

Optimasi Metode Support Vector Machine (SVM) Menggunakan Particle Swarm Optimization pada Permasalahan Klasifikasi Diabetes

Anak Agung Gde Agung Pranandita^{a1}, I Made Widiartha^{a2}

^aProgram Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana

Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia

¹Pranandita.2308561106@student.unud.ac.id

² madewidiartha@unud.ac.id

Abstract

Diabetes mellitus is a chronic disease that requires accurate early detection. This study presents a diabetes classification system by integrating Support Vector Machine (SVM) with Particle Swarm Optimization (PSO) to automatically optimize model parameters. The dataset used was obtained from Kaggle, consisting of 100,000 entries and nine medical attributes. Data preprocessing included cleaning, encoding, Min-Max normalization, and undersampling to balance class distribution. Model performance was evaluated using 5-Fold Cross Validation. The results showed that the SVM- PSO achieved an average accuracy of 83.60% which is higher than the conventional SVM with 83.39% accuracy. These findings demonstrate that PSO effectively enhances the classification performance of SVM and is recommended for machine learning-based medical diagnosis, especially in diabetes prediction.

Keywords: *Diabetes, Support Vector Machine, Particle Swarm Optimization, Parameter Optimization, Medical Classification*

1. Pendahuluan

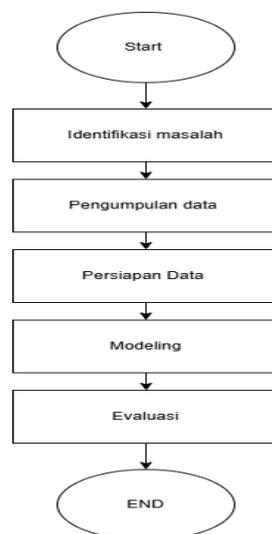
Penyakit tidak menular seperti diabetes mellitus merupakan tantangan serius bagi sistem kesehatan global. Diabetes, sebagai penyakit metabolik kronis, ditandai oleh gangguan produksi atau efektivitas insulin sehingga menyebabkan peningkatan kadar glukosa darah yang berkelanjutan. Dampaknya mencakup komplikasi serius seperti kardiovaskular, gangguan ginjal, neuropati, dan retinopati. Deteksi dini dan akurasi diagnosis menjadi aspek krusial dalam pengelolaan penyakit ini [1]. Seiring perkembangan teknologi informasi, pemanfaatan *machine learning* dalam diagnosis medis telah menunjukkan performa yang menjanjikan. Salah satu algoritma yang banyak digunakan adalah Support Vector Machine (SVM) karena kemampuannya dalam menangani data berdimensi tinggi serta memberikan hasil klasifikasi yang akurat. Namun, efektivitas SVM sangat bergantung pada pemilihan parameter yang optimal, seperti nilai C, gamma, serta jenis fungsi kernel yang digunakan. Ketidak tepatan dalam pemilihan parameter tersebut dapat menyebabkan penurunan akurasi model [2]. Untuk mengatasi kelemahan tersebut, algoritma optimasi seperti Particle Swarm Optimization (PSO) dapat digunakan. PSO merupakan *metaheuristic* berbasis populasi yang terinspirasi dari perilaku sosial kawanan burung, efektif dalam pencarian solusi global optimum pada ruang parameter yang kompleks. Pada penelitian yang dilakukan oleh Maulana *et al.* [3] mengatakan bahwa Kombinasi Particle Swarm Optimization (PSO) dengan Support Vector Machine (SVM) telah terbukti menjadi pendekatan yang efektif untuk meningkatkan kinerja klasifikasi. Dalam berbagai penelitian Keunggulan utama PSO adalah kemampuannya untuk menjelajahi ruang pencarian parameter secara efisien guna menemukan kombinasi yang lebih optimal dibandingkan menggunakan parameter default. Optimalisasi ini secara konsisten menghasilkan peningkatan pada berbagai metrik evaluasi, termasuk akurasi, presisi, recall, dan F1-score. Berbagai penelitian telah menggabungkan PSO dengan SVM untuk meningkatkan akurasi klasifikasi, di antaranya pada diagnosis penyakit jantung, ginjal kronis, dan diabetes. Kombinasi ini terbukti mampu

