



## Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film

Styawati<sup>1</sup>, Andi Nurkholis<sup>2</sup>, Zaenal Abidin<sup>3</sup>, Heni Sulistiani<sup>4</sup>

<sup>1,3</sup>Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

<sup>2</sup>Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

<sup>4</sup>Sistem Informasi Akuntansi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia

<sup>1</sup>styawati@teknokrat.ac.id, <sup>2</sup>andinh@teknokrat.ac.id, <sup>3</sup>zabin@teknokrat.ac.id, <sup>4</sup>henisulistiani@teknokrat.ac.id

### Abstract

The Support Vector Machine (SVM) method is a method that is widely used in the classification process. The success of the classification of the SVM method depends on the soft margin coefficient  $C$ , as well as the parameter  $\sigma$  of the kernel function. The SVM parameters are usually obtained by trial and error, but this method takes a long time because they have to try every combination of SVM parameters, therefore the purpose of this study is to find the optimal SVM parameter value based on accuracy. This study uses the Firefly Algorithm (FA) as a method for optimizing SVM parameters. The data set used in this study is data on public opinion on several films. Class labels used in data classification are positive class labels and negative class labels. The amount of data used in this study is 2179 data, with the distribution of 436 data as test data and 1743 data as training data. Based on this data, an evaluation process was carried out on the Firefly Algorithm-Support Vector Machine (FA-SVM). The results of this study indicate that the Firefly Algorithm can obtain the optimal combination of SVM parameters based on accuracy, so there is no need for trial and error to get that value. This is evidenced by the results of the FA-SVM evaluation using a value range of  $C=1.0-3.0$  and  $\sigma=0.1-1.0$  resulting in the highest accuracy of 87.84%. The next evaluation using a range of values  $C=1.0-3.0$  and  $\sigma=1.0-2.0$  resulted in the highest accuracy of 87.15%.

Keywords: SVM, FA-SVM, Classification, Optimization, Public Opinion

### Abstrak

Metode Support Vector Machine (SVM) merupakan metode yang banyak digunakan dalam proses klasifikasi. Keberhasilan klasifikasi metode SVM bergantung pada koefisien *soft margin*  $C$ , serta parameter  $\sigma$  dari fungsi kernel. Parameter SVM tersebut biasanya didapatkan dengan cara *trial and error*, namun cara tersebut membutuhkan waktu yang cukup lama karena harus mencoba setiap kombinasi parameter SVM, maka dari itu tujuan dari penelitian ini adalah mencari nilai parameter SVM yang optimal berdasarkan akurasi. Penelitian ini menggunakan *Firefly Algorithm* (FA) sebagai metode optimasi parameter SVM. Data set yang digunakan dalam penelitian ini adalah data opini masyarakat terhadap beberapa film. Label kelas yang digunakan dalam klasifikasi data yaitu label kelas positif dan label kelas negatif. Banyaknya data yang digunakan dalam penelitian ini yaitu 2179 data, dengan pembagian data sebanyak 436 sebagai data pengujian dan 1743 data sebagai data pelatihan. Berdasarkan data tersebut dilakukan proses evaluasi terhadap *Firefly Algorithm-Support Vector Machine* (FA-SVM). Hasil penelitian ini menunjukkan bahwa Algoritma *Firefly* mampu mendapatkan kombinasi parameter SVM yang optimal berdasarkan akurasi, sehingga tidak diperlukan cara *trial and error* untuk mendapatkan nilai tersebut. Hal ini dibuktikan dengan hasil evaluasi FA-SVM menggunakan rentang nilai  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$  menghasilkan akurasi tertinggi yaitu 87.84%. Evaluasi berikutnya menggunakan rentang nilai  $C=1.0-3.0$  dan  $\sigma=1.0-2.0$  menghasilkan akurasi tertinggi 87.15%.

Kata kunci: SVM, FA-SVM, Klasifikasi, Optimasi, Opini masyarakat.

### 1. Pendahuluan

Optimasi merupakan proses mencari solusi optimal untuk masalah tertentu yang menarik, dan proses pencarian ini dapat dilakukan dengan menggunakan beberapa agen yang pada dasarnya membentuk sistem

agen yang berkembang [1]. Terdapat banyak teknik optimasi diantaranya yaitu Particle Swarm Optimization (PSO), *Firefly* (FA), Bat Algorithm (BA), Cuckoo Search (CS), Ant Colony Optimization (ACO), Flower Pollination Algorithm (FPA), Weighted Superposition

Attraction (WSA) [2]. Teknik optimasi juga banyak digunakan diberbagai bidang penelitian diantaranya yaitu bidang transportasi, telekomunikasi, *image processing, engineering design, Vehicle routing* [3]. Selain bidang yang telah disebutkan, teknik optimasi juga banyak digunakan untuk mencari nilai optimal parameter metode klasifikasi diantaranya yaitu optimasi parameter SVM [4][5]. SVM merupakan metode yang banyak digunakan dalam proses klasifikasi [6].

