

Optimasi Parameter Support Vector Machine Dengan Particle Swarm Optimization Untuk Prediksi Tunggakan Iuran Sekolah

Optimization of Support Vector Machine Parameter Using Particle Swarm Optimization For Prediction of School Fees Arrangements

Syahril¹, Reny Medikawati Taufiq², Taslim^{*3}, Dafwen Toresa⁴, Fajrizal⁵, Susi Handayani⁶

^{1,2}Program Studi Sistem Informasi, Universitas Muhammadiyah Riau,

^{3,4,5,6}Program Studi Teknik Informatika, Universitas Lancang Kuning,

*Penulis Korespondensi

Email : taslim@unilak.ac.id*

Abstrak. Penelitian ini berfokus pada pengembangan model prediktif untuk mengidentifikasi tunggakan iuran siswa di Sekolah Madrasah Ibtidaiyah Muhammadiyah Pekanbaru. Model ini memanfaatkan metode Support Vector Machine (SVM) dengan optimasi parameter menggunakan algoritma Particle Swarm Optimization (PSO) untuk mencapai akurasi prediksi yang tinggi. Data iuran siswa tahun 2022 digunakan sebagai sumber data, dan proses pra-pemrosesan data dilakukan untuk membersihkan noise dan outlier serta mengatasi ketidakseimbangan kelas dengan teknik SMOTE. Model yang diusulkan menggabungkan SVM dengan PSO untuk mencari nilai C dan gamma yang optimal pada model SVM. Hasil penelitian menunjukkan bahwa SVM dengan kernel radial dan 80% data latih mencapai akurasi sebesar 86,39%. Pengujian SVM+PSO dilakukan dengan dua ukuran populasi (10 dan 15) dan 1000 generasi evolusi. Penggunaan 10 populasi menghasilkan akurasi sebesar 89,20%, sedangkan 15 populasi mencapai akurasi 90,64% dengan nilai optimal $C=1,4$ dan $\gamma=2,6$. Penelitian ini menekankan pentingnya penyesuaian parameter dan strategi evolusi seperti PSO dalam meningkatkan kinerja model SVM untuk tugas mengklasifikasikan tunggakan iuran siswa. Hasil ini menunjukkan bahwa model yang diusulkan mampu memberikan prediksi yang lebih akurat dibandingkan SVM standar, sehingga memberikan kontribusi penting dalam identifikasi dan pengambilan keputusan terkait tunggakan iuran siswa di sekolah.

Kata kunci: Prediksi, Tunggakan, Pelajar, Optimasi, PSO, Naïve Bayes, Akurasi

Abstract. This research focuses on developing a predictive model to identify student tuition arrears at the Madrasah Ibtidaiyah Muhammadiyah Pekanbaru School. This model utilizes the Support Vector Machine (SVM) method with parameter optimization using the Particle Swarm Optimization (PSO) algorithm to achieve high prediction accuracy. 2022 student contribution data is used as a data source, and data pre-processing is carried out to clean noise and outliers and overcome class imbalances with the SMOTE technique. The proposed model combines SVM with PSO to find optimal C and gamma values in the SVM model. The research results show that SVM with a radial kernel and 80% training data achieves an accuracy of 86.39%. SVM+PSO testing was carried out with two population sizes (10 and 15) and 1000 generations of evolution. The use of 10 populations produces an accuracy of 89.20%, while 15 populations achieve an accuracy of 90.64% with optimal values of $C=1.4$ and $\gamma=2.6$. This research emphasizes the importance of parameter adjustments and evolutionary strategies such as PSO in improving the performance of SVM models for the task of classifying student fees arrears. These results show that the proposed model is able to provide more accurate predictions than standard SVM, thereby providing an

important contribution in identifying and making decisions regarding student tuition arrears in schools.

