

Penerapan *Linear Discriminant Analysis* Untuk Meningkatkan Kinerja Algoritma *Support Vector Machine*

Fenly*¹, Gusrianty*², Deny Jollyta³ Erlin⁴, Ramalia Noratama Putri⁵, Dwi Oktarina⁶,

^{1,2,3,4} *Teknik Informatika, Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No 78-88, Pekanbaru, 28127, Indonesia*

^{5,6} *Sistem Informasi, Institut Bisnis dan Teknologi Pelita Indonesia, Jl. Jend. Ahmad Yani No 78-88, Pekanbaru, 28127, Indonesia*

Info Artikel

Riwayat Artikel :

Received 2025-05-16

Revised 2025-08-27

Accepted 2025-08-29

Abstract – Obesity is a complex chronic disease influenced by various factors, such as genetic, environmental, and lifestyle, which is characterized by excess body weight due to the excessive accumulation of body fat. With the rapid advancement of technology and digitalization across all sectors, data has become increasingly vital, as large datasets generate valuable information. However, a key challenge in data analysis is addressing redundancy, noise, and high dimensionality, which can affect the performance of machine learning algorithms. This study aims to investigate the effectiveness of combining *Linear Discriminant Analysis (LDA)* and *Support Vector Machine (SVM)* in enhancing the accuracy and efficiency of high-dimensional data classification, particularly in predicting obesity levels. *LDA* is employed to reduce data dimensionality while retaining the most relevant features, whereas *SVM* is utilized as the classification algorithm to predict obesity levels based on patterns identified within the dataset. The research was conducted using a dataset consisting of 779 training samples and 195 testing samples. The results reveal that the combination of *LDA* and *SVM* achieved a classification accuracy of up to 99%, with a 50% reduction in data dimensionality and a computation speed of 0,0696 second. Moreover, computation time was significantly reduced, indicating that *LDA* not only facilitates data simplification but also improves the overall efficiency of the classification process.

Keywords: Classification, Dimensionality Reduction, Linear Discriminant Analysis, Obesity, Support Vector Machine

Corresponding Author:

Gusrianty

Email:

gusrianty@lecturer.pelitaindonesia.ac.

id

Abstrak - Obesitas adalah adalah penyakit kronis yang bersifat kompleks dan dipengaruhi oleh berbagai faktor, baik genetik, lingkungan, maupun gaya hidup yang ditandai dengan kelebihan berat badan akibat akumulasi lemak tubuh yang berlebihan. Seiring dengan pesatnya kemajuan teknologi dan digitalisasi di segala sektor, data kini memegang peranan sangat penting, karena dari kumpulan data menghasilkan informasi yang berguna. Namun terdapat tantangan utama dalam analisis data adalah mengatasi masalah redundansi, noise, dan tingginya dimensi data yang dapat memengaruhi kinerja algoritma machine learning. Penelitian ini bertujuan untuk mengeksplorasi efektivitas kombinasi *Linear Discriminant Analysis (LDA)* dan *Support Vector Machine (SVM)* dalam meningkatkan akurasi dan efisiensi klasifikasi data berdimensi tinggi, khususnya pada prediksi tingkat obesitas. *LDA* digunakan untuk mereduksi dimensi data dengan mempertahankan fitur-fitur yang paling relevan, sedangkan *SVM* diterapkan sebagai algoritma klasifikasi untuk memprediksi tingkat obesitas berdasarkan pola yang ditemukan dalam data. Penelitian dilakukan menggunakan dataset yang terdiri dari 779 data latih dan 195 data uji. Hasil penelitian menunjukkan bahwa kombinasi *LDA* dan *SVM* mampu meningkatkan akurasi klasifikasi hingga 99% dengan pengurangan dimensi data sebesar 50% dengan kecepatan komputasi 0,0696 detik. Selain itu, waktu komputasi berkurang secara signifikan, menunjukkan bahwa *LDA* tidak hanya membantu dalam penyederhanaan data tetapi juga meningkatkan efisiensi proses klasifikasi.

Kata Kunci: Klasifikasi, Linear Discriminant Analysis, Obesitas, Reduksi Dimensi, Support Vector Machine.



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

I. PENDAHULUAN

Obesitas adalah penyakit kronis yang bersifat kompleks dan dipengaruhi oleh berbagai faktor, baik genetik, lingkungan, maupun gaya hidup. Kondisi ini ditandai dengan kelebihan berat badan akibat akumulasi lemak tubuh yang berlebihan. Penyebab utama obesitas adalah ketidakseimbangan energi, yaitu ketika asupan energi dari makanan lebih besar daripada energi yang dikeluarkan melalui aktivitas fisik maupun proses metabolisme. Ketidakseimbangan ini menyebabkan kelebihan energi disimpan dalam bentuk lemak, sehingga berat badan meningkat secara signifikan melebihi kisaran berat badan ideal [1].

Pesatnya kemajuan teknologi dan digitalisasi di semua sektor, data telah menjadi aset yang sangat berharga, karena kumpulan data menghasilkan informasi yang bermanfaat. Salah satu sektor yang mengalami perkembangan yaitu pada sektor kesehatan. Data adalah sesuatu yang belum jelas (fakta mentah), dan harus diolah agar memiliki arti [2]. Data dapat berbentuk angka, teks, atau simbol lainnya, yang semuanya dapat diolah

atau dipahami . Kebenaran dan ketepatan data berdampak pada hasil yang diinginkan[3]. Namun, dengan bertambahnya volume dan dimensi data, tantangan dalam analisis data semakin kompleks, terutama dalam mengatasi masalah *noise*, redundansi, dan dimensi tinggi. Dalam penelitian ini pemanfaatan dataset Kesehatan tentang prediksi obesitas. Analisis data menjadi sangat penting karena melibatkan banyak faktor yang dapat mempengaruhi tingkat obesitas individu. Tantangan utama dalam reduksi dimensi data obesitas adalah mengidentifikasi fitur-fitur yang paling relevan tanpa kehilangan informasi penting. Data obesitas sering kali mencakup berbagai variabel seperti indeks massa tubuh (BMI), aktivitas fisik, dan jenis kelamin. Mengelola dan menganalisis data dengan banyak variabel ini memerlukan teknik yang efektif untuk mengurangi kompleksitas sambil tetap mempertahankan informasi yang berharga.

