

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339016305>

Penerapan Algoritma Genetika pada Support Vector Machine Sebagai Pengoptimasi Parameter untuk Memprediksi Kesuburan

Article · August 2019

CITATION

1

READS

170

2 authors, including:



Hani Harafani

Sekolah Tinggi Manajemen Informatika an Komputer Nusa Mandiri, Jakarta

11 PUBLICATIONS 11 CITATIONS

SEE PROFILE

Penerapan Algoritma Genetika pada *Support Vector Machine* Sebagai Pengoptimasi Parameter untuk Memprediksi Kesuburan

Hani Harafani¹, Andry Maulana²

Abstract— *Fertility rates in various countries have decreased. The result of the WHO study found 50% the causes of infertility were men caused by a decrease in the quality of semen. In this study, Genetic Algorithm and SVM Methods are used to predict the quality of semen in the Fertility dataset. Based on experiments with 10 iterations, the highest level of accuracy know is SVM+GA(dot kernel) of 89%, then SVM of 88%, followed by Decision Tree 84%, Neural Network 82%, and Naïve Bayes 82%. In Conclusion, GA is proven to increase the accuracy value of SVM with kernel dot which shows a significant difference, although 2 kernel of SVM shows insignificant differences.*

Intisari— Tingkat kesuburan di berbagai Negara mengalami penurunan, Hasil riset WHO mendapatkan 50% penyebab infertilitas adalah pihak pria yang disebabkan oleh menurunnya kualitas semen. Pada penelitian ini Algoritma Genetika dan metode SVM digunakan untuk memprediksi kualitas semen pada dataset *Fertility*. Berdasarkan eksperimen dengan 10 iterasi, didapatkan tingkat akurasi paling tinggi adalah SVM+GA(kernel dot) sebesar 89%, kemudian SVM sebesar 88%, disusul Decision Tree 84%, Neural Network 83%, dan Naïve Bayes 82%. Kesimpulannya GA terbukti dapat meningkatkan akurasi pada SVM dengan kernel dot yang menunjukkan perbedaan yang signifikan, meskipun 2 kernel dari SVM menunjukkan perbedaan yang tidak signifikan.

Kata Kunci— Letakkan 4-8 kata kunci Anda di sini, kata kunci dipisahkan dengan koma.

I. PENDAHULUAN

Tingkat kesuburan menurut Bloom & Souzapoza dalam [1] di berbagai negara mengalami penurunan seperti di Eropa, Italia[2], bahkan di Indonesia[3]. Di Indonesia, Angka infertilitas pasangan suami-istri yang mengalami kesulitan untuk mendapatkan anak sekitar 10%. Hasil riset WHO mendapatkan 50% penyebab infertilitas adalah pihak pria dan yang terbesar adalah karena faktor semen.

Banyak faktor yang menentukan dalam masalah infertilitas produksi semen diantaranya faktor lingkungan dan gaya hidup [4] seperti penggunaan alkohol dan rokok[5], obesitas [6], dan perilaku seksual yang tidak aman[7]

Masalah infertilitas pada sperma juga dapat disebabkan oleh masalah kesehatan alat reproduksi [8] seperti kanker [9]. Pola gaya hidup, lingkungan, dan kesehatan tentu memiliki pengaruh terhadap konsentrasi sperma yang merupakan faktor yang dapat mempengaruhi kualitas sperma[3].

Banyak cara yang dapat ditempuh untuk memperbaiki masalah infertilitas seperti menjaga pola hidup yang sehat. Namun, untuk dapat mengevaluasi tingkat kesuburan pria, menurut Koletis dalam [10] para dokter membutuhkan data yang diperoleh dari analisis semen.

Seiring perkembangan penelitian analisis semen, banyak metode komputasi yang digunakan untuk memprediksi kualitas semen diantaranya Artificial Neural Network, Support Vector Machine, Decision Tree [10], Particle Swarm Optimization, Multilayer Perceptron, *Support Vector Machine* [8], K- Nearest Neighbor [3], dan Auto Regressive Integrater Moving Average [2].

Pada penelitian [10] menunjukkan SVM unggul dibandingkan ANN, dan DT, sementara pada penelitian [8] SVM unggul dibandingkan MLP dan PSO, sedangkan [3] dapat membuktikan bahwa K-NN juga dapat digunakan untuk memprediksi kualitas sperma.

Multilayer perceptron sebagai salah satu model yang paling populer dari Artificial Neural Network [11], memiliki kelebihan untuk menemukan pola dari data yang terlalu rumit untuk diketahui oleh manusia atau dengan teknik komputasi lainnya[12]. Selain itu MLP memiliki kekurangan yaitu menemukan pola bila data berdimensi tinggi atau sering disebut dengan “kutukan dimensionalitas” [13], dan *Overfitting* [14].

Support Vector Machine (SVM) memiliki keunggulan dibandingkan metode MLP yaitu : SVM dapat mengatasi masalah klasifikasi dan regresi linier maupun nonlinier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi [15], dan baik untuk mengatasi kutukan dimensionalitas [16]. SVM juga memiliki akurasi tinggi dan tingkat kesalahan yang relative kecil, kemampuan untuk mengatasi *overfitting*, tidak membutuhkan data yang terlalu besar, dan dapat digunakan untuk melakukan prediksi. Berdasarkan beberapa kelebihan SVM pada ulasan yang telah disebutkan, maka SVM dirasa sangat cocok diterapkan untuk memprediksi kualitas semen.

