

Optimasi Parameter pada Support Vector Machine Berbasis Algoritma Genetika untuk Estimasi Kebakaran Hutan

Hani Harafani

Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri
haniharafani@gmail.com

Romi Satria Wahono

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
romi@romisatriawahono.net

Abstract: Kebakaran hutan merupakan salah satu masalah lingkungan yang mengancam hutan, menimbulkan dampak negatif pada lingkungan, menciptakan masalah ekonomi, dan kerusakan ekologis, serta menyebabkan kerugian penting di seluruh dunia setiap tahunnya. Estimasi area yang terbakar penting dilakukan, karena area yang terbakar dapat mencerminkan berapa kuat radiasi api pada vegetasi disekitarnya. SVM dapat mengatasi masalah klasifikasi dan regresi linier ataupun nonlinier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi. Namun, SVM juga memiliki kelemahan yaitu sulitnya menentukan nilai parameter yang optimal. Untuk menyelesaikan permasalahan tersebut algoritma genetika diusulkan untuk diterapkan sebagai algoritma pencarian nilai parameter yang efisien pada SVM. Beberapa eksperimen dilakukan untuk menghasilkan estimasi yang akurat. Awalnya percobaan dilakukan pada kernel –kernel SVM (dot, RBF, polynomial) untuk menentukan kernel mana yang akan digunakan, kemudian model SVM+GA juga dibandingkan dengan model regresi lainnya seperti Linear Regression, k-NN, dan Neural Network. Berdasarkan eksperimen dengan 10 kombinasi parameter pada metode SVM dan SVM+GA dengan kernel dot, RMSE terkecil dihasilkan oleh model SVM+GA sebesar 1.379, sementara pada percobaan SVM dan SVM+GA dengan kernel polynomial RMSE terkecil diperoleh model SVM+GA sebesar 1.379, sedangkan pada percobaan SVM dan SVM+GA dengan kernel RBF diperoleh RMSE terkecil pada model SVM+GA sebesar 1.379. Selanjutnya berdasarkan perbandingan rata-rata RMSE, kernel RBF unggul dengan nilai RMSE terkecil yaitu 1.432 pada SVM, dan 1.418 pada SVM+GA. Pada perbandingan nilai rata-rata RMSE antara SVM(RBF)+GA dengan model lainnya, RMSE terkecil dihasilkan oleh SVM(RBF)+GA yaitu sebesar 1.418, disusul dengan model SVM(RBF) sebesar 1.432, kemudian Linear Regression sebesar 1.459, dilanjutkan oleh model k-NN sebesar 1.526 dan yang terakhir adalah NN dengan nilai RMSE sebesar 1.559. maka dapat disimpulkan bahwa optimasi parameter yang dilakukan GA pada model SVM terbukti dapat mengurangi tingkat error pada model SVM tanpa optimasi parameter pada dataset *forestfire*, selain model SVM(RBF)+GA pada penelitian ini juga terbukti lebih baik dari model regresi lainnya.

Keywords: Estimasi, Kebakaran Hutan, Support Vector Machine, Algoritma Genetika, Optimasi Parameter.

1 PENDAHULUAN

Kebakaran hutan merupakan salah satu masalah lingkungan yang mengancam hutan, menimbulkan dampak

negatif pada lingkungan, menciptakan masalah ekonomi, dan kerusakan ekologis (Özbayoğlu & Bozer, 2012), serta menyebabkan kerugian penting di seluruh dunia setiap tahunnya (Brun, Margalef, & Cortés, 2013). Kebakaran hutan terjadi karena beberapa hal diantaranya: pembakaran hutan yang disengaja (Denham, Wendt, Bianchini, Cortés, & Margalef, 2012), petir (Cortez & Morais, 2007), dan perubahan cuaca yang ekstrim (Eastaugh & Hasenauer, 2014), serta beberapa penyebab lainnya.

Estimasi area yang terbakar penting dilakukan, karena area yang terbakar dapat mencerminkan berapa kuat radiasi api pada vegetasi disekitarnya (Quintano, Fernández-Manso, Stein, & Bijker, 2011), sehingga dapat memberikan informasi mengenai kerusakan lahan yang terjadi. Namun, metode estimasi konvensional yang dilakukan oleh banyak peneliti berdasarkan *Thresholding* menghasilkan nilai estimasi yang akurat.

Ada beberapa studi yang dilakukan untuk mengestimasi lahan yang terbakar pada kebakaran hutan dengan menggunakan metode komputasi antara lain: *support vector machine* (SVM) (Cortez & Morais, 2007), dan *multi layer perceptron* (MLP) (Özbayoğlu & Bozer, 2012). Selain itu banyak juga metode regresi yang digunakan para peneliti dunia pada berbagai permasalahan estimasi seperti k-NN (Lee, Kang, & Cho, 2014), *linear regression* (LR) (Lira, Da Silva, Alves, & Veras, 2014), dan *neural network* (NN) (Tiryaki, Öz, & Y, 2014). Metode-metode tersebut sangat direkomendasikan oleh banyak peneliti di dunia.

Multilayer perceptron sebagai salah satu model yang paling populer dari *artificial neural network* (ANN) (Singh & Borah, 2014) memiliki kelebihan untuk menemukan pola dari data yang terlalu rumit untuk diketahui oleh manusia atau dengan teknik komputasi lainnya (Yilmaz & Kaynar, 2011). Selain itu MLP memiliki kekurangan yaitu sulit menemukan pola bila data berdimensi tinggi atau sering disebut dengan “kutukan dimensionalitas” (Pan, Iplikci, Warwick, & Aziz, 2012), dan *overfitting* (Rynkiewicz, 2012).

Support vector machine (SVM) memiliki keunggulan dibandingkan metode MLP yaitu: SVM dapat mengatasi masalah klasifikasi dan regresi dengan linier ataupun nonlinier kernel yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi (Maimon & Rokach, 2010), dan baik untuk mengatasi kutukan dimensionalitas (Wang, Wen, Zhang, & Wang, 2014). 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 cocok diterapkan untuk memprediksi kebakaran hutan. Selain SVM

memiliki banyak kemampuan, SVM juga memiliki kelemahan yaitu sulitnya menentukan nilai parameter yang optimal (Ilhan & Tezel, 2013; Raghavendra. N & Deka, 2014; M. Zhao, Fu, Ji, Tang, & Zhou, 2011).

Beberapa algoritma pun banyak direkomendasikan oleh peneliti dunia untuk mengoptimasi parameter pada machine learning, seperti: *particle swarm optimization* (PSO)(Wang et al., 2014), *simulated annealing* (SA) (Z.-Y. Jia, Ma, Wang, & Liu, 2010), dan *genetic algorithm* (GA) (Guo, Li, & Zhang, 2012).

Simulated Annealing (SA) efektif pada pemuatan masalah optimasi pola, namun SA memiliki kecenderungan untuk terjebak dalam minimum lokal ketika suhu anil rendah (tingkat anil cepat) dan semakin tidak konvergen ketika suhu anil tinggi (tingkat anil lambat) (Zameer, Mirza, & Mirza, 2014), selain itu PSO juga sulit mendapatkan nilai yang optimum dalam mengoptimasi lebih dari sepuluh parameter. *Genetic algorithm* atau algoritma genetika dapat mengatasi masalah yang nonlinier dengan diskontinuitas dan minima lokal secara efisien, serta GA juga lebih efisien dalam mengoptimasi lebih dari sepuluh parameter (Machairas, Tsangrassoulis, & Axarli, 2014).