meningkatkan metrik evaluasi model seperti akurasi, presisi, recall, dan F1-score, bahkan hingga lebih dari 5–7% dibandingkan metode SVM konvensional. Keunggulan PSO terletak pada kemampuannya untuk menghindari jebakan lokal optimum yang sering menjadi masalah pada metode pencarian tradisional. Sejumlah studi sebelumnya telah membuktikan efektivitas PSO dalam mengoptimalkan kinerja model klasifikasi. Purwaningsih dan Nurelasari [4] menunjukkan bahwa penerapan PSO pada metode SVM dalam klasifikasi penyakit ginjal kronis mampu meningkatkan akurasi dari 92,25% menjadi 99,50%, atau terjadi peningkatan sebesar 7,25% setelah dilakukan optimasi parameter. Hal ini membuktikan bahwa PSO efektif dalam membantu pemilihan parameter optimal dan seleksi fitur yang berdampak langsung pada performa model klasifikasi. Selain itu, PSO juga mampu meningkatkan nilai precision dan recall secara signifikan. Dengan hasil-hasil tersebut, PSO menjadi salah satu pendekatan yang unggul dan banyak digunakan dalam optimasi algoritma machine learning, termasuk dalam domain klasifikasi penyakit medis lainnya seperti kanker dan jantung. Keunggulan utama PSO terletak pada kemampuannya menjelajahi ruang pencarian parameter secara efisien dan menghindari jebakan lokal optimum, yang sering kali menjadi kendala dalam metode optimasi konvensional. Dalam penelitian ini, dilakukan pengembangan sistem klasifikasi penyakit diabetes berbasis kombinasi PSO-SVM, dengan tujuan utama untuk mengoptimasi parameter SVM secara otomatis guna memperoleh kinerja klasifikasi terbaik. Dataset yang digunakan adalah dataset penyakit diabetes dari Kaggle, yang mencakup berbagai parameter medis seperti kadar glukosa, tekanan darah, insulin, BMI, hingga jumlah kehamilan. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem deteksi dini diabetes berbasis kecerdasan buatan, serta menunjukkan efektivitas PSO dalam meningkatkan performa SVM.

2. Metode Penelitian

2.1. Tahap penelitian

Penelitian ini terdiri dari lima tahapan utama yang digunakan untuk menyelesaikan permasalahan yang dianalisis. Tahapan-tahapan tersebut meliputi: perumusan masalah, pengumpulan data, pengolahan data, pembangunan model, serta evaluasi kinerja model. Seluruh rangkaian proses penelitian ini divisualisasikan pada Gambar 1.



Gambar 1. Tahap Penelitian

a. Identifikasi masalah

Penelitian ini membandingkan kinerja dua algoritma, yaitu Support Vector Machine (SVM) dan Particle Swarm Optimization (PSO), dalam klasifikasi penyakit diabetes. SVM dikenal efektif untuk data berdimensi tinggi, namun kinerjanya sangat bergantung pada pemilihan parameter seperti C dan γ . Pemilihan parameter secara manual atau grid search seringkali tidak efisien dan kurang optimal. PSO digunakan untuk mengoptimalkan parameter SVM

secara otomatis, namun PSO standar masih memiliki kelemahan seperti konvergensi prematur dan jebakan solusi lokal. Oleh karena itu, perlu dievaluasi efektivitas PSO dalam meningkatkan akurasi dan generalisasi model SVM untuk klasifikasi diabetes.

b. Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Diabetes Dataset, yang diperoleh melalui situs web Kaggle [5]. Dataset ini tersedia dalam format comma-separated values (.csv) dan terdiri atas 100.000 entri sebelum dilakukan pra-proses. Dataset ini memuat sembilan atribut fitur, yaitu: Gender, Age, Hypertension, Heart Disease, Smoking History, BMI, HbA1c_level, Blood Glucose Level, dan label target Diabetes. Mengingat distribusi kelas yang tidak seimbang (91.500 data non-diabetes dan 8.500 data diabetes), dilakukan proses undersampling untuk menyeimbangkan jumlah antar kelas. Setelah proses ini, diperoleh total 16.964 entri data seimbang yang digunakan untuk pelatihan dan pengujian model. Visualisasi atribut disajikan dalam Tabel 1.