Salah satu teknik reduksi dimensi yang paling umum digunakan adalah *Linear Discriminant Analysis* (LDA). LDA bekerja dengan memproyeksikan data ke ruang berdimensi lebih rendah yang memaksimalkan pemisahan antara kelas-kelas[4]. Metode LDA akan membentuk sekumpulan dimensi baru yang kemudian diranking berdasarkan varian datanya, sehingga tercipta kumpulan data dengan fitur yang lebih sederhana. LDA dapat menghasilkan representasi yang lebih sederhana dan mudah diinterpretasikan dari data asli melalui proyeksi data ke sumbu-sumbu utama variabilitas.

Dalam konteks ini, kinerja algoritma dalam menangani data obesitas ini menjadi fokus yang penting, terutama dalam konteks klasifikasi. Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dalam memprediksi tingkat obesitas. Kinerja kemampuan algoritma ini untuk menghasilkan prediksi yang akurat, efisien, dan relevan meskipun dihadapkan pada data dengan dimensi yang sangat besar. Algoritma *machine learning* seperti SVM merupakan salah satu metode klasifikasi populer dalam *machine learning*, di mana pendekatan pembelajaran terawasi digunakan untuk memprediksi kelas berdasarkan pola yang diperoleh selama proses pelatihan [5].

Cara kerja SVM melibatkan pencarian hyperplane, yaitu batas pemisah yang memisahkan data dari dua kelas dengan margin terbesar. SVM secara teoritis memiliki kemampuan yang kuat dalam menangani data berdimensi tinggi. Namun, ketika jumlah atribut atau fitur dalam data meningkat secara signifikan, SVM dapat menghadapi masalah kinerja yang serius yang dikenal sebagai *curse of dimensionality*, di mana peningkatan jumlah dimensi dapat menyebabkan performa algoritma menurun [6].

Seiring bertambahnya dimensi data, SVM sering menghadapi masalah kinerja yang tidak konsisten. Hal ini terjadi karena ketika jumlah atribut atau fitur terlalu besar, SVM dapat menghasilkan model yang terlalu kompleks dan kurang bermakna, sehingga memperlambat proses pengolahan data dan menurunkan akurasi klasifikasi sehingga keandalan model tersebut menjadi rendah atau dengan kata lain tidak dapat diandalkan. Ketika jumlah fitur terlalu besar, SVM memerlukan lebih banyak waktu komputasi untuk mencari *hyperplane* optimal, yang dapat memperlambat proses pengolahan data dan meningkatkan risiko *overfitting* [7].

Untuk mengatasi tantangan ini, salah satu pendekatan yang bisa digunakan adalah dengan melakukan reduksi dimensi atribut sebelum menerapkan metode klasifikasi seperti SVM. Reduksi data dapat membantu menyederhanakan analisis dengan mengurangi jumlah kemungkinan nilai yang harus dipertimbangkan [8]. Dalam penelitian ini, penulis menggunakan algoritma LDA dalam proses reduksi data. LDA bekerja dengan menghitung matriks *scatter* antar kelas dan dalam kelas untuk memahami distribusi data berdasarkan label kelas. Setelah itu, dilakukan perhitungan *eigenvector* dan *eigenvalue* dari matriks diskriminan yang menentukan arah terbaik untuk memisahkan kelas. *Eigenvector* dengan *eigenvalue* terbesar dipilih sebagai representasi fitur yang paling informatif untuk klasifikasi. Dengan memilih sejumlah komponen utama yang paling berkontribusi dalam separasi kelas, LDA dapat mereduksi dimensi data tanpa kehilangan informasi penting untuk analisis klasifikasi. LDA lebih unggul dalam tugas klasifikasi karena mampu mengoptimalkan pemisahan antar kelas. Dengan demikian, LDA menjadi pilihan yang tepat untuk meningkatkan efisiensi dan kinerja analisis klasifikasi.

Sebelum menerapkan LDA, dimensi data yang besar dapat direduksi secara signifikan, sehingga mengurangi kompleksitas model yang dibangun oleh SVM. Pemanfaatan LDA membantu memaksimalkan separasi antar kelas dalam data, sehingga fitur yang paling diskriminatif lebih menonjol, dan model yang dihasilkan oleh SVM menjadi lebih akurat serta lebih mudah untuk diinterpretasikan. Pengurangan dimensi ini tidak hanya meningkatkan efisiensi komputasi tetapi juga dapat membantu mengurangi risiko *overfitting*, karena model SVM menjadi lebih fokus pada fitur-fitur yang paling relevan untuk klasifikasi. LDA dapat mengatasi keterbatasan teknik konvensional yang memerlukan waktu komputasi yang lama dan memiliki tingkat kesalahan yang tinggi [9]. penelitian ini memiliki tujuan untuk membandingkan skenario tanpa reduksi dengan LDA menunjukkan bahwa LDA mampu mempercepat waktu komputasi sekaligus meningkatkan akurasi.

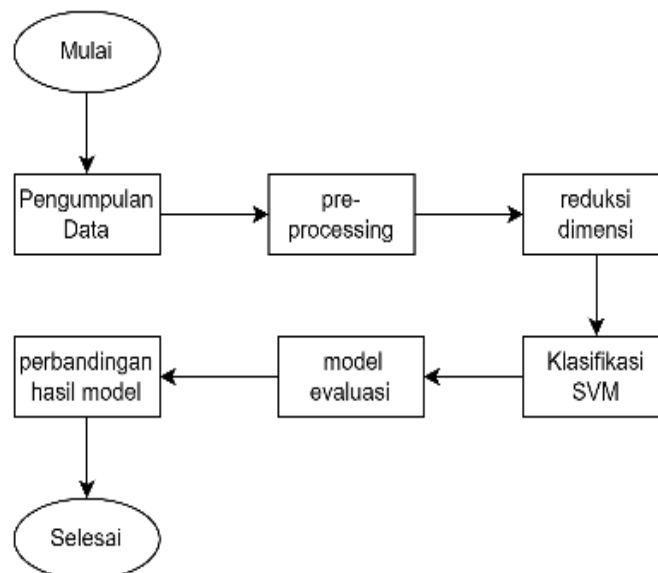
Penelitian sebelumnya telah membahas berbagai kombinasi teknik LDA dengan algoritma *machine learning* lainnya. Misalnya, penelitian [10] yang membahas perbandingan model *Decision Tree*, *Random Forest*, dan LDA dalam prediksi kanker payudara. Penelitian ini berfokus untuk merancang model prediksi kanker payudara menggunakan tiga metode tersebut dan membandingkan kinerjanya berdasarkan akurasi dan kemampuan generalisasi. Sementara itu, [11] dalam penelitiannya mengembangkan sistem deteksi berita palsu

berbahasa Indonesia menggunakan kernel SVM dan n-gram. Tujuannya adalah mengidentifikasi model dengan kinerja terbaik. Hasil penelitian menunjukkan bahwa algoritma SVM kernel linier menghasilkan akurasi terbaik, yaitu 0,974. [12] juga meneliti klasifikasi ekspresi wajah dengan memadukan *Linear Discriminant Analysis* (LDA) dan SVM tiga lipatan (*three-fold SVM*) guna mengurangi label palsu dan meningkatkan presisi klasifikasi. Kombinasi ini digunakan secara khusus untuk mengatasi tantangan dalam pengenalan ekspresi wajah melalui pendekatan gabungan teknik statistik dan machine learning. Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi efektivitas kombinasi LDA dan SVM dalam konteks analisis tingkat obesitas.

