyek baru berdasarkan atribut dan sampel dari training data [21]. Cara kerja algoritma ini dengan melakukan perhitungan jarak antara data baru dengan data latih, lalu mengklasifikasikan data tersebut berdasarkan data yang paling dekat. Dalam algoritma K-Nearest Neighbor, perhitungan jarak biasanya menggunakan *Euclidean Distance*. Pada perhitungan *Euclidean Distance*, semakin kecil nilai yang dihasilkan, maka semakin dekat kemiripan antara data uji tersebut dengan data latih. Penentuan jumlah tetangga pada algoritma K-Nearest Neighbor sebaiknya menggunakan angka ganjil, untuk mengurangi resiko kebingungan dalam pengklasifikasian data. Perhitungan jarak menggunakan rumus *Euclidean Distance* seperti ditunjukkan pada persamaan 2 [22].

$$\text{Euclidean Distance} = \sqrt{\sum_{i=0}^n (x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

#### 2.4.2 Support Vector Machine

Support Vector Machine merupakan algoritma yang bekerja dengan cara mencari *hyperplane* yang optimal sebagai pemisah antar kelas yang diklasifikasi [23]. Sebagai algoritma klasifikasi, SVM menerima kumpulan data pelatihan yang berisi data ukuran baju yang telah ditandai berdasarkan kategori ukuran tertentu, seperti S, M, L, dan lainnya. Algoritma SVM kemudian membangun model yang mampu memprediksi ukuran baju yang sesuai untuk data baru, berdasarkan fitur-fitur seperti jenis kelamin, bentuk tubuh, tinggi, berat badan, dan lingkar dada. Pada SVM, terdapat dua jenis *hyperplane* yang penting: *hyperplane linear* dan *hyperplane non-linear*. Ketika data bisa dipisahkan secara sempurna dengan *hyperplane linear*, algoritma ini disebut SVM linear. Namun, jika data tidak dapat dipisahkan secara linear, SVM akan menggunakan transformasi kernel untuk memetakan data ke dalam dimensi fitur yang lebih tinggi. Di dimensi ini, *hyperplane linear* bisa dibuat untuk memisahkan data. SVM dapat menggunakan berbagai jenis fungsi kernel, seperti kernel linear, polinomial, atau Gaussian, untuk menangani data non-linear. Fungsi kernel ini membantu mengubah data ke dimensi lebih tinggi, sehingga mempermudah proses pemisahan antar kelas [24]. Berikut merupakan formula umum dari perhitungan SVM pada persamaan 3.

$$f(x) = \text{sign}(w \cdot x + b) \quad (3)$$

$f(x)$  merupakan fungsi prediksi yang menentukan kelas data berdasarkan posisi relatif terhadap *hyperplane*. Fungsi sign memberikan hasil positif atau negatif untuk menunjukkan kelas.  $w$  merupakan vektor normal *hyperplane* yang mengatur arah dan orientasi *hyperplane*. Ini adalah parameter utama yang dioptimalkan untuk memisahkan kelas.  $x$  merupakan vektor fitur input yang mewakili data, seperti jenis kelamin, tinggi dalam klasifikasi baju.  $b$  merupakan bias yang menentukan posisi *hyperplane* terhadap titik asal, memungkinkan pemisah data yang lebih baik.

### 2.5. Evaluasi

Evaluasi model ini dilakukan untuk menganalisa kinerja dari algoritma K-Nearest Neighbor dan Support Vector Machine dalam mengklasifikasikan Ukuran Baju. Pada penelitian ini dilakukan beberapa eksperimen pembagian data latih dan data uji untuk mendapatkan akurasi yang terbaik dan mengukur kinerja algoritma K-Nearest Neighbor dan Support Vector Machine dalam menentukan rekomendasi ukuran baju.

### 2.6. Implementasi

Pada tahap implementasi, dilakukan pembuatan website yang berfungsi memberikan rekomendasi ukuran baju kepada pengguna. Website ini akan menggunakan model klasifikasi dengan akurasi terbaik yang

telah diidentifikasi melalui eksperimen, sehingga dapat memberikan rekomendasi ukuran yang paling akurat berdasarkan data input pengguna.