Dalam penelitian ini kami mengusulkan algoritma genetika (GA) untuk melakukan optimasi parameter pada *support vector machine* untuk meningkatkan akurasi dalam mengestimasi kebakaran hutan.

Paper ini disusun sebagai berikut: pada bagian 2 paper-paper terkait dijelaskan. Pada bagian 3, metode yang diusulkan disajikan. Hasil percobaan perbandingan antara metode yang diusulkan dengan metode lainnya disajikan pada bagian 4. Akhirnya, kesimpulan dari penelitian kami disajikan pada bagian terakhir.

2 PENELITIAN TERKAIT

Support Vector Machine (SVM) secara konseptual adalah mesin linier yang dilengkapi dengan fitur special (Gorunescu, 2011), dan didasarkan pada metode minimalisasi resiko struktural (Dua, 2011), serta teori pembelajaran statistik. 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 (Gorunescu, 2011).

Untuk permasalahan 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. Sementara untuk permasalahan regresi, Vapnik juga memperkenalkan fungsi $\varepsilon - insensitive loss$ yang disebut sebagai SVM untuk regresi.

Support Vector Regression (SVR) adalah metode untuk mengestimasi sebuah fungsi yang dipetakan dari objek input ke jumlah riil berdasarkan data pelatihan. Serupa dengan pengklasifikasian SVM, SVR memiliki properti yang sama tentang memaksimalkan margin dan trik kernel untuk memetakan data yang nonlinier. Secara singkat sekumpulan data training untuk regresi SVM digambarkan sebagai dataset $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ Dimana x_i adalah vector n-dimensional, sedangkan y adalah jumlah real untuk setiap x_i .

Tugas dari regresi adalah untuk menemukan fungsi antara x_i dengan y_i (W. Zhao, Tao, & Zio, 2015) yang dalam kasus linier dapat ditulis:

$$y_i = f(x) = w \cdot x + b \quad (1)$$

Dimana w adalah vector beban dan b adalah bias. Kedua parameter ini adalah parameter yang perlu ditentukan nilainya agar dapat memberikan fungsi yang terbaik untuk memetakan data input ke data output.

Pada kasus nonlinier, pemetaan nonlinier: $R^1 \rightarrow F$, dimana F merupakan ruang fitur dari ϕ yang diperkenalkan untuk menerjemahkan kerumitan masalah regresi nonlinier pada R^1 untuk sebuah masalah sederhana regresi linier pada F . Fungsi regresi setelah transformasi menjadi seperti berikut:

$$y_i = f(x) = w * \phi(x) + b \quad (2)$$

Untuk mengevaluasi seberapa baik fungsi regresi, fungsi $\varepsilon - insensitive loss$ digunakan:

$$L_\varepsilon(y, f(x)) = \begin{cases} 0 & \text{untuk } |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \varepsilon & \text{dan sebaliknya} \end{cases} \quad (3)$$

Fungsi $\varepsilon - insensitive loss$ digunakan untuk mengukur resiko empiris, resiko empiris diukur berdasarkan persamaan (3), selisih output/target dengan hasil estimasi. Oleh karenanya parameter ε harus diatur. Kemudian, prosedur diatur untuk meminimalisir resiko empiris dengan memperkenalkan variable slack ξ, ξ^* yang menggambarkan simpangan dari data pelatihan diluar zona $\varepsilon - insensitive$.

Disamping meminimalisir kesalahan empiris dengan fungsi $\varepsilon - insensitive loss$, kita juga harus meminimalisir norma Euclidean dari beban yang linier $\|w\|$ yang mana berhubungan dengan kemampuan generalisasi dari model SVR yang dilatih (W. Zhao et al., 2015). Tujuannya adalah untuk memperlebar (*maximize*) margin sehingga kelandaian kurva beserta kompleksitas model dapat dipastikan (Suganyadevi & Babulal, 2014). Sehingga permasalahan regresi dapat dinyatakan seperti masalah optimasi kuadrat berikut ini:

$$L(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_i (\xi_{2i}, \xi'_{2i}), c > 0$$

$$\text{subject to } \begin{cases} y_i - w * \phi(x_i) - b \leq \varepsilon + \xi_i \\ w * \phi(x_i) + b - y_i \leq \varepsilon + \xi'_i \\ \xi_i, \xi'_i \geq 0 \end{cases} \quad (4)$$

Dimana C menyatakan koefisien penalti yang mendeterminasikan *trade-off* antara keempirisan dengan kesalahan generalisasi yang mana nilai C tersebut perlu diatur (W. Zhao et al., 2015). Untuk menyelesaikan permasalahan pada optimasi kuadrat pada persamaan (4), kita dapat menggunakan dual Lagrangian:

$$f(x_i) = (w \cdot \phi(x_i) + b) = \sum_{j=1}^n \alpha_j K(x_i, x_j) + b \quad (5)$$

Dimana $K(x_i, x_j)$ merupakan fungsi kernel yang memuaskan kondisi Mercer. Fungsi kernel yang digunakan pada penelitian ini adalah kernel RBF dengan parameter γ yang juga perlu diatur (W. Zhao et al., 2015).

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (6)$$

Berdasarkan uraian mengenai SVR dapat dilihat bahwa terdapat tiga parameter bebas $C, \varepsilon, \text{ dan } \gamma$ yang penting bagi performa metode SVR, kebanyakan peneliti masih mengikuti prosedur yang standar dalam pemilihan parameter (*trial and error*), yaitu dengan membangun model SVR dengan parameter yang berbeda-beda, kemudian mengujinya pada set validasi untuk menghasilkan parameter yang optimal. Namun

prosedur ini sangat memakan waktu (Chen, 2007) dan tergantung factor keberuntungan. Oleh karenanya parameter-parameter tersebut perlu di atur nilainya. Beberapa penelitian telah dilakukan untuk mengatur nilai parameter C , ϵ , dan γ pada SVM dengan beberapa metode metaheuristik.

Penelitian yang dilakukan oleh (Wang et al., 2014) terfokus pada peningkatan akurasi prediksi SVM untuk mengatasi kinerja SVM yang terpengaruh akibat pemilihan parameter yang tidak tepat untuk memprediksi harga *real estate* di China. Pada penelitiannya (Wang et al., 2014) menerapkan *Particle swarm optimization* (PSO) untuk menentukan nilai σ , ϵ , dan c pada SVM. Selain itu, sebagai perbandingan (Wang et al., 2014) juga membandingkan hasil prosentase MAPE SVM, dan SVM+PSO dengan MAPE BPNN juga. Hasil penelitian dapat disimpulkan bahwa akurasi prediksi PSO+SVM lebih tinggi dibandingkan dengan model SVM dan BPNN.

