

Implementasi *Support Vector Machine-Genetic Algorithm* (SVM-GA) dalam Mengklasifikasikan Kelulusan Tepat Waktu Mahasiswa Statistika Unisba

Indah Siti Rahmawati*, Ilham Faisahl Mahdy

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*indahrahma387@gmail.com, ilham.faishal@gmail.com

Abstract. Support Vector Machine (SVM) is a classification method that finds the optimal hyperplane for separating two data classes. However, SVM has weaknesses in determining hyperparameter values, necessitating parameter optimization. The Genetic Algorithm (GA) is an optimization method proven effective in handling complex optimization problems and irregular, complicated solutions. The classification system for the timely graduation of Unisba Statistics students is important for improving the quality and accreditation of study programs. Therefore, this study aims to obtain the best SVM model for classifying the timely graduation of Unisba Statistics students with the help of a Genetic Algorithm (GA). The data used in this study includes the graduation records of Unisba Statistics students from January 24, 2019, to August 29, 2023. By doing SVM-GA modeling using 9 combinations of GA, 10-Cross Validation, and RBF kernels, the best performance model was obtained from the genetic algorithm parameter of a population of 200, the value of 0,8, value of 0,05, an SVM parameter of 2242,1776, and 0,0002. Model performance evaluation showed that the resulting model achieved an average accuracy of 78.2664%, an average recall of 78,3287%, an average precision of 86.2500%, and an average f1-score of 81.92%.

Keywords: *Genetic Algorithm, Students Graduate on Time, Classification.*

Abstrak. *Support Vector Machine* (SVM) merupakan metode klasifikasi yang menemukan *hyperplane* optimal dalam memisahkan dua kelas data. SVM memiliki kelemahan dalam menentukan nilai *hyperparameter*, sehingga perlu dilakukan optimasi parameter. *Genetic Algorithm* (GA) merupakan metode optimasi yang terbukti efektif dalam menangani masalah optimasi yang kompleks dan solusi yang tidak teratur serta rumit. Sistem klasifikasi dalam kasus kelulusan tepat waktu mahasiswa Statistika Unisba menjadi hal penting untuk meningkatkan kualitas dan akreditasi program studi. Maka dari itu, penelitian ini bertujuan untuk memperoleh model SVM terbaik dalam mengklasifikasikan kelulusan tepat waktu mahasiswa Statistika Unisba dengan bantuan *Genetic Algorithm* (GA). Data yang digunakan pada penelitian ini merupakan data kelulusan mahasiswa Statistika Unisba pada periode kelulusan sejak tanggal 24 Januari 2019 sampai 29 Agustus 2023. Dengan melakukan pemodelan SVM-GA menggunakan 9 kombinasi parameter GA, 10-Cross Validation, dan kernel RBF, diperoleh model dengan kinerja terbaik dari parameter algoritma genetika berupa populasi sebanyak 200, nilai P_c sebesar 0,8, nilai P_m sebesar 0,05, dengan parameter SVM berupa parameter C sebesar 2242,1776 dan γ sebesar 0,0002. Evaluasi kinerja model menunjukkan bahwa model yang dihasilkan memperoleh rata-rata akurasi sebesar 78,2664%, Rata-rata *recall* sebesar 78,3287%, rata-rata *precision* sebesar 86,2500%, dan rata-rata *f1-score* sebesar 81,9225%.

Kata Kunci: *Genetic Algorithm, Kelulusan Tepat Waktu Mahasiswa, Klasifikasi.*

A. Pendahuluan

Support Vector Machine (SVM) memiliki cara kerja memisahkan kedua kelas dengan menggunakan *hyperplane* optimal. Dalam prinsip kerjanya, SVM menerapkan prinsip *Structural Risk Minimization* (SRM) yang terbukti lebih unggul dibandingkan prinsip *Empirical Risk Minimization* (ERM) (Gunn, 1998). Prinsip SRM merupakan upaya untuk meminimalkan kesalahan pada data pelatihan dan meminimalkan faktor kedua yaitu dimensi vektor. SVM mewujudkan prinsip SRM ini ketika memilih *hyperplane* optimal yang memiliki *margin* terbesar. Dengan prinsip SRM tersebut, SVM memiliki kemampuan generalisasi yang jauh lebih baik dibandingkan dengan metode lainnya (Handayani et al., 2024).