Keywords: *Predict, Arrear, student, optimization, PSO, Naïve Bayes, Accuracy*

1. Pendahuluan

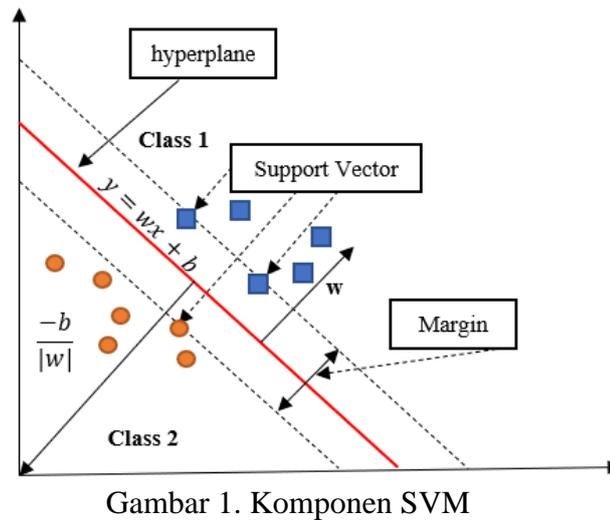
Muhammadiyah merupakan gerakan Islam modernis terbesar di Indonesia dan menjadi pionir dalam pemurnian dan pembaharuan Islam di tanah air. Organisasi ini didirikan pada tanggal 8 Dzulhijjah 1330 atau 18 November 1912 oleh Kyai Haji Ahmad Dahlan di Kauman, Yogyakarta. Salah satu fokus utamanya adalah untuk mencerahkan dan mencerdaskan kehidupan bangsa, terbukti dari “Statuten Muhammadiyah” atau peraturan dasar pertama pada tahun 1912, yang bertujuan untuk memajukan dan memajukan ajaran dan ilmu agama di Hindia Belanda dan memajukan dan memajukan cara hidup yang sesuai dengan kehendak Islam kepada para anggotanya. (Muhammadiyah dan Semangat Hidup Berkemajuan (Materi OSDI 2014), 2014)

Muhammadiyah berperan penting dalam mewujudkan pendidikan modern yang berhasil melahirkan generasi-generasi umat Islam yang terdidik. Pendidikan yang dikelola oleh Muhammadiyah mulai dari pendidikan anak usia dini hingga pendidikan tinggi, tersebar di seluruh wilayah NKRI bahkan luar negeri. Lembaga pendidikan dan sekolah tersebut dikelola secara mandiri oleh Muhammadiyah melalui upaya sukarela.

Sebagai sekolah mandiri, biaya sekolah merupakan bentuk dukungan terhadap keberlanjutan fiskal infrastruktur pendidikan. Biaya tersebut dapat berupa iuran pembangunan, biaya masuk, biaya seragam, dan lain-lain. Permasalahan umum yang muncul dalam proses pendidikan adalah terjadinya tunggakan biaya sekolah, yang dapat berdampak signifikan terhadap kelancaran kegiatan belajar mengajar serta acara sekolah yang telah direncanakan sebelumnya. Permasalahan ini juga banyak ditemui pada lembaga pendidikan atau sekolah yang dikelola oleh Muhammadiyah, antara lain Madrasah Ibtidaiyah Muhammadiyah 02 Pekanbaru.

Adanya tunggakan biaya sekolah tentunya berdampak pada kelancaran keadaan keuangan sekolah. Oleh karena itu, perlu dilakukan klasifikasi data mengenai tunggakan, dimana hasil klasifikasi tersebut dapat digunakan untuk memprediksi biaya sekolah, sehingga memberikan wawasan yang berharga bagi sekolah dalam pengambilan keputusan terkait pengelolaan keuangan sekolah. Prediksi ini dapat membantu memastikan bahwa pengajaran dan kegiatan sekolah lainnya dapat berjalan dengan lancar, sehingga memungkinkan berjalannya proses pembelajaran dan acara terkait sekolah lainnya dengan baik.

Klasifikasi merupakan masalah utama dalam aplikasi dunia nyata (Gupta & Gupta, 2022), dan salah satu algoritma klasifikasi terawasi yang populer adalah *Support Vector Machine* (Wang *et al.*, 2020) (Alcaraz *et al.*, 2022). SVM merupakan salah satu metode pembelajaran klasik yang telah dipelajari dan dibahas secara ekstensif selama hampir setengah abad, dimana prinsip utamanya adalah mencari pemisah *hyperplane* yang optimal pada ruang input atau fitur dengan memaksimalkan pemisahan antara dua sisi *hyperplane* (Liu *et al.*, 2022). SVM dapat bersifat linier atau non-linier, dan SVM linier selanjutnya dapat dikategorikan menjadi *hard margin* dan *soft margin*. Dalam SVM *hard margin*, data dapat dipisahkan sepenuhnya secara linier oleh *hyperplane*. *Soft margin*, sebaliknya, digunakan ketika data tidak sepenuhnya dapat dipisahkan. Komponen-komponen SVM dapat dilihat pada Gambar 1 di bawah ini (Rani *et al.*, 2022).