**2.7. Pengujian**

Pada tahap pengujian, dilakukan evaluasi terhadap website yang telah dibuat untuk memastikan bahwa fungsionalitasnya berjalan sesuai dengan yang diharapkan. Pengujian ini bertujuan untuk memverifikasi bahwa website mampu memberikan rekomendasi ukuran baju dengan benar dan sesuai kebutuhan pengguna.

**3. HASIL DAN DISKUSI**

Dataset yang digunakan dalam penelitian ini berjumlah 129 data yang didapatkan melalui kuesioner yang disebar melalui Whatsapp dan wawancara langsung kepada responden. Dataset berisi beberapa atribut meliputi Jenis Kelamin, Bentuk Tubuh, Tinggi Badan, Berat Badan, Lingkar Dada, Serta Ukuran Baju meliputi ukuran S, M, L, XL, XXL, XXXL sebagai Label atau Targetnya. Berikut merupakan cuplikan dari dataset yang digunakan untuk penelitian ini.

**Tabel 1. Contoh Dataset**

| No  | Jenis Kelamin | Bentuk Tubuh      | Tinggi Badan | Berat Badan | Lingkar Dada | Ukuran Baju |
|-----|---------------|-------------------|--------------|-------------|--------------|-------------|
| 1   | Laki-Laki     | Rounded           | 170          | 110         | 130          | XXXL        |
| 2   | Laki-Laki     | Rounded           | 167          | 100         | 128          | XXXL        |
| 3   | Perempuan     | Triangle          | 159          | 60          | 99           | L           |
| 4   | Perempuan     | Rectangle         | 158          | 46          | 85           | S           |
| 5   | Laki-Laki     | Rounded           | 165          | 85          | 116          | XXXL        |
| ... | .....         | .....             | .....        | .....       | .....        | .....       |
| 125 | Perempuan     | Hourglass         | 160          | 70          | 107          | XL          |
| 126 | Laki-Laki     | Triangle          | 166          | 74          | 104          | L           |
| 127 | Perempuan     | Inverted Triangle | 160          | 60          | 101          | L           |
| 128 | Laki-Laki     | Triangle          | 167          | 80          | 104          | XL          |
| 129 | Perempuan     | Rectangle         | 165          | 50          | 93           | M           |

**3.1. Preprocessing**

Proses pertama yang dilakukan pada tahap ini yakni melakukan pelabelan data dengan mengubah data yang berbentuk string kedalam bentuk numerik. Pada atribut Jenis Kelamin dilakukan pelabelan dimulai dari jenis kelamin Laki-laki dengan angka 0 dan jenis kelamin Perempuan dengan angka 1. Pada atribut Ukuran Baju dilakukan pelabelan dimulai dengan ukuran 'S': 0, 'M': 1, 'L': 2, 'XL': 3, 'XXL': 4, 'XXXL': 5. Setelah dilakukan proses labelling ini maka data siap untuk diproses ke tahap selanjutnya. Berikut merupakan cuplikan hasil dari proses labelling data.

**Tabel 2. Proses Labeling**

| No  | Jenis Kelamin | Bentuk Tubuh      | Tinggi Badan | Berat Badan | Lingkar Dada | Ukuran Baju |
|-----|---------------|-------------------|--------------|-------------|--------------|-------------|
| 1   | 0             | Rounded           | 170          | 110         | 130          | 5           |
| 2   | 0             | Rounded           | 167          | 100         | 128          | 5           |
| 3   | 1             | Triangle          | 159          | 60          | 99           | 2           |
| 4   | 1             | Rectangle         | 158          | 46          | 85           | 0           |
| 5   | 0             | Rounded           | 165          | 85          | 116          | 5           |
| ... | .....         | .....             | .....        | .....       | .....        | .....       |
| 125 | 1             | Hourglass         | 160          | 70          | 107          | 3           |
| 126 | 0             | Triangle          | 166          | 74          | 104          | 2           |
| 127 | 1             | Inverted Triangle | 160          | 60          | 101          | 2           |
| 128 | 0             | Triangle          | 167          | 80          | 104          | 3           |
| 129 | 1             | Rectangle         | 165          | 50          | 93           | 1           |