Sedangkan (Z. Jia, Ma, Wang, & Liu, 2011) menerapkan *Simulated annealing* (SA) untuk mencari nilai parameter σ , ϵ , dan c yang optimal pada SVM. Percobaan yang dilakukan oleh (Z. Jia et al., 2011) yaitu dengan mengatur batasan minimal dan batasan maksimal pada masing-masing parameter seperti parameter C dengan batasan minimal 100 dan maksimal 1000, parameter ϵ dengan nilai minimal 0,001 dan maksimal 0,01, dan parameter σ dengan nilai minimal 0,5 dan maksimal 1,5. Kemudian Jia melakukan hal yang sama dengan Wang, yaitu membandingkan hasil RMSE dan MAPE metode yang diusulkan dengan metode yang lainnya yaitu (Z. Jia et al., 2011) membandingkan hasil RMSE dan MAPE antara SA+SVM, G-ANFIS, dan ANN. Berdasarkan hasil penelitian dapat disimpulkan bahwa nilai RMSE SA+SVM hanya sedikit lebih tinggi dibandingkan dengan hasil RMSE dan MAPE G-ANFIS, namun SA+SVM berhasil membuat jarak akurasi yang jauh jika dibandingkan dengan ANN.

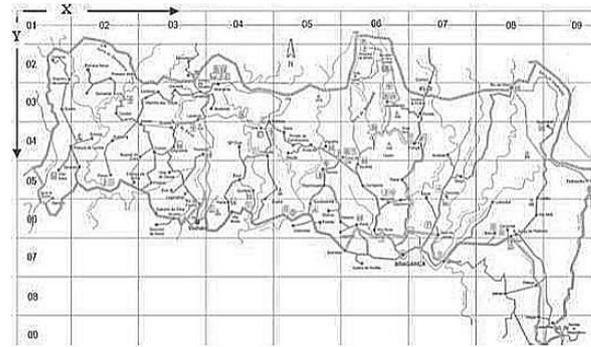
Sementara untuk mengatasi masalah pencarian nilai parameter pada SVM, (Gu, Zhu, & Jiang, 2011) menerapkan *Genetic Algorithm* (GA) dan juga membandingkan hasil prediksi dengan metode *Grey model* (GM). Berdasarkan hasil penelitian dapat disimpulkan bahwa GA+SVM menghasilkan nilai MAPE yang lebih superior dari pada GM.

Pada penelitian ini kami menggunakan GA sebagai algoritma untuk mengatur nilai parameter pada SVR dengan kombinasi kernel yaitu (RBF, dot, polynomial). Kemudian untuk membuktikan kehandalan metode yang telah diusulkan, dengan metode-metode regresi lainnya untuk memprediksi kebakaran hutan. Pemilihan metode-metode regresi tersebut sesuai dengan beberapa penelitian terbaru yang telah dilakukan terhadap permasalahan regresi seperti penelitian yang telah dilakukan oleh (Tiryaki et al., 2014) yang mencari metode regresi yang tepat untuk dapat mendeteksi efek perlakuan pada pada ikatan kayu. Tiryaki membandingkan dua metode diantaranya ANN dan MLR. Berdasarkan percobaan, dapat disimpulkan ANN telah terbukti menjadi metode yang sukses dan *sufficient* untuk memodelkan karakteristik kekuatan ikatan jenis kayu. Selain ANN & MLR ada juga *Linear regression* (LR) (Lira et al., 2014), dan k-NN (Lee et al., 2014).

3 METODE YANG DIUSULKAN

Pada penelitian ini, data yang digunakan adalah dataset *forestfire* yang diambil dari laman <https://archive.ics.edu/ml/datasets/Forest+Fires>. Data *forest fire* merupakan data kebakaran hutan sejak Januari tahun 2000 sampai dengan Desember tahun 2003. Data ini berasal dari taman alam Mountesinho, Portugal yang terdiri dari 13 atribut

(12 atribut/*input* dan 1 target estimasi/*output*), dan 517 record yang dapat dilihat pada Tabel 1.



Gambar 1. Peta Taman Alam Mountesinho

Pada Gambar 1 sumbu x dan y merupakan koordinat terjadinya kebakaran hutan, kemudian hari dan bulan terjadinya kebakaran hutan, data meteorologi yang terdiri dari: temperature, kelembaban relatif (RH), angin, dan hujan, selanjutnya terdapat empat komponen dari *Forest Fire Weather Index* (FWI) yaitu: *Fine Fuel Moisture Code* (FFMC) yang merupakan indikator untuk bahan yang mudah terbakar dan mudah membentuk pengapian di dalam hutan, *Duff Moisture Code* (DMC) yang merupakan indikasi dari konsumsi bahan bakar pada kayu, dan *Drought Code* (DC) yang merupakan indikator dari efek kemarau pada bahan bakar hutan, *Initial Spread Index* (ISI) yaitu indikator tingkat penyebaran api dan indikator kesulitan pengendalian api. Berdasarkan data yang tersedia, telah ditentukan area yang terbakar sebagai *output* dan 12 atribut lainnya merupakan *input*. Dataset *forest fire* dapat dilihat pada Tabel 1.

Tabel 1. Dataset Forest Fire

X	Y	Month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
7	5	Mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
7	5	Mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
7	4	Oct	tue	90.6	35.4	669.1	6.7	18	33	0.9	0	0
7	4	Oct	sat	90.6	43.7	686.9	6.7	14.6	33	1.3	0	0
8	6	Mar	fri	91.7	33.3	77.5	9	8.3	97	4	0.2	0
8	6	Mar	sun	89.3	51.3	102.2	9.6	11.4	99	1.8	0	0
8	6	Aug	sun	92.3	85.3	488	14.7	22.2	29	5.4	0	0
8	6	Aug	mon	92.3	88.9	495.6	8.5	24.1	27	3.1	0	0
8	6	Aug	mon	91.5	145.4	608.2	10.7	8	86	2.2	0	0
8	6	Sep	tue	91	129.5	692.6	7	13.1	63	5.4	0	0
7	5	Sep	sat	92.5	88	698.6	7.1	22.8	40	4	0	0
7	5	Sep	sat	92.5	88	698.6	7.1	17.8	51	7.2	0	0
7	5	Sep	sat	92.8	73.2	713	22.6	19.3	38	4	0	0
6	5	Aug	fri	63.5	70.8	665.3	0.8	17	72	6.7	0	0
6	5	Sep	mon	90.9	126.5	686.5	7	21.3	42	2.2	0	0
6	5	Sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0	0
6	5	Sep	fri	93.3	141.2	713.9	13.9	22.9	44	5.4	0	0
5	5	Mar	sat	91.7	35.8	80.8	7.8	15.1	27	5.4	0	0
8	5	Oct	mon	84.9	32.8	664.2	3	16.7	47	4.9	0	0
6	4	Mar	wed	89.2	27.9	70.8	6.3	15.9	35	4	0	0
6	4	Apr	sat	86.3	27.4	97.1	5.1	9.3	44	4.5	0	0
6	4	Sep	tue	91	129.5	692.6	7	18.3	40	2.7	0	0
5	4	Sep	mon	91.8	78.5	724.3	9.2	19.1	38	2.7	0	0
7	4	Jun	sun	94.3	96.3	200	56.1	21	44	4.5	0	0
7	4	Aug	sat	90.2	110.9	537.4	6.2	19.5	43	5.8	0	0
7	4	Aug	sat	93.5	139.4	594.2	20.3	23.7	32	5.8	0	0
7	4	Aug	sun	91.4	142.4	601.4	10.6	16.3	60	5.4	0	0
7	4	Sep	fri	92.4	117.9	668	12.2	19	34	5.8	0	0
7	4	Sep	mon	90.9	126.5	686.5	7	19.4	48	1.3	0	0
6	3	Sep	sat	93.4	145.4	721.4	8.1	30.2	24	2.7	0	0
6	3	Sep	sun	93.5	149.3	728.6	8.1	22.8	39	3.6	0	0
6	3	Sep	fri	94.3	85.1	692.3	15.9	25.4	24	3.6	0	0
6	3	Sep	mon	88.6	91.8	709.9	7.1	11.2	78	7.6	0	0
6	3	Sep	fri	88.6	69.7	706.8	5.8	20.6	37	1.8	0	0
6	3	Sep	sun	91.7	75.6	718.3	7.8	17.7	39	3.6	0	0
6	3	Sep	mon	91.8	78.5	724.3	9.2	21.2	32	2.7	0	0