SVM selain berkemampuan tinggi dalam memprediksi, SVM juga memiliki kelemahan, yaitu sulitnya dalam menentukan nilai parameter yang optimal [17]–[19]. Beberapa algoritma pun banyak direkomendasikan oleh peneliti dunia untuk mengoptimasi parameter pada machine learning, seperti: *particle swarm optimization*(PSO)[20],

¹ Jurusan Teknik Informatika STMIK Nusa Mandiri Jakarta, Jl. Damai No.8, Warung Jati Barat, Margasatwa, Ragunan, Pasar Minggu, Jakarta Selatan 12450 Indonesia (telp: 021-78839513; e-mail: haniharafani@gmail.com)

²Jurusan Sistem Informasi STMIK Nusa Mandiri Jakarta, Jl. Damai No.8, Warung Jati Barat, Margasatwa, Ragunan, Pasar Minggu, Jakarta Selatan 12450 Indonesia (telp: 021-78839513; e-mail: andry.ayz@nusamandiri.ac.id)

simulated annealing (SA)[21], dan *Genethic Algorithm* (GA)[22].

Simulated Annealing (SA) efektif pada pemuatan masalah optimasi pola, namun SA memiliki kecenderungan untuk terjebak dalam minimum local ketika suhu anil rendah (tingkat anil cepat) dan semakin tidak konvergen ketika suhu anil tinggi (tingkat anil lambat) [23], selain itu PSO juga sulit mendapatkan nilai yang optimum dalam mengoptimasi lebih dari sepuluh parameter. *Genethic Algorithm* (GA) dapat mengatasi masalah yang nonlinier dengan diskontinuitas dan minima local secara efisien, serta lebih efisien dalam mengoptimasi lebih dari sepuluh parameter [4]

Dalam penelitian ini kami mengusulkan algoritma genetika (GA) untuk melakukan optimasi parameter pada SVM untuk meningkatkan akurasi dalam memprediksi kualitas semen.

II. KAJIAN LITERATUR

A. Dataset Kesuburan

Dataset kesuburan diambil dari situs *UCI learning repositories* yang mana penelitian pada dataset tersebut telah dilakukan oleh [4]. Dataset ini didapatkan dari 100 relawan muda di Universitas Alicante yang mana umur relawan berkisar antara 18-36 Tahun. Pengambilan sampel air mani yang dilakukan oleh WHO dari relawan ialah dalam keadaan yang sama yaitu sebelum pengambilan sampel, dalam waktu 3-6 hari relawan dilarang untuk melakukan hubungan sexual. Bagi relawan yang ketahuan menderita varises pada kemaluannya (varicocele) maka dikeluarkan dari analisa statistik. Variable pada dataset ini terdiri dari 4 jenis variable yaitu variable numerik (umur, tahun analisis, jumlah rata-rata penggunaan rokok per hari), variable biner (penyakit dari kecil, asident atau trauma serius, intervensi bedah), variable tiga atribut bebas (panas tinggi di tahun terakhir, kebiasaan merokok, frekuensi penggunaan alcohol, berapa jam duduk per hari), dan variable dengan 4 atribut bebas (musim pada saat analisis berlangsung).

Musim pada saat analisis berlangsung diwakili dengan range (-1,-0,33,0,33,1) dimana urutannya adalah musim salju, musim semi, musim panas, dan musim gugur.

Penyakit masakecil (cacar air, campak, gondok, polio) jika ada maka bernilai 0, jika tidak ada bernilai 1. asident atau trauma serius jika ada bernilai 0, jika tidak bernilai 1. Intervensi bedah jika ada bernilai 0, jika tidak bernilai 1. Demam tinggi pada Tahun terakhir diberi range (-1,0,1) dengan urutan 1) kurang dari tiga bulan yang lalu, 2) lebih dari tiga bulan yang lalu, 3) tidak demam. Kebiasaan merokok diwakili oleh range (-1,0,1) dengan urutan 1) tidak pernah, 2) jarang, 3) setiap hari.

Untuk variable umur, frekuensi konsumsi alcohol, dan jumlah jam yang dihabiskan untuk duduk terdapat perhitungan sendiri seperti misalnya maksimal konsumsi

rokok perhari adalah 25 maka seseorang yang mengonsumsi rokok sebanyak 25 per hari diberi nilai 1, jika seseorang mengonsumsi rokok 14 per hari maka perhitungannya menjadi $14/25 = 0.56$ Seperti yang ditunjukkan pada Tabel 1.

B. Support Vector Machine

Support Vector Machine (SVM) secara konseptual adalah mesin linear yang dilengkapi dengan fitur spesial [24], dan didasarkan pada metode minimalisasi resiko struktural [25], serta teori pembelajaran statistik. SVM telah banyak diteliti dalam komunitas data mining dan pembelajaran mesin (*Machine Learning*) selama sepuluh tahun terakhir [26]. Dua sifat khusus dari SVM yaitu (1) mencapai generalisasi yang tinggi dengan memaksimalkan margin, dan (2) mendukung pembelajaran yang efisien dari fungsi nonlinier pada trik kernel sehingga membuat kinerja generalisasinya baik dalam menyelesaikan masalah pengenalan pola [24].

Untuk masalah klasifikasi SVM mencoba untuk mencari garis pemisah yang optimal yang diekspresikan sebagai kombinasi linier dari subset data pelatihan dengan menyelesaikan masalah keterbatasan linier pemrograman kuadrat (QP) dengan margin maksimum antara dua kelas.