Proses selanjutnya dilakukan perubahan data kategorikal pada atribut Bentuk Badan menjadi numerik dengan teknik one hot encoder. Pada proses ini dilakukan perubahan data dari kategorikal kedalam numerik dengan cara membuat kolom biner baru untuk setiap kategori unik dalam kolom aslinya. Setiap baris dalam data asli akan memiliki nilai 1 pada kolom yang sesuai dengan kategorinya dan 0 pada kolom lainnya. Pada atribut Bentuk Tubuh terdapat 6 jenis bentuk tubuh yakni Trapezoid, Rectangle, Triangle, Inverted Triangle, Hourglass, dan Rounded. Dari keenam jenis tersebut akan diubah menjadi numerik dengan teknik One Hot Encoder, berikut representasi dari perubahan bentuk tubuh menjadi numerik dengan teknik One Hot Encoder ditunjukkan pada tabel 3.

**Tabel 3.** One Hot Encoder Bentuk Tubuh

| Trapezoid | Rectangle | Triangle | Inverted Triangle | Hourglass | Rounded |
|-----------|-----------|----------|-------------------|-----------|---------|
| 1         | 0         | 0        | 0                 | 0         | 0       |
| 0         | 1         | 0        | 0                 | 0         | 0       |
| 0         | 0         | 1        | 0                 | 0         | 0       |
| 0         | 0         | 0        | 1                 | 0         | 0       |
| 0         | 0         | 0        | 0                 | 1         | 0       |
| 0         | 0         | 0        | 0                 | 0         | 1       |

Proses terakhir dalam tahap ini dilakukan normalisasi pada atribut Jenis Kelamin, Tinggi Badan, Berat Badan, dan Lingkar dada menggunakan teknik standar scaler. Berikut merupakan cuplikan proses normalisasi dari data, ditunjukkan pada tabel 4.

**Tabel 4.** Cuplikan Normalisasi Data

| No  | Jenis Kelamin | Tinggi Badan | Berat Badan | Lingkar Dada |
|-----|---------------|--------------|-------------|--------------|
| 1   | -0.769800     | 0.594307     | 2.880520    | 3.606638     |
| 2   | -0.769800     | 0.254075     | 2.192141    | 3.343261     |
| 3   | 1.299038      | -0.653211    | -0.561376   | -0.475713    |
| ..  | ...           | ...          | ...         | ...          |
| 127 | 1.299038      | -0.539800    | -0.561376   | -0.212335    |
| 128 | -0.769800     | 0.254075     | 0.815382    | 0.182731     |
| 129 | 1.299038      | 0.027254     | -1.249755   | -1.265845    |

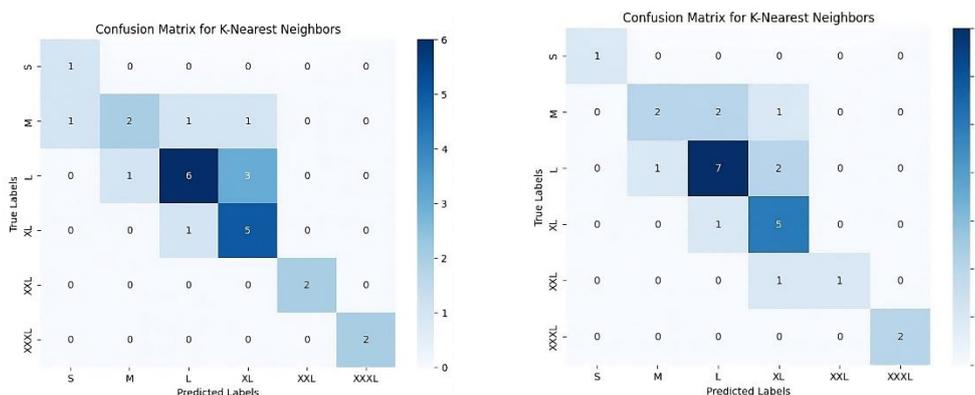
### 3.2. Evaluasi Klasifikasi K-Nearest Neighbor dan Support Vector Machine