Tabel 1. Atribut Dataset Penelitian

No	Atribut	Tipe	Deskripsi
1	Gender	Kategorikal	Jenis kelamin merupakan salah satu faktor yang memengaruhi risiko terkena diabetes. Perempuan dengan riwayat diabetes gestasional memiliki kemungkinan lebih besar mengalami diabetes tipe 2, meskipun sejumlah studi juga menemukan bahwa pria cenderung memiliki risiko yang sedikit lebih tinggi.
2	Age	Numerik	Usia individu dinyatakan dalam satuan tahun. Seiring bertambahnya usia, risiko terkena diabetes cenderung meningkat, khususnya pada kelompok usia lanjut.
3	Hypertension	Numerik	Menyatakan kondisi tekanan darah tinggi pada subjek. Hipertensi dan diabetes tipe 2 memiliki keterkaitan yang erat, di mana masing-masing dapat meningkatkan risiko terjadinya yang lain sebagai bagian dari sindrom metabolik.
4	Heart Disease	Numerik	Menunjukkan apakah subjek memiliki riwayat penyakit jantung. Penyakit jantung berkorelasi dengan risiko diabetes dan komplikasi kardiovaskular.
5	Smoking History	Kategorikal	Riwayat merokok subjek (seperti "never", "former", "current", "No Info"). Merokok merupakan faktor risiko yang memperburuk resistensi insulin dan komplikasi diabetes.
6	BMI	Numerik	Indeks massa tubuh. BMI tinggi menunjukkan obesitas, yang secara signifikan meningkatkan risiko diabetes tipe 2.
7	HbA1c_level	Numerik	Kadar HbA1c (persentase) sebagai indikator kadar glukosa darah rata-rata dalam 2–3 bulan terakhir. Nilai >6.5% sering dijadikan sebagai ambang diagnosis diabetes.
8	Blood Glucose Level	Numerik	Tingkat glukosa dalam darah (mg/dL). Kadar tinggi menunjukkan hiperglikemia, gejala utama dari diabetes.
9	Diabetes	Numerik (Target)	Menunjukkan apakah subjek menderita diabetes (1) atau tidak (0). Digunakan sebagai label dalam pemodelan klasifikasi.

c. Persiapan Data

Dalam penelitian ini Proses persiapan data dilakukan dalam empat tahap utama. Tahap pertama adalah pembersihan data untuk menghapus nilai yang tidak valid atau tidak konsisten. Selanjutnya dilakukan proses encoding serta penyesuaian data guna

mengonversi data bertipe string dan menangani ketidakseimbangan data. Tahap ketiga adalah normalisasi, yang bertujuan untuk menyamakan skala nilai setiap atribut. Terakhir, data dibagi menjadi dua bagian: data pelatihan untuk membangun model dan data pengujian untuk mengevaluasi performa model tersebut.

d. Data Cleaning

Untuk memastikan kualitas data yang akan digunakan, tahap awal dilakukan dengan menelusuri dataset secara teliti guna memastikan tidak terdapat nilai kosong pada atribut mana pun. Selain itu, dilakukan pengecekan terhadap keberadaan entri yang berulang, yang kemudian dihapus agar tidak memengaruhi hasil analisis.

e. Encoding Penyesuaian Data

Tahap encoding dan penyesuaian data bertujuan untuk mengubah nilai kategori yang berbentuk string menjadi representasi numerik, dengan memanfaatkan metode seperti one-hot encoding atau label encoding. Di samping itu, proses ini juga mencakup penanganan ketidakseimbangan distribusi data antar kelas, misalnya dengan menerapkan teknik *oversampling* atau *undersampling* guna memastikan proporsi data yang lebih seimbang.

f. Normalisasi

Normalisasi Min-Max merupakan metode transformasi linier yang diterapkan pada data mentah. Tujuan utama dari proses ini adalah untuk mengubah nilai setiap variabel agar berada dalam skala yang konsisten, yaitu antara 0 hingga 1. Dengan demikian, seluruh variabel memiliki kontribusi yang seimbang dalam perhitungan tingkat kemiripan (*similarity*) [7]. Proses normalisasi ini dapat dijelaskan melalui persamaan matematika berikut:

$$x' = \frac{x - \text{nilai}_{\min}}{\text{nilai}_{\max} - \text{nilai}_{\min}} \quad (1)$$