Oleh karena itu, penelitian ini difokuskan pada penerapan *Linear Discriminant Analysis* (LDA) untuk meningkatkan kinerja algoritma *Support Vector Machine* (SVM) dalam klasifikasi data berdimensi tinggi. Objek penelitian adalah prediksi tingkat obesitas berdasarkan data kesehatan individu. Pendekatan yang digunakan adalah mereduksi dimensi data menggunakan LDA untuk mempertahankan fitur paling relevan, kemudian menerapkan SVM sebagai metode klasifikasi, dengan tujuan memperoleh model prediksi obesitas yang lebih akurat dan efisien dari sisi komputasi.

II. METODE

Penelitian ini dilakukan dengan menggunakan berbagai metode yang dijelaskan dalam diagram proses penelitian, seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian.

A. Pengumpulan Data

Proses pengumpulan data dilakukan untuk dapat memperoleh informasi yang dibutuhkan selama melakukan penelitian dalam rangka untuk mencapai tujuan yang diinginkan. Dalam penelitian ini, data yang digunakan dikumpulkan dari platform *Kaggle* oleh *Mrsimple07* yang tersedia di tautan : <https://www.kaggle.com/datasets/mrsimple07/obesity-prediction> [13], yang dikenal sebagai sumber data terbuka yang menyediakan berbagai dataset untuk analisis dan penelitian.

B. Data Preprocessing

Tahap ini bertujuan untuk membersihkan data dari *noise* agar memudahkan pemrosesan machine learning, dan juga memastikan kualitas dan keakuratan data [14]. Berikut ini adalah tahapan yang umumnya dilakukan dalam pra-proses data.

1. Data cleaning

Data Cleaning: Proses mengeluarkan data yang tidak konsisten, terduplikasi, atau memiliki nilai yang hilang demi menjaga integritas dataset. Pertama, data duplikat diperiksa menggunakan fungsi “*duplicated()*” dan seluruh baris yang terdeteksi duplikat dihapus untuk menghindari bias. Kedua, data diperiksa adanya *missing value* dengan “*isnull().sum()*”, dan hasilnya tidak ditemukan nilai kosong sehingga tidak diperlukan imputasi. Ketiga, deteksi dan penghapusan *outlier* dilakukan pada fitur numerik (*Age*, *Height*, *Weight*, *BMI*, dan

PhysicalActivityLevel) menggunakan metode *Interquartile Range (IQR)*. Data dengan nilai di bawah *lower bound* ($Q1 - 1.5 \times IQR$) atau di atas *upper bound* ($Q3 + 1.5 \times IQR$) dianggap sebagai *outlier* dan dihapus dari dataset. Keempat, variabel kategorikal seperti *Gender* dan *ObesityCategory* dikonversi ke bentuk numerik menggunakan *Label Encoding*, sehingga dapat diproses oleh algoritma klasifikasi. Dengan tahapan ini, jumlah data berkurang dari 1000 baris menjadi 974 baris, dan dataset siap digunakan pada tahap reduksi dimensi (LDA) serta klasifikasi (SVM).

2. Labeling Data

Labeling Data: Proses pemberian label atau anotasi pada data, khususnya dalam pembelajaran terbimbing, untuk menandai variabel yang menjadi target yang akan diprediksi oleh model.

3. Data Scaling

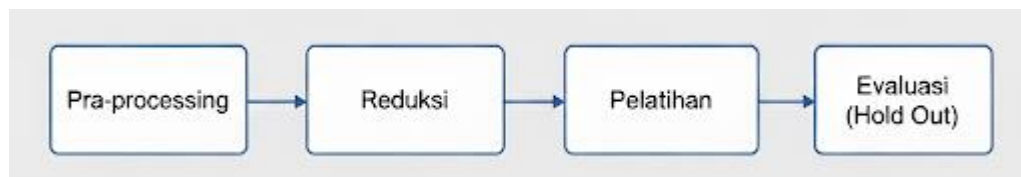
Data Scaling: Proses penskalaan ulang fitur numerik sehingga berada dalam rentang yang sama, dan ini membuat algoritma lebih efektif dan efisien.

C. Reduksi Dimensi

Reduksi dimensi merupakan teknik mengurangi dimensi atau kolom dari data yang digunakan tanpa menghilangkan informasi yang terdapat didalam data tersebut [15]. Pada penelitian ini, dilakukan dua skenario dalam mengevaluasi pengaruh reduksi terhadap kinerja SVM yaitu sebagai berikut.

- 1) Tanpa Reduksi: Model ini dikembangkan menggunakan semua fitur asli dalam kumpulan data tanpa proses reduksi apa pun, sebagai dasar untuk membandingkan perkembangan yang dihasilkan oleh teknik reduksi lainnya.
- 2) Reduksi LDA *Hold-Out* (80:20): Memilih fitur yang memaksimalkan pemisahan antar kelas [16]. Keunggulan LDA adalah kemampuannya untuk memisahkan data dengan cara memaksimalkan jarak antar kelas sambil meminimalkan variasi di dalam setiap kelas [17].

Adapun gambar Diagram Alur dalam memproses 2 skenario di atas Adalah sebagai berikut.



Gambar 2. Diagram Alur Pemrosesan.

Alur tersebut menggambarkan proses standar dalam machine learning, yang diawali dengan tahap pra-proses untuk membersihkan dan mempersiapkan data mentah. Setelah bersih, data tersebut melalui tahap reduksi untuk menyederhanakan jumlah fitur agar lebih efisien dan fokus pada informasi penting. Selanjutnya, model dilatih pada data yang sudah siap ini dalam tahap pelatihan, dan terakhir, performa model diukur secara objektif pada tahap evaluasi menggunakan metode seperti hold-out untuk memastikan keandalannya pada data baru.