#### 3.2.1. K-Nearest Neighbor

Pada tahap klasifikasi dengan K-Nearest Neighbor, digunakan proporsi data latih dan data uji sebesar 70:30, 80:20, dan 60:40, serta parameter k bernilai 3,5, dan 7 untuk menemukan kombinasi terbaik. Evaluasi model dilakukan menggunakan confusion matrix guna menilai performa klasifikasi berdasarkan akurasi, presisi, recall, dan nilai F1-score. Selanjutnya, perbandingan akurasi dari berbagai konfigurasi K-NN akan disajikan dalam bentuk tabel untuk memudahkan analisis performa model pada tiap proporsi data dan nilai k yang digunakan. Dalam penelitian ini, digunakan library *scikit-learn* dan bahasa pemrograman *Python* untuk implementasi klasifikasi. Berikut merupakan tabel perbandingan akurasi dari tiap eksperimen pada tabel 5 dan gambar 2.

**Tabel 5.** Hasil Eksperimen K-NN

| No | Proporsi | K   | Akurasi | Presisi | Recall | F1-score |
|----|----------|-----|---------|---------|--------|----------|
| 1  | 70:30:00 | K=3 | 51%     | 47%     | 51%    | 48%      |
|    |          | K=5 | 64%     | 65%     | 64%    | 62%      |
|    |          | K=7 | 61%     | 58%     | 61%    | 57%      |
| 2  | 80:20:00 | K=3 | 57%     | 55%     | 57%    | 55%      |
|    |          | K=5 | 69%     | 71%     | 69%    | 68%      |
|    |          | K=7 | 69%     | 71%     | 69%    | 68%      |
| 3  | 60:40:00 | K=3 | 55%     | 53%     | 55%    | 53%      |
|    |          | K=5 | 63%     | 66%     | 63%    | 62%      |
|    |          | K=7 | 65%     | 68%     | 65%    | 63%      |

**Gambar 2.** Confusion Matrix k=5 dan k=7

Berdasarkan hasil eksperimen yang telah dilakukan terdapat 2 kombinasi eksperimen yang menghasilkan akurasi sama dan tertinggi sebesar 69%, namun berdasarkan hasil confusion matrix pada masing-masing eksperimen, model dengan k=7 ditunjukkan pada gambar 2 bagian kanan menunjukkan performa yang lebih baik dalam mengklasifikasikan ukuran L dengan lebih sedikit kesalahan. Dengan k=7, model lebih konsisten dalam mengenali kategori ukuran menengah, terutama ukuran L, yang memiliki lebih banyak instance yang benar tanpa salah klasifikasi yang signifikan. Oleh karena itu, meskipun akurasinya sama, kombinasi k=7 dianggap lebih optimal untuk klasifikasi pada dataset ini karena distribusi kesalahan yang lebih rendah dan prediksi yang lebih akurat untuk ukuran menengah.

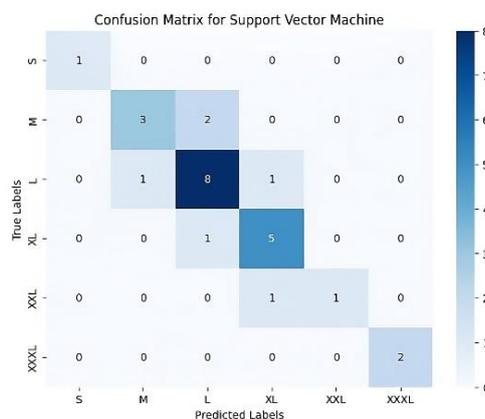
### 3.2.2. Support Vector Machine

Pada tahap klasifikasi menggunakan Support Vector Machine (SVM), eksperimen dilakukan dengan proporsi data latih dan data uji sebesar 70:30, 80:20, dan 60:40. Model SVM ini menggunakan kernel RBF serta pendekatan *One-vs-Rest*, yang dipilih berdasarkan penelitian [25] karena memberikan akurasi tertinggi dalam klasifikasi yang serupa. Dalam penelitian ini, digunakan library *scikit-learn* dan bahasa pemrograman *Python* untuk implementasi klasifikasi. Berikut merupakan tabel perbandingan akurasi tiap proporsi data terdapat pada tabel 6.