Keterangan:

X = merupakan data perkolom

nilai_{\min} = minimum nilai data perkolom

nilai_{\max} = maksimal nilai data perkolom

Selain normalisasi, penanganan ketidak seimbangan kelas juga menjadi tahap penting dalam preprocessing data, khususnya dalam konteks klasifikasi biner seperti pada diagnosis penyakit. Dalam penelitian ini, digunakan teknik *undersampling* untuk mengatasi distribusi kelas yang tidak seimbang. Teknik ini bekerja dengan mengurangi jumlah sampel dari kelas mayoritas sehingga proporsi antara kelas positif dan negatif menjadi lebih seimbang. Di kutip dari Penelitian yang dilakukan oleh Bekkar et al. (2021) menunjukkan bahwa *undersampling* memberikan kinerja yang lebih stabil pada model SVM dan mengurangi kompleksitas pelatihan secara signifikan [6].

g. Pembagian Data

Dataset yang digunakan dalam proses machine learning terlebih dahulu dipisahkan menjadi dua subset, yaitu data pelatihan dan data pengujian. Pembagian ini dilakukan dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian, guna memastikan model dapat belajar dari sebagian besar data sekaligus dievaluasi performanya menggunakan data yang belum pernah dilihat sebelumnya.

2.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma supervised learning yang efektif untuk klasifikasi, termasuk diagnosis penyakit seperti diabetes. SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan dua kelas data dengan margin terbesar [8]. Untuk data yang tidak separabel secara linear, SVM menggunakan *kernel trick* untuk memetakan data ke

ruang berdimensi lebih tinggi . Secara matematis Support Vector Machine (SVM) memecahkan permasalahan optimasi secara matematis sebagai berikut :

$$\frac{1}{2} ||w||^2 \text{ dengan kendala } y_i (w^t x_i + b) \geq 1 \quad (2)$$

Keterangan:

- w = vector bobot
- b = bias
- x_i = vector fitur
- $y_i \in \{-1, +1\}$ = Label kelas

Proses klasifikasi dengan Support Vector Machine (SVM) dilakukan melalui serangkaian tahapan sistematis yang dapat dijelaskan sebagai berikut [9]:

- a. Proses klasifikasi dengan Support Vector Machine (SVM) dilakukan melalui serangkaian tahapan sistematis yang dapat dijelaskan sebagai berikut:

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\} \quad (3)$$

- b. Membangun fungsi pemisah linear:

$$f(x) = w^T x + b \quad (4)$$

- c. Optimasi margin

Dengan meminimalkan $\frac{1}{2} ||w||^2$ dan memenuhi kendala $y_i (w^t x_i + b) \leq 1$ (5)

- d. Soft margin Svm

Untuk mengatasi data yang tidak sepenuhnya dapat dipisahkan secara linear, ditambahkan slack variable ξ_i dan parameter regulasi C

$$\frac{1}{2} ||w||^2 + C \sum_{i=1}^n \xi_i, y_i (w^t x_i + b) \leq 1 \quad (6)$$

2.3. Penggunaan kernel terbaik untuk di optimasi

a. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) adalah algoritma optimasi yang terinspirasi dari perilaku sosial kawanan burung dalam mencari makanan. Diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995, PSO efektif digunakan untuk menemukan solusi optimal pada berbagai permasalahan, termasuk tuning parameter dalam model Support Vector Machine (SVM). Setiap solusi dalam PSO disebut partikel, yang bergerak dalam ruang pencarian berdasarkan dua hal: pengalaman terbaiknya sendiri (*personal best/pbest*) dan terbaik dari seluruh populasi (*global best/gbest*). Posisi dan kecepatan partikel diperbarui secara iteratif dengan rumus:

$$v_i = w \cdot v_i + c_1 \cdot r_1 \cdot (pbest_i - x_i) + c_2 \cdot r_2 \cdot (gbest - x_i) \quad (7)$$