Klasifikasi merupakan proses mengelompokkan data ke dalam label kelas yang telah ditentukan [7]. SVM dan ekstensinya merupakan metode yang paling sukses dalam melakukan klasifikasi [8]. SVM juga memiliki kemampuan generalisasi yang baik dalam kasus dengan sample yang kecil [9]. SVM merupakan metode yang memiliki kinerja lebih baik dibandingkan dengan metode *machine learning* lainnya tetapi sensitif terhadap pengaturan parameter dan data latih [6]. Menurut [9] SVM tidak dapat memilih parameter yang sesuai sehingga penggunaan parameter menjadi tidak optimal. Dengan penggunaan parameter yang sesuai diharapkan dapat meningkatkan akurasi SVM [10]. Pentingnya penggunaan nilai parameter yang sesuai juga dijelaskan oleh [11] dikatakan bahwa keberhasilan dari model SVM bergantung pada koefisien soft-margin C, serta pada parameter dari fungsi kernel. Penelitian yang dilakukan oleh [12] juga menyatakan bahwa pengaturan parameter SVM sangat penting dilakukan karena berkaitan dengan akurasi dan efisiensi. Berdasarkan pada penelitian terdahulu dapat disimpulkan bahwa memilih parameter optimal untuk SVM adalah salah satu langkah penting ketika menggunakan SVM sebagai metode klasifikasi. Penentuan parameter optimal dapat dilakukan dengan metode optimasi.

Terdapat banyak metode optimasi diantaranya yaitu particle swarm optimization algorithm (PSO) [13], genetic algorithm (GA) [14] dan *firefly* (FA) [15]. Menurut penelitian yang dilakukan oleh [16] menggunakan dataset heart, diabetes, liver, iris, dan cancer dengan menggunakan algoritma PSO-SVM, APSO-SVM, dan FA-SVM diperoleh hasil akurasi tertinggi dari kombinasi algoritma FA-SVM. Penelitian yang dilakukan oleh [17] menggunakan metode FA-SVM untuk melakukan klasifikasi penyakit kanker dengan data microarray ekspresi gen juga menghasilkan akurasi yang tinggi.

Berdasarkan penelitian terdahulu, maka pada penelitian ini dilakukan optimasi parameter SVM. Parameter yang dioptimasi adalah C dan  $\sigma$ . Proses optimasi menggunakan algoritma *Firefly*. Optimasi ini bertujuan untuk mencari nilai parameter SVM yang optimal berdasarkan akurasi.

## 2. Metode Penelitian

Penelitian ini dilakukan dengan metode *Firefly* (FA) sebagai metode optimasi dan metode *Support Vector Machine* (SVM) sebagai metode klasifikasi. Penggunaan metode FA sebagai metode optimasi karena metode ini berpotensi lebih cepat dalam menemukan nilai optimal jika dibandingkan dengan algoritma Genetika dan PSO [3]. Sedangkan metode SVM digunakan pada penelitian ini karena metode ini berbasis statistik sehingga klasifikasi yang dihasilkan lebih baik dibandingkan dengan metode lain [18]. Selain itu SVM juga bekerja dengan sangat baik pada data berdimensi tinggi. SVM dengan menggunakan teknik kernel dapat memetakan data asli yang berdimensi rendah ke dimensi yang lebih tinggi [19]. SVM juga memiliki kinerja yang sangat baik dalam memprediksi kelas dari data baru [20].

### 2.1. Algoritma FA

FA merupakan metode metaheuristik dalam kelompok Swarm Intelligence (SI) yang mengadopsi perilaku sosial dan cara berkomunikasi sekelompok kunang-kunang melalui cahaya di bagian ekornya [21]. Algoritma FA sebagai berikut

```
Algoritma Firefly
Fungsi objektif  $f(x)$ ,  $x = (x_1, \dots, x_d)^T$ 
Inisialisasi populasi kunang-kunang  $x_i$  ( $i = 1, 2, \dots, n$ )
Definisikan koefisien light absorption ( $\gamma$ )
while ( $t < \text{MaksGenerasi}$ )
  for  $i = 1$  to  $n$  //untuk semua  $n$  kunang-kunang
    for  $j = 1$  to  $n$  //untuk semua  $n$  kunang-kunang
      if ( $I_j > I_i$ )
        Gerakkan kunang-kunang  $i$  menuju  $j$ 
        untuk semua dimensi  $d$ 
      end
      Attractiveness bervariasi dengan jarak  $r$  melalui  $\exp[-\gamma r]$ 
      Evaluasi solusi-solusi baru dan perbarui intensitas cahaya (light intensity)
    end for  $j$ 
  end for  $i$ 
  Buat peringkat kunang-kunang dan cari yang terbaik saat ini (current best)
end while
Pascaproses hasil dan visualisasi
```