**Tabel 6.** Hasil Eksperimen SVM

| No | Proporsi | Akurasi | Presisi | Recall | F1-score |
|----|----------|---------|---------|--------|----------|
| 1  | 70:30    | 71%     | 74%     | 71%    | 70%      |
| 2  | 80:20    | 76%     | 78%     | 76%    | 76%      |
| 3  | 60:40    | 73%     | 69%     | 73%    | 69%      |

Pada tabel 6. Menunjukkan semua hasil eksperimen yang telah dilakukan, dan dapat dilihat bahwa pada proporsi data dengan 80% untuk data uji dan 20% untuk data latih mendapatkan akurasi tertinggi diantara proporsi dataset lainnya dengan akurasi sebesar 76%, presisi 78%, recall 76%, dan f1 score 76%. Dapat dilihat pada confusion matrix untuk proporsi data 80:20 pada gambar 3.



**Gambar 3.** Confusion Matrix SVM Terbaik

Berdasarkan confusion matrix pada gambar 4, hasil menunjukkan bahwa model cukup efektif dalam mengklasifikasikan sebagian besar ukuran, terutama ukuran M, L, dan XL, yang memiliki jumlah prediksi benar lebih tinggi. Namun, terdapat beberapa kesalahan klasifikasi, terutama pada ukuran M yang sering diklasifikasikan sebagai L, serta pada ukuran L yang diklasifikasikan sebagai M atau XL. Ukuran S dan XXXL memiliki hasil yang akurat tanpa kesalahan, menunjukkan performa yang cukup baik dalam mengidentifikasi ukuran pakaian, meskipun masih ada ruang untuk peningkatan akurasi pada kelas menengah seperti M, L, dan XL.

Dari kedua algoritma yang digunakan, yaitu SVM dan K-NN, algoritma SVM dengan kernel RBF dan pendekatan *One-vs-Rest* pada proporsi data 80:20 menunjukkan hasil yang paling unggul. Dengan akurasi sebesar 76%, presisi 78%, recall 76%, dan F1 score 76%, SVM memberikan kinerja yang lebih baik dibandingkan K-NN dalam klasifikasi ukuran pakaian. Oleh karena itu, model SVM dipilih sebagai model yang akan diimplementasikan pada website rekomendasi ukuran pakaian, karena memiliki akurasi terbaik yang diharapkan dapat meningkatkan keakuratan dan keandalan dalam memberikan rekomendasi sesuai data pengguna.

### 3.3. Implementasi

Pada tahap implementasi, sebuah website dikembangkan sebagai platform untuk memberikan rekomendasi ukuran pakaian kepada pengguna. Website ini memanfaatkan model machine learning terbaik yang dihasilkan dari eksperimen sebelumnya, yaitu SVM dengan kernel RBF. Model tersebut disimpan dalam format pipeline, yang memungkinkan integrasi yang mudah ke dalam aplikasi web untuk memberikan rekomendasi ukuran baju secara efisien. Dengan pendekatan ini, pengguna dapat dengan cepat memperoleh saran ukuran yang akurat berdasarkan data input yang diberikan, meningkatkan pengalaman berbelanja dan meminimalkan kesalahan dalam pemilihan ukuran pakaian. Berikut merupakan bagian-bagian halaman atau menu yang terdapat pada Website rekomendasi ukuran baju dewasa.

Halaman Menu Rekomendasi Baju ditunjukkan pada gambar 4 terdapat form untuk menginputkan kriteria yang digunakan untuk merekomendasikan ukuran baju, form rekomendasi berisikan input untuk jenis kelamin, gambar bentuk tubuh berdasarkan jenis kelamin yang diinputkan oleh pengguna, pilihan bentuk tubuh, input tinggi badan, input berat badan, lingkar dada, dan hasil dari rekomendasi kemeja ditunjukkan pada gambar 5.

**Gambar 5.** Halaman Menu Rekomendasi Baju

**Gambar 6.** Form Hasil Rekomendasi Ukuran Baju

### 3.4. Pengujian

Pengujian Black Box adalah algoritma pengujian yang mengevaluasi hasil eksekusi aplikasi berdasarkan masukan yang diberikan (data uji) untuk memastikan bahwa fungsi aplikasi sesuai dengan persyaratan yang ditetapkan. Algoritma ini berfokus pada antarmuka dan fungsionalitas aplikasi, serta kesesuaian alur fungsi yang diinginkan oleh pengguna. Dalam pengujian ini, tidak dilakukan pemeriksaan terhadap kode sumber program [26]. Berdasarkan penelitian oleh [26], pengujian black box testing dilakukan dengan fokus pada evaluasi fungsionalitas aplikasi. Pada website rekomendasi ukuran pakaian ini, teknik State Transition Testing akan diterapkan, yang memungkinkan pengujian alur transisi antar status dalam aplikasi untuk memastikan bahwa setiap keadaan berfungsi sebagaimana mestinya. Pendekatan ini diharapkan dapat