Konsep SVM merupakan upaya untuk menemukan *hyperplane* optimal yang bertindak sebagai pemisah antara dua kelas dalam ruang input, yang disebut juga dengan *linear separability*. Gambar 1 menunjukkan beberapa pola yang termasuk dalam dua kelas, positif (kotak biru) dan negatif (lingkaran kuning). Proses pembelajaran pada masalah klasifikasi dimaknai sebagai upaya menemukan *hyperplane* yang memisahkan kedua kelompok secara optimal. *Hyperplane* pemisah optimal antara kedua kelas dapat dicari dengan mengukur tepi *hyperplane* dan mencari titik maksimumnya. Jarak adalah jarak antara *hyperplane* dengan titik data terdekat untuk setiap kelas. Subset data pelatihan berikutnya disebut vektor dukungan. Misalnya nilai *hyperplane* akan dicari untuk N titik data pelatihan.

$((x_1, y_1), (x_2, y_2) \dots (x_n, y_n))$, dimana $x_1, \in R^d$ dan $y, \in [+1, -1]$.

$$f(x) = \text{sign}(w \cdot x - b) \quad (1)$$

$$\text{minimize}_{(w,b)} 1/2w^T w \quad (2)$$

$$y_i(w \cdot x_i - b) \geq 1 \quad (3)$$

$$\text{Maximize } L(w, b, \alpha) 1/2w^T w - \sum_{i=1}^N \alpha_i y_i (w x_i - b) + \sum_i \alpha_i \quad (4)$$

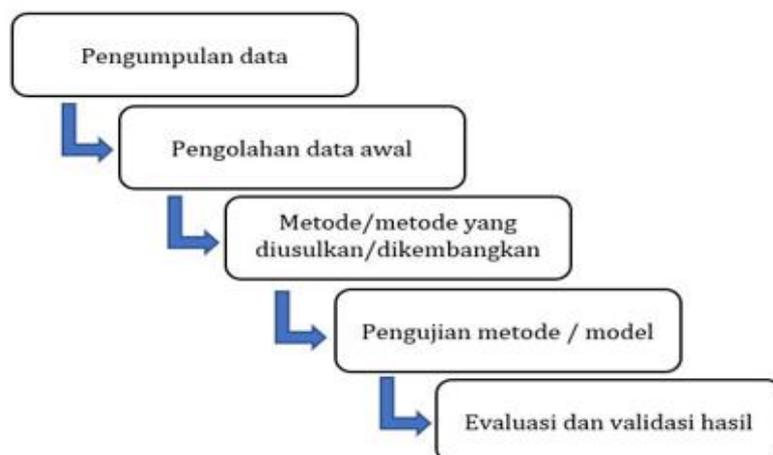
$$f(x) = \text{sign}(w x + b) = \text{sign}\left(\left(\sum_i \alpha_i x_i(x_i, x)\right) + b\right) \quad (5)$$

Beberapa penelitian telah diajukan oleh peneliti untuk meningkatkan kinerja SVM. Misalnya, Adankon dan Cheriet menyarankan metode berdasarkan estimasi gradien kesalahan empiris (Adankon & Cheriet, 2007). Zhang mengusulkan metode hybrid (Zhang *et al.*, 2014), Huang dan Wang memperkenalkan metode untuk pemilihan fitur dan optimasi (Huang & Wang, 2022), sementara Hitam mengusulkan SVM yang dioptimalkan menggunakan *particle swarm optimization* (Hitam *et al.*, 2019).

Berdasarkan uraian di atas kami tertarik untuk melakukan penelitian mengenai tunggakan biaya sekolah di Madrasah Ibtidaiyah Muhammadiyah 02 Pekanbaru dengan melakukan optimasi parameter pada algoritma SVM menggunakan *Particle Swarm Optimization* (PSO).