Berdasarkan pada algoritma FA, fungsi objektif didapatkan dari persamaan berikut

$$\text{maks } f(x), x = (x_1, \dots, x_d)^T \quad (1)$$

Selain menentukan fungsi objektif, algoritma FA juga harus menentukan intensitas cahaya dari setiap kunang-kunang. Rumus untuk menentukan intensitas cahaya adalah sebagai berikut

$$I(x) = f(x) \quad (2)$$

Setelah mendapatkan intensitas cahaya, selanjutnya mencari jarak antar kunang-kunang. Jarak antar kunang diperoleh dari rumus berikut

$$r_{1,2} = \sqrt{(m_1^1 - m_2^1)^2 + (m_1^2 - m_2^2)^2} \quad (3)$$

Selain jarak, algoritma *Firefly* juga membutuhkan nilai daya tarik setiap kunang-kunang (*attractiveness*). Nilai *attractiveness* dapat dicari dengan rumus

$$\beta = \beta_0 e^{-\gamma r^2} \quad (4)$$

Perpindahan *firefly* adalah penentu dari algoritma *Firefly*. *Firefly* yang memiliki intensitas cahaya lebih kecil akan berpindah menuju *firefly* yang memiliki intensitas cahaya lebih tinggi. Rumus untuk menghitung perpindahan *firefly* sebagai berikut

$$m_{i\_baru}^k = m_{i\_lama}^k + \beta_0 e^{-\gamma r^2} (m_{j\_lama}^k - m_{i\_lama}^k) + \alpha \left( \text{rand} - \frac{1}{2} \right) \quad (5)$$

## 2.2. Algoritma Support Vector Machine (SVM)

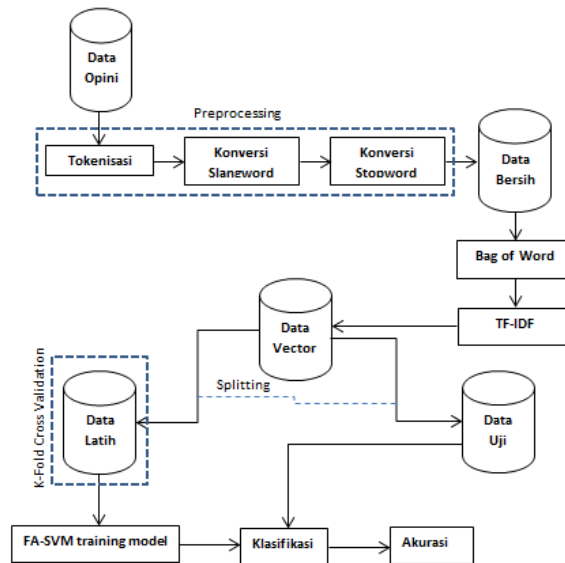
SVM merupakan metode data mining yang masuk dalam kelas supervised learning. Tujuan utama dari teknik ini adalah mencari *hyperplane* terbaik dari beberapa *hyperplane* yang ada. Seperti yang dikatakan oleh [22] metode SVM berusaha mendapatkan fungsi pemisah yang optimal untuk dapat memisahkan dua set data dari dua kelas yang berbeda.

Pada teknik SVM, *hyperplane* terbaik adalah *hyperplane* yang terletak di tengah-tengah antara dua set data dari dua kelas yang berbeda. Mencari *hyperplane* terbaik adalah dengan cara memaksimalkan margin. Margin adalah jarak tegak lurus antara *hyperplane* dengan obyek terdekat. Obyek terdekat disebut juga support vector [20].

## 2.3. Rancangan Proses Penelitian

Proses penelitian ini secara umum memiliki lima proses utama, yaitu pengumpulan data, *preprocessing* data, pembobotan, klasifikasi menggunakan metode FA-SVM, dan pengujian. Proses penelitian dapat dilihat pada gambar 1.

Berdasarkan gambar 1, tahap pertama dari penelitian ini adalah pengumpulan data. Pengumpulan data dilakukan dengan teknik *scraping* yang telah dilakukan oleh peneliti sebelumnya yaitu [23]. Data yang dihasilkan oleh peneliti sebelumnya dilakukan *preprocessing* dengan metode *case folding* dan normalisasi fitur. Pada penelitian ini ditambahkan proses tokenisasi, konversi *slangword*, dan konversi *stopword*. Penambahan proses tersebut dikarenakan masih terdapat kata yang tidak memiliki makna seperti “yaitu” dan kesalahan penulisan seperti “gak jelas”. Setelah proses *preprocessing* kemudian data tersebut dipetakan ke dalam model vektor dengan cara *bag of word*. Kemudian dilakukan pembobotan dengan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*.



Gambar 1. Proses Penelitian

Pembobotan kata pada penelitian ini dilakukan untuk menentukan bobot kata pada data opini film. Contoh data yang digunakan pada penelitian ini dapat dilihat pada tabel 1

Tabel 1. Contoh data *tweet*

Komentar	Sentimen
membosankan	Negatif
film aneh	Negatif
sedikit membosankan	Negatif
saya bosan lihat film ini	Negatif
kamu akan kecewa	Negatif
bagus sekali	Positif
tidak rugi nonton film ini	Positif
gak nyangka filmnya bakal seru	Positif
sumpah keren	Positif

Data pada tabel 1 dilakukan pembobotan, sehingga didapatkan bobot kata yang dapat merepresentasikan dasar informasi yang akan digunakan untuk menentukan klasifikasi pada data tersebut. Setelah proses pembobotan, kemudian dilakukan pembagian data. Data dibagi menjadi dua yaitu data latih dan data uji. Data latih dibagi menjadi dua yaitu data latih dan data validasi. Metode yang digunakan dalam pembagian data latih dan data uji adalah metode *Splitting*, sedangkan metode untuk membagi data latih dan data validasi adalah metode *K-Fold Cross Validation*. Penelitian ini menggunakan 10 *fold*. Data latih digunakan untuk membuat model klasifikasi FA-SVM. Sedangkan data validasi untuk mencari parameter yang paling baik. Kemudian dilakukan pengujian model dengan data uji yang bertujuan untuk mengukur sejauh mana *classifier* berhasil melakukan proses klasifikasi.