Meskipun demikian, SVM sering kali mengalami permasalahan dalam penentuan nilai *hyperparameter* (Yenaeng et al., 2014). Padahal, penentuan parameter yang optimal dapat berperan penting dalam membangun model yang memiliki akurasi dan stabilitas yang tinggi (Min & Lee, 2005). Berdasarkan hal itu, penelitian ini akan memanfaatkan metode *Genetic Algorithm* (GA) untuk memperoleh parameter SVM yang optimal. GA dikenal sebagai alat yang ampuh untuk memecahkan masalah optimasi non-linear skala besar (Gallagher S et al., 1994). Selain itu, GA telah digunakan untuk menyelesaikan sejumlah masalah optimasi yang kompleks dan dapat menangani solusi yang tidak teratur dan rumit (Kusumaningrum, 2017).

Menurut BAN-PT, lama studi dan persentase mahasiswa lulus tepat waktu menjadi dua faktor yang menentukan kualitas dan akreditasi program studi. BAN-PT mengatur bahwa jika persentase kelulusan tepat waktu $\geq 50\%$, maka program studi akan diberi skor penilaian tertinggi, yaitu 4. Sebaliknya, jika persentase kelulusan tepat waktu $< 50\%$, program studi akan diberi skor penilaian 1-3 (BAN-PT, 2021). Berdasarkan data Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung, diperoleh informasi bahwa persentase mahasiswa yang lulus tepat waktu adalah sebesar 59%. Pada periode tersebut, kriteria untuk skor penilaian tertinggi sudah terpenuhi. Namun, masih perlu upaya untuk mempertahankan dan meningkatkan angka tersebut pada mahasiswa periode selanjutnya guna meningkatkan kualitas dan akreditasi Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung (Maryanto Rompon et al., 2023).

Penggunaan sistem klasifikasi dengan SVM-GA ini dapat membantu program studi meningkatkan persentase kelulusan tepat waktu mahasiswa dengan mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu (Yuniar & Kismiantini, 2023). Dengan mengetahui kelulusan mahasiswa sejak dini, program studi dapat menetapkan strategi untuk mengurangi jumlah mahasiswa yang tidak lulus tepat waktu, sehingga program studi dapat meningkatkan persentase mahasiswa yang lulus tepat waktu. Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana implementasi serta evaluasi kinerja metode *Support Vector Machine-Genetic Algorithm* (SVM-GA) dalam mengklasifikasikan kelulusan tepat waktu mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung?”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Mengetahui hasil klasifikasi kelulusan tepat waktu mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung dengan menggunakan *Support Vector Machine-Genetic Algorithm* (SVM-GA).
2. Mengetahui hasil kinerja metode *Support Vector Machine-Genetic Algorithm* (SVM-GA) dalam mengklasifikasikan kelulusan tepat waktu mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung.

B. Metodologi Penelitian

Data penelitian ini merupakan data kelulusan mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung yang berjumlah 433 data, dengan data yang digunakan selama periode kelulusan sejak tanggal 24 Januari 2019 sampai 29 Agustus 2023. Variabel penelitian yang digunakan yaitu 6 variabel independen (X) berupa variabel jenis kelamin (X_1), asal daerah (X_2), usia saat masuk (X_3), nilai pesantren (X_4), nilai toefl (X_5), dan IPK (X_6), dan 1 variabel dependen (Y) berupa status kelulusan mahasiswa. Penelitian ini menggunakan metode SVM-GA yang dilakukan dengan *software* R.

Support Vector Machine (SVM)

SVM adalah teknik pembelajaran mesin yang menggunakan strategi untuk memisahkan data menjadi dua kelas berbeda dengan menggunakan *hyperplane* (Vapnik dalam Kusumaningrum, 2017). *Hyperplane* merupakan garis batas keputusan yang digunakan untuk memisahkan dua buah kelas data. SVM memiliki konsep untuk mencari *hyperplane* optimal. Memaksimalkan jarak antara *hyperplane* dengan pola terdekat dari masing-masing kelas akan menghasilkan *hyperplane* optimal (Kusumaningrum, 2017). *Hyperplane* optimal dalam SVM adalah *hyperplane* yang memaksimalkan kemampuan generalisasi. Namun, ketika data pelatihan tidak dapat dipisahkan secara linear, klasifikasi yang dihasilkan mungkin belum mencapai kemampuan generalisasi maksimal (Kusumaningrum, 2017). Maka dari itu, untuk mengatasi permasalahan yang tidak dapat dipisahkan secara linear, SVM menggunakan parameter *cost* (C) yang berguna untuk memberikan penalti antara menyeimbangkan *margin* dan *error* (Fatmawati & Rifai, 2023). Selain itu, SVM juga melakukan modifikasi dengan menambahkan fungsi *kernel* (Nugroho et al., 2013).