2. Metode

Dalam penelitian ini akan dilakukan beberapa langkah atau tahapan seperti yang digambarkan pada Gambar 2 berikut.



Gambar 2. Metodologi Penelitian

1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data biaya siswa Madrasah Ibtidaiyah Muhammadiyah 02 Pekanbaru tahun 2022. Data ini berupa biaya siswa kelas 1 sampai dengan kelas 6 dan terdiri dari 16 variabel.

2. Pengolahan Data

Data yang digunakan dalam penelitian ini berasal dari data pembayaran siswa Madrasah Ibtidaiyah Muhammadiyah 02 Pekanbaru pada tahun 2022. Terdapat total 350 catatan data yang terdiri dari 16 variabel yang akan digunakan dalam analisis dan pengembangan model klasifikasi pada penelitian ini. Pengolahan data awal dilakukan untuk mendapatkan data yang bersih dan bebas dari noise atau outlier. Beberapa tahapan yang dilakukan antara lain:

a. Validasi data

Validasi data dilakukan untuk mengidentifikasi adanya noise, outlier, data yang tidak lengkap, atau tidak konsisten.

b. Transformasi data

Karena dalam penelitian ini terdapat data kategorikal, maka data tersebut diubah ke dalam bentuk numerik.

c. Normalisasi data

Normalisasi bertujuan untuk mencapai nilai atribut atau data yang seimbang dalam rentang tertentu. Rentang nilai yang tidak seimbang pada setiap atribut dapat mempengaruhi kualitas hasil data mining.

3. Metode yang Diusulkan

Penelitian ini akan dilakukan dengan menggunakan metode support vector machine dengan optimasi variabel menggunakan algoritma PSO.

4. Evaluasi dan Validasi

Validasi merupakan tahapan krusial dalam pemodelan, untuk menilai reliabilitas model yang akan digunakan.

3. Hasil dan Pembahasan

1. Pemrosesan Awal Data

Data yang digunakan dalam penelitian ini adalah data pembayaran siswa Madrasah Ibtidiyah Muhammadiyah 02 Pekanbaru tahun 2022. Terdapat 350 catatan data dengan 16 variabel. Data pembayaran siswa dari kelas 1 sampai dengan 6 dapat dilihat pada Tabel 1.

Tabel 1. Data Pembayaran Siswa

No	Tunggakan Tahun Lalu	Bayar Tahun Lalu	Sisa Tahun Lalu	Penagihan KT	Pembayaran KT	Sisa KT	Juli (Masuk PPDB)	Juni	Tunggakan
1	0	0	0	0	0	0	400	...	0	400
2	0	0	0	0	0	0	400	400	0
3	0	0	0	0	0	0	400	400	0
4	800	800	0	0	0	0	400	400	800
5	0	0	0	0	0	0	400	0	800
6	0	0	0	0	0	0	400	400	0
7	0	0	0	0	0	0	400	400	0
8	0	0	0	0	0	0	400	400	0
9	1,600,000	1,600,000	0	0	0	0	400	0	0
10	0	0	0	0	0	0	400	400	0
11	1,200,000	1,200,000	0	0	0	0	400	400	- 1,200,000
12	0	0	0	0	0	0	0	400	6,400,000
13	0	0	0	0	0	0	400	400	0
14	0	0	0	0	0	0	400	400	0
15	3,600,000	3,600,000	0	0	0	0	400	0	- 1,200,000
16	0	0	0	0	0	0	400	400	0

Selanjutnya pada data pembayaran siswa diberi label pada setiap record data yang terdiri dari dua label yaitu “lunas” dan “tertunggak”. Label lunas diberikan bila tidak ada tunggakan pembayaran pada setiap variabel pembayaran, sedangkan label tertunggak diberikan bila ada variabel pembayaran yang belum dibayar.