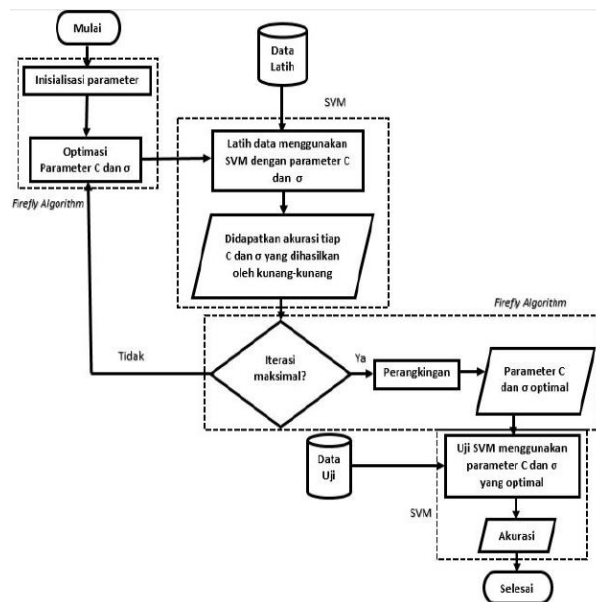
## 3. Hasil dan Pembahasan

Pada penelitian ini dilakukan optimasi terhadap parameter SVM dengan menggunakan metode *Firefly*.

Parameter yang dioptimasi adalah  $C$  dan  $\sigma$ . Parameter  $C$  pada SVM berfungsi untuk mengontrol pertukaran antara margin dan kesalahan klasifikasi. Sedangkan  $\sigma$  pada SVM digunakan untuk menemukan nilai optimal dalam setiap dataset.

### 3.1 Rancangan Pelatihan Metode FA-SVM

Klasifikasi data opini film menggunakan metode FA-SVM diawali dengan melakukan inialisasi parameter yang dibutuhkan untuk proses pencarian dengan *firefly* yaitu menentukan jumlah populasi *firefly* (*number\_of\_fireflies*), banyaknya generasi (*maximun\_generation*), koefisien *attractiveness* awal ( $\beta_0$ ), koefisien *light absorption* ( $\gamma$ ), dan koefisien parameter random ( $\alpha$ ). Setelah diinisialisasi parameter yang dibutuhkan, kemudian lakukan optimasi terhadap parameter  $C$  dan  $\sigma$  menggunakan metode *Firefly*. Selanjutnya nilai  $C$  dan  $\sigma$  yang diperoleh dari masing-masing *firefly* digunakan untuk melatih data. Setelah data dilatih menggunakan model SVM, kemudian hitung akurasi tiap  $C$  dan  $\sigma$  yang dihasilkan oleh kunang-kunang. Setelah akurasi dari setiap *firefly* dan setiap generasi diketahui maka dilakukan perangkingan untuk mengetahui nilai parameter  $C$  dan  $\sigma$  yang paling optimal. Selanjutnya nilai  $C$  dan  $\sigma$  digunakan untuk memodelkan classifier SVM kemudian model tersebut diuji dengan menggunakan data uji. Dari hasil pengujian akan didapatkan akurasi berdasarkan data uji. Proses klasifikasi dapat dilihat pada gambar 2.



Gambar 2. Proses klasifikasi FA-SVM

Proses klasifikasi FA-SVM berdasarkan gambar 2 yaitu pertama menentukan fungsi objektif. Fungsi objektif didapatkan dari persamaan (1).

$$\text{maks } f(x) = \frac{TP(x)+TN(x)}{TP(x)+TN(x)+(FP(x)+FN(x))}$$

Setelah menentukan fungsi objektif, kemudian menentukan konstanta yang akan digunakan dalam optimasi. Konstanta tersebut adalah banyaknya *firefly*,  $\beta_0 = 1$ ,  $\gamma = 0.23$ ,  $\text{rand} = 0.2$ , dan  $\alpha = 0.2$ . Jika nilai konstanta telah ditentukan, selanjutnya membangkitkan populasi awal *Firefly*. Populasi awal *firefly* diberikan secara random dengan interval nilai [0,1]. Nilai populasi awal *firefly* dapat dilihat pada tabel 2.