Pada tahap awal pengolahan data (*preprocessing*), kami melakukan penghapusan beberapa atribut yang tidak diperlukan seperti dalam (Cortez & Morais, 2007) sekaligus menghapus *record* yang berisi data ganda, sehingga dari ke 13 atribut hanya tersisa 9 atribut saja (8 estimator dan 1 target estimasi), dan dari 517 record tersisa 513 data yang unik. Kemudian kami melakukan transformasi pada target estimasi (atribut *area*) atau label dengan rumus $y = \ln(x + 1)$ untuk menghilangkan kecenderungan positif pada dataset (Cortez & Morais, 2007), hasil dari tahapan *preprocessing* menghasilkan data baru yang dapat dilihat pada Tabel 2.

Tabel 2. Dataset Forest Fire Setelah Dilakukan *Preprocessing*

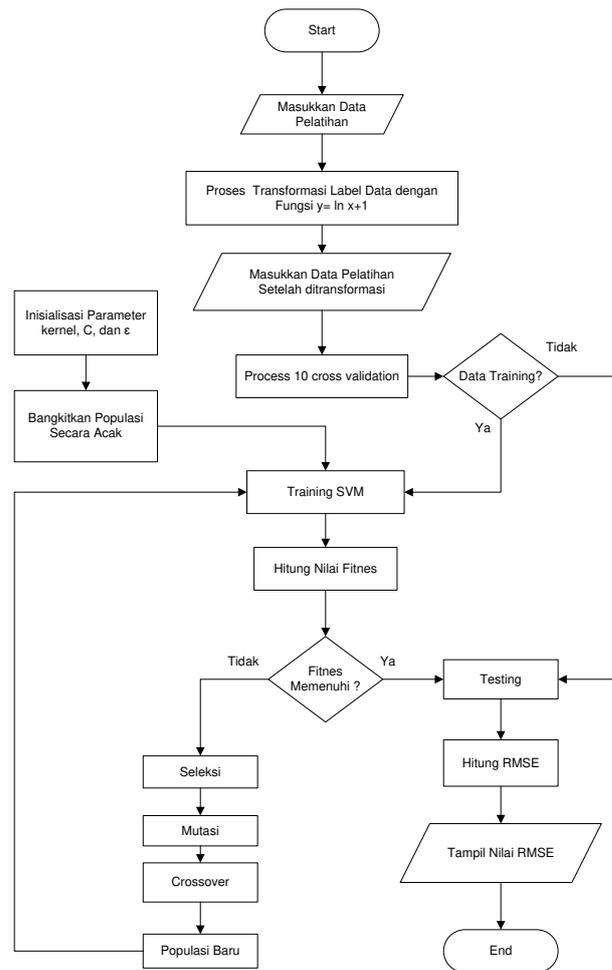
FFMC	DMC	DC	ISI	temp	RH	wind	rain	ln(area+1)
86.2	26.2	94.3	5.1	8.2	51	6.7	0	0
90.6	35.4	669.1	6.7	18	33	0.9	0	0
90.6	43.7	686.9	6.7	14.6	33	1.3	0	0
91.7	33.3	77.5	9	8.3	97	4	0.2	0
89.3	51.3	102.2	9.6	11.4	99	1.8	0	0
92.3	85.3	488	14.7	22.2	29	5.4	0	0
92.3	88.9	495.6	8.5	24.1	27	3.1	0	0
91.5	145.4	608.2	10.7	8	86	2.2	0	0
91	129.5	692.6	7	13.1	63	5.4	0	0
92.5	88	698.6	7.1	22.8	40	4	0	0
92.5	88	698.6	7.1	17.8	51	7.2	0	0
92.8	73.2	713	22.6	19.3	38	4	0	0
63.5	70.8	665.3	0.8	17	72	6.7	0	0
90.9	126.5	686.5	7	21.3	42	2.2	0	0
92.9	133.3	699.6	9.2	26.4	21	4.5	0	0
93.3	141.2	713.9	13.9	22.9	44	5.4	0	0
91.7	35.8	80.8	7.8	15.1	27	5.4	0	0
84.9	32.8	664.2	3	16.7	47	4.9	0	0
89.2	27.9	70.8	6.3	15.9	35	4	0	0
86.3	27.4	97.1	5.1	9.3	44	4.5	0	0
91	129.5	692.6	7	18.3	40	2.7	0	0
91.8	78.5	724.3	9.2	19.1	38	2.7	0	0
94.3	96.3	200	56.1	21	44	4.5	0	0
90.2	110.9	537.4	6.2	19.5	43	5.8	0	0
93.5	139.4	594.2	20.3	23.7	32	5.8	0	0
91.4	142.4	601.4	10.6	16.3	60	5.4	0	0
92.4	117.9	668	12.2	19	34	5.8	0	0
90.9	126.5	686.5	7	19.4	48	1.3	0	0
93.4	145.4	721.4	8.1	30.2	24	2.7	0	0
93.5	149.3	728.6	8.1	22.8	39	3.6	0	0

Selanjutnya kami mengusulkan metode yang disebut SVM+GA yang mana GA digunakan untuk mengoptimasi parameter pada SVM untuk mendapatkan hasil estimasi area yang terbakar akibat kebakaran hutan yang lebih akurat seperti yang dapat dilihat pada Gambar 2.

Pada Gambar 2. Dataset asli *forest fire* akan melalui tahapan *preprocessing data* yang mana pada tahapan ini terdapat dua proses yaitu *deletion attribute* dan *label transformation*, kemudian setelah tahap *preprocessing* akan terbentuk dataset yang baru. Sebelum dataset baru dilatih (*training*) dan diuji (*testing*), dataset akan dipecah terlebih dahulu dengan menerapkan *10-fold cross validation* untuk membagi data menjadi dua yaitu 90% data *training* dan 10% data *testing*. Kemudian data dilatih dan diuji dengan metode SVM yang mana nilai parameter-parameter ($c, \epsilon, \text{ dan } \gamma$) dari kernel-kernel SVM (dot, polynomial, RBF) telah diatur oleh GA sebelumnya. Selanjutnya kernel dengan nilai RMSE yang terkecil akan dipergunakan pada model SVM+GA untuk dibandingkan dengan model regresi lainnya.