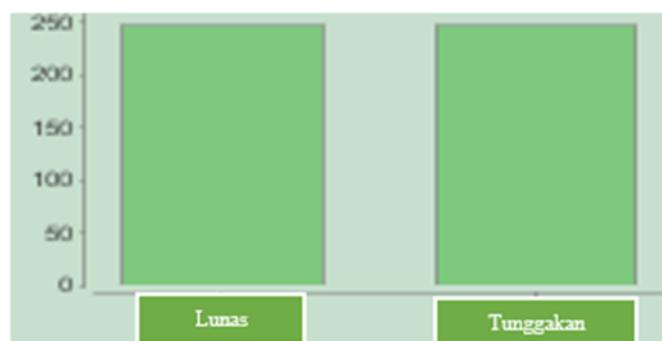
Dari data yang diberi label, terlihat adanya data yang tidak seimbang, dimana jumlah data yang berlabel lunas lebih banyak dibandingkan dengan jumlah data yang berlabel tertunggak. Dalam pembelajaran mesin, kinerja dapat terpengaruh secara negatif ketika belajar dari kumpulan data yang tidak seimbang, sehingga pembelajaran dari data tersebut menjadi masalah yang mendesak. Kelas yang tidak seimbang dapat menyebabkan prediksi yang bias dan menurunkan kinerja model (Gu *et al.*, 2008).

Dalam penelitian ini, untuk mengatasi masalah ketidakseimbangan kelas, teknik *oversampling* diterapkan pada data pelatihan. *Oversampling* dilakukan dengan menambahkan data duplikat dari kelas minoritas menggunakan metode SMOTE (*Synthetic Minority Over-sampling*

Technique). Jumlah label data sebelum dan sesudah penerapan SMOTE dapat dilihat pada Gambar 3 dan 4 di bawah ini.



Gambar 3. Jumlah Label Data Awal

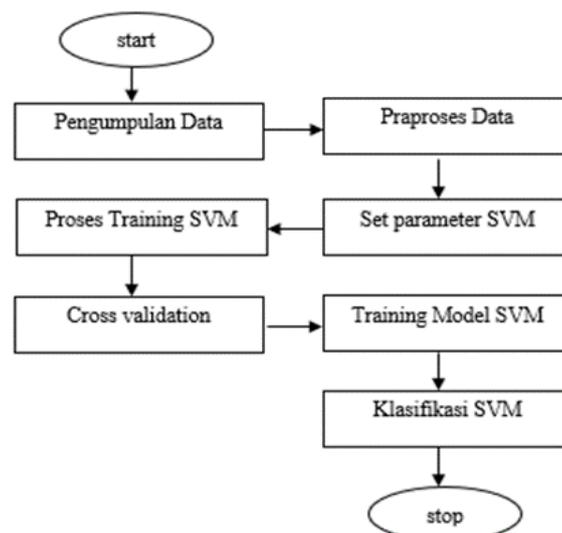


Gambar 4. Jumlah Label Setelah SMOTE

Pengujian dilakukan dengan dua cara yaitu SVM tanpa PSO dan SVM+PSO untuk menilai sejauh mana optimasi PSO dapat berdampak pada algoritma SVM.

2. SVM Tanpa PSO

Percobaan awal dilakukan dengan menggunakan algoritma SVM tanpa optimasi PSO. Flowchart SVM tanpa PSO dapat dilihat pada Gambar 5 dibawah ini.



Gambar 5. SVM Tanpa PSO

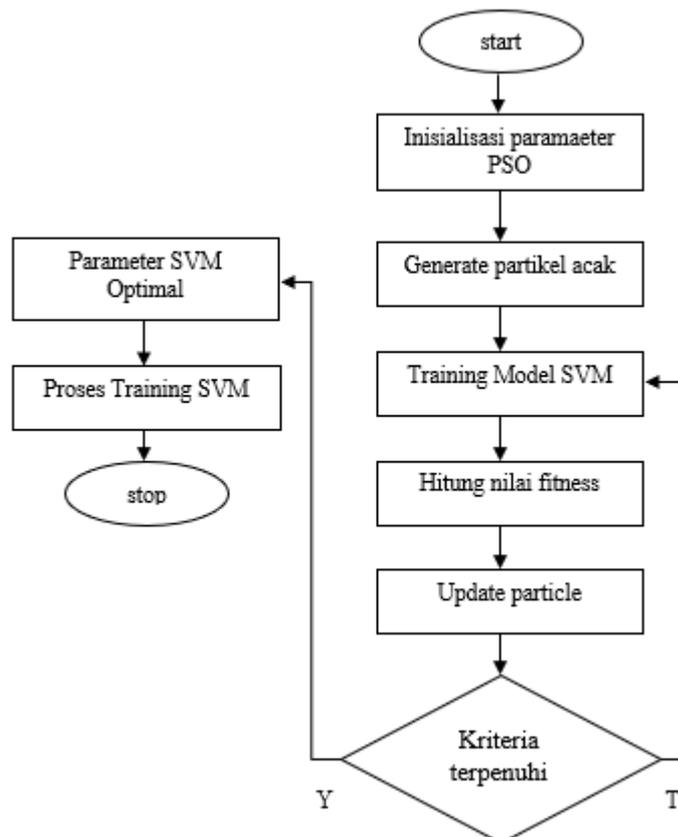
Dari hasil pengujian menggunakan model kernel radial dan data latih 80% diperoleh tingkat akurasi sebesar 86,39% seperti terlihat pada matriks konfusi pada Tabel 2.