Dengan menggunakan fungsi pemetaan x yaitu, $\phi(x) = (\phi_1(x), \dots, \phi_l(x))^T$ berupa fungsi *kernel* yang ditentukan oleh pengguna yang memetakan vektor masukan x berdimensi m ke dalam ruang fitur berdimensi l , diperoleh solusi *dual problem* pada ruang fitur sebagai berikut (Abe, 2010).

$$\max_{\alpha} L_D = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (1)$$

dengan batasan, $0 \leq \alpha_i \leq C$, $i, j = 1, 2, \dots, M$ dan $\sum_{i=1}^M \alpha_i y_i = 0$. Setelah menyelesaikan persamaan (1), maka kelas dari data *testing* yang akan diprediksi ditentukan berdasarkan persamaan berikut.

$$D(x) = \sum_{i=1}^S \alpha_i y_i K(x_i, x_{testing}) + b \quad (2)$$

Dimana nilai b diperoleh berdasarkan rumus berikut.

$$b = \frac{1}{|U|} \sum_{j=1}^U (y_j - \sum_{i=1}^S \alpha_i y_i K(x_i, x_j)) \quad (3)$$

Dalam penelitian ini, model SVM akan menggunakan fungsi *kernel* RBF. Fungsi *kernel* ini disarankan karena kemampuannya dalam menangani non-linearitas, memiliki kesulitan numerik yang lebih sedikit daripada *kernel* lainnya, serta lebih tahan terhadap *outlier* (Hsu et al., 2003). *Kernel* RBF dirumuskan sebagai berikut.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (4)$$

Genetic Algorithm (GA)

Genetic algorithm (GA) pertama dirancang oleh John Holland pada tahun 1975. GA sering kali digunakan untuk menyelesaikan permasalahan optimasi. GA mencari solusi terbaik dengan menggunakan kriteria kinerja (*fitness*) (Kusumaningrum, 2017). Salah satu keunggulan GA adalah kemampuannya untuk menangani berbagai fungsi objektif dari kromosom yang berbeda. GA juga fleksibel dan dapat diintegrasikan dengan pendekatan lain (Gen & Cheng, 1997). Selain itu, GA cocok untuk memanipulasi model dengan berbagai resolusi dan struktur secara bersamaan karena kemampuannya untuk mengeksplorasi ruang solusi non-linear tanpa memerlukan pengetahuan sebelumnya tentang karakteristik model (McCall, J., & Petrovski, A dalam Wu et al., 2007).

GA terdiri dari tujuh komponen, seperti yang dijelaskan secara rinci oleh (Suyanto dalam Rindrasari, 2017) sebagai berikut.

1. Skema Pengkodean. Skema pengkodean merupakan cara pengkodean isi kromosom yang menghasilkan nilai tertentu yang mana nilai tersebut mewakili atribut atau parameter-parameter dari masalah yang sedang diselesaikan. Skema pengkodean yang digunakan yaitu *real number encoding*, dimana nilai gen diwakili oleh bilangan bilangan