D. Klasifikasi SVM

Algoritma pemodelan *Support Vector Machine* adalah algoritma klasifikasi yang bekerja dengan mencari titik (*hyperplanes*) untuk mengklasifikasikan dua set data. Untuk mendapatkan hasil perhitungan SVM dibutuhkan langkah-langkah yang harus dilalui sebagai berikut:

1) Margin,

Margin merupakan jarak antara hyperplane dengan titik data terdekat dari masing-masing kelas, dan margin maksimum menunjukkan pemisahan terbaik antar kelas. Nilai margin ditentukan melalui rumus berikut:

$$\text{Margin} = \frac{1}{w} \quad (1)$$

Di mana w adalah vektor bobot dari *hyperplane*. Semakin kecil nilai w , semakin besar margin yang terbentuk, sehingga model memiliki generalisasi yang lebih baik (lihat Persamaan (1)). Untuk menentukan margin maksimum, digunakan fungsi kernel dalam bentuk berikut:

$$d(X^T) = \sum_{i=1}^l y_i a_i X_1 X^t + b_0 \quad (2)$$

dengan:

$d(X^T)$: margin maksimal
y_i	: kelas label
X^t	: data latih
a_i	: nilai bobot setiap data
b_0	: parameter/bias SVM

Persamaan (2) digunakan untuk menghitung jarak margin maksimum dari *hyperplane* terhadap titik-titik data terdekat.

2) Hyperplane

Hyperplane adalah bidang pemisah linier yang membagi dua kelas. Untuk memperoleh *hyperplane* optimal, digunakan fungsi objektif minimisasi seperti pada Persamaan (3):

$$\frac{1}{2} \|w\|^2 = \frac{1}{2} (w_1^2 + w_2^2) \quad (3)$$

Persamaan (3) digunakan dalam optimasi *margin*, dengan mengurangi norm $\|w\|$ seminimal mungkin. *Hyperplane* secara umum dapat dituliskan dalam bentuk linier sebagai:

$$wx_1 + b = 0 \quad (4)$$

atau dalam dua dimensi sebagai:

$$w_1 x_1 + w_2 x_2 + b = 0 \quad (5)$$

dengan:

w	: bobot <i>margin</i>
x	: titik bias
b	: bias

Persamaan (4) dan (5) mewakili kondisi keseimbangan data pada *hyperplane*.

3) Class

Data latih yang digunakan diklasifikasikan ke dalam dua kelas, yaitu positif dan negatif, berdasarkan posisi relatifnya terhadap *hyperplane*. Untuk setiap titik data x_i , berlaku syarat:

$$y_i (wx_1 + b) \geq 1 \quad i = 1, 2, 3, \dots, N \quad (6)$$

Syarat ini (Persamaan (6)) memastikan bahwa semua data diklasifikasikan dengan *margin* minimal sebesar 1. Maka gabungan dari batas *hyperplane* untuk kelas positif dan negatif adalah:

$$y_i (w_1 x_1 + w_2 x_2 + b) \geq 1 \quad (7)$$

Untuk data dengan label kelas positif persamaan *hyperplane* menjadi:

$$y_i (w_1 x_1 + w_2 x_2 + b) \geq 1 \quad \text{for } y_1 = +1 \quad (8)$$

Sebaliknya, untuk data dengan label kelas negatif:

$$y_i (w_1 x_1 + w_2 x_2 + b) \leq -1 \quad \text{for } y_1 = -1 \quad (9)$$

Persamaan (8) dan (9) membentuk batas pemisah antara dua kelas yang sejajar terhadap *hyperplane* utama, masing-masing merepresentasikan sisi *margin*.

E. Model Evaluasi

Hasil dari model SVM yang menggunakan LDA dan tanpa reduksi dibandingkan untuk mengevaluasi apakah penerapan LDA dapat meningkatkan performa model SVM. Evaluasi model dilakukan dengan menggunakan *confusion matrix*.

F. Perbandingan Hasil Model

Model yang telah dibangun kemudian diuji menggunakan data uji. Performanya dievaluasi dengan menggunakan *confusion matrix* untuk mengukur *accuracy*, *precision*, *recall*, dan metrik lainnya guna mengetahui efektivitas model. Dari ke tiga scenario tersebut maka dilaporkan hasil rata-rata dari setiap model.

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengevaluasi apakah pengurangan dimensionalitas melalui *Linear Discriminant Analysis* dapat meningkatkan kemampuan klasifikasi *Support Vector Machine* tanpa mengorbankan integritas data. Penelitian ini meliputi tahap pengumpulan data (data set), *implementasi preprocessing data*, *implementasi reduksi data*, evaluasi dan sistem prediksi obesitas LDA-SVM *implementasi klasifikasi SVM*

A. Pengumpulan Data

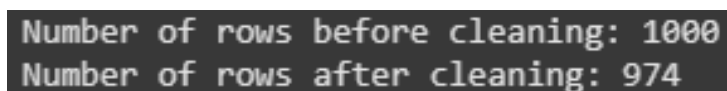
Penelitian ini menggunakan 1000 data, terdiri dari 371 data terkategori normal weight, 191 data terkategori obese, 295 data terkategori *overweight*, dan 143 data terkategori *underweight*. Data obesitas dapat dilihat pada tabel 1.

TABEL 1
 DATASET OBESITAS

Age	Gender	Height	Weight	BMI	PhysicalActivityLevel	ObesityCategory
56	Male	173.5752624	71.98205082	23.89178262	4	Normal weight
69	Male	164.1273058	89.95925553	33.39520945	2	Obese
46	Female	168.0722021	72.93062927	25.81773746	4	Overweight
32	Male	168.4596328	84.88691247	29.91224698	3	Overweight
60	Male	183.5685677	69.0389455	20.48790306	3	Normal weight
25	Female	166.4056273	61.14586775	22.08162832	4	Normal weight
78	Male	183.5663338	92.20852083	27.36434056	3	Overweight
...
...
38	Male	142.8750951	59.35974568	29.07896647	1	Overweight
56	Male	183.4785578	75.15767235	22.32557668	4	Normal weight

B. Implementasi preprocessing data

Langkah pertama dalam penyiapan data adalah *preprocessing data*, yang dilakukan dari kumpulan data yang diperoleh. Selain menjamin keakuratan kumpulan data untuk analisis selanjutnya, pemrosesan data awal ini berupaya meminimalkan gangguan dalam data yang dapat menurunkan kualitasnya. Pembersihan data yang tidak lengkap, duplikasi, *outlier*, dan pembagian kategori semuanya telah dilakukan selama proses ini. Kumpulan data telah berhasil disiapkan untuk analisis selanjutnya. Dari 1000 data setelah *cleaning* menjadi 974 ditunjukkan pada gambar 3. Hasil Preprocessing