Tabel 2. Confusion Matrix Kebingungan SVM

	lunas	Tunggakan	Class Precision
Pred. Lunas	48	10	82,76%
Pred. Tunggakan	1	39	1 97,50%
Class Recall	97,96%	79,59%	2

3. Optimasi SVM dengan PSO

Optimasi adalah proses menemukan solusi terbaik untuk suatu masalah tertentu dalam serangkaian batasan. Di bidang pembelajaran mesin, pengoptimalan sering kali digunakan untuk menemukan parameter terbaik suatu model guna mencapai akurasi tertinggi. Pengoptimalan sangat penting dalam pembelajaran mesin karena dapat meningkatkan akurasi dan efisiensi model. *Particle Swarm Optimization* (PSO) adalah salah satu metode optimasi yang digunakan dalam pembelajaran mesin. PSO merupakan algoritma metaheuristik yang dapat mencari solusi terbaik dalam ruang solusi yang besar tanpa membuat asumsi terhadap masalah yang dioptimalkan. Algoritma optimasi SVM dengan PSO dapat dilihat pada Gambar 6 dibawah ini.



Gambar 6. Optimasi SVM+PSO

Inisialisasi parameter PSO dilakukan dengan membuat range parameter yang terdiri dari nilai *Complexity* (C) dan kernel *gamma* (γ). Batasan partikel ini selanjutnya akan digunakan dalam metode SVM. Kisaran partikel PSO dapat dilihat pada Tabel 3.

Tabel 3. Rentang Partikel

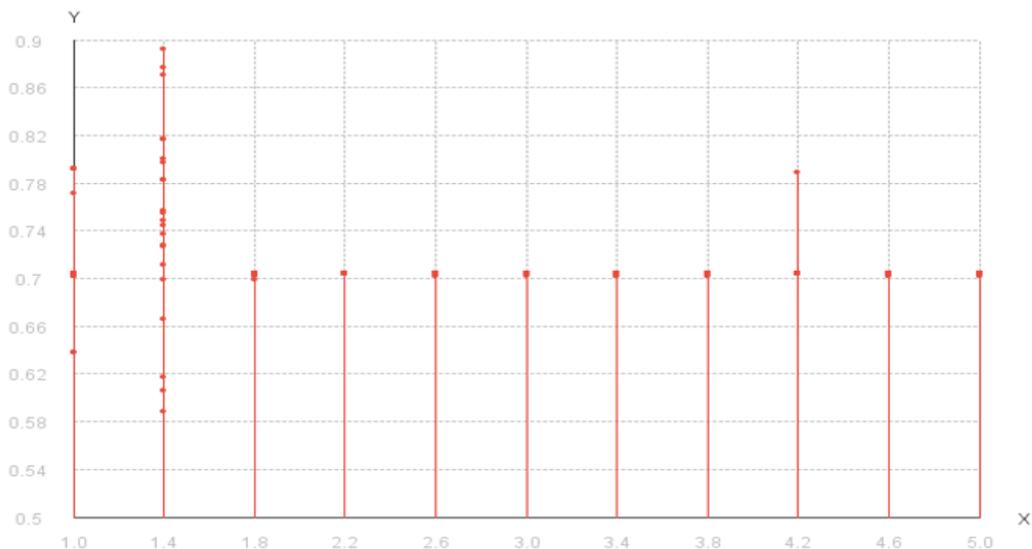
Parameters	C	γ
Upper Limit	1	1
Step	10	10
Lower Limit	5	5

Dalam pengujian SVM+PSO, proses SMOTE pertama kali diterapkan untuk mengatasi dataset yang tidak seimbang. Pengujian dilakukan dengan menggunakan dua ukuran populasi yang berbeda yaitu 10 dan 15, serta menjalankan proses evolusi selama 1000 generasi untuk setiap konfigurasi populasi. Untuk mengukur performa model secara objektif, digunakan pendekatan validasi silang k-fold dengan k=10. Dari keseluruhan dataset, 80% digunakan sebagai data latih.