Algoritma genetika (GA) merupakan algoritma evolusioner yang paling populer (Yang, 2014) 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 (Gorunescu, 2011).



Gambar 2. Metode Penelitian yang Diusulkan

Algoritma genetika memiliki tiga operator genetik utama yaitu *crossover* (proses penukaran kromosom), *mutasi* (proses penggantian salah satu solusi untuk meningkatkan keragaman populasi), *seleksi* (penggunaan solusi dengan nilai fitness yang tinggi untuk lulus ke generasi berikutnya). Terdapat langkah-langkah 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 3.

Terdapat beberapa kelebihan dari algoritma genetika dibandingkan algoritma optimasi tradisional lainnya, dua diantaranya yaitu kemampuan untuk menangani permasalahan kompleks dan parallel. Algoritma genetika dapat menangani 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 (Zhang, Liu, Wang, & Deng, 2011) ditunjukkan pada persamaan (7).

$$fitness = \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2} \tag{7}$$

Dimana \hat{y}_i merupakan nilai prediksi, dan y_i merupakan nilai asli dan dari sampel dataset. N adalah jumlah sampel total.

Akurasi prediksi keseluruhan percobaan pada penelitian ini ditunjukkan pada Gambar 6. Metrik yang kami gunakan untuk mengukur akurasi predikai adalah nilai *root mean square error* (RMSE) yang didefinisikan pada persamaan 8.

$$RMSE = \sqrt{\frac{\sum_{j=i}^n (y' - y)^2}{n}} \quad (8)$$

RMSE sangat populer untuk menilai algoritma mesin pembelajaran, termasuk algoritma yang jauh lebih canggih dari regresi linier (Conway & White, 2012). Nilai RMSE digunakan untuk membedakan kinerja model dalam periode kalibrasi dengan periode validasi serta untuk membandingkan kinerja model individual dengan model prediksi lainnya (Hosseini, Javaherian, & Movahed, 2014).

4 HASIL EKSPERIMEN

Eksperimen dilakukan menggunakan komputer personal Intel Core i3, 4GB RAM, 500GB HDD, sistem operasi Windows 8.1, dan Rapidminer 5.0.

Penelitian ini dilakukan dalam tiga tahapan. Tahap pertama untuk mendapatkan hasil estimasi kebakaran hutan yang lebih akurat, kami membandingkan hasil eksperimen antara data yang diproses menggunakan metode SVM yang parameternya diatur secara manual dengan data yang diproses menggunakan metode SVM yang parameternya telah diatur oleh GA pada masing-masing kernel SVM. Kemudian kami melakukan uji beda untuk mengukur signifikansi keakuratan estimasinya.

Tahap kedua, untuk mengidentifikasi kernel terbaik pada SVM, kami melakukan perbandingan rata-rata RMSE dari 10 kali percobaan antara metode SVM dengan SVM+GA pada masing-masing kernel.

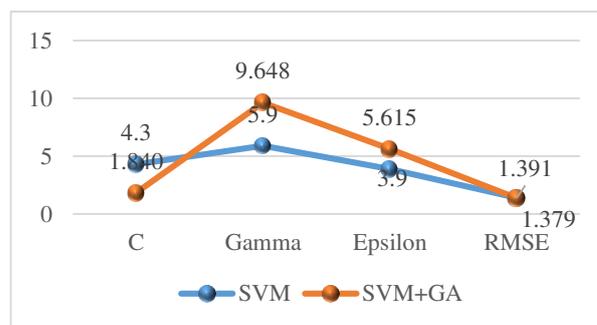
Tahap ketiga untuk mengetahui kehandalan model SVM+GA, kami melakukan perbandingan akurasi antara metode SVM+GA dengan metode regresi lainnya seperti k-NN, LR, dan NN.

Pada tahapan pertama, eksperimen pada masing-masing kernel SVM dilakukan sebanyak 10 kali. Pada kernel dot, pada tahapan inialisasi populasi, kami memasukkan 10 kombinasi jangkauan nilai input parameter seperti yang dapat dilihat pada Tabel 3.

Tabel 3. Hasil Eksperimen dengan Kernel Dot

γ			C			ϵ		
MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL
0	0.5	0.308	-1	-0.5	-0.890	0	0.5	0.263
0	1	0.939	-1	1	-0.888	0	1	0.685
0	5	4.663	-1	5	-0.618	0	5	3.443
0	7	4.363	-1	7	5.906	0	7	3.952
0	10	1.840	-1	10	9.648	0	10	5.615
0	100	18.261	-1	100	96.812	0	100	56.201
0.001	0.1	0.001	-1	0.1	0.002	0.001	0.1	0.057
0.01	0.1	0.035	-1	0.1	0.014	0.01	0.1	0.100
0.5	10	1.143	-0.5	10	1.821	0.5	10	5.089
0.5	100	18.670	-0.5	100	96.828	0.5	100	56.420

Nilai RMSE terkecil yaitu 1.379 diperoleh dari percobaan SVM (kernel dot) dan GA dengan pencarian nilai parameter $c = 1.840$, $\epsilon = 5.615$, dan $\gamma = 9.648$ seperti yang dapat dilihat pada Gambar 3.



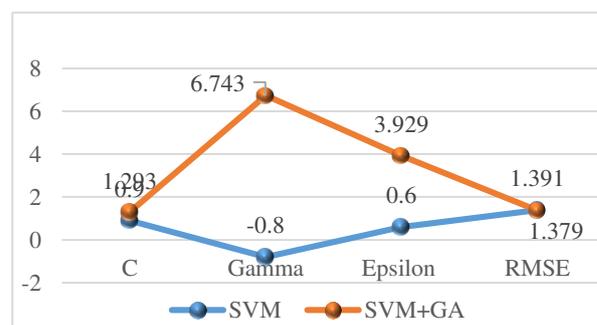
Gambar 3. Perbandingan Estimasi SVM dengan SVM+GA pada Kernel Dot

Pada kernel polynomial Pada kernel dot, pada tahapan inialisasi populasi, kami memasukkan 10 kombinasi jangkauan nilai input parameter seperti yang dapat dilihat pada Tabel 4.

Tabel 4. Hasil Eksperimen dengan Kernel Polynomial

γ			C			ϵ		
MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL
0	0.5	0.311	-1	-0.5	-0.513	0	0.5	0.292
0	1	0.953	-1	1	-0.881	0	1	0.687
0	5	0.369	-1	5	0.316	0	5	2.405
0	7	1.293	-1	7	6.743	0	7	3.929
0	10	1.840	-1	10	9.648	0	10	5.615
0	100	18.261	-1	100	96.812	0	100	56.201
0.001	0.1	0.089	-1	0.1	-0.940	0.001	0.1	0.082
0.01	0.1	0.090	-1	0.1	-0.940	0.01	0.1	0.085
0.5	10	2.249	-0.5	10	9.664	0.5	10	5.834
0.5	100	18.670	-0.5	100	96.828	0.5	100	56.420

Nilai RMSE terkecil yaitu 1.379 diperoleh dari percobaan SVM+GA dengan hasil pencarian nilai parameter $c = 1.293$, $\epsilon = 6.743$, dan $\gamma = 3.929$ seperti yang dapat dilihat pada Gambar 4.