- real positif.
2. Nilai *fitness*. Nilai *fitness* merupakan ukuran kinerja individu yang menentukan kemungkinan mereka untuk bertahan hidup. Pada penelitian ini sudah ditetapkan bahwa fungsi *fitness* yang digunakan adalah akurasi, maka nilai *fitness* yang diperoleh merupakan nilai akurasi dari kombinasi parameter yang terdapat dalam kromosom.
 3. Seleksi calon orang tua. Seleksi calon orang tua berupaya memberikan peluang bereproduksi kepada individu-individu dalam populasi yang memiliki nilai *fitness* lebih rendah. Pada penelitian ini, yang akan digunakan metode seleksi *roulette wheel selection*. Kelebihan menggunakan metode ini adalah menjamin kesempatan yang sama bagi semua kromosom untuk dipilih.
 4. *Crossover* (pindah silang). Proses ini terdiri dari pertukaran bagian antara dua kromosom yang berasal dari proses reproduksi sebelumnya. Probabilitas *crossover* mewakili frekuensi terjadinya proses pindah silang antara dua kromosom orang tua. Nilai P_c optimal terletak pada rentang 0,75-0,95 (Schaffer, dkk dalam Kusumaningrum, 2017). Pada penelitian ini, *crossover* yang akan digunakan yaitu *local arithmetic crossover*. Teknik *crossover* ini sering digunakan untuk kasus GA yang menggunakan bilangan *real*, serta memiliki kelebihan dalam mencegah konvergensi dini dan dapat memperoleh kombinasi yang lebih baik dari sifat-sifat yang diwarisi kedua orang tua, sehingga meningkatkan peluang untuk menemukan solusi yang lebih baik.
 5. Mutasi. Pada saat seleksi dan *crossover* dapat menyebabkan suatu informasi dari individu terbaik hilang karena tidak terpilih saat seleksi dan *crossover*. Nilai P_m yang sering digunakan dalam penerapan GA berada pada rentang 0,001 dan 0,05 (Davis dalam Ismail & Irhamah, 2008). Pada penelitian ini, mutasi yang akan digunakan yaitu *uniform mutation*. Dalam hal ini nilai-nilai gen pada keturunannya diambil secara acak dan seragam dari domain definisi yang telah ditentukan (Grosan & Abraham, 2011).
 6. Elitisme. Elitisme menggantikan kromosom berkualitas rendah pada keturunan dengan kromosom berkualitas tertinggi pada populasi lama, yang berjumlah 10%-20% dari seluruh populasi.
 7. Penggantian populasi. Penggantian populasi pada GA mengacu pada proses menggantikan seluruh individu lama dengan N individu baru yang diperoleh dari tahapan *crossover*, mutasi, dan elitisme pada setiap generasi. Individu baru tersebut akan diseleksi untuk memperoleh solusi terbaik dari permasalahan GA.

K-Fold Cross Validation

K-Fold Cross Validation adalah salah satu algoritma yang digunakan untuk mengevaluasi kinerja model. Model dibangun dengan *subset k-1* sebagai data pelatihan dan diuji dengan *subset* sisanya sebagai data pengujian. Prosedur ini dilakukan sebanyak k kali, dengan masing-masing k *subset* hanya digunakan satu kali sebagai data pengujian. Hasil akurasi yang diperoleh merupakan rata-rata dari hasil perhitungan akurasi setiap data pelatihan dan pengujian (Han et al., 2012).

Confusion Matrix

Confusion matrix merupakan tabel yang berguna untuk menganalisis kualitas hasil pengklasifikasian (Han et al., 2012). *Confusion matrix* disajikan sebagai berikut:

Tabel 1. Confusion Matrix

<i>Actual</i>	<i>Predicted</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Dengan menggunakan *confusion matrix* dapat dihitung parameter evaluasi model sebagai berikut:

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$F1-Score = \frac{2 \times precision \times recall}{precision + recall} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

C. Hasil Penelitian dan Pembahasan

Sebelum melakukan tahapan pemodelan, perlu dilakukan terlebih dahulu tahapan *pre-processing* data. *Pre-processing* merupakan tahapan penting yang bertujuan untuk membersihkan dan mempersiapkan data agar dapat digunakan secara efektif dalam melakukan pemodelan. Tahapan yang dilakukan adalah sebagai berikut.

Pengecekan Data Duplikat

Pengecekan data duplikat perlu dilakukan untuk menghindari adanya *noise* pada data. Hasil dari pengecekan data duplikat menggunakan software R menunjukkan bahwa tidak terdapat data duplikat dalam dataset.

Pengecekan Missing Value

Berdasarkan pengecekan *missing value* menggunakan software R, diperoleh hasil bahwa terdapat sebanyak 45 data *missing value* pada variabel nilai toefl (X_5). Maka, untuk membangun keandalan model data, *missing value* tersebut perlu dilakukan penanganan. Pada penelitian ini akan dilakukan imputasi atau penggantian nilai *missing value* dengan menggunakan metode imputasi median. Berdasarkan hasil perhitungan median variabel nilai toefl (X_5) diperoleh median sebesar 497, maka nilai *missing value* pada 45 data akan diganti dengan nilai sebesar 497.

Pengkodean Variabel Kategorik

Pengkodean variabel kategorik yang digunakan penelitian ini adalah *label encoding*. *Label encoding* yang dilakukan dengan mengubah kategori ke dalam bentuk numerik 0 dan 1. *Label encoding* akan dilakukan pada variabel jenis kelamin (X_1), asal daerah (X_2), dan status kelulusan (X_7) menggunakan software R.