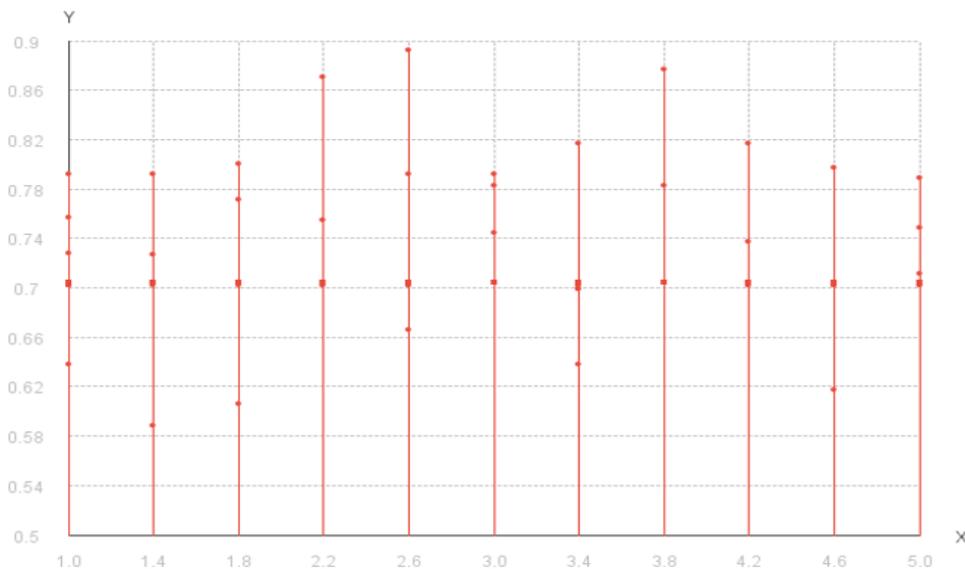
Hasil pengujian pertama dengan menggunakan ukuran populasi 10 dan 1000 generasi menunjukkan bahwa model yang dihasilkan mencapai tingkat akurasi sebesar 89,20%. Penentuan performa model melibatkan parameter penting yaitu C dan Gamma. Dari pengujian diperoleh nilai C sebesar 1,4, sedangkan nilai Gamma sebesar 2,6. Informasi detail mengenai performa dan kemampuan prediksi model dapat dilihat pada matriks konfusi pada Tabel 4. Selain itu, hasil percobaan ini juga didukung oleh grafik C dan Gamma yang ditunjukkan pada Gambar 7 dan Gambar 8.

Tabel 4. Confusion Matrix SVM+PSO Dengan 10 Populasi Dan 1000 Generasi

	True Lunas	True Menunggak	Class Precision
Pred. Lunas	195	41	82,63%
Pred. Menunggak	2	156	98,73%
Class Recall	98,98%	79,19%	



Gambar 7. Grafik Nilai SVM+PSO C Dengan 10 Populasi Dan 1000 Generasi



Gambar 8. Grafik Nilai Gamma SVM+PSO Dengan 10 Populasi dan 1000 Generasi.

Pengujian kedua dilakukan dengan ukuran populasi 15 dan 1000 generasi. Pada pengujian kedua diperoleh tingkat akurasi sebesar 90,64% dengan nilai C dan Gamma yang sama seperti pada pengujian SVM+PSO pertama yaitu 1,4 untuk C dan 2,6 untuk Gamma. Nilai Confusion Matrix untuk SVM+PSO dengan 15 populasi dapat dilihat pada Tabel 5 dibawah ini.