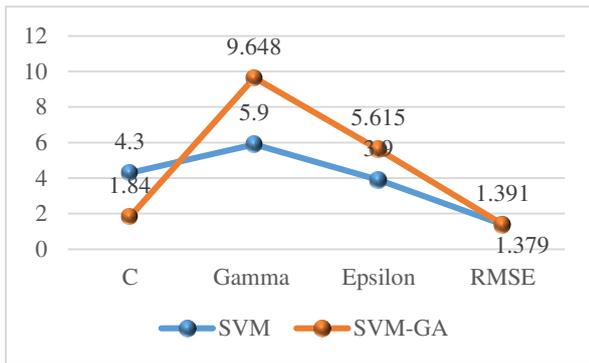


Gambar 4. Perbandingan Estimasi SVM dengan SVM+GA pada Kernel Polynomial

Pada kernel RBF Pada kernel dot, pada tahapan inialisasi populasi, kami memasukkan 10 kombinasi jangkauan nilai input parameter seperti yang dapat dilihat pada Tabel 5. Hasil terbaik juga diperoleh dari percobaan SVM+GA dengan nilai RMSE=1.379 dengan hasil pencarian parameter $c = 1.840$, $\epsilon = 9.648$, dan $\gamma = 5.615$ seperti yang dapat dilihat pada Gambar 5.

Tabel 5. Hasil Eksperimen dengan Kernel RBF

γ			C			ϵ		
MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL	MIN	MAX	OPTIMAL
0	0.5	0.091	-1	-0.5	-0.503	0	0.5	0.297
0	1	0.952	-1	1	-0.885	0	1	0.679
0	5	1.216	-1	5	1.410	0	5	0.939
0	7	4.370	-1	7	5.914	0	7	3.944
0	10	1.840	-1	10	9.648	0	10	5.615
0	100	18.261	-1	100	96.812	0	100	56.201
0.001	0.1	0.027	-1	0.1	-0.735	0.001	0.1	0.053
0.01	0.1	0.031	-1	0.1	-0.558	0.01	0.1	0.026
0.5	10	2.249	-0.5	10	9.664	0.5	10	5.834
0.5	100	18.670	-0.5	100	96.828	0.5	100	56.420



Gambar 5. Perbandingan Estimasi SVM dengan SVM+GA pada Kernel RBF

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 RMSE dari 10 kali percobaan. Uji beda dilakukan untuk menguji hipotesa:

H_0 : Tidak ada perbedaan nilai rata-rata RMSE 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

Berdasarkan hasil uji t sampel berpasangan yang telah dilakukan pada ketiga kernel, kernel dot menunjukkan bahwa tidak ada perbedaan yang signifikan antara metode SVM dengan metode SVM+GA yang dapat dilihat pada Tabel 3, namun kernel polynomial dan kernel RBF menunjukkan bahwa terdapat perbedaan yang signifikan antara metode SVM dengan SVM+GA yang dapat dilihat pada Tabel 4, dan 5.

Berdasarkan Tabel 3. Diketahui nilai t hitung yang diwakili oleh t stat sebesar 1.710268, sedangkan nilai t tabel diwakili oleh t *critical two tail* sebesar 2.262157, maka dapat dipastikan nilai t hitung < t tabel yang artinya H_0 diterima dan H_1 ditolak, sedangkan nilai probabilitas yang ditunjukkan oleh nilai $P(T \leq t)$ two tail sebesar 0.121382 lebih besar dari 0.05 yang artinya tidak terdapat perbedaan yang signifikan dari rata-rata RMSE model SVM dengan SVM+GA menggunakan kernel dot.

Pada Tabel 4. Diketahui nilai t hitung yang diwakili oleh t stat sebesar 2.429748, dan nilai t tabel yang diwakili oleh nilai 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 diketahui nilai probabilitas sebesar 0.037998 yang mana nilai probabilitas < 0.05 yang artinya terdapat perbedaan yang signifikan dari rata-rata RMSE model SVM dengan SVM+GA menggunakan kernel polynomial.

Tabel 3. Hasil Uji Beda Statistik SVM dan SVM+GA Menggunakan Kernel Dot

t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	1.9109	1.8428
Variance	0.643116	0.481873
Observations	10	10
Pearson Correlation	0.996192	
Hypothesized Mean Difference	0	
df	9	
t Stat	1.710268	
P(T<=t) one-tail	0.060691	
t Critical one-tail	1.833113	
P(T<=t) two-tail	0.121382	
t Critical two-tail	2.262157	

Tabel 4. Hasil Uji Beda Statistik SVM dengan SVM+GA Menggunakan Kernel Polynomial

t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	1.6293	1.4762
Variance	0.106602	0.018734
Observations	10	10
Pearson Correlation	0.958101	
Hypothesized Mean Difference	0	
df	9	
t Stat	2.429748	
P(T<=t) one-tail	0.018999	
t Critical one-tail	1.833113	
P(T<=t) two-tail	0.037998	
t Critical two-tail	2.262157	

Berdasarkan Tabel 5. Diketahui nilai t hitung yang diwakili oleh t stat sebesar 2.537358, dan nilai t tabel yang diwakili oleh nilai t *critical two tail* sebesar 2.262157. Sehingga dapat dipastikan nilai t hitung > t tabel yang artinya H_0 ditolak dan H_1 diterima, sedangkan diketahui nilai probabilitasnya sebesar 0.031849 yang mana nilai probabilitas ini lebih kecil dari pada 0.05 yang artinya terdapat perbedaan yang signifikan dari rata-rata RMSE model SVM dengan SVM+GA menggunakan kernel RBF.

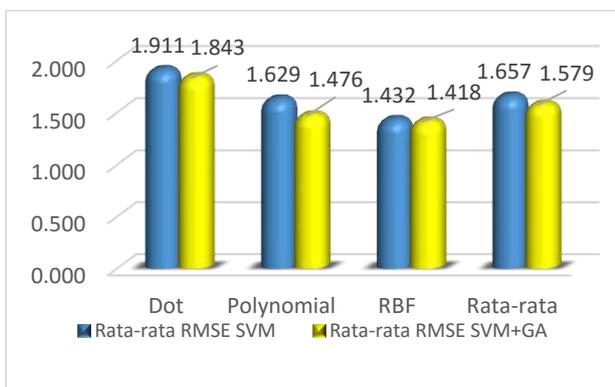
Pada tahapan kedua kami mengambil nilai rata-rata RMSE dari 10 kali percobaan. Berdasarkan hasil perhitungan rata-rata kernel RBF terbukti memiliki nilai rata-rata RMSE yang terkecil baik pada metode SVM maupun pada metode SVM+GA. Sementara kernel dot memiliki nilai rata-rata RMSE yang paling besar diantara yang lainnya pada metode SVM dan SVM+GA. Perolehan nilai rata-rata RMSE dapat dilihat pada Gambar 6.

Pada Gambar 6 dapat disimpulkan bahwa nilai rata-rata RMSE pada model SVM+GA dengan kernel (dot, polynomial, dan RBF) lebih kecil dibandingkan dengan nilai rata-rata RMSE pada model SVM dengan kernel (dot, polynomial, RBF) tanpa optimasi parameter.