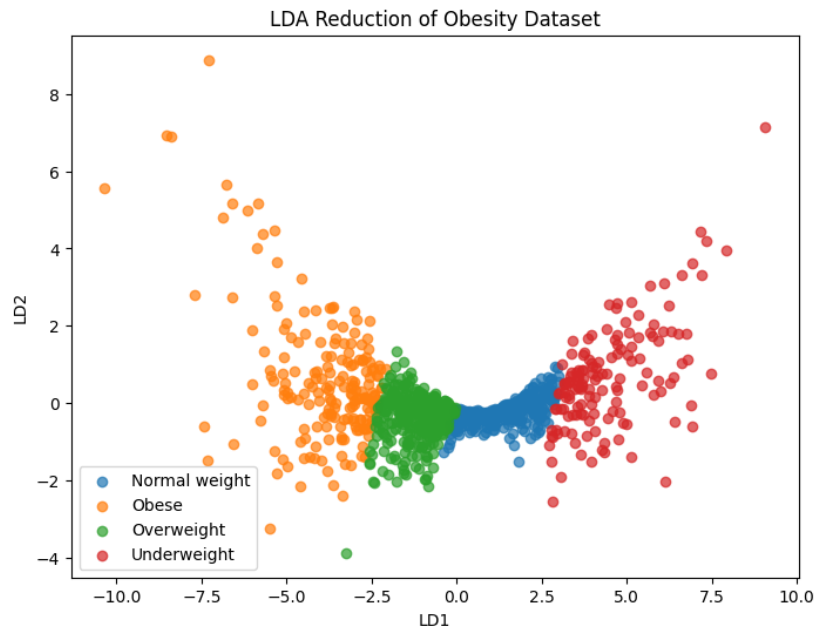


Gambar 3. Hasil *Preprocessing*

C. Implementasi Reduksi Data

Proses ini bertujuan untuk menyederhanakan analisis dengan mengurangi jumlah kemungkinan nilai yang perlu dipertimbangkan. Selama proses Reduksi LDA ini, data dipadatkan menjadi sejumlah kecil komponen diskriminan (LD1, LD2, dll.) yang terutama menjelaskan variasi di antara kategori obesitas. Distribusi data

dalam ruang berdimensi rendah yang baru, di mana setiap titik data yang diklasifikasikan lebih terpisah antara kelas obesitas, ditampilkan dalam visualisasi hasil LDA pada Gambar 4.