Tabel 5. Confusion Matrix SVM+PSO Dengan 15 Populasi Dan 1000 Generasi

	True Lunas	True Menunggak	Class Precision
Pred. Lunas	196	36	84,45%
Pred. Menunggak	1	161	99,38%
Class Recall	99,49%	81,73%	

4. Kesimpulan

Hasil penelitian menunjukkan tingkat akurasi SVM tanpa PSO menggunakan model kernel radial dan 80% data latih mencapai 86,39%. Pada pengujian SVM+PSO, dua ukuran populasi berbeda yaitu 10 dan 15 diuji dengan proses evolusi 1000 generasi untuk setiap konfigurasi populasi. Parameter C dan gamma ditetapkan dalam rentang 1 hingga 5. Untuk mengukur kinerja model, digunakan validasi silang 10 kali lipat. Pengujian dengan 10 populasi memperoleh tingkat akurasi sebesar 89,20%, dan pengujian dengan 15 populasi memperoleh tingkat akurasi sebesar 90,64% dengan nilai optimal $C=1,4$ dan $\gamma=2,6$.

Konsistensi nilai C dan gamma antara kedua skenario pengujian menunjukkan bahwa parameter tersebut berperan penting dalam membentuk model yang akurat. Hasil pengujian menyoroti pentingnya penyetelan parameter dan strategi evolusi dalam meningkatkan kinerja model dalam tugas klasifikasi dengan SVM

Daftar Pustaka

Adankon, M. M., & Cheriet, M. (2007). *Optimizing resources in model selection for support vector machine*. 40, 953–963. <https://doi.org/10.1016/j.patcog.2006.06.012>

- Alcaraz, J., Labbé, M., & Landete, M. (2022). Support Vector Machine with feature selection: A multiobjective approach. *Expert Systems with Applications*, 204, 117485. <https://doi.org/https://doi.org/10.1016/j.eswa.2022.117485>
- Gupta, U., & Gupta, D. (2022). Bipolar fuzzy based least squares twin bounded support vector machine. *Fuzzy Sets and Systems*. <https://doi.org/https://doi.org/10.1016/j.fss.2022.06.009>
- Hitam, N. A., Ismail, A. R., & Saeed, F. (2019). ScienceDirect ScienceDirect 16th Support 2019 on Particle An Optimized Vector (SVM) based Swarm Optimization (PSO) for Cryptocurrency Forecasting An Optimized Support Vector Machine (SVM) based on Particle Azidah Hitam (PSO) Amelia for Optimizatio. *Procedia Computer Science*, 163, 427–433. <https://doi.org/10.1016/j.procs.2019.12.125>
- Huang, C., & Wang, C. (2022). A GA-based feature selection and parameters optimization for support vector machines A GA-based feature selection and parameters optimization for support vector machines. August 2006. <https://doi.org/10.1016/j.eswa.2005.09.024>
- Liu, B., Huang, R., Xiao, Y., Liu, J., Wang, K., Li, L., & Chen, Q. (2022). Adaptive robust Adaboost-based twin support vector machine with universum data. *Information Sciences*, 609, 1334–1352. <https://doi.org/https://doi.org/10.1016/j.ins.2022.07.155>
- Muhammadiyah dan Semangat Hidup Berkemajuan (*Materi OSDI 2014*). (2014). <https://lppi.umy.ac.id/muhammadiyah-dan-semangat-hidup-berkemajuan-materi-osdi-2014/#:~:text=Muhammadiyah merupakan organisasi dakwah Islam,Ngayogyakarta%2C yaitu K.H. Ahmad Dahlan.>
- Rani, A., Kumar, N., Kumar, J., Kumar, J., & Sinha, N. K. (2022). Chapter 6 - Machine learning for soil moisture assessment. In R. C. Poonia, V. Singh, & S. R. Nayak (Ed.), *Deep Learning for Sustainable Agriculture* (hal. 143–168). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-323-85214-2.00001-X>
- Wang, S., Dong, L., & Hua, H. (2020). Parameter optimization of support vector machine based on improved grid algorithm Parameter optimization of support vector machine based on improved grid algorithm. <https://doi.org/10.1088/1742-6596/1693/1/012108>
- Zhang, X., Qiu, D., & Chen, F. (2014). Neurocomputing Support vector machine with parameter optimization by a novel hybrid method and its application to fault diagnosis. *Neurocomputing*, 1–11. <https://doi.org/10.1016/j.neucom.2014.08.010>