$x_{i+1} = x_i + v_i$

Pada Penelitian ini PSO digunakan untuk mencari parameter Optmal SVM (seperti C dan gamma) guna meningkatkan akurasi klasifikasi diabetes. Parameter C mengatur keseimbangan antara margin maksimal dan kesalahan klasifikasi, sedangkan γ menentukan seberapa jauh pengaruh satu data terhadap pembentukan batas keputusan. PSO bekerja dengan membentuk populasi partikel, PSO bekerja dengan membentuk sekelompok solusi yang disebut partikel. Masing-masing partikel bergerak dalam ruang pencarian untuk menemukan solusi terbaik berdasarkan pengalaman pribadi dan pengalaman kolektif. Algoritma PSO dalam konteks optimasi parameter SVM dijalankan melalui urutan langkah-langkah sebagai berikut [10]:

- Inisialisasi Populasi Partikel
 Tentukan jumlah partikel lalu inisialisasi posisi awal x_i dan kecepatan v_i secara acak. Setiap partikel merepresentasikan kombinasi parameter dari Support Vector Machine (SVM):

$$x_i(C, \gamma) \tag{8}$$

- Evaluasi Fitness
 Hitung fitness setiap partikel berdasarkan performa akurasi model SVM pada parameter x_i menggunakan validasi silang (cross-validation).

- Update Personal Best (pbest)
 Simpan posisi terbaik tiap partikel secara individu:

$$Pbest_i = \arg \max \{fitness_{i,1}, fitness_{i,2} \dots \dots \} \tag{9}$$

- Update Global Best (gbest)
 Tentukan partikel dengan fitness terbaik secara global:

$$gbest_i = \arg \max_i \{fitness_i\} \tag{10}$$

- Update Kecepatan dan Posisi Partikel

$$v_i(t+1) = wv_i(t) + c_1 \cdot r_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot r_2 \cdot (gbest - x_i(t)) \tag{11}$$

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

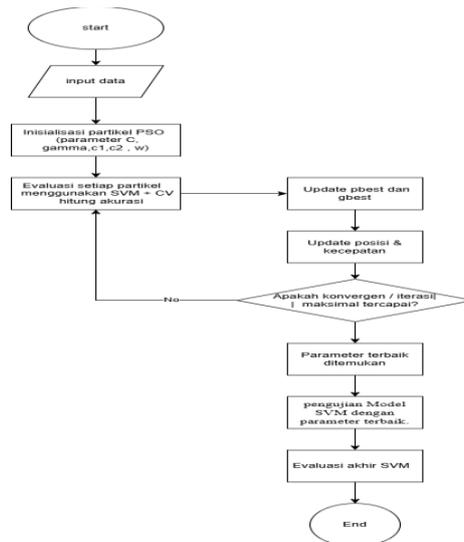
Keterangan:

w = Faktor inersia
 $c_1 c_2$ = koefisien percepatan
 $r_1 r_2$ = bilangan acak [0,1]

- Iterasi
 Ulangi langkah 2- 5 hingga jumlah iterasi maksimum tercapai atau Solusi konvergen.

b. Hibridisasi PSO-SVM

Dalam penelitian ini, dilakukan hibridisasi antara PSO dan SVM, di mana PSO berperan sebagai metode pencarian parameter terbaik untuk SVM. Tiap partikel dalam PSO mewakili pasangan parameter SVM C, γ dan fitness-nya ditentukan berdasarkan akurasi klasifikasi. Setelah iterasi selesai, parameter terbaik (global best) digunakan untuk membangun model SVM akhir. Untuk menggambarkan alur proses ini secara visual, disediakan flowchart hibridisasi PSO-SVM pada Gambar 2.