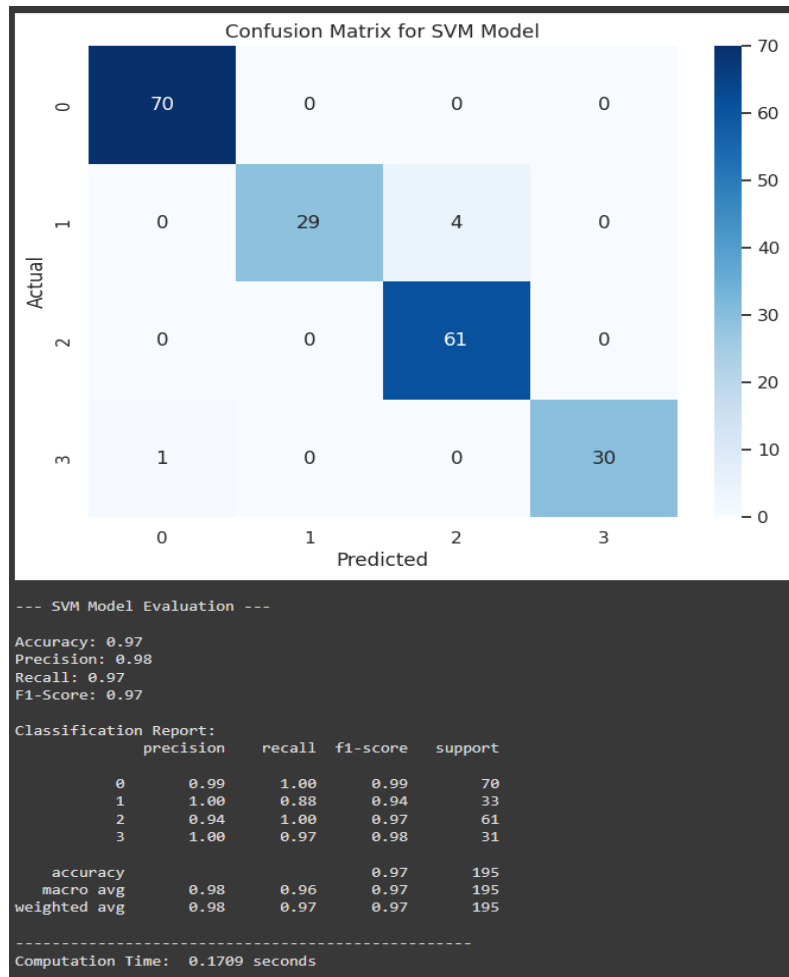


Gambar 4. Visualisasi data setelah reduksi LDA

D. Implementasi Klasifikasi SVM

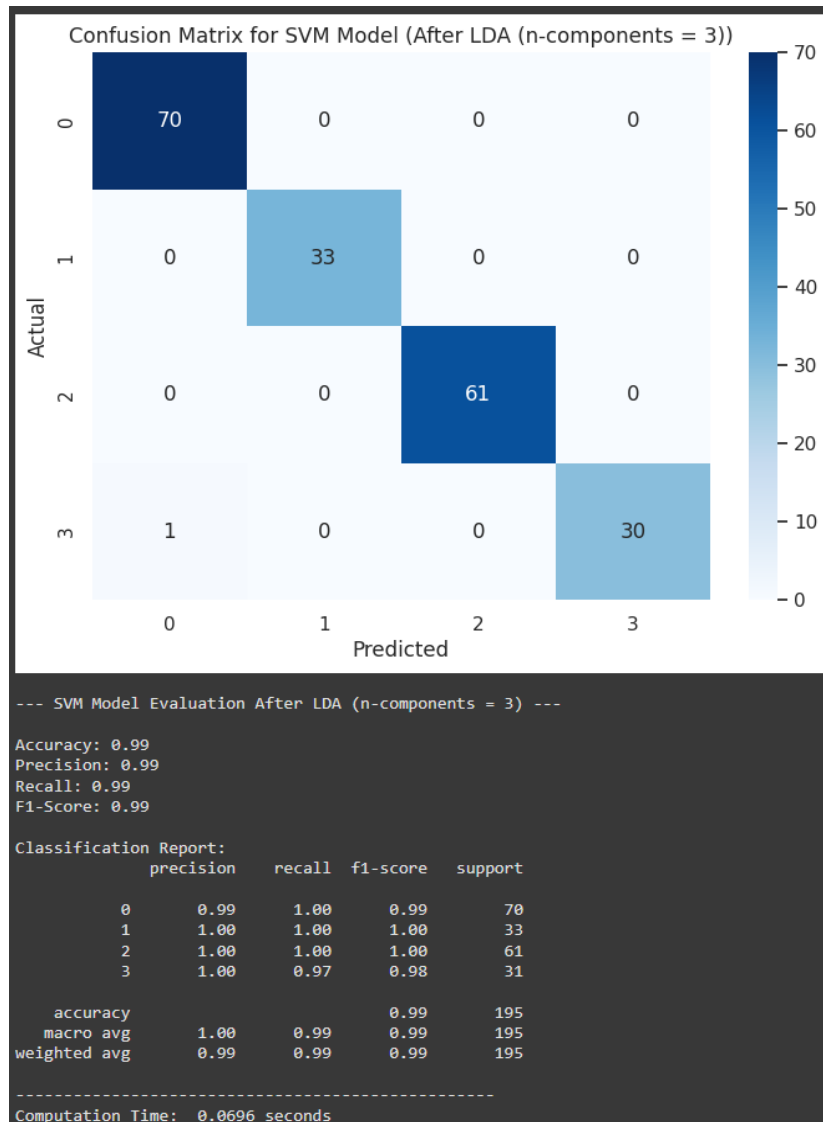
Setelah dataset melalui tahap pembersihan (*cleaning*) dan reduksi dimensi, dilakukan implementasi Algoritma SVM terhadap 2 skenario. Adapun hasil implementasi dari 2 skenario tersebut adalah sebagai berikut.

1. Tanpa Reduksi: Hasil analisis menunjukkan bahwa model SVM, yang dibangun berdasarkan *confusion matrix*, dapat mengklasifikasikan data dengan akurasi 97%, presisi 98%, recall 97%, dan skor F1 sebesar 0,97. menunjukkan bahwa, dengan sangat sedikit kesalahan prediksi, mayoritas data berhasil diklasifikasikan dengan benar di setiap kelas. Lebih jauh, model tersebut tidak bias terhadap satu kelas, sebagaimana dibuktikan oleh nilai presisi, recall, dan skor F1 yang tinggi dan seimbang untuk setiap kelas dan waktu komputasi yang singkat yaitu 0,1789 detik seperti yang ditunjukkan pada Gambar 5. *Confusion matrix SVM tanpa reduksi*



Gambar 5. Confusion matrix SVM tanpa reduksi

2. Reduksi LDA: Performa model SVM meningkat setelah transformasi data menggunakan tiga komponen LDA. Model tersebut menunjukkan akurasi 99%, presisi 99%, recall 99%, dan skor F1 sebesar 0,99 berdasarkan *confusion matrix*. Hampir tidak ada kesalahan prediksi dan hampir semua data dikategorikan dengan tepat di setiap kelas. Model tersebut tidak hanya akurat secara umum tetapi juga seimbang dalam mengklasifikasikan antarkelas, sebagaimana dibuktikan oleh nilai presisi, recall, dan skor F1 yang hampir sempurna di setiap kelas. Lebih jauh, setelah menggunakan LDA, waktu komputasi berkurang menjadi hanya 0,0606 detik, yang menunjukkan peningkatan efisiensi pemrosesan data. Hasilnya, penggunaan LDA telah berhasil meningkatkan performa model dalam hal kecepatan dan akurasi prediksi. Seperti yang terlihat pada Gambar 6.



Gambar. 6 Confusion matrix SVM setelah reduksi LDA

E. Evaluasi

Perbandingan Evaluasi Reduksi LDA, Dalam pengujian yang berbeda, model SVM dievaluasi menggunakan LDA setelah reduksi dimensi dengan mempertahankan beberapa komponen utamanya (1LD, 2LD, dan 3LD). Hasil evaluasi menunjukkan bahwa, meskipun menggunakan lebih sedikit komponen diskriminan, kinerja model SVM tetap tinggi dan stabil. Hal ini menunjukkan bagaimana LDA dapat meningkatkan efisiensi waktu komputasi sambil mempertahankan informasi diskriminatif yang relevan untuk klasifikasi obesitas. Tabel 2 memberikan gambaran umum hasil evaluasi kinerja algoritma dengan berbagai tingkat reduksi LDA.

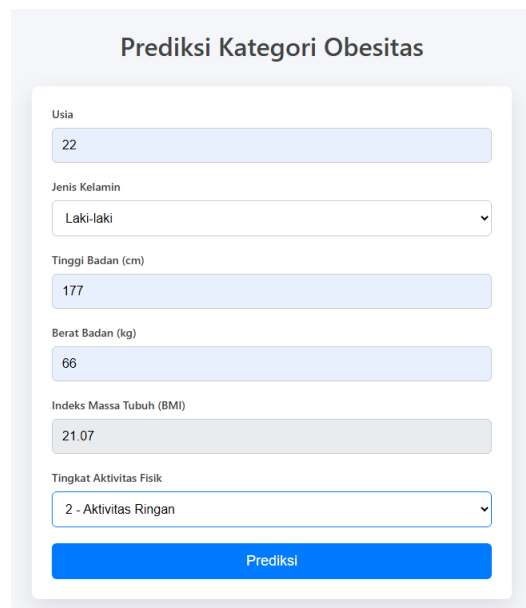
TABEL 2
 EVALUASI PERFORMA ALGORITMA DENGAN LDA

Reduksi	Accuracy	Precision	Recall	F1-Score
3LD	0.99	0.99	0.99	0.99
2LD	0.98	0.98	0.98	0.98
1LD	0.97	0.98	0.97	0.97
70%	0.97	0.98	0.97	0.97

F. Sistem Prediksi Obesitas LDA-SVM

Berdasarkan faktor-faktor seperti usia, jenis kelamin, tinggi badan, berat badan, BMI, dan tingkat aktivitas fisik, dirancang sistem aplikasi web berbasis *Python* untuk memperkirakan kategori obesitas. Aplikasi ini menggunakan algoritma SVM untuk memproses data melalui normalisasi, pengodean kategoris, dan