Tujuan utama dari SVM adalah memperkirakan fungsi klasifikasi dengan menggunakan data pelatihan input-output dari dua kelas $(x_1, y_1), \dots, (x_n, y_n) \in R^m \times \{\pm 1\}$ [27]. Tujuan dari fungsi klasifikasi adalah untuk membentuk persamaan hyperplane yang membagi data pelatihan dan meninggalkan semua titik dari kelas yang sama pada sisi yang sama sambil memaksimalkan jarak minimum antara hyperplane dan masing-masing dari dua kelas (w, b). Dimana w mewakili vector bobot yang mewujudkan margin fungsional 1 pada titik positif x^+ serta titik negatif x^- dan matrik geometric dapat dihitung dengan rumus berikut:

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1, \dots, m \quad (1)$$

Hyperplane optimal $w \cdot x + b = 0$ secara geometris setara dengan memaksimalkan margin yaitu jarak antara dua bidang sejajar $w \cdot x + b = 1$ dan $w \cdot x + b = -1$. Jarak panjang Euclidean dari margin adalah $2/\|w\|^2$, dimana $\|w\|^2 = \sum_{i=1}^m w_i^2$. Margin maksimum juga merupakan 2-norm $\|w\|^2$ tergantung pada kendalanya. Oleh karena itu permasalahan ini dapat diformulasikan dengan rumus:

$$\min_{w, b} \frac{\|w\|^2}{2} \quad (2)$$

Tergantung pada $y_i(w \cdot x_i + b) \geq 1$

Dikarenakan class jarang sekali dapat dipisahkan secara linier, generalisasi permasalahan bidang yang optimal dibutuhkan. Dengan demikian, satu set variable n yang mengukur variasi kendala

ditambahkan untuk setiap titik. Formulasi mutasi akhir adalah

$$\min_{w,b,\xi} \frac{\|w\|^2}{2} + \frac{c}{m} \sum_{i=1}^m \xi_i \quad (3)$$

Tergantung pada $y_i(w \cdot x_i + b) + \xi_i \geq 1$

$$\xi_i \geq 0 \quad i = 1, \dots, m$$

Dimana b merupakan bias[28], parameter w dan b adalah parameter yang perlu ditentukan nilainya agar dapat memberikan fungsi yang terbaik untuk memetakan input ke data output. Untuk kasus pemisahan yang tidak linier, tidak ada hyperplane yang dapat digunakan secara sempurna untuk memisahkan dua bidang[29]. Oleh karena itu variable slack diperkenalkan. Rumus yang harus dipecahkan menjadi:

$$\max_{w,b} \frac{2}{\|w\|} + C \sum_{i=1}^m \xi_i \quad (4)$$

Tergantung pada $y_i(w \cdot x_i + b) - \xi_i \geq 0$

Dimana C merupakan parameter penalty

Untuk kasus terpisah tidak linier, ide dasarnya adalah untuk memproyeksikan dataset x_1 pada ruang fitur berdimensi tinggi cara yang nonlinier menggunakan fungsi kernel. Sampai saat ini banyak fungsi kernel yang disugestikan untuk mengatasi permasalahan ini diantaranya, *Radial Basis Function* (RBF) yang paling banyak digunakan dan dilakukan pada banyak kasus. Rumus dari RBF adalah:

$$k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (5)$$

Dimana parameter γ menunjukkan lebar dari kernel Gaussian.

TABEL I
POTONGAN DATA DARI DATASET KESUBURAN

Seas on	Age	Chil d ish dis eases	Acci dent or seri ous trau ma	Su rg ic al in te rv en t io n	Hi gh fe ve rs in th e st ye ar	Fre que ncy of alco hol cons ump tion	S m ok in g ha bi t	Nu mbe r of hou rs spen t sitti ng per day	Di agno sis
-0,3	0,69	0	1	1	0	0	0,88	N	88

-0,3	0,94	1	0	1	0	0,8	1	0,31	O
-0,3	0,5	1	0	0	0	1	-1	0,5	N
-0,3	0,75	0	1	1	0	1	-1	0,38	N
-0,3	0,67	1	1	0	0	0,8	-1	0,5	O
-0,3	0,67	1	0	1	0	0,8	0	0,5	N
-0,3	0,67	0	0	0	-1	0,8	-1	0,44	N
-0,3	1	1	1	1	0	0,6	-1	0,38	N
1	0,64	0	0	1	0	0,8	-1	0,25	N
1	0,61	1	0	0	0	1	-1	0,25	N

C. Algoritma Genetika

Algoritma genetika merupakan algoritma evolusioner yang paling populer [30] yang mana algoritma ini menggunakan prinsip dasar dari seleksi alam yang diperkenalkan oleh Charles Darwin. Algoritma genetika diterapkan sebagai pendekatan untuk mengidentifikasi pencarian nilai dan solusi bagi berbagai permasalahan optimasi [24]. Terdapat beberapa kelebihan dari algoritma genetika dibandingkan algoritma optimasi tradisional lainnya, dua diantaranya yaitu kemampuan untuk menangani permasalahan kompleks dan paralel. Algoritma genetika dapat menganalisis berbagai macam optimasi tergantung pada fungsi objektifnya (*fitness*) apakah seimbang atau tidak seimbang, linier atau tidak linier, berkesinambungan atau tak berkesinambungan, atau dengan *random noise*. Fungsi *fitness* [31] ditunjukkan pada persamaan 6

$$fitness = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6)$$

Dimana $\hat{y}_i$ merupakan nilai prediksi, dan y_i merupakan nilai asli dan dari sampel dataset. N adalah jumlah sampel total. Intisari dari algoritma genetika mencakup pengkodean/penyandian fungsi optimasi sebagai array berisi bit-bit atau karakter berupa sting untuk menggambarkan kromosom, operasi manipulasi string dengan operator genetik, dan seleksi sesuai dengan *fitness*, dengan tujuan untuk menemukan solusi yang baik dan optimal terhadap masalah yang sedang dihadapi. Algoritma genetika memiliki tiga operator genetik utama yaitu *crossover* (proses penukaran kromosom), mutasi (proses penggantian salah satu solusi dengan nilai fitness yang tinggi untuk lulus ke generasi berikutnya). Terdapat langkah-langka yang sering dilakukan untuk menyelesaikan permasalahan-permasalahan dalam optimasi:

1. Inisialisasi Populasi
2. Evaluasi Populasi
3. Seleksi populasi
4. Proses penyilangan kromosom (*crossover*)
5. Evaluasi populasi baru
6. Selama syarat belum terpenuhi ulangi dari langkah