Normalisasi Data

Pada penelitian ini akan dilakukan normalisasi data karena pada variabel usia saat masuk (X_3), nilai pesantren (X_4), nilai toefl (X_5), dan IPK (X_6) memiliki satuan yang berbeda. Maka dari itu, untuk menyamakan satuan dari variabel-variabel tersebut perlu dilakukan normalisasi data. Normalisasi data yang dilakukan yaitu normalisasi *z-score*. Normalisasi *z-score* menormalkan nilai suatu atribut berdasarkan pada rata-rata dan standar deviasinya. Hasil tahapan normalisasi *z-score* dapat dilihat pada tabel berikut.

Tabel 2. Hasil Normalisasi Data Z-Score

X_3	X_4	X_5	X_6
1,3526	-0,3698	2,1817	-3,0116
-0,3649	-2,5805	-5,2638	-2,9465
-0,3649	-1,0014	-3,1524	-2,4914
-2,0824	-0,3698	-0,9187	-1,3859
⋮	⋮	⋮	⋮

-0,3649	-1,3172	0,1814	0,3697
---------	---------	--------	--------

Pembagian Dataset 10-Fold Cross Validation

Setelah dilakukan *pre-processing* data, tahapan yang perlu dilakukan yaitu membagi data dengan 10-fold cross validation sebagai berikut.

Tabel 3. Pembagian Dataset 10-Fold Cross Validation

Fold	Observasi Ke-											
	1	2	3	4	5	6	7	8	9	10	...	44
1	13	26	37	44	80	83	106	114	115	121	...	
2	2	3	7	17	23	48	51	81	88	119	...	426
3	20	30	35	52	54	69	82	87	90	97	...	
4	8	15	31	38	40	45	65	76	102	116	...	
5	5	6	22	24	39	50	53	55	71	103	...	415
6	16	21	29	36	41	42	61	67	72	74	...	
7	25	27	28	47	49	56	57	58	64	78	...	
8	4	9	12	18	34	46	60	62	70	77	...	
9	10	33	43	59	63	66	68	86	104	107	...	
10	1	11	14	19	32	73	79	96	108	109	...	430

Setelah diperoleh pembagian *dataset* untuk setiap *fold*, selanjutnya akan terbentuk sebanyak 10 model dengan masing-masing pembagian data *training* dan data *testing* sebagai berikut.

1. Model ke-1 yaitu *fold* 2-10 sebagai data *training* dan *fold* 1 sebagai data *testing*.
2. Pembagian data *training* dan data *testing* dilakukan sampai model ke-10 dengan *fold* ke-k sebagai data *testing* dan sisanya sebagai data *training*.

Hasil Pemodelan SVM-GA

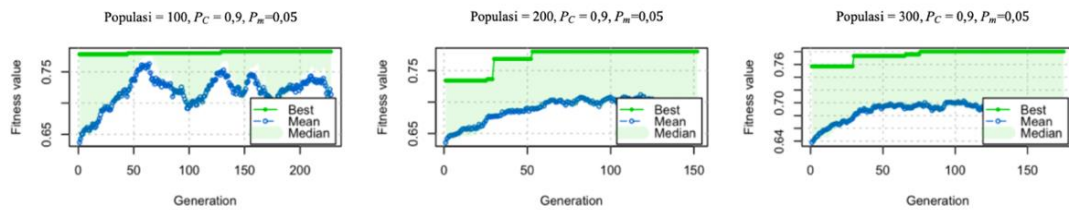
Setelah semua tahapan *pre-processing* dan pembagian *dataset* dilakukan, tahapan selanjutnya adalah melakukan pemodelan SVM-GA. Pemodelan SVM-GA ini digunakan untuk memperoleh model SVM dengan parameter yang optimal dengan bantuan *Genetic Algorithm* (GA). Dalam pemodelan tersebut, parameter SVM yang akan dioptimasi merupakan parameter C dan γ . Parameter tersebut dibatasi dengan rentang parameter C sebesar $[2^{-5}, 2^{15}]$ dan rentang parameter γ sebesar $[2^{-15}, 2^3]$. Pemodelan SVM-GA ini akan dilakukan sebanyak 9 kombinasi populasi, P_c , dan P_m , dengan jumlah populasi sebanyak 100;200;300, P_c sebesar 0,75;0,8;0,9 dan P_m sebesar 0,05. Hasil dari pemodelan SVM-GA adalah sebagai berikut.