Tabel 2. Nilai populasi awal *firefly*

<i>Firefly</i>	Random nilai $C$ dan $\sigma$	
	$C$	$\sigma$
$m1$	0.2	0.6
$m2$	0.3	0.5
$m3$	0.5	1

Berdasarkan pada tabel 1, nilai  $C$  dan  $\sigma$  digunakan untuk mencari nilai akurasi yang nantinya digunakan sebagai nilai evaluasi setiap *firefly*. Hasil dari evaluasi *firefly* dengan nilai random yang telah ditentukan sebelumnya dapat dilihat pada tabel 3.

Tabel 3. Evaluasi awal

	$m1$	$m2$	$m3$
$f(m)$	0.71	0.76	0.76

Setelah mendapatkan nilai evaluasi awal dari masing-masing *firefly*, kemudian mencari nilai intensitas cahaya untuk masing-masing *firefly*. Nilai intensitas cahaya dapat dihitung menggunakan persamaan (2). Hasil intensitas cahaya masing-masing *firefly* terlihat pada tabel 4.

Tabel 4. Intensitas cahaya masing-masing *firefly*

	$m1$	$m2$	$m3$
$I(m)$	0.71	0.76	0.76

Berdasarkan pada tabel 4, nilai intensitas cahaya digunakan untuk membandingkan setiap *firefly*. Jika intensitas cahaya suatu *firefly* lebih kecil dari pada intensitas cahaya *firefly* lainnya, maka *firefly* tersebut akan melakukan perpindahan menuju *firefly* yang memiliki intensitas cahaya lebih besar.  $I(m1)$  dibandingkan dengan  $I(m2)$ , karena  $I(m1) < I(m2)$  maka  $m1$  berpindah menuju ke arah  $m2$ . Perpindahan *firefly* dapat dilakukan dengan terlebih dahulu menghitung jarak (*distance*) antara *firefly* 1 dan 2 untuk masing-masing elemen real pada *firefly*. Perhitungan jarak antar *firefly* dapat dilakukan dengan persamaan (3).

$$r_{1,2} = \sqrt{(m_1^1 - m_2^1)^2 + (m_1^2 - m_2^2)^2}$$

$$r_{1,2} = \sqrt{(0.2 - 0.3)^2 + (0.6 - 0.5)^2}$$

$$r_{1,2} = 0.141$$

Setelah mendapatkan jarak antara *firefly* 1 dengan *firefly* 2, kemudian menghitung *attractiveness firefly* tersebut. *Attractiveness firefly* 1 dan 2 dapat dihitung dengan persamaan (4).

$$\beta = \beta_0 e^{-\gamma r^2}$$

$$\beta = 1 \times e^{-0.23(0.141)^2}$$

$$\beta = 0.995$$

Nilai *attractiveness firefly* digunakan sebagai salah satu acuan perpindahan *firefly*. Perpindahan *firefly* per elemen dapat dilakukan menggunakan persamaan (5)

Perpindahan *firefly* pada elemen 1

$$m_{i\_baru}^k = m_{i\_lama}^k + \beta_0 e^{-\gamma r^2} (m_{j\_lama}^k - m_{i\_lama}^k) + \alpha(\text{rand} - \frac{1}{2})$$

$$m_{1\_baru}^1 = 0.2 + 0.995(0.3 - 0.2) + 0.2(0.2-0.5)$$

$$m_{1\_baru}^1 = 0.4405$$

Perpindahan *firefly* pada elemen 2

$$m_{1\_baru}^2 = 0.6 + 0.995(0.5 - 0.6) + 0.2(0.2 - 0.5)$$

$$m_{1\_baru}^2 = 0.4405$$

Posisi baru *firefly m1* adalah  $C=0.2395$  dan  $\sigma=0.4405$ .

Kemudian menghitung fungsi objektif baru pada *firefly* 1. Fungsi objektif diperoleh dari nilai akurasi yang dihasilkan oleh algoritma SVM. Sehingga diperoleh fungsi objektif baru  $f(m1) = 0.74$ . Setelah itu hitung intensitas cahayanya  $I(m1) = 0.74$ . Perbandingan dilanjutkan pada *firefly* 1 terhadap *firefly* 3.  $I(m1)$  dibandingkan dengan  $I(m3)$ , karena  $I(m1) < I(m3)$  maka  $m1$  berpindah menuju ke arah  $m3$ . Perpindahan *firefly* dapat dilakukan dengan langkah f dan g. Hasil dari perhitungan langkah f dan g, didapatkan posisi baru dari  $m1$  yaitu  $C=0.4232$  dan  $\sigma = 0.9176$ . Fungsi tujuan baru  $f(m1) = 0.80$  dan  $I(m1) = 0.80$ . Kemudian, perbandingan dilanjutkan pada *firefly* 2 terhadap *firefly* 3.  $I(m2)$  dibandingkan dengan  $I(m3)$ , karena  $I(m2)=I(m3)$  maka tidak terjadi perpindahan dan seterusnya. Proses *movement* menghasilkan parameter baru dan intensitas cahaya baru dari masing-masing *firefly*. Dari hasil perpindahan yang telah dilakukan, kemudian menentukan *firefly* terbaik. *Firefly* terbaik dapat ditentukan berdasarkan pada intensitas cahayanya  $I(m)$ . Intensitas cahaya tertinggi dari masing-masing *firefly* pada iterasi 1 adalah  $I(m1) = 0.80$ . Berdasarkan pada intensitas cahaya tersebut maka dapat disimpulkan bahwa *firefly* 1 ( $m1$ ) adalah *firefly* terbaik. Berdasarkan pada nilai *firefly* terbaik dapat ditentukan global best sementara. Hasil *movement* pada iterasi 1 diperoleh satu *firefly* dengan intensitas cahaya tertinggi yaitu *firefly*1, maka dari itu *firefly* 1 digunakan sebagai  $g\text{-best} = m1$ . Setelah itu dilakukan *movement* pada *firefly* terbaik. Pergerakan pada *firefly* terbaik atau  $g\text{-best}$  sementara dilakukan agar tidak terjebak pada optimum lokal dan untuk membentuk populasi baru. Proses perpindahan  $g\text{-best}$  sementara pada elemen 1 sebagai berikut