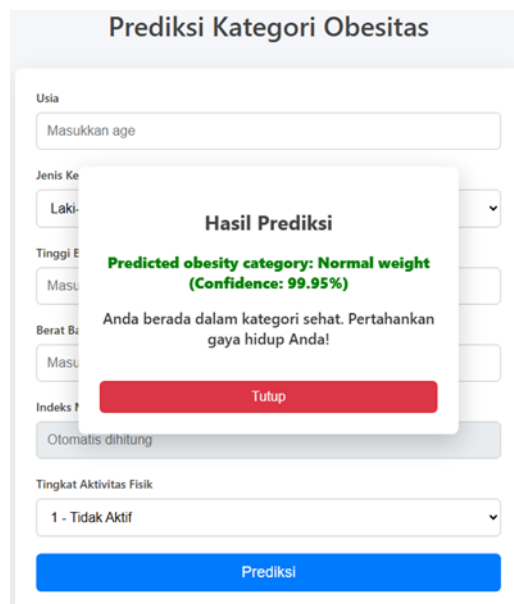
pengurangan dimensionalitas berbasis LDA. Pengguna dapat memasukkan data ke dalam antarmuka pengguna berbasis HTML, dan hasil prediksi yang menampilkan kategori obesitas berdasarkan input ditampilkan seperti yang diilustrasikan pada Gambar 7.



The screenshot shows a web form titled "Prediksi Kategori Obesitas". It contains several input fields: "Usia" with the value 22, "Jenis Kelamin" with a dropdown menu showing "Laki-laki", "Tinggi Badan (cm)" with the value 177, "Berat Badan (kg)" with the value 66, "Indeks Massa Tubuh (BMI)" with the value 21.07, and "Tingkat Aktivitas Fisik" with a dropdown menu showing "2 - Aktivitas Ringan". A blue button labeled "Prediksi" is located at the bottom of the form.

Gambar. 7 Input Data

Pengguna dapat mengklik tombol Prediksi setelah semua data dimasukkan. Setelah itu, hasil prediksi ditampilkan sebagai kartu *pop-up* dengan detail tentang kategori obesitas pengguna dengan mencantumkan tingkat probabilitas. Penggunaan warna yang berbeda untuk setiap kategori—merah untuk obesitas, oranye untuk kelebihan berat badan, hijau untuk berat badan normal, dan biru untuk kekurangan berat badan—membuat tampilan hasil lebih jelas. Selain itu, setiap kategori menyertakan saran kesehatan yang relevan, seperti menyarankan orang yang obesitas untuk meningkatkan aktivitas fisik atau orang yang kekurangan berat badan untuk meningkatkan asupan kalori. Sebagaimana diperlihatkan pada Gambar 8.



The screenshot shows the same web form as in Gambar 7, but with a pop-up window overlaid. The pop-up window is titled "Hasil Prediksi" and contains the text: "Predicted obesity category: Normal weight (Confidence: 99.95%)". Below this, there is a message: "Anda berada dalam kategori sehat. Pertahankan gaya hidup Anda!". A red button labeled "Tutup" is located at the bottom of the pop-up window.

Gambar. 8 Output sistem

Dengan menggunakan teknik *machine learning*, khususnya LDA dan SVM, aplikasi ini memberikan hasil prediksi yang akurat dan tepat waktu. LDA mengurangi dimensi data tanpa mengorbankan informasi penting, sehingga memungkinkan SVM untuk mengklasifikasikan kategori obesitas secara optimal. Secara keseluruhan,

aplikasi ini efektif dan efisien dalam mengidentifikasi kategori obesitas, sehingga mengurangi masalah dan komplikasi terkait obesitas.

F. Uji Black Box Testing Sistem

Metode *Black Box Testing* merupakan pendekatan pengujian yang mengevaluasi fungsionalitas sebuah program berdasarkan spesifikasinya. Dalam proses ini, penguji (*tester*) menyiapkan berbagai skenario masukan data pada setiap formulir untuk memastikan perangkat lunak beroperasi sesuai dengan tujuannya dan memenuhi semua kebutuhan pengguna. Berikut Adalah pengujian *Black Box Testing* terdapat pada sistem

TABEL 3
PENGUJIAN *BLACK BOX TESTING*

No	Skenario Pengujian	Input	Proses Sistem	Output yang Diharapkan	Hasil Uji
1	Input data normal	Umur=25, Tinggi=170cm, Berat=60kg, Aktivitas=Sedang	Sistem menghitung BMI & prediksi obesitas	Hasil: Normal	✓ Berhasil
2	Input data obesitas	Umur=30, Tinggi=165cm, Berat=95kg, Aktivitas=Ringan	Sistem menghitung BMI & prediksi obesitas	Hasil: Obesitas	✓ Berhasil
3	Input data underweight	Umur=20, Tinggi=172cm, Berat=45kg, Aktivitas=Tinggi	Sistem menghitung BMI & prediksi obesitas	Hasil: Underweight	✓ Berhasil
4	Input data overweight	Umur=28, Tinggi=160cm, Berat=70kg, Aktivitas=Sedang	Sistem menghitung BMI & prediksi obesitas	Hasil: Overweight	✓ Berhasil
5	Input field kosong	Tinggi= (tidak diisi), Berat=65kg, Umur=24	Sistem validasi input	Pesan error: "Data tidak boleh kosong"	✓ Berhasil
6	Input format salah	Tinggi=abc, Berat=xyz	Sistem validasi input	Pesan error: "Format data tidak valid"	✓ Berhasil

IV. SIMPULAN

Penerapan *Linear Discriminant Analysis* (LDA) sebagai teknik reduksi dimensi berhasil secara signifikan meningkatkan performa dan efisiensi model *Support Vector Machine* (SVM) di dibandingkan dengan tanpa teknik reduksi. Peningkatan ini terbukti dari pencapaian akurasi, presisi, dan recall yang nyaris sempurna sebesar 99%, yang menunjukkan bahwa model mampu mengklasifikasikan data dengan kesalahan yang sangat minim. Selain itu, efisiensi model juga meningkat drastis, yang ditandai dengan penurunan waktu komputasi menjadi hanya 0,0696 detik. Dengan demikian, dapat disimpulkan bahwa transformasi data menggunakan LDA merupakan metode yang sangat efektif untuk mengoptimalkan model SVM pada kasus ini, baik dari segi kecepatan pemrosesan maupun ketepatan prediksi..