Tabel 4. Hasil Pemodelan SVM-GA

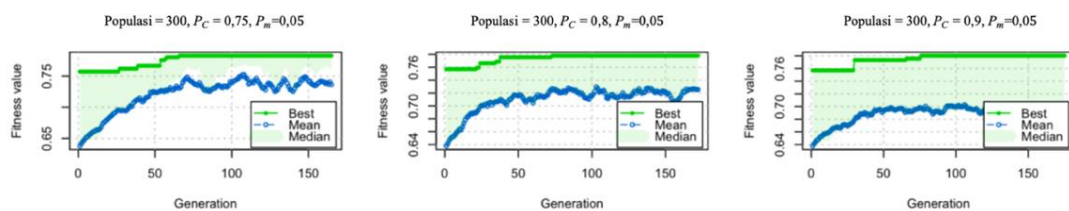
Populasi	P_m	P_c	Parameter Optimal		Rata-Rata Akurasi (%)
			C	γ	
100	0,05	0,75	15851,40241	0,0035	78,2505
		0,8	17283,4915	0,0034	78,2505
		0,9	12512,1145	0,0036	78,0233
200	0,05	0,75	1138,1449	0,0046	78,2611
		0,8	2242,1776	0,0002	78,2664
		0,9	1,1141	0,0346	78,0391
300	0,05	0,75	1442,2231	0,0040	78,2611
		0,8	830,5641	0,0113	77,7907

		0,9	8686,1026	0,0042	78,0233
--	--	-----	-----------	--------	---------

Berdasarkan Tabel 4, dapat dilihat bahwa terjadi penurunan akurasi ketika nilai P_c yang digunakan sebesar 0,9 dan ukuran populasi sebanyak 300. Untuk mengetahui dengan jelas peningkatan solusi dalam setiap generasi pada nilai P_c yang digunakan sebesar 0,9 dan ukuran populasi sebanyak 300 disajikan plot hasil SVM-GA sebagai berikut.



Gambar 1. Plot Hasil SVM-GA Untuk Nilai $P_c = 0,9$



Gambar 2. Plot Hasil SVM-GA Untuk Populasi Sebanyak 300

Gambar 1 menunjukkan bahwa pada probabilitas *crossover* (P_c) sebesar 0,9, nilai *fitness* terbaik pada setiap generasi yang ditunjukkan pada grafik berwarna hijau mengalami peningkatan pada rentang 20-50, namun setelah mengalami peningkatan tersebut terjadi peningkatan yang lambat dan cenderung tetap konstan hingga mencapai konvergen. Hal ini menunjukkan bahwa setelah beberapa generasi, algoritma tidak mampu untuk menemukan solusi yang lebih baik. Berdasarkan gambar tersebut juga dapat dilihat bahwa nilai *fitness* yang terbentuk pada setiap generasi hingga mencapai konvergen tidak memiliki nilai *fitness* yang beragam. Pada Gambar 2 terlihat bahwa dengan ukuran populasi sebesar 300, solusi yang dihasilkan oleh algoritma genetika tidak mengalami peningkatan pada generasi 50-80 dan tetap konstan hingga generasi ke-150. Dengan awal konvergensi yang lebih cepat serta tanpa adanya peningkatan yang terjadi lagi, maka pada kombinasi parameter dengan populasi sebesar 300 mengindikasikan terjadinya konvergensi dini.

Konvergensi dini dapat terjadi karena hilangnya keragaman dalam populasi. Hilangnya keragaman dalam setiap populasi tersebut mengakibatkan proses pencarian solusi yang optimal yang dilakukan algoritma genetika gagal untuk mengeksplorasi solusi secara menyeluruh, sehingga algoritma genetika tidak mampu untuk menemukan solusi yang lebih baik. Maka, untuk menjaga keragaman dan meningkatkan eksplorasi pencarian perlu dilakukan penentuan parameter GA berupa populasi, P_c dan P_m yang optimal. Berdasarkan Tabel 4 diperoleh hasil model terbaik dengan menggunakan 200 populasi, P_m sebesar 0,05, dan P_c sebesar 0,8 dengan parameter C sebesar 2242,1776 dan parameter γ sebesar 0,0002, serta akurasi sebesar 78,2664%. Dengan menggunakan parameter optimal tersebut, diperoleh hasil pengklasifikasian kelulusan tepat waktu mahasiswa Prodi Statistika FMIPA Unisba menggunakan metode SVM-GA adalah sebagai berikut.