III. METODOLOGI PENELITIAN

Penelitian telah didefinisikan oleh pendidikan tinggi dewan pendanaan Inggris sebagai penyelidikan yang dilakukan untuk mendapatkan pengetahuan dan pemahaman [32]. Ada empat metode penelitian yang paling umum digunakan yaitu: *action research*, *experiment*, *case study*, dan *survey*. Pada penelitian ini metodologi yang digunakan adalah *experiment*, dimana penelitian ini menyelidiki beberapa variabel menggunakan tes tertentu yang dikendalikan sendiri oleh peneliti. Metode yang digunakan untuk memprediksi adalah *Support Vector Machine* dengan Algoritma genetika sebagai pengoptimasi parameter dengan langkah-langkah sebagai berikut:

1. Pengumpulan Data

Data diambil dari situs *UCI learning repositories* dalam format file .txt, kemudian disalin ke dalam format .xls

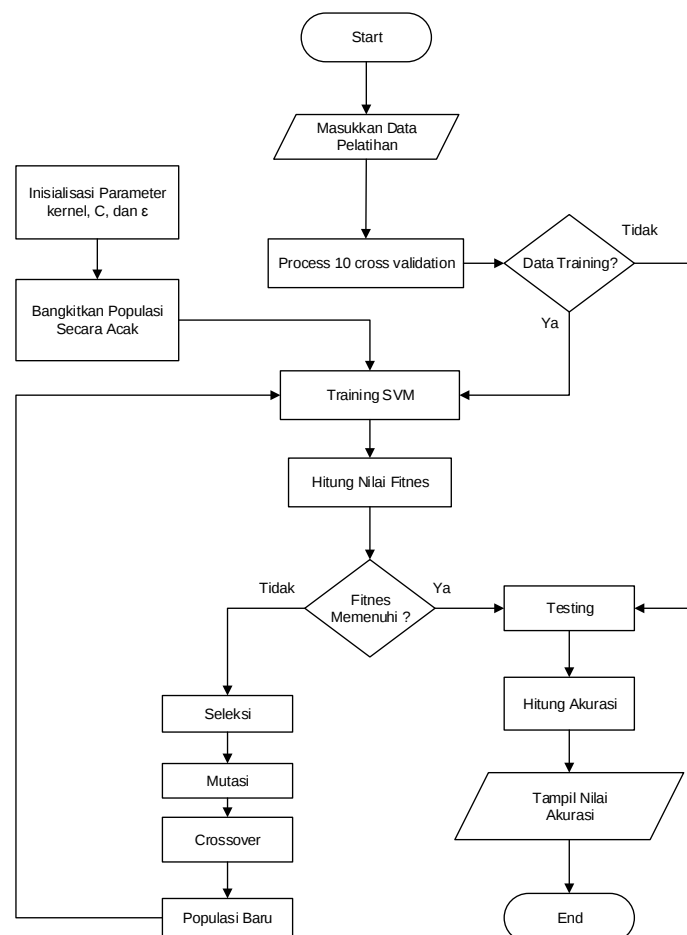
2. Metode yang diusulkan

Metode yang diusulkan yaitu dengan menerapkan algoritma genetika untuk pencarian nilai parameter untuk diterapkan pada proses training SVM. Bermula dengan menerapkan *10-fold cross validation* untuk membagi dataset menjadi data *training* dan data *testing*,

kemudian menerapkan algoritma genetika untuk mencari nilai parameter c , ϵ , dan γ untuk diterapkan pada proses *training*, dan *testing*, kemudian model prediksi yang dihasilkan oleh SVM pada proses training diterapkan untuk menguji data *testing*, kemudian hasil evaluasi model dengan cara membandingkan nilai Accuracy.

Akurasi prediksi keseluruhan percobaan pada penelitian ini ditunjukkan pada Gbr 5. Metrik yang kami gunakan untuk mengukur akurasi prediksi adalah Accuracy yang didefinisikan pada persamaan 7.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (7)$$



Gbr 1. Metode Penelitian

3. Eksperiment dan Pengujian Model

Tahapan eksperimen pada penelitian ini adalah:

- Menyiapkan dataset untuk eksperimen
- Pembagian dataset menjadi dua, yaitu dataset *training* dan dataset *testing* dengan menerapkan 10-fold *cross validation* untuk membagi data menjadi 90 Training dan 10 Testing
- Merancang arsitektur SVM dengan nilai parameter yang berbeda.
- Melakukan training dan testing terhadap SVM dengan kernel yang berbeda (dot, polynomial, radial) dan mencatat nilai Accuracy.
- Melakukan training dan testing terhadap metode klasifikasi lainnya

IV. HASIL PENELITIAN DAN PEMBAHASAN

Eksperimen ini dilakukan menggunakan komputer personal Intel Core i3, 4 GB RAM, 500 GB HDD, Sistem Operasi Windows 10, dan Rapidminer 6.0.

Penelitian ini dilakukan dalam tiga tahapan. Tahap pertama untuk mendapatkan hasil akurasi prediksi kualitas sperma yang lebih akurat, kami membandingkan hasil eksperimen antara data yang diproses menggunakan metode SVM yang menggunakan parameter SVM secara manual dengan data yang diproses menggunakan metode SVM yang parameternya telah diatur dengan GA pada masing-masing kernel.

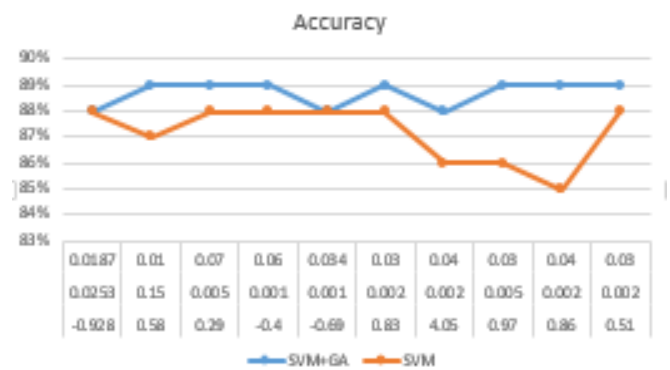
Tahap kedua, untuk mengidentifikasi kernel terbaik pada SVM, kami melakukan perbandingan rata-rata akurasi dari 10 iterasi.