Tabel 5. Hasil Uji Beda Statistik SVM dengan SVM+GA Menggunakan Kernel RBF

t-Test: Paired Two Sample for Means

	Variable 1	Variable 2
Mean	1.4318	1.4179
Variance	0.002526	0.002326
Observations	10	10
Pearson Correlation	0.938956	
Hypothesized Mean Difference	0	
df	9	
t Stat	2.537358	
P(T<=t) one-tail	0.015924	
t Critical one-tail	1.833113	
P(T<=t) two-tail	0.031849	
t Critical two-tail	2.262157	



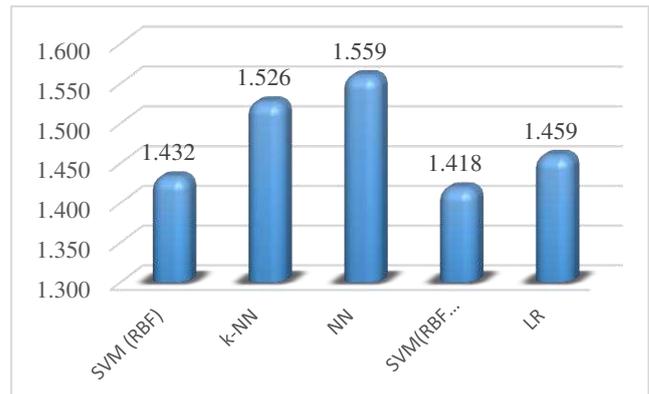
Gambar 6. Perbandingan Nilai Rata-rata RMSE SVM dengan SVM+GA

Pada tahapan ketiga, dataset *forest fire* juga dipercobakan pada metode regresi lainnya. hal ini dilakukan untuk mengetahui kehandalan model SVM+GA jika dibandingkan dengan model-model regresi lainnya. Metode yang akan dibandingkan dengan model SVM+GA pada tahap ketiga ini antara lain SVM, k-NN, LR, dan NN dengan 6 kali percobaan. Hasil perbandingan metode-metode regresi akan ditunjukkan pada Gambar 7. Pada metode k-NN parameter k di atur secara manual yaitu dengan memasukkan nilai parameter 15, 10, 50, 100, dan 150 pada 6 kali percobaan. Sementara pada metode LR nilai *forward alpha* dan *backward alpha* juga diinput secara manual. Begitu pula pada NN, kami memasukkan variasi nilai parameter untuk *learning rate* dan *momentum* secara manual.

k-NN memiliki algoritma yang sederhana dan kinerja prediksi yang tinggi terhadap berbagai aplikasi, karena kelebihanannya k-NN dianggap sebanding dengan model yang lebih kompleks seperti ANN atau SVM (Lee et al., 2014), Namun Gambar 6 menjelaskan bahwa perbandingan nilai rata-rata RMSE k-NN jauh lebih besar dibandingkan metode SVM. Oleh karena itu hasil penelitian ini secara otomatis bertolak belakang dengan pendapat (Lee et al., 2014) mengingat nilai akurasi k-NN yang buruk terhadap estimasi kebakaran hutan.

Penggunaan NN menjadi semakin populer dibanyak model prediksi (Kaytez, Taplamacioglu, Cam, & Hardalac, 2015).

Pada penelitian (Tiryaki et al., 2014) ANN juga terbukti dapat menyelesaikan permasalahan estimasi melebihi MLR.



Gambar 7. Perbandingan Rata-rata RMSE Kelima Metode Regresi

Namun Gambar 6 pada penelitian ini menunjukkan NN menghasilkan akurasi yang paling buruk diantara metode lainnya dalam mengestimasi kebakaran hutan. Sementara metode LR seperti dalam (Lira et al., 2014) menghasilkan nilai akurasi yang baik ketika diterapkan pada dataset *forest fire*. Hal ini dapat dilihat berdasarkan perbandingan nilai rata-rata RMSE yaitu nilai rata-rata RMSE LR jauh lebih kecil jika dibandingkan dengan nilai rata-rata RMSE NN.

Perbandingan nilai RMSE NN dengan SVM(RBF)+GA juga sangat timpang, nilai rata-rata RMSE NN sangat jauh lebih besar dari pada nilai rata-rata RMSE SVM(RBF)+GA. Selain itu nilai rata-rata RMSE LR hamper sebanding dengan nilai rata-rata SVM(RBF) dan SVM(RBF)+GA dengan selisih nilai RMSE hanya sebesar 0.027 dan 0.041 saja. Oleh karena itu penelitian ini secara otomatis mendukung penelitian yang dilakukan oleh (Lira et al., 2014) sekaligus bertolak belakang dengan hasil penelitian (Tiryaki et al., 2014).

Berdasarkan hasil percobaan yang telah dilakukan, metode yang telah diusulkan jauh lebih unggul jika dibandingkan dengan hasil percobaan (Cortez & Morais, 2007). Pada penelitian sebelumnya, penerapan *sequential minimize optimization algorithm* (SMO) untuk mengoptimasi parameter C , ϵ , dan γ pada SVM berhasil melebihi hasil prediksi metode regresi lainnya yaitu *naïve predictor* (NP), *multiple regression* (MR), *descicion tree* (DT), *random forest* (RF), dan *neural network* (NN) dengan nilai RMSE SVM(RBF)+SMO sebesar 12.71. oleh karena itu Gambar 6 pada penelitian ini sesuai dengan pernyataan (Cortez & Morais, 2007) bahwa SVM dengan optimasi parameter dapat mengungguli metode-metode regresi lainnya.

5 KESIMPULAN

SVM dapat mengatasi masalah klasifikasi dan regresi dengan kernel linier ataupun kernel nonlinier nonlinier yang dapat menjadi satu kemampuan algoritma pembelajaran untuk klasifikasi serta regresi. Namun, dibalik keunggulannya SVM juga memiliki kelemahan yaitu sulitnya menentukan nilai parameter yang optimal. Pada penelitian ini SVM digunakan untuk memprediksi area yang terbakar pada dataset *forest fire* dengan fungsi regresinya. Penerapan algoritma genetika (GA) pada metode SVM diusulkan untuk mengoptimasi nilai parameter C , ϵ , dan γ pada kernel-kernel SVM (dot, polynomial, RBF) untuk mendapatkan akurasi yang terbaik dan untuk mengidentifikasi kernel yang terbaik pula.

Dikarenakan permasalahan pada data *forest fire* merupakan tugas regresi, Beberapa metode regresi pun diusulkan untuk membuktikan kehandalan metode yang telah diusulkan. Hasil eksperimen membuktikan bahwa metode yang diusulkan SVM (RBF)+GA memiliki nilai akurasi estimasi yang lebih baik dari pada metode regresi lainnya. Untuk penelitian dimasa mendatang, kami percaya kombinasi SVM dengan algoritma metaheuristik lainnya serta penambahan metode spatial sebagai pendeteksian *outlier* dapat meningkatkan akurasi lebih signifikan.