Gambar 2. Flowchart Hybridisasi PSO-SVM

3. Hasil dan Diskusi

3.1 Proses Data

Dataset yang digunakan terlebih dahulu diperiksa untuk memastikan tidak terdapat nilai yang hilang. Hasil pemeriksaan menunjukkan bahwa tidak ada nilai yang hilang, namun ditemukan adanya data yang duplikat. Oleh karena itu, data duplikat tersebut dihapus dari dataset. Pada tahap awal, dilakukan proses encoding untuk menangani atribut yang masih bertipe string dalam dataset, seperti pada atribut *gender* dan *smoking_history*. Nilai-nilai kategori ini diubah ke dalam format numerik menggunakan metode one hot encoding, agar dapat diproses oleh model machine learning secara optimal. Proses ini menghasilkan representasi biner untuk setiap kategori yang relevan dan memastikan bahwa data bersifat numerik sepenuhnya. Selanjutnya, dilakukan analisis terhadap distribusi kelas pada data target. Hasil analisis menunjukkan adanya ketidak seimbangan jumlah data antar kelas yang cukup signifikan. Untuk mengatasi permasalahan ini, diterapkan teknik undersampling guna menyamakan jumlah sampel antara kelas mayoritas dan minoritas. Setelah proses undersampling, diperoleh distribusi data yang seimbang dengan total masing-masing 4.000 data untuk kelas 0 dan kelas 1. Setelah distribusi data diseimbangkan, tahap selanjutnya adalah normalisasi fitur. Normalisasi dilakukan menggunakan teknik Min-Max Normalization untuk menjaga konsistensi skala antar fitur, sehingga model tidak bias terhadap atribut yang memiliki nilai numerik lebih besar. Setelah normalisasi, dataset dibagi menjadi data latih dan data uji dengan rasio 80%:20%. Pembagian ini bertujuan untuk menyediakan data pelatihan yang cukup bagi model, sekaligus memastikan adanya data uji independen untuk mengevaluasi performa model secara objektif [7].

3.2 Pengujian Model

a. Support Vector Machine (SVM)

Metode support vector machine memiliki beberapa jenis kernel yang dapat digunakan. Pada penelitian ini terdapat pengujian dengan 4 kernel yaitu LINEAR, POLY, RBF, SIGMOID.

Tabel 2. Hasil pegujian kernel

Kernel	Akurasi
LINEAR	0.8822
POLY	0.8835

Kernel	Akurasi
RBF	0.8810
SIGMOID	0.5911

Berdasarkan hasil akurasi kernel pada tabel 2 diatas kernel Polynomial memberikan akurasi tertinggi sebesar 0.8835, sedikit lebih unggul dibanding Linear 0.8822 dan RBF 0.8810. Sementara itu, kernel Sigmoid menunjukkan performa terendah dengan akurasi 0.5911. Dengan demikian, kernel Polynomial dipilih sebagai kernel terbaik karena mampu menangkap hubungan non-linear antar fitur dan memberikan akurasi paling optimal pada dataset yang digunakan. Setelah menjadi kernel terbaik kernel polynomial diuji Kembali dengan metode 5-fold cross-validation. Nilai setiap fold nya dapat dilihat pada tabel 3.

Tabel 3. Nilai dari setiap fold support vector machine (SVM)

Fold	Rentang Index Data	Akurasi
Fold 1	1 – 1599	0.8256
Fold 2	1600 – 3199	0.8331
Fold 3	3200 – 4799	0.8931
Fold 4	4800 – 6399	0.8037
Fold 5	6400 – 7999	0.8137
Rata-rata (5-Fold)	0.8339	

Dari hasil pengujian di setiap foldnya kernel Polynomial mendapatkan hasil rata-rata 83,19% dimana Kernel polynomial mampu menangani pola data non-linear dengan baik, sehingga sangat cocok digunakan pada kasus klasifikasi diabetes yang memiliki kompleksitas hubungan antar fitur. Hasil ini menunjukkan bahwa pemilihan kernel yang tepat berpengaruh terhadap kinerja dari model SVM.

b. Particle Swarm Optimization (PSO)

Setelah dilakukan proses optimasi parameter Support Vector Machine (SVM) menggunakan algoritma Particle Swarm Optimization (PSO), diperoleh nilai parameter terbaik berupa $C = 49.3643$ dan $\gamma = 0.9496$ Kombinasi ini kemudian digunakan untuk melakukan pengujian dan pelatihan dari model particle Particle Swarm Optimization (PSO) menggunakan skema 5-fold cross-validation Hasil evaluasi menunjukkan bahwa akurasi model bervariasi antar fold, dengan nilai terendah 81,70% dan tertinggi 89,92%, serta rata-rata akurasi mencapai 83,99%. Fold ketiga mencatatkan akurasi tertinggi, yang mengindikasikan bahwa model berhasil mengidentifikasi pola dalam data pelatihan dengan baik pada pembagian tersebut. Meskipun ada sedikit perbedaan antar fold, hasil ini menunjukkan bahwa model SVM yang dioptimalkan dengan PSO memiliki performa klasifikasi yang stabil dan konsisten. nilai dari setiap fold dapat dilihat pada tabel 5.