Tahap ketiga untuk mengetahui kehandalan model SVM+GA kami melakukan perbandingan akurasi antara SVM+GA dengan metode klasifikasi lainnya seperti *Neural Network*, *Naïve Bayes*, *Decision Tree*, dan juga metode yang hybrid yang telah digunakan oleh peneliti sebelumnya dalam memprediksi data Seminal Quality dari UCI Learning Repositories.

TABEL II
OPTIMASI PARAMETER PADA KERNEL DOT

C			Epsilon			Gamma			accuracy
Min	Max	Optimal	Min	Max	Optimal	Min	Max	Optimal	
-	0.0	-	0.0	0.0	0.02	0.0	0.0	0.01	88%
1.0	0.1	0.928	1	5	53	1	5	87	
-	2	0.58	0.0	0.5	0.15	0.0	0.0	0.01	89%
0.9			2			1	5		
-	1	0.29	0.0	0.0	0.00	0.0	0.1	0.07	89%
0.5			0.1	0.5	5	5			
-	-	-0.4	0.0	0.0	0.00	0.0	0.0	0.06	89%
0.5	0.4		0.1	0.2	1	5	9		
-	-	-0.69	0	0.0	0.00	0.0	0.0	0.03	88%
0.9	0.5		5	1	09	5	4		

-	1	0.83	0.0	0.0	0.00	0.0	0.0	0.03	89%
0.2			0.1	0.2	2	1	5		
1	5	4.05	0.0	0.0	0.00	0.0	0.0	0.04	88%
			0.2	0.5	2	1	5		
-	2	0.97	0.0	0.0	0.00	0.0	0.0	0.03	89%
0.1			0.2	0.5	5	1	5		
-	1	0.86	0.0	0.0	0.00	0.0	0.0	0.04	89%
0.1			0.2	0.5	2	1	5		
-	1	0.51	0.0	0.0	0.00	0.0	0.0	0.03	89%
0.0			0.2	0.5	2	1	5		
1									



Gbr 2. Perbandingan Akurasi SVM dengan SVM+GA pada kernel Dot

Nilai akurasi tertinggi yaitu 89% diperoleh dari percobaan SVM (kernel dot) dan GA dengan pencarian range nilai parameter $c = -0.9 \frac{s}{d} 2$, $\epsilon = 0.02 \frac{s}{d} 0.05$, dan $\gamma = 0.01$ s/d 0.09 Seperti yang dapat dilihat pada Gambar 2.

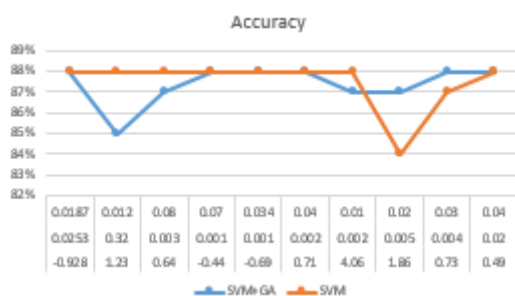
TABEL III
OPTIMASI PARAMETER PADA KERNEL POLYNOMIAL

C			Epsilon			Gamma			accuracy
Min	Max	Optimal	Min	Max	Optimal	Min	Max	Optimal	
-	0.0	-	0.0	0.0	0.025	0.0	0.0	0.018	88%
1.0	0.1	0.928	1	5		1	5	7	
-	2	1.23	0.0	0.5	0.32	0.0	0.0	0.012	85%
0.9			2			1	5		
-	1	0.64	0.0	0.0	0.003	0.0	0.1	0.08	87%
0.5			0.1	0.5		5			
-	-	-0.44	0.0	0.0	0.001	0.0	0.0	0.07	88%
0.5	0.4		0.1	0.2		5	9		
-	-	-0.69	0	0.0	0.001	0.0	0.0	0.034	88%
0.9	0.5			5		0.09	5		
-	1	0.71	0.0	0.0	0.002	0.0	0.0	0.04	88%
0.2			0.1	0.2		1	5		
1	5	4.06	0.0	0.0	0.002	0.0	0.0	0.01	87%
			0.2	0.5		1	5		
-	2	1.86	0.0	0.0	0.005	0.0	0.0	0.02	87%
0.1			0.2	0.5		1	5		
-	1	0.73	0.0	0.0	0.004	0.0	0.0	0.03	88%
0.9			0.2	0.5		1	5		

1	1	0.49	0.0	0.0	0.02	0.0	0.0	0.04	88%
0.			02	05		1	5		
01									

Nilai akurasi tertinggi yaitu 88% diperoleh dari percobaan SVM (kernel polynomial) dan GA dengan pencarian range nilai parameter $c = -0.5 \frac{s}{d} 1, \varepsilon = 0.01 \frac{s}{d} 0.05$, dan $\gamma = 0.01s/d0.09$ Seperti yang dapat dilihat pada Gbr 3.

Nilai akurasi tertinggi yaitu 89% diperoleh dari percobaan SVM (kernel Radial) dan GA dengan pencarian range nilai parameter $c = -0.1 \frac{s}{d} 0.001, \varepsilon = 0.01 \frac{s}{d} 0.05$, dan $\gamma = 0.01s/d0.05$ Seperti yang dapat dilihat pada Gambar 4.