mengidentifikasi potensi kesalahan dalam alur interaksi pengguna dengan sistem, sehingga meningkatkan keandalan dan akurasi dari rekomendasi yang diberikan. Pengujian black box ditunjukkan pada gambar 7.

**Tabel 7.** Pengujian Black Box

| Transisi | Menu                         | Aksi   | Tujuan                              | Skenario  | Output                                     | Kesimpulan |
|----------|------------------------------|--|-------------------------------------|---|--|------------|
| 1        | Registrasi                   | Klik tombol registrasi   | Halaman login                       | User melakukan registrasi ketika belum memiliki akun              | User berhasil melakukan registrasi         | Berhasil   |
| 2        | Login                        | Klik tombol login  | Halaman utama                       | User melakukan login dengan email dan password benar              | User berhasil melakukan login              | Berhasil   |
| 3        | Rekomendasi Ukuran Baju      | Klik tombol menu rekomendasi baju  | Halaman Rekomendasi Ukuran Baju     | User memasuki halaman rekomendasi                                 | User berhasil memasuki halaman rekomendasi | Berhasil   |
| 4        | Form Rekomendasi Ukuran Baju | Input jenis kelamin, bentuk tubuh, tinggi badan, berat badan, lingkar dada | Input data rekomendasi              | User input data untuk rekomendasi ukuran baju                     | User berhasil input data                   | Berhasil   |
| 5        | Tombol Get Recommendation    | Klik tombol Get Recommendation   | Mendapatkan rekomendasi ukuran baju | User klik tombol Get Recommendation setelah mengisi data          | Hasil Rekomendasi Ukuran baju              | Berhasil   |
| 6        | Tombol reset                 | Klik tombol reset  | Mengosongkan form rekomendasi       | User klik tombol reset setelah mendapat rekomendasi ukuran        | Form kosong dari inputan                   | Berhasil   |
| 7        | Halaman utama                | Klik tombol menu utama   | Halaman utama                       | User berada di menu rekomendasi baju dan kembali ke halaman utama | User berhasil ke halaman utama             | Berhasil   |
| 8        | Logout                       | Klik tombol logout   | Keluar akun                         | User keluar akun dari website                                     | User berhasil keluar akun dari website     | Berhasil   |

Berdasarkan hasil pengujian yang dilakukan, website rekomendasi ukuran pakaian berhasil memenuhi seluruh fungsi yang diujikan, termasuk proses registrasi pengguna, login, akses ke halaman rekomendasi, input data rekomendasi, memberi hasil rekomendasi ukuran, hingga logout akun. Keseluruhan pengujian menunjukkan bahwa setiap fungsi berjalan sesuai dengan yang diharapkan dan menghasilkan output yang diinginkan. Hasil ini menegaskan bahwa website telah memenuhi persyaratan fungsional dan siap digunakan oleh pengguna.

#### 4. KESIMPULAN

Berdasarkan hasil eksperimen dan pengujian yang dilakukan, penelitian ini menunjukkan bahwa algoritma SVM dengan kernel RBF dan pendekatan *One-vs-Rest* pada proporsi data 80:20 memberikan performa terbaik dalam klasifikasi ukuran pakaian, dengan akurasi tertinggi sebesar 76%, presisi 78%, recall 76%, dan F1 score 76%. Dibandingkan dengan K-NN, yang memiliki akurasi tertinggi sebesar 69%, presisi 71%, recall 69%, f1-score 68%, SVM lebih unggul dalam memberikan prediksi yang akurat dan konsisten untuk ukuran pakaian, terutama pada kelas menengah seperti M, L, dan XL. Oleh karena itu, model SVM dipilih untuk diterapkan pada website rekomendasi ukuran pakaian. Selain itu, hasil pengujian black box testing menunjukkan bahwa website telah berhasil memenuhi semua persyaratan fungsional, mulai dari proses registrasi, login, akses ke halaman rekomendasi, input data rekomendasi, memberi hasil rekomendasi ukuran baju, hingga logout akun. Setiap fungsi pada website berjalan sesuai dengan yang diharapkan dan menghasilkan output yang diinginkan, memastikan website siap digunakan oleh pengguna. Untuk penelitian selanjutnya, disarankan untuk meningkatkan jumlah data yang digunakan agar hasil klasifikasi menjadi lebih akurat dan andal. Selain itu, eksplorasi terhadap algoritma lain dapat dipertimbangkan untuk mengidentifikasi