$$m_i^k = m_i^k + \alpha(\text{rand} - 0.5)$$

$$m_1^1 = 0.4 + 0.2(0.2 - 0.5)$$

$$m_1^1 = 0.34$$

Proses perpindahan  $g\text{-best}$  sementara pada elemen 2

$$m_1^1 = 0.9 + 0.2(0.2 - 0.5)$$

$$m_1^1 = 0.84$$

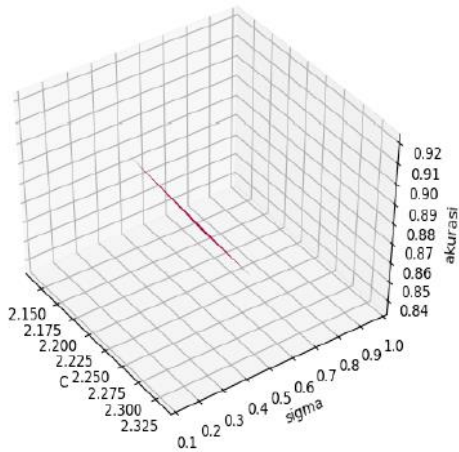
Proses *movement g-best* menghasilkan nilai parameter  $C=0.34$  dan  $\sigma=0.84$  dengan fungsi tujuan  $f(m1) = 0.80$  dan intensitas cahaya  $I(m1) = 0.80$ . Perbandingan dilakukan antara  $g\text{-best}$  sementara dengan  $f(m1)$  yang diperoleh dari hasil *movement*, karena nilai  $g\text{-best}$  sementara  $I(g\text{-best})=I(m1)$  maka nilai  $g\text{-best}$  sementara menjadi nilai  $g\text{-best}$  tetap pada iterasi 1. Setelah didapatkan nilai  $g\text{-best}$  pada iterasi 1, kemudian melakukan pengecekan iterasi, jika proses pencarian telah mencapai maksimum iterasi maka  $m_i$  pada iterasi terakhir akan dibandingkan dengan  $g\text{-best}$ , jika  $I(m1) > g\text{-best}$  maka  $g\text{-best} = m_i$ . Maksimum iterasi pada perhitungan manual ini adalah 1. Hasil pencarian menunjukkan bahwa  $I(g\text{-best}) = I(m_i)$ , maka  $g\text{-best} = g\text{-best}$ .

### 3.2. Evaluasi Metode FA-SVM

Evaluasi kinerja metode FA-SVM dilakukan dengan menggunakan pendekatan *k-fold cross validation*. Proses *k fold cross validation* dalam penelitian ini menggunakan 10 *fold* ( $k=10$ ) dengan ukuran partisi data yang sama dan dilakukan secara acak. Data di dalam *fold* akan dibagi menjadi 10 *subset* sehingga setiap *subset* memiliki ukuran yang sama dan memiliki data yang berbeda. Penggunaan 10 *fold* terhadap 2179 data mendapatkan pembagian sebanyak 436 data sebagai data pengujian dan 1743 data sebagai data pelatihan. Proses tersebut akan diulang sebanyak 10 *fold* dengan distribusi data yang berbeda. Sebelum dilakukan evaluasi terhadap metode FA-SVM terlebih dahulu dilakukan pencarian parameter SVM menggunakan metode *Firefly*. Parameter SVM yang dicari dalam rentang  $C=1.0-3.0$ ,  $\sigma= 0.1-1.0$  dan  $C=1.0-3.0$ ,  $\sigma=1.0-2.0$ . Jumlah *firefly* yang digunakan sebanyak 30, jumlah generasi yang digunakan adalah 50. Nilai random  $\alpha =0.2$  dan nilai  $\gamma=1.0$ .

Hasil evaluasi metode FA-SVM menggunakan data opini film dengan rentang nilai parameter  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$  memberikan hasil akurasi tertinggi yaitu 87.84%. Hasil tersebut ditunjukkan pada Gambar 3.