Gbr 3. Perbandingan Akurasi SVM dengan SVM+GA pada kernel Polynomial

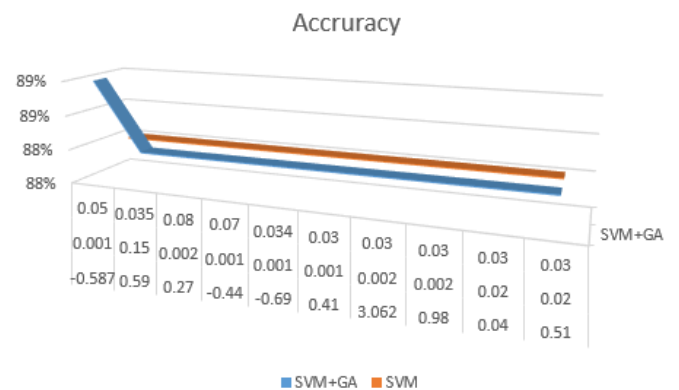
TABEL II
OPTIMASI PARAMETER PADA KERNEL RADIAL

C			Epsilon			Gamma			accuracy
Mi	Ma	Opti	Mi	Ma	Opti	Mi	Ma	Opti	
n	x	mal	n	x	mal	n	ax	mal	
-	0.0	-	0.0	0.0	0.00	0.0	0.	0.05	89%
1.0	01	0.58	1	5	1	1	05		
-	2	0.59	0.0	0.5	0.15	0.0	0.	0.03	88%
0.9			2			1	05	5	
-	1	0.27	0.0	0.0	0.00	0.0	0.	0.08	88%
0.5			01	05	2	5	1		
-	-	-0.44	0.0	0.0	0.00	0.0	0.	0.07	88%
0.5	0.4		01	02	1	5	09		
-	-	-0.69	0	0.0	0.00	0.0	0.	0.03	88%
0.9	0.5		5	1		09	05	4	
--									
0.5									
-	1	0.41	0.0	0.0	0.00	0.0	0.	0.03	88%
0.2			01	02	1	1	05		
1	5	3.06	0.0	0.0	0.00	0.0	0.	0.03	88%
		2	02	05	2	1	05		
-	2	0.98	0.0	0.0	0.00	0.0	0.	0.03	88%
0.1			02	05	2	1	05		
-	1	0.04	0.0	0.0	0.02	0.0	0.	0.03	88%
0.1			02	05		1	05		
-	1	0.51	0.0	0.0	0.02	0.0	0.	0.03	88%
0.0			02	05		1	05		

Pada Tahap kedua, untuk membuktikan signifikansi antara metode SVM dengan SVM+GA pada masing-masing kernel, kami melakukan uji t sampel berpasangan dengan membandingkan nilai rata-rata masing-masing akurasi dari 10 iterasi. Uji beda dilakukan untuk menguji hipotesa:

H_0 = Tidak ada perbedaan nilai rata-rata Akurasi antara model SVM yang dioptimasi secara manual dengan model SVM yang telah dioptimasi dengan GA

H_1 = Terdapat perbedaan nilai rata-rata RMSE antara model SVM yang dioptimasi secara manual dengan model SVM yang telah dioptimasi dengan GA



Gbr 4. Perbandingan Akurasi SVM dengan SVM+GA pada kernel Radial

Setelah melakukan uji beda pada dua sample antara akurasi SVM dengan akurasi SVM+GA pada kernel dot, diketahui nilai t hitung yang diwakili oleh t stat sebesar -3,73705, sedangkan nilai t table yang diwakili oleh t critical two tail sebesar 2,262157, maka dapat dipastikan nilai t hitung < t tabel yang artinya H_0 ditolak dan H_1 diterima, sedangkan nilai probabilitas yang ditunjukkan oleh $P(T \leq t)$ two tail sebesar 0.004647 lebih kecil dari pada 0.05 yang artinya ada perbedaan yang signifikan dari rata-rata Akurasi model SVM dengan SVM+GA menggunakan kernel dot. Seperti pada Tabel 5

TABEL V
HASIL UJI BEDA STATISTIK SVM DENGAN SVM+GA PADA KERNEL DOT

	Variable 1	Variable 2
Mean	0,872	0,887
Variance	0,000129	2,33E-05
Observations	10	10
Pearson Correlation	-0,08104	
Hypothesized Mean Difference	0	

df	9
t Stat	-3,73705
P(T<=t) one-tail	0,002323
t Critical one-tail	1,833113
P(T<=t) two-tail	0,004647
t Critical two-tail	2,262157

Setelah melakukan uji beda pada dua sample antara akurasi SVM dengan akurasi SVM+GA pada kernel polynomial, diketahui nilai t hitung yang diwakili oleh t stat sebesar 0,207514, sedangkan nilai t table yang diwakili oleh t critical two tail sebesar 0.840228, maka dapat dipastikan nilai t hitung < t tabel yang artinya H_0 ditolak dan H_1 diterima, sedangkan nilai probabilitas yang ditunjukkan oleh P(T<=t) two tail sebesar 0.840228 yang mana lebih besardari pada 0.05 yang artinya ada perbedaan yang tidak signifikan dari rata-rata Akurasi model SVM dengan SVM+GA menggunakan kernel polynomial. Seperti pada Tabel 6.