Tabel 5. Hasil Klasifikasi Dengan Menggunakan Parameter Optimal Hasil SVM-GA

X_1	X_2	X_3	X_4	X_5	X_6	Y	$\hat{Y}$
Laki-Laki	Luar Jawa	19	83	570	2,61	Tidak Lulus Tepat Waktu	Tidak Lulus Tepat Waktu

Laki-Laki	Luar Jawa	18	76	347	2,63	Tidak Lulus Tepat Waktu	Tidak Lulus Tepat Waktu
Laki-Laki	Jawa	18	81	410	2,77	Tidak Lulus Tepat Waktu	Tidak Lulus Tepat Waktu
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Laki-Laki	Luar Jawa	18	86	517	3,6	Lulus Tepat Waktu	Lulus Tepat Waktu

Evaluasi Hasil Kinerja Model

Dengan menggunakan parameter optimal yang diperoleh, akan dilakukan perhitungan rata-rata *recall*, *precision*, dan *f1-score* dari model ke-1 sampai model ke-10 berdasarkan *confusion matrix* hasil klasifikasi pada data *testing* setiap model. Hasil dari perhitungan tersebut adalah sebagai berikut.

Tabel 6. Hasil Perhitungan Rata-Rata *Recall*, *Precision*, dan *F1-Score*

Model	<i>Recall</i>	<i>Precision</i>	<i>F1-Score</i>
1	77,7778	84,0000	80,7692
2	85,1852	92,0000	88,4615
3	84,0000	87,5000	85,7143
4	76,9231	80,0000	78,4314
5	77,4194	96,0000	85,7143
6	82,6087	76,0000	79,1667
7	80,7692	84,0000	82,3529
8	67,7419	87,5000	76,3636
9	75,0000	87,5000	80,7692
10	75,8621	88,0000	81,4815
Rata-rata	78,3287	86,2500	81,9225

Dengan rata-rata akurasi pada Tabel 4 yang menggunakan parameter optimal SVM-GA yaitu, populasi sebesar 200, P_m sebesar 0,05, dan P_c sebesar 0,8, diperoleh rata-rata akurasi sebesar 78,2664%, yang menunjukkan bahwa dengan variabel independen berupa variabel jenis kelamin (X_1), asal daerah (X_2), usia saat masuk (X_3), nilai pesantren (X_4), nilai toefl (X_5), dan IPK (X_6), dapat membantu model dalam mengidentifikasi kelulusan tepat waktu mahasiswa Prodi Statistika FMIPA Unisba dengan cukup baik (*fair classification*). Berdasarkan Tabel 6, rata-rata *recall* sebesar 78,3287% menunjukkan kemampuan model dalam memanfaatkan informasi dari variabel independen (X) dengan baik. Nilai rata-rata *precision* sebesar 86,2500% menggambarkan tingkat akurasi antara data yang diminta dengan nilai prediksi model. Selain itu, nilai rata-rata *f1-score* sebesar 81,9225% menggambarkan kemampuan model memiliki keseimbangan yang baik antara *precision* dan *recall*.

D. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan menggunakan metode *Support Vector Machine-Genetic Algorithm* (SVM-GA), diperoleh hasil sebagai berikut.

1. Dengan menggunakan algoritma 10-Cross Validation, kernel RBF, dan mengkombinasikan parameter algoritma genetika dengan populasi sebanyak [100;200;300], nilai P_c sebesar [0,75;0,8;0,9], dan nilai P_m sebesar [0,05], dihasilkan 9 model SVM-GA. Dari 9 model SVM-GA tersebut, model terbaik dihasilkan pada model ke-5 dengan parameter algoritma genetika berupa populasi sebanyak 200, nilai P_c sebesar 0,8, nilai P_m sebesar 0,05. Dengan parameter optimal yang dihasilkan dari kombinasi parameter GA tersebut, diperoleh parameter SVM yang optimal yaitu parameter C

sebesar 2242,1776 dan γ sebesar 0,0002.

2. Berdasarkan parameter optimal yang diperoleh dari model SVM-GA, dihasilkan rata-rata akurasi sebesar 78,2664% yang menunjukkan bahwa kemampuan model dalam mengklasifikasikan kelulusan tepat waktu mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Bandung dengan menggunakan variabel jenis kelamin (X_1), asal daerah (X_2), usia saat masuk (X_3), nilai pesantren (X_4), nilai toefl (X_5), dan IPK (X_6), sudah cukup baik (*fair classification*). Selain itu, model memiliki kinerja dengan nilai rata-rata *recall* sebesar 78,3287%, rata-rata *precision* sebesar 86,2500% dan rata-rata *f1-score* sebesar 81,9225%.