algoritma yang lebih optimal dalam memberikan rekomendasi ukuran baju, khususnya pada kelas ukuran menengah yang memerlukan peningkatan akurasi.

## REFERENSI

- [1] A. P. Afrianto and I. Irwansyah, “Eksplorasi Kondisi Masyarakat Dalam Memilih Belanja Online Melalui Shopee Selama Masa Pandemi Covid-19 Di Indonesia,” *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 3, no. 1, pp. 10–29, 2021.
- [2] R. D. Setianingsih, D. I. Wiranata, J. F. Putra, M. M. Al-Afif, N. A. Setiawan, and M. A. Zahwa, “Analisis Tingkat Belanja Online Saat Pandemi dan Sesudah Pandemi pada Mahasiswa Fakultas Ekonomi Universitas Negeri Semarang,” *Jurnal Majemuk*, vol. 2, no. 3, pp. 339–348, 2023.
- [3] F. A. Oktavirahani and R. Maharesi, “Implementasi Algoritma Decision Tree Cart Untuk Merekomendasikan Ukuran Baju,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 1, pp. 138–147, 2022.
- [4] L. N. Aziza, R. Y. Astuti, B. A. Maulana, and N. Hidayati, “Penerapan Algoritma K-Nearest Neighbor untuk Klasifikasi Ketahanan Pangan di Provinsi Jawa Tengah: Application of The K-Nearest Neighbor Algorithm for Food Security Classification In Central Java Province,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 404–412, 2024.
- [5] A. D. Putro and H. Tantyoko, “Hybrid Algoritma Vgg16-Net Dengan Support Vector Machine Untuk Klasifikasi Jenis Buah dan sayuran,” *JTIM: Jurnal Teknologi Informasi dan Multimedia*, vol. 5, no. 2, pp. 56–65, 2023.
- [6] S. A. Utiahman and A. M. M. Pratama, “Analisis Perbandingan K-NN, SVM, Decision Tree dan Regresi Logistik Untuk Klasifikasi Obesitas Multi Kelas,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 3137–3146, 2024.
- [7] B. V. Jayadi, T. Handhayani, and M. D. Lauro, “Perbandingan K-NN Dan Svm Untuk Klasifikasi Kualitas Udara Di Jakarta,” *Jurnal Ilmu Komputer Dan Sistem Informasi*, vol. 11, no. 2, 2023.
- [8] Y. Mufflihah, “Sistem Rekomendasi Ukuran Baju Pada E-commerce Dengan Algoritma Naive Bayes,” *Jurnal Mahasiswa Teknik Informatika*, vol. 3, no. 1, pp. 10–15, 2024.
- [9] F. Rizky, H. Fitriyah, and R. Primananda, “Implementasi K-NN untuk Sistem Klasifikasi Ukuran Baju Pria berdasarkan Pengukuran Badan menggunakan Algoritma Pengolahan Citra Digital berbasis Raspberry Pi,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 7, pp. 3046–3053, 2022.
- [10] R. Fazlia, A. Anwar, and H. T. Hidayat, “Augmented Reality Virtual Dressing Room Berbasis Android,” *Jurnal Teknologi Rekayasa Informasi dan Komputer*, vol. 5, no. 2, 2022.
- [11] R. Aqshal and F. Nurzaman, “Rancang Bangun Sistem Rekomendasi Ukuran Baju Berdasarkan Ukuran Tubuh Menggunakan Algoritma K-Nearest Neighbor (K-NN) Berbasis Internet Of Things,” *Jurnal Ilmiah Teknik Informatika (TEKINFO)*, vol. 23, no. 2, pp. 76–87, 2022.
- [12] A. Fauziah and Y. Saragih, “Sistem Identifikasi Pengukuran Baju Menggunakan Human Body Estimation Dataset Mediapipe Dengan Algoritma Euclidean Distance,” *Aisyah Journal Of Informatics and Electrical Engineering (AJIEE)*, vol. 5, no. 2, pp. 127–134, 2023.
- [13] I. Kurniasari, Y. B. Utomo, and A. D. Evasari, “Perancangan Aplikasi Pengukur Pakaian Berbasis Mobile,” *INFORMASI (Jurnal Informatika dan Sistem Informasi)*, vol. 14, no. 2, pp. 170–179, 2022.
- [14] A. Damayunita, R. S. Fuadi, and C. Juliane, “Comparative Analysis of Naive Bayes, K-Nearest Neighbors (K-NN), and Support Vector Machine (SVM) Algorithms for Classification of Heart Disease Patients,” *Jurnal Online Informatika*, vol. 7, no. 2, pp. 219–225, 2022.
- [15] S. A. Laga, “Perbandingan Algoritma K-NN dan SVM Berdasarkan Kinerja Pegawai,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 3, pp. 420–425, 2023.
- [16] M. D. Purbolaksono, M. I. Tantowi, A. I. Hidayat, and A. Adiwijaya, “Perbandingan support vector machine dan modified balanced random forest dalam deteksi pasien penyakit diabetes,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 393–399, 2021.
- [17] M. J. Aufa and A. Qoiriah, “Analisis Sentimen Pengguna Platform Belajar Online Coursera menggunakan Random Forest dengan Algoritma Ekstraksi Fitur Word2vec,” *Journal of Informatics and Computer Science (JINACS)*, pp. 244–255, 2022.
- [18] S. H. Anggela, L. W. Santoso, and J. Andjarwirawan, “Sistem Rekomendasi Pembelian Laptop dengan K-Nearest Neighbor (K-NN),” *Jurnal Infra*, vol. 10, no. 2, pp. 254–260, 2022.
- [19] A. T. Akbar, N. Yudistira, and A. Ridok, “Identifikasi Gagal Ginjal Kronis dengan Mengimplementasikan Algoritma Support Vector Machine beserta K-Nearest Neighbour (SVM-K-NN),” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 10, no. 2, 2023.
- [20] A. H. Permana, F. R. Umbara, and F. Kasyidi, “Klasifikasi Penyakit Jantung Tipe Kardiovaskular Menggunakan Adaptive Synthetic Sampling dan Algoritma Extreme Gradient Boosting,” *Building of Informatics, Technology and Science (BITS)*, vol. 6, no. 1, pp. 499–508, 2024.