Pengujian metode FA-SVM dilakukan dengan mencari nilai  $C$  dan  $\sigma$  terbaik berdasarkan nilai akurasi tertinggi yang diperoleh dari setiap *Firefly*. Pencarian dengan metode *Firefly* menghasilkan 50 kombinasi  $C$  dan  $\sigma$ . Dari 50 kombinasi tersebut terdapat tiga kombinasi terbaik berdasarkan akurasi tertinggi yaitu 87.84%. Pengujian ini membutuhkan waktu eksekusi selama 2330 detik. Nilai  $C$ ,  $\sigma$ , akurasi dan waktu eksekusi disajikan pada Tabel 3



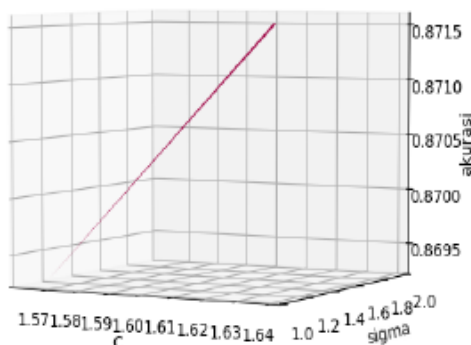
Gambar 3. Hasil Evaluasi FA-SVM  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$

Tabel 3. Hasil evaluasi FA-SVM  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$

C	Sigma	Akurasi %	Waktu/s
2.33	0.45	87.84	
2.25	0.46	87.84	2330
2.13	0.50	87.84	

Berdasarkan Tabel 3 dapat dilihat bahwa dengan metode *Firefly* dihasilkan tiga kombinasi nilai  $C$  dan  $\sigma$  terbaik berdasarkan akurasi tertinggi yaitu 87.84%. Kombinasi tersebut didapatkan dari rentang nilai  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$ . Hasil akurasi 87.84% didapatkan dari proses evaluasi menggunakan *Confusion Matrix*. Evaluasi tersebut menghasilkan 266 data true positive, 117 data true negative, 39 data false positive, dan 14 data false negative.

Pengujian selanjutnya masih menggunakan 436 data uji namun menggunakan rentang yang berbeda dengan pengujian sebelumnya. Rentang parameter FA-SVM yang digunakan adalah  $C=1.0-3.0$  dan  $\sigma=1.0-2.0$ . Hasil dari pengujian ini dapat dilihat pada gambar 4.



Gambar 4. Hasil Evaluasi FA-SVM  $C=1.0-3.0$  dan  $\sigma=1.0-2.0$

Berdasarkan Gambar 4 dapat dilihat bahwa pencarian dengan metode *Firefly* menghasilkan 50 kombinasi  $C$  dan  $\sigma$ . Dari kombinasi tersebut didapatkan akurasi tertinggi yaitu 87.15% dengan nilai  $C=1.63$  dan  $\sigma=1.08$ . Akurasi 87.15% diperoleh dari proses evaluasi

menggunakan *Confusion Matrix*. Evaluasi tersebut menghasilkan 265 data true positive, 115 data true negative, 41 data false positive, dan 15 data false negative. Pengujian ini membutuhkan waktu eksekusi selama 2388 detik. Berdasarkan pada percobaan dengan nilai  $C=1.0-3.0$ ,  $\sigma=0.1-1.0$  dan  $C=1.0-3.0$ ,  $\sigma=1.0-2.0$  dapat disimpulkan bahwa algoritma *firefly* mampu memperoleh nilai parameter SVM yang optimal berdasarkan akurasi.

#### 4. Kesimpulan

Hasil penelitian ini menunjukkan bahwa *Firefly* mampu mendapatkan kombinasi parameter SVM yang sesuai berdasarkan akurasi, sehingga tidak diperlukan cara *trial and error* untuk mendapatkan nilai tersebut. Kesimpulan ini dibuktikan dengan hasil evaluasi metode FA-SVM dengan rentang nilai  $C=1.0-3.0$  dan  $\sigma=0.1-1.0$  menghasilkan akurasi tertinggi sebesar 87.84%. Evaluasi berikutnya dilakukan dengan rentang nilai  $C=1.0-3.0$  dan  $\sigma=1.0-2.0$  menghasilkan akurasi tertinggi yaitu 87.15%.

#### Ucapan Terimakasih

Ucapan terima kasih kami sampaikan kepada Universitas Teknokrat Indonesia yang telah memberikan hibah Penelitian Pembinaan Kapasitas (PPK) tahun pelaksanaan 2021.