Tabel 4. Nilai dari setiap fold Particle Swarm Optimization (PSO)

Fold	Rentang Index Data	Akurasi
Fold 1	1– 1599	0.8244
Fold 2	1600 – 3199	0.8331
Fold 3	3200 – 4799	0.8994
Fold 4	4800 – 6399	0.8081
Fold 5	6400 – 7999	0.8150

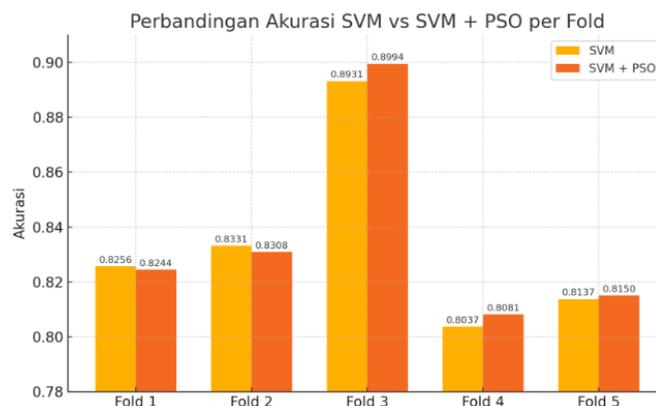
Fold	Rentang Index Data	Akurasi
Rata-rata (5-Fold)	0.8360	

3.3 Akurasi Perbandingan

Tabel 5. Tabel hasil akurasi perbandingan SVM dan SVM+ PSO

Fold	Akurasi SVM	Akurasi SVM + PSO
Fold 1	0.8256	0.8244
Fold 2	0.8331	0.8308
Fold 3	0.8931	0.8994
Fold 4	0.8037	0.8081
Fold 5	0.8137	0.8150
Rata-Rata	0.8339	0.8360

Berdasarkan Tabel 5, perbandingan akurasi antara dua model klasifikasi, yaitu SVM dengan kernel polynomial tanpa optimasi dan SVM + PSO, menunjukkan perbedaan yang relatif kecil. Analisis per fold mengungkapkan bahwa peningkatan akurasi tidak terjadi secara konsisten. Pada Fold 1, akurasi mengalami penurunan dari 0,8256 menjadi 0,8244 (turun 0,15%), dan pada Fold 2 juga menurun dari 0,8331 menjadi 0,8308 (turun 0,28%). Sebaliknya, Fold 3 menunjukkan peningkatan dari 0,8931 menjadi 0,8994 (naik 0,71%), yang merupakan peningkatan tertinggi. Fold 4 mengalami kenaikan dari 0,8037 menjadi 0,8081 (naik 0,55%), dan Fold 5 meningkat dari 0,8137 menjadi 0,8150 (naik 0,16%). Secara keseluruhan, hasil pengujian menggunakan 5-Fold Cross Validation menunjukkan bahwa rata-rata akurasi model SVM + PSO adalah 0,8360, lebih tinggi dibandingkan SVM tanpa optimasi yang memperoleh 0,8339, dengan rata-rata peningkatan sebesar 0,0021 atau 0,20%. Untuk memastikan apakah perbedaan ini signifikan, dilakukan uji t berpasangan (paired t-test), yang menghasilkan nilai $t = 1,044$ dengan $p\text{-value} = 0,355 (> 0,05)$. Nilai ini menunjukkan bahwa perbedaan akurasi antara kedua model tidak signifikan secara statistik pada tingkat kepercayaan 95%. Hasil ini menunjukkan bahwa meskipun penerapan PSO pada parameter SVM memberikan peningkatan rata-rata akurasi, dampaknya sangat kecil dan tidak konsisten pada seluruh fold. Selain itu, karena perbedaan tidak signifikan secara statistik, penerapan PSO pada penelitian ini belum memberikan pengaruh yang bermakna terhadap kinerja klasifikasi. Temuan ini mengindikasikan perlunya penelitian lanjutan dengan pendekatan optimasi yang lebih adaptif, seperti penyesuaian parameter PSO (inertia weight, jumlah partikel), penggunaan hybrid PSO dengan metode lain, atau eksperimen pada dataset yang lebih besar untuk mengevaluasi apakah metode ini dapat memberikan peningkatan yang signifikan. Visualisasi perbandingan akurasi pada setiap fold ditampilkan pada Gambar 4