- [21] F. Kusumah and M. Ardhiansyah, *Analisis Sistem Pendeteksi Wajah Pada Gambar Dengan Algoritma K-Nearest Neighbor*. Pascal Books, 2022. [Online]. Available: <https://books.google.co.id/books?id=ylypEAAAQBAJ>
- [22] S. Simare-mare and H. Pandia, "Rekomendasi Komoditas Ekspor Menggunakan K-Nearest Neighbor," *JSiI (Jurnal Sistem Informasi)*, vol. 10, no. 2, pp. 150–156, 2023.
- [23] O. H. Rahman, G. Abdillah, and A. Komarudin, "Klasifikasi Ujaran Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine," *jurnal resti (rekayasa sistem dan teknologi informasi)*, vol. 5, no. 1, pp. 17–23, 2021.
- [24] S. Rabbani, D. Safitri, N. Rahmadhani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM: Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 2, pp. 153–160, 2023.
- [25] R. W. Pratama, Y. H. Chrisnanto, and G. Gunawan, "Klasifikasi Efek Kerusakan Gempa Bumi Berdasarkan Skala Modified Mercalli Intensity Menggunakan Algoritma Multiclass Support Vector Machine," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1739–1745, 2024.
- [26] M. Mintarsih, "Pengujian Black Box Dengan Teknik Transition Pada Sistem Informasi Perpustakaan Berbasis Web Dengan Algoritma Waterfall Pada SMC Foundation," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 1, pp. 33–35, 2023.