#### Daftar Rujukan

- [1] X. Yang, "Metaheuristic Algorithms for Self-Organizing Systems: A Tutorial," *2012 IEEE Sixth Int. Conf. Self-Adaptive Self-Organizing Syst.*, vol. 40, no. 3, pp. 28–29, 2012, doi: 10.1109/SASO.2012.40.
- [2] X. S. Yang, S. Fong, X. He, S. Deb, and Y. Zhao, "Swarm Intelligence: Today and Tomorrow," *Proc. - 2016 3rd Int. Conf. Soft Comput. Mach. Intell. ISCM 2016*, pp. 219–223, 2017, doi: 10.1109/ISCM.2016.34.
- [3] X. Yang, "From Swarm Intelligence to Metaheuristics: Nature-Inspired Optimization Algorithms," *Computer (Long. Beach. Calif.)*, vol. 49, no. 9, 2016.
- [4] N. Almugren and H. Alshamlan, "FF-SVM: New FireFly-based Gene Selection Algorithm for Microarray Cancer Classification," *2019 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2019*, 2019, doi: 10.1109/CIBCB.2019.8791236.
- [5] W. Jiao, Z. Liu, and Y. Zhang, "Fault Diagnosis of Modular Multilevel Converter with FA-SVM Algorithm," *Chinese Control Conf.*, pp. 5093–5098, 2019.
- [6] A. Lawi and F. Aziz, "Classification of credit card default clients using LS-SVM ensemble," *Proc. 3rd Int. Conf. Informatics Comput. ICIC 2018*, pp. 1–4, 2018, doi: 10.1109/IAC.2018.8780427.
- [7] S. Styawati and K. Mustofa, "A Support Vector Machine-Firefly Algorithm for Movie Opinion Data Classification," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 13, no. 3, p. 219, 2019, doi: 10.22146/ijccs.41302.
- [8] X. Wu, W. Zuo, L. Lin, W. Jia, and D. Zhang, "F-SVM: Combination of Feature Transformation and SVM Learning via Convex Relaxation," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 11, pp. 5185–5199, 2018, doi: 10.1109/TNNLS.2018.2791507.

- [9] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion Mining of Movie Review Using Hybrid Method of Support Vector Machine and Particle Swarm Optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013, doi: 10.1016/j.proeng.2013.02.059.
- [10] N. Devi K and J. P., "Sentiment Classification Using SVM And PSO," *Int. J. Adv. Eng. Technol. E- Int J Adv Engg Tech VII/Issue II*, pp. 411–413, 2016.
- [11] E. Tuba, L. Mrkela, and M. Tuba, "Support Vector Machine Parameter Tuning using Firefly Algorithm Eva Tuba , Lazar Mrkela and Milan Tuba University of Belgrade Faculty of Mathematics Belgrade , Serbia," *2016 26th Int. Conf. Radioelektronika*, pp. 20–25, 2016.
- [12] C. F. Chao and M. H. Horng, "The construction of support vector machine classifier using the firefly algorithm," *Comput. Intell. Neurosci.*, vol. 2015, 2015, doi: 10.1155/2015/212719.
- [13] M. A. R. Khalid, M. Alwaqdani, and M. A. H. Farquad, "Comparative Analysis of Support Vector Machine: Employing Various Optimization Algorithms," *Proc. - 2015 14th Int. Conf. Inf. Technol. ICIT 2015*, no. 1, pp. 171–174, 2016, doi: 10.1109/ICIT.2015.52.
- [14] P. D. Windha Mega and Haryoko, "Optimization of parameter support vector machine (SVM) using genetic algorithm to review go-jek's services," *2019 4th Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2019*, vol. 6, pp. 301–304, 2019, doi: 10.1109/ICITISEE48480.2019.9003894.
- [15] Y. D. N. H, B. A. Pramudita, S. Wibirama, L. I. Izhar, and N. A. Setiawan, "EEG Motor Imagery Signal Classification Using Firefly Support Vector Machine," *2018 Int. Conf. Intell. Adv. Syst.*, 2018.
- [16] A. Sharma, A. Zaidi, R. Singh, S. Jain, and A. Sahoo, "Optimization of SVM classifier using Firefly algorithm," *2013 IEEE 2nd Int. Conf. Image Inf. Process. IEEE ICIIP 2013*, pp. 198–202, 2013, doi: 10.1109/ICIIP.2013.6707582.
- [17] N. Almugren and H. M. Alshamlan, "New Bio-Marker Gene Discovery Algorithms for Cancer Gene Expression Profile," *IEEE Access*, vol. 7, pp. 136907–136913, 2019, doi: 10.1109/ACCESS.2019.2942413.
- [18] H. Huang and Z. Wang, "Efficient Parameter Selection for SVM : The Case of Business Intelligence Categorization," *2017 IEEE Int. Conf. Intell. Secur. Informatics*, pp. 158–160, 2017.
- [19] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. CV Andi Offset, 2012.
- [20] B. Santoso and A. Umam, *Data Mining Dan Big Data Analytics*. Penebar Media Pustaka : Yogyakarta, 2018.
- [21] S. Suyanto, *Swarm Intelligence Komputasi Modern untuk Optimasi dan Big Data Mining*. Informatika Bandung, 2017.
- [22] K. Kr, A. R. Kv, and A. Pillai, "An Improved Feature Selection and Classification of Gene Expression Profile using SVM," *Int. Conf. Intell. Comput. Instrum. Control Technol.*, 2019.
- [23] S. Ernawati, "Penerapan Particle Swarm Optimization Untuk Seleksi Fitur Pada Analisis Sentimen Review Perusahaan Penjualan Online Menggunakan Naïve Bayes," *J. Evolusi*, vol. 4, no. 2015, pp. 8–14, 2016.