UCAPAN TERIMAKASIH

Penulis mengucapkan rasa terima kasih kepada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Institut Bisnis dan Teknologi Pelita Indonesia atas kesempatan yang diberikan dalam penelitian ini. Penulis juga berterima kasih kepada Ibu Gusrianty, M.Kom., selaku pembimbing, atas arahan, dukungan, serta masukan berharga yang sangat membantu dalam penyusunan jurnal ini. Selain itu rasa terima kasih disampaikan kepada keluarga serta teman-teman atas dukungan moral, semangat, dan motivasi yang diberikan selama proses penelitian dan penulisan jurnal ini.

Meskipun penelitian ini berhasil, terdapat batasan berupa penggunaan satu jenis dataset dengan pembagian data tunggal (*hold-out*). Oleh karena itu, penelitian selanjutnya disarankan untuk menguji model pada dataset yang lebih beragam, termasuk data yang tidak seimbang (*imbalanced*), untuk menguji generalisasi model. Selain itu, penerapan validasi silang seperti *K-Fold* dan proses tuning hyperparameter pada model SVM dianjurkan untuk mendapatkan evaluasi performa yang lebih robust serta menemukan konfigurasi model yang paling optimal.

DAFTAR PUSTAKA

- [1] R. Dianah, E. A. Andari, Elvira Anjani Putri, Cahya Chita Dwinanti, and D. N. Nafisah, "Penyuluhan Cara Mencegah Obesitas Pada Remaja dengan Pola Makan Yang Sehat," *J. Abdimas ADPI Sains dan Teknol.*, vol. 3, no. 3, pp. 41–50, 2022, doi: 10.47841/saintek.v3i3.220.
- [2] O. Dammann, "Data, Information, Evidence, and Knowledge: A Proposal for Health Informatics and Data Science," *Online J. Public Health Inform.*, vol. 10, no. 3, Mar. 2019, doi: 10.5210/ojphi.v10i3.9631.
- [3] Willy Fernando, D. Jollyta, Dadang Priyanto, and Dwi Oktarina, "The Influence Of Data Categorization And Attribute Instances Reduction Using The Gini Index On The Accuracy Of The Classification Algorithm Model," *J. Ilm. Kursor*, vol. 12, no. 3, pp. 111–122, May 2024, doi: 10.21107/kursor.v12i3.372.

- [4] A. Setiawan and Sumijan, "Penerapan Metode *Linear Discriminant Analysis* Dalam Mendeteksi Kematangan Buah Tomat," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 6, no. 1, pp. 1–11, Jan. 2025, doi: <https://doi.org/10.30645/kesatria.v6i1.539.g534>.
- [5] S. Nur Rismanah, R. Astuti, and F. M. Basysyar, "Penerapan Algoritma *Support Vector Machine* Dalam Menganalisis Sentimen Ulasan Pelanggan Shopeefood Berdasarkan Twitter," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 406–412, Feb. 2024, doi: 10.36040/jati.v8i1.8401.
- [6] V. Berisha *et al.*, "Digital medicine and the curse of dimensionality," *npj Digit. Med.*, vol. 4, no. 1, p. 153, Oct. 2021, doi: 10.1038/s41746-021-00521-5.
- [7] S. Abimanyu, N. Bahtiar, and E. Adi Sarwoko, "Implementasi Metode *Support Vector Machine* (SVM) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) untuk Klasifikasi Depresi," *J. Masy. Inform.*, vol. 14, no. 2, pp. 146–158, 2023, doi: 10.14710/jmasif.14.2.59513.
- [8] D. Jollyta, Prihandoko, A. Hajjah, E. Haerani, and M. Siddik, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish, 2023.
- [9] M. H. Ramdani, I. G. P. S. Wijaya, and R. Dwiyanaputra, "Optimalisasi Pengenalan Wajah Berbasis Linear Discriminant Analysis Dan K-Nearest Neighbor menggunakan Particle Swarm Optimization," *J. Teknol. Informasi, Komput. dan Apl.*, vol. 4, Mar. 2022, [Online]. Available: <http://jtika.if.unram.ac.id/index.php/JTIKA/>
- [10] R. Wang, "Comparison of Decision Tree, Random Forest and *Linear Discriminant Analysis* Models in Breast Cancer Prediction," *J. Phys. Conf. Ser.*, vol. 2386, no. 1, p. 012043, Dec. 2022, doi: 10.1088/1742-6596/2386/1/012043.
- [11] D. Jollyta, G. Gusrianty, P. Prihandoko, and D. Sukrianto, "N-gram and Kernel Performance Using *Support Vector Machine* Algorithm for Fake News Detection System," *Ilk. J. Ilm.*, vol. 15, no. 3, pp. 398–404, Dec. 2023, doi: 10.33096/ilkom.v15i3.1770.398-404.
- [12] J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Facial expressions classification and false label reduction using LDA and threefold SVM," *Pattern Recognit. Lett.*, vol. 139, pp. 166–173, Nov. 2020, doi: 10.1016/J.PATREC.2017.06.021.
- [13] MrSimple, "Obesity prediction," *Kaggle*, 2024. <https://www.kaggle.com/datasets/mrsimple07/obesity-prediction> (accessed Jan. 09, 2025).
- [14] Y. R. Wulan, N. Susanto, A. Larasati, and V. E. Darmawan, "The Sentiment Analysis of User Perception on The Peduli Lindungi Application Using *Support Vector Machine* Algorithm," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Michigan, USA: IEOM Society International, Sep. 2022, pp. 832–842. doi: 10.46254/AP03.20220165.
- [15] A. Dewan, D. Wibiyanto, and A. Wibowo, "Penerapan Algoritma Multiclass *Support Vector Machine* dan TF-IDF Untuk Klasifikasi Topik Tugas Akhir," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 1, pp. 42–51, 2023.
- [16] D. Jollyta, A. Hajjah, E. Haerani, and M. Siddik, *Algoritma Klasifikasi untuk Pemula Solusi Python dan RapidMiner*. Deepublish, 2023. [Online]. Available: <https://books.google.co.id/books?id=y84TEQAAQBAJ>
- [17] R. Destriana, D. Numaningsih, D. Alamsyah, and A. A. J. Sinlae, "Implementasi Metode *Linear Discriminant Analysis* (LDA) Pada Klasifikasi Tingkat Kematangan Buah Nanas," *Build. Informatics, Technol. Sci.*, vol. 3, no. 1, pp. 56–63, 2021, doi: 10.47065/bits.v3i1.1007.