Acknowledge

Dalam proses penyusunan penelitian ini, penulis menerima banyak bantuan dan dukungan dari berbagai pihak. Penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak atas segala bentuk bantuan dan dukungan yang telah diberikan. Khususnya kepada Bapak Ilham Faishal Mahdy, S.Stat., M.Stat., sebagai dosen pembimbing yang telah memberikan waktu, ilmu, pengetahuan, kesempatan, dan kemudahan bagi penulis, juga kepada orang tua, keluarga, dan teman yang selalu mendoakan serta memberikan dukungan moral dan materi kepada penulis.

Daftar Pustaka

- [1] Abe, S. (2010). *Support Vector Machines for Pattern Classification* (2nd ed.). Springer-Verlag. https://doi.org/10.1007/978-1-84996-098-4_4
- [2] BAN-PT. (2021). *Pedoman Penilaian Pemantauan dan Evaluasi Peringkat Akreditasi*.
- [3] Fatmawati, & Rifai, N. A. K. (2023). Klasifikasi Penyakit Diabetes Retinopati Menggunakan Support Vector Machine dengan Algoritma Grid Search Cross-validation. *Jurnal Riset Statistika*, 79–86. <https://doi.org/10.29313/jrs.v3i1.1945>
- [4] Gallagher S, Kerry, & Malcolm. (1994). Genetic algorithms: A powerful tool for large-scale nonlinear optimization problems. *Computers & Geosciences*, 20(7), 1229–1236.
- [5] Gen, M., & Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. John Wiley & Sons, Inc.
- [6] Grosan, C., & Abraham, A. (2011). *Computational intelligence : collaboration, fusion and emergence* (Vol. 17). Springer.
- [7] Gunn, S. (1998). *Support Vector Machines for Classification and Regression*.
- [8] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Morgan Kuafmann Publishers.
- [9] Handayani, V. A., Sulistyono, E., Sunarsono, H., Araf, A., & Harahap, D. S. (2024). ARIMA (p, d, q) Modeling for Predicting Exports of Fresh and Chilled Fish Based on Market Conditions and Main Destination Countries : The Case of Indonesia 2012-2022. *Statistika*, 24(1), 83–92. <https://doi.org/10.29313/statistika.v24i1.3376>
- [10] Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2003). *A Practical Guide to Support Vector Classification*. <http://www.csie.ntu.edu.tw/~cjlin>
- [11] Ismail, Z., & Irhamah. (2008). Adaptive permutation-based genetic algorithm for solving VRP with stochastic demands. *Journal of Applied Sciences*, 8(18), 3228–3234. <https://doi.org/10.3923/jas.2008.3228.3234>
- [12] Kusumaningrum, A. P. (2017). Optimasi Parameter Support Vector Machine Menggunakan Genetic Algorithm Untuk Klasifikasi Microarray Data. In *Surabaya: Departemen Statistika FMKSD ITS*. Institut Teknologi Sepuluh Nopember.
- [13] Maryanto Rompon, Hamim Tsalis Soblia, Putri Monika, Atje Setiawan Abdullah, & Budi Nurani Ruchjana. (2023). Identifikasi Autokorelasi Spasial Warisan Budaya Tak Benda di Indonesia Menggunakan Indeks Moran. *Statistika*, 23(2), 156–163.

- <https://doi.org/10.29313/statistika.v23i2.2675>
- [14] Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28(4), 603–614. <https://doi.org/10.1016/j.eswa.2004.12.008>
 - [15] Nugroho, A. S., Witarto, A. B., & Handoko, D. (2013). *Support Vector Machine-Teori dan Aplikasinya dalam Bioinformatika 1*. <http://asnugroho.net>
 - [16] Rindrasari, R. (2017). *Klasifikasi Kelas Risiko Pasien Pneumonia Menggunakan Metode Support Vector Machine-Genetic Algorithm (SVM-GA) Hybrid*. Institut Teknologi Sepuluh Nopember.
 - [17] Wu, C. H., Tzeng, G. H., Goo, Y. J., & Fang, W. C. (2007). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Systems with Applications*, 32(2), 397–408. <https://doi.org/10.1016/j.eswa.2005.12.008>
 - [18] Yenaeng, S., Saelee, S., & Samai, W. (2014). Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms. *International Journal of Information and Education Technology*, 4(2), 132–137. <https://doi.org/10.7763/ijiet.2014.v4.384>
 - [19] Yuniar, P., & Kismiantini. (2023). Analisis Sentimen Ulasan pada Gojek Menggunakan Metode Naive Bayes. *Statistika*, 23(2), 164–175. <https://doi.org/10.29313/statistika.v23i2.2353>