TABEL VI
HASIL UJI BEDA STATISTIK SVM DENGAN SVM+GA PADA KERNEL POLYNOMIAL

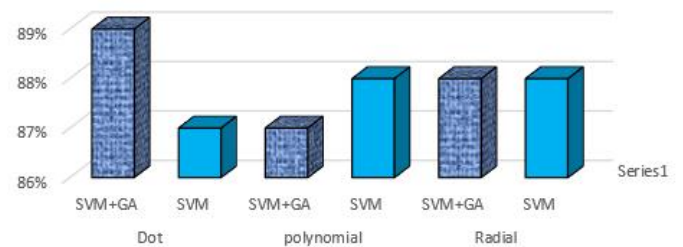
	Variable 1	Variable 2
Mean	0,875	0,874
Variance	0,000161	9,33E-05
Observations	10	10
Pearson Correlation	0,09061	
Hypothesized Mean Difference	0	
df	9	
t Stat	0,207514	
P(T<=t) one-tail	0,420114	
t Critical one-tail	1,833113	
P(T<=t) two-tail	0,840228	
t Critical two-tail	2,262157	

TABEL VII
HASIL UJI BEDA STATISTIK SVM DENGAN SVM+GA PADA KERNEL RADIAL

	Variable 1	Variable 2
Mean	0,88	0,881
Variance	1,37E-32	0,00001
Observations	10	10
Pearson Correlation	0	
Hypothesized Mean Difference	0	

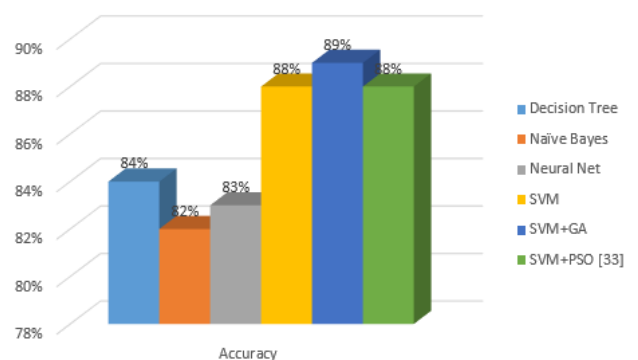
df	9
t Stat	-1
P(T<=t) one-tail	0,171718
t Critical one-tail	1,833113
P(T<=t) two-tail	0,343436
t Critical two-tail	2,262157

Sedangkan Setelah melakukan uji beda pada dua sample antara akurasi SVM dengan akurasi SVM+GA pada kernel Radial, diketahui nilai t hitung yang diwakili oleh t stat sebesar -1, sedangkan nilai t table yang diwakili oleh t critical two tail sebesar 2,262157, maka dapat dipastikan nilai t hitung < t tabel yang artinya H_0 ditolak dan H_1 diterima, sedangkan nilai probabilitas yang ditunjukkan oleh P(T<=t) two tail sebesar 0.343436 yang mana lebih besardari pada 0.05 yang artinya ada perbedaan yang tidak signifikan dari rata-rata Akurasi model SVM dengan SVM+GA menggunakan kernel radial. Seperti pada Tabel 7.



Gbr 5. Perbandingan Nilai Rata-rata SVM dengan SVM+GA

Pada Tahap ketiga, dataset Fertility juga dipercobakan pada metode klasifikasi yang lainnya, metode yang akan dibandingkan adalah metode SVM, SVM+GA, Decision Tree, Naïve Bayes, dan Neural Network, serta hasil akurasi dari penelitian sebelumnya SVM+PSO [33] yang dapat dilihat pada Gbr 6.



Gbr 5. Perbandingan Nilai Akurasi SVM+GA dengan metode klasifikasi lainnya

Pada Gbr 6. Menunjukkan nilai akurasi tertinggi didapat dengan kombinasi SVM+GA, diposisi kedua oleh hybrid metode yaitu SVM+PSO[33] yang mana setara dengan metode SVM tanpa optimasi parameter, kemudian disusul oleh Descision Tree, Neural Network, dan Naïve Bayes.

V. PENUTUP

Berdasarkan seluruh percobaan pada penelitian ini kami menemukan bahwa SVM mampu mengatasi masalah klasifikasi bahkan pada data yang binomial. Pada kasus prediksi kualitas seminal, kernel dot mempunyai tingkat perbedaan akurasi yang signifikan dibandingkan dengan kernel yang lain, sedangkan kernel polynomial dan radia tidak memiliki perbedaan akurasi yang signifikan setelah dilakukan optimasi parameter. Algoritma genetika pada penelitian ini juga terbukti dapat memberikan referensi parameter yang optimal bagi SVM untuk dapat meningkatkan akurasi prediksi pada dataset *fertility*. Perbedaan akurasi yang tidak signifikan pada kedua kernel SVM menjadi kekurangan dalam penelitian ini yang mana kami percaya, dengan mengoptimalkan mutasi dan seleksi pada algoritma genetika diyakini dapat turut serta dalam meningkatkan optimasi parameter pada SVM, selain itu algoritma yang terinspirasi dari alam kami percaya dapat menjadi isu dalam optimasi parameter SVM selanjutnya.

UCAPAN TERIMAKASIH

Akhir kata penulis mengucapkan terimakasih pada para peneliti sebelumnya, yang banyak menginspirasi kami dalam melakukan penelitian ini, tak lupa kami ucapkan terimakasih pada grup penelitian Intelligent Systems dan kepada semua tokoh pada referensi kami atas kontribusi tidak langsung pada penelitian ini, tanpa referensi ini, kami tidak mungkin memiliki pengetahuan dan ide mengenai penelitian ini

REFERENSI

- [1] N. Novianti, S. Fauziati, and I. Hidayah, "Analisis Pengaruh Seleksi Fitur pada Klasifikasi Konsentrasi ... (Novianti dkk.)," in *SNST*, 2015, pp. 160–165.
- [2] S. De Iaco and S. Maggio, "A dynamic model for age-specific fertility rates in Italy," *Spat. Stat.*, vol. 17, pp. 105–120, 2016.
- [3] R. Yepriyanto, Kustanto, and Y. R. W. Utami, "SISTEM DIAGNOSA KESUBURAN SPERMA DENGAN METODE K-NEAREST NEIGHBOR (K-NN)," *SINUS*, pp. 33–44, 2015.
- [4] D. Gil, J. Luis, J. De Juan, M. J. Gomez-torres, and M. Johnsson, "Expert Systems with Applications Predicting seminal quality with artificial intelligence methods," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12564–12573, 2012.
- [5] A. C. Martini, R. I. Molina, D. Estofan, D. Senestrari, M. Fiol De Cuneo, and R. D. Ruiz, "Effects of alcohol and cigarette consumption on human seminal quality," *Fertil. Steril.*, vol. 82, no. 2, pp. 374–377, 2004.
- [6] A. C. Martini *et al.*, "Overweight and seminal quality: A study of 794 patients," *Fertil. Steril.*, vol. 94, no. 5, pp. 1739–1743, 2010.
- [7] W. Ombelet, N. Dhont, A. Thijssen, E. Bosmans, and T. Kruger, "Semen quality and prediction of IUI success in male subfertility: A systematic review," *Reprod. Biomed. Online*, vol. 28, no. 3, pp. 300–309, 2014.
- [8] A. J. Sahoo and Y. Kumar, "Seminal quality prediction using data mining methods," vol. 22, pp. 531–545, 2014.
- [9] M. C. Lindgren, "Male Infertility," *Physician Assist. Clin.*, vol. 3, no. 1, pp. 139–147, 2018.
- [10] D. Gil, J. L. Girela, J. De Juan, M. J. Gomez-torres, and M. Johnsson, "Expert Systems with Applications Predicting seminal quality with artificial intelligence methods," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12564–12573, 2012.
- [11] P. Singh and B. Borah, "International Journal of Approximate Reasoning Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization," *Int. J. Approx. Reason.*, vol. 55, no. 3, pp. 812–833, 2014.
- [12] I. Yilmaz and O. Kaynar, "Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5958–5966, May 2011.
- [13] S. Pan, S. Iplikci, K. Warwick, and T. Z. Aziz, "Parkinson's Disease tremor classification – A comparison between Support Vector Machines and neural networks," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10764–10771, Sep. 2012.
- [14] J. Rynkiewicz, "General bound of overfitting for MLP regression models," *Neurocomputing*, vol. 90, pp. 106–110, Aug. 2012.
- [15] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*. 2010.
- [16] X. Wang, J. Wen, Y. Zhang, and Y. Wang, "Optik Real estate price forecasting based on SVM optimized by PSO," *Opt. - Int. J. Light Electron Opt.*, vol. 125, no. 3, pp. 1439–1443, 2014.
- [17] I. Ilhan and G. Tezel, "A genetic algorithm-Support Vector Machine method with parameter optimization for selecting the tag SNPs," *J. Biomed. Inform.*, vol. 46, no. 2, pp. 328–40, Apr. 2013.
- [18] S. Raghavendra. N and P. C. Deka, "Support Vector Machine applications in the field of hydrology: A review," *Appl. Soft Comput.*, vol. 19, pp. 372–386, Jun. 2014.
- [19] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5197–5204, May 2011.
- [20] L. Wang, B. Yang, and Y. Chen, "Improving particle swarm optimization using multi-layer searching strategy," *Inf. Sci. (Ny)*, vol. 274, pp. 70–94, Aug. 2014.
- [21] Z. Jia, J. Ma, F. Wang, and W. Liu, "Hybrid of simulated annealing and SVM for hydraulic valve characteristics prediction," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8030–8036, Jul. 2011.
- [22] X. Guo, D. Li, and A. Zhang, "Improved Support Vector Machine Oil Price Forecast Model Based on Genetic Algorithm Optimization Parameters," *AASRI Procedia*, vol. 1, pp. 525–530, Jan. 2012.

- [23] A. Zameer, S. M. Mirza, and N. M. Mirza, "Core loading pattern optimization of a typical two-loop 300MWe PWR using Simulated Annealing (SA), novel crossover Genetic Algorithms (GA) and hybrid GA(SA) schemes," *Ann. Nucl. Energy*, vol. 65, pp. 122–131, Mar. 2014.
- [24] F. Gorunescu, *Intelligent Systems Reference Library*. 2011.
- [25] Dua, *Data Mining and Machine Learning in Cybersecurity*. 2011.
- [26] H. Yu and S. Kim, *SVM Tutorial: Classification, Regression, and Ranking*. 2012.
- [27] J.-S. Chou, M.-Y. Cheng, Y.-W. Wu, and A.-D. Pham, "Optimizing parameters of Support Vector Machine using fast messy genetic algorithm for dispute classification," *Expert Syst. Appl.*, vol. 41, no. 8, pp. 3955–3964, Jun. 2014.
- [28] W. Zhao, T. Tao, and E. Zio, "System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection," *Appl. Soft Comput.*, vol. 30, pp. 792–802, 2015.
- [29] S. Zhou *et al.*, "A prediction model of drug-induced ototoxicity developed by an optimal Support Vector Machine (SVM) method," *Comput. Biol. Med.*, vol. 51C, pp. 122–127, May 2014.
- [30] X. Yang, *Nature-Inspired Optimization Algorithms*. Elsevier, 2014.
- [31] D. Zhang, W. Liu, A. Wang, and Q. Deng, "Parameter Optimization for Support Vector Regression Based on Genetic Algorithm with Simplex Crossover Operator," *J. Inf. Comput. Sci.*, vol. 6, no. June, pp. 911–920, 2011.
- [32] C. W. Dawson, *Projects in Computing and Information Systems*. 2009.
- [33] E. Nurelasari, "Komparasi Algoritma Naive Bayes Dengan Support Vector Machine Berbasis Particle Swarm Optimization untuk Prediksi Kesuburan," *Bina Insa. ICT J.*, vol. 5, no. 1, pp. 61–70, 2018.



Hani Harafani. Jakarta, 12 Juli 1989. Menyelesaikan kuliah S2 pada Tahun 2015. Minat penelitian saat ini pada bidang Data Mining, dan telah menerbitkan 3 judul penelitian yang terkait dengan data mining khususnya, SVM, dan GA



Andry Maulana. Jakarta, 8 Oktober 1990. Mendapatkan Gelar M.Kom dari Pascasarjana STMIK Nusamandiri. Aktif sebagai peneliti dibidang network.