REFERENSI

- Brun, C., Margalef, T., & Cortés, A. (2013). Coupling Diagnostic and Prognostic Models to a Dynamic Data Driven Forest Fire Spread Prediction System. *Procedia Computer Science*, 18, 1851–1860. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S1877050913004973>
- Chen, K.-Y. (2007). Forecasting systems reliability based on support vector regression with genetic algorithms. *Reliability Engineering & System Safety*, 92(4), 423–432.
- Conway, D., & White, J. M. (2012). *Machine Learning for Hackers*. (J. Steele, Ed.).
- Cortez, P., & Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data.
- Denham, M., Wendt, K., Bianchini, G., Cortés, A., & Margalef, T. (2012). Dynamic Data-Driven Genetic Algorithm for forest fire spread prediction. *Journal of Computational Science*, 3(5), 398–404. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S187750312000658>
- Dua. (2011). *Data Mining and Machine Learning in Cybersecurity*. (Dua, Ed.).
- Eastaugh, C. S., & Hasenauer, H. (2014). Deriving forest fire ignition risk with biogeochemical process modelling. *Environmental Modelling & Software*, 55, 132–142. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S1364815214000280>
- Gorunescu, F. (2011). *Intelligent Systems Reference Library*. (Gorunescu, Ed.).
- Gu, J., Zhu, M., & Jiang, L. (2011). Housing price forecasting based on genetic algorithm and support vector machine. *Expert Systems with Applications*, 38(4), 3383–3386. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0957417410009310>
- Guo, X., Li, D., & Zhang, A. (2012). Improved Support Vector Machine Oil Price Forecast Model Based on Genetic Algorithm Optimization Parameters. *AASRI Procedia*, 1, 525–530. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S2212671612000832>
- Hosseini, M., Javaherian, A., & Movahed, B. (2014). Determination of permeability index using Stoneley slowness analysis, NMR models, and formation evaluations: a case study from a gas reservoir, south of Iran. *Journal of Applied Geophysics*, 109, 80–87.
- Ilhan, I., & Tezel, G. (2013). A genetic algorithm-support vector machine method with parameter optimization for selecting the tag SNPs. *Journal of Biomedical Informatics*, 46(2), 328–40. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23262450>
- Jia, Z., Ma, J., Wang, F., & Liu, W. (2011). Hybrid of simulated annealing and SVM for hydraulic valve characteristics prediction. *Expert Systems with Applications*, 38(7), 8030–8036.
- Jia, Z.-Y., Ma, J.-W., Wang, F.-J., & Liu, W. (2010). Characteristics forecasting of hydraulic valve based on grey correlation and ANFIS. *Expert Systems with Applications*, 37(2), 1250–1255. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S09574174090005624>
- Kaytez, F., Taplamacioglu, M. C., Cam, E., & Hardalac, F. (2015). Electrical Power and Energy Systems Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *International Journal of Electrical Power and Energy Systems*, 67, 431–438. Retrieved from
- Lee, S., Kang, P., & Cho, S. (2014). Neurocomputing Probabilistic local reconstruction for k-NN regression and its application to virtual metrology in semiconductor manufacturing. *Neurocomputing*, 131, 427–439. Retrieved from
- Lira, M. A. T., Da Silva, E. M., Alves, J. M. B., & Veras, G. V. O. (2014). Estimation of wind resources in the coast of Ceará, Brazil, using the linear regression theory. *Renewable and Sustainable Energy Reviews*, 39, 509–529. Retrieved from
- Machairas, V., Tsangrassoulis, A., & Axarli, K. (2014). Algorithms for optimization of building design: A review. *Renewable and Sustainable Energy Reviews*, 31(1364), 101–112. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S1364032113007855>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*.
- Özbayoğlu, a. M., & Bozer, R. (2012). Estimation of the Burned Area in Forest Fires Using Computational Intelligence Techniques. *Procedia Computer Science*, 12, 282–287.
- Pan, S., Iplikci, S., Warwick, K., & Aziz, T. Z. (2012). Parkinson's Disease tremor classification – A comparison between Support Vector Machines and neural networks. *Expert Systems with Applications*, 39(12), 10764–10771.
- Quintano, C., Fernández-Manso, A., Stein, A., & Bijker, W. (2011). Estimation of area burned by forest fires in Mediterranean countries: A remote sensing data mining perspective. *Forest Ecology and Management*, 262(8), 1597–1607. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S0378112711004385>
- Raghavendra, N. S., & Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, 19, 372–386. Retrieved from <http://linkinghub.elsevier.com/retrieve/pii/S1568494614000611>
- Rynkiewicz, J. (2012). General bound of overfitting for MLP regression models. *Neurocomputing*, 90, 106–110.
- Singh, P., & Borah, B. (2014). International Journal of Approximate Reasoning Forecasting stock index price based on M-factors fuzzy time series and particle swarm optimization. *International Journal of Approximate Reasoning*, 55(3), 812–833.
- Suganyadevi, M. V., & Babulal, C. K. (2014). Support Vector Regression Model for the prediction of Loadability Margin of a Power System. *Applied Soft Computing Journal*, 24, 304–315.
- Tiryaki, S., Öz, Ş., & Y, İ. (2014). International Journal of Adhesion & Adhesives Comparison of arti fi cial neural network and multiple linear regression models to predict optimum bonding strength of heat treated woods, 55, 29–36.
- Wang, X., Wen, J., Zhang, Y., & Wang, Y. (2014). Optik Real estate price forecasting based on SVM optimized by PSO. *Optik - International Journal for Light and Electron Optics*, 125(3), 1439–1443.
- Yang, X. (2014). *Nature-Inspired Optimization Algorithms*. Elsevier. doi:10.1016/B978-0-12-416743-8.00005-1
- Yilmaz, I., & Kaynar, O. (2011). Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. *Expert Systems with Applications*, 38(5), 5958–5966.
- Zameer, A., Mirza, S. M., & Mirza, N. M. (2014). 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. *Annals of Nuclear Energy*, 65, 122–131.

- Zhang, D., Liu, W., Wang, A., & Deng, Q. (2011). Parameter Optimization for Support Vector Regression Based on Genetic Algorithm with Simplex Crossover Operator. *Journal of Information & Computational Science*, 6(June), 911–920. Retrieved from http://www.joics.com/publishedpapers/2011_8_6_911_920.pdf
- Zhao, M., Fu, C., Ji, L., Tang, K., & Zhou, M. (2011). Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm with feature chromosomes. *Expert Systems with Applications*, 38(5), 5197–5204.
- Zhao, W., Tao, T., & Zio, E. (2015). System reliability prediction by support vector regression with analytic selection and genetic algorithm parameters selection. *Applied Soft Computing*, 30, 792–802.

BIOGRAFI PENULIS



Hani Harafani. Memperoleh gelar M.Kom dari Sekolah Tinggi Manajemen Ilmu Komputer Nusa Mandiri, Jakarta. Staff pengajar di salah satu Perguruan Tinggi Swasta. Minat penelitian saat ini pada bidang data mining.



Romi Satria Wahono. Memperoleh gelar B.Eng dan M.Eng pada bidang ilmu komputer di Saitama University, Japan, dan Ph.D pada bidang software engineering di Universiti Teknikal Malaysia Melaka. Menjadi pengajar dan peneliti di Fakultas Ilmu Komputer, Universitas Dian Nuswantoro. Merupakan pendiri dan CEO PT Brainmatics, sebuah perusahaan yang bergerak di bidang pengembangan software. Minat penelitian pada bidang software engineering dan machine learning. Profesional member dari asosiasi ilmiah ACM, PMI dan IEEE Computer Society.