Gambar 4. Grafik perbandingan antar fold

4. Kesimpulan

Berdasarkan hasil evaluasi menggunakan metode 5-Fold Cross Validation, diperoleh bahwa model Support Vector Machine (SVM) yang dioptimasi dengan algoritma Particle Swarm Optimization (PSO) menunjukkan kinerja yang lebih baik dibandingkan dengan model SVM menggunakan kernel polynomial (POLY) tanpa optimasi. Hal ini terlihat dari rata-rata akurasi SVM + PSO yang mencapai 0.8399, lebih tinggi dibandingkan dengan SVM kernel POLY yang memiliki rata-rata akurasi sebesar 0.8312. Selain itu, pada setiap fold, model SVM + PSO secara konsisten menghasilkan akurasi yang lebih tinggi dibandingkan model tanpa optimasi, meskipun selisihnya relatif kecil. Hasil ini menunjukkan bahwa penerapan algoritma PSO dalam proses optimasi parameter SVM mampu meningkatkan performa model secara keseluruhan, baik dari segi akurasi maupun konsistensi prediksi. Oleh karena itu, SVM + PSO dapat dianggap sebagai pendekatan yang lebih unggul dan layak dipertimbangkan untuk digunakan dalam implementasi akhir model klasifikasi.

Daftar Pustaka

- [1] Z. Shaukat *et al.*, "Revolutionizing diabetes diagnosis: Machine learning techniques unleashed," *Healthcare*, vol. 11, no. 21, p. 2864, Oct. 2023.
- [2] C. A. Tsai and Y. J. Chang, "Efficient selection of Gaussian kernel SVM parameters for imbalanced data," *Genes*, vol. 14, no. 3, p. 583, Feb. 2023.
- [3] M. R. Maulana, A. Sucipto, and H. M. Mulyo, "Optimisasi parameter Support Vector Machine dengan Particle Swarm Optimization untuk peningkatan klasifikasi diabetes," *J. Inform. Teknol. dan Sains (JINTEKS)*, vol. 6, no. 4, pp. 802–812, 2024.
- [4] E. Purwaningsih and E. Nurelasari, "Peningkatan akurasi metode Support Vector Machine melalui Particle Swarm Optimization pada penyakit ginjal kronis," *Inf. Manaj. Educ. Prof.*, vol. 9, no. 1, pp. 61–70, 2024.
- [5] H. Eldo, A. Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, "Penggunaan algoritma Support Vector Machine (SVM) untuk deteksi penipuan pada transaksi online," *J. Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, 2024.
- [6] M. Bekkar, H. K. Djemaa, and T. A. Alitouche, "Evaluation measures for models' assessment over imbalanced data sets," *J. Inf. Eng. Appl.*, vol. 3, no. 10, pp. 27–38, 2013.
- [7] P. Dhawas, M. A. Ramteke, A. Thakur, P. V. Polshetwar, R. V. Salunkhe, and D. Bhagat, "Big data analysis techniques: Data preprocessing techniques, data mining techniques, machine learning algorithm, visualization," in *Big Data Analytics Techniques for Market Intelligence*, IGI Global Scientific Publishing, 2024, pp. 183–208.
- [8] K. L. Du, B. Jiang, J. Lu, J. Hua, and M. N. S. Swamy, "Exploring kernel machines and support vector machines: Principles, techniques, and future directions," *Mathematics*, vol. 12, no. 24, Art. no. 3935, 2024.
- [9] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Oct. 2020.
- [10] Z. Xiong, S. Huang, S. Ren, Y. Lin, Z. Li, D. Li, and F. Deng, "Particle Swarm Optimization Support Vector Machine-Based Grounding Fault Detection Method in Distribution Network," *Algorithms*, vol. 18, no. 5, p. 